Integrating Population-based Patterns with Personal Routine to Re-engage Fitbit Use

Chia-Fang Chung
Human Centered Design & Engineering
DUB Group, University of Washington
cfchung@uw.edu

Catalina M. Danis
Computational Health Behavior & Decision Science
IBM T.J. Watson Research Center
danis@us.ibm.com

ABSTRACT

In this paper, we explore user reactions to prototypes that integrate population fitness data with personal practice to bolster motivation and help decrease pragmatic barriers to incorporating exercise in daily life. We conducted a study in a major United States based company that makes wearable devices available to their employees through its wellness program. We interviewed each of 26 employees to understand their exercise and tracking habits. Each expressed an interest in improving or maintaining his or her fitness level but was frustrated by not getting appropriate support from the current Fitbit application. Based on the interview findings, we designed four use cases and two prototypes to address the main problems revealed through analyses of the participants' interviews and their Fitbit data. In a second interview probing reactions to they prototype, the same group of users reported a desire for Fitbit to support their daily, dynamic routines and to help them make adaptive, weekly plans. In addition, the participants perceived benefits to leveraging at least one of three different types of personalized reference groups (self, friends, or a population of similar individuals) to increase their motivation and to help incorporate activities into their daily or weekly plans. We discuss design consideration for researchers and designers of personal informatics tools and organizational wellness plans.

CCS Concepts

• **Human-centered computing→Human computer interaction (HCI)→Empirical studies in HCI**

Keywords

fitness tracking; wearable devices; personal informatics; physical activity; planning; reference group; goal adaptation

1. INTRODUCTION

Well into the second decade after the introduction of consumer-grade wearable fitness devices [9], their value for helping guide users towards healthier lifestyles is still unclear. Many people stop using wearable devices within six months of starting [5][20]. Although self-tracking has shown to increase or maintain the frequency of many healthy behaviors [23], simply wearing the device, does not necessarily indicate that users are actively using the data to guide their behavior. In two studies, participants reported that they did not get support for achieving their fitness goals through these devices and lost motivation after a while [5][20]. Nonetheless, large employer organizations (i.e., those with over 200 employees), which are influential in the health insurance marketplace, are increasingly promoting use of fitness devices through their corporate wellness programs [13]. Thus it is important to continue to explore ways of improving the effectiveness of such devices for helping users manage their health related activities.

One approach used to engage users of fitness devices with their behavioral data is to provide them with suggestions relevant to personal routines. *Habito* provided feedback messages based on user historical and current activity levels and found that this increased users' frequency of checking tracking data [17]. *MyBehavior* learned user physical activity behavior patterns to provide actionable suggestions, which were shown to increase user adherence rate to these suggestions in the long run [27]. These efforts to provide context-relevant recommendations have been shown to be important for engaging users with and sustaining their tracking behavior, and have proven effective in sustaining physical activity [23]. We wonder what other dimensions people would consider *relevant* when receiving fitness recommendations pertinent to their personal routine? And, how can the information be integrated into people's daily routines to help guide behavior?

Another approach is through addressing barriers to and providing support for incorporating activity into daily life. Planning support solicited from groups of strangers or friends has been shown to produce benefits in health and other behavior change domains [2]. Online platforms such as PatientsLikeMe [33] help people manage their disease trajectory by providing opportunities to glimpse the adaptations of people drawn from a population similar to them. Population-based analyses based on the increasing number of people using wearable trackers and generating self-tracking data, may be able to extract valid insights and provide the basis for actionable suggestions to guide user behavior [28]. One open question is how these population-based data might be leveraged to engage an individual user to sustain her use of a device or motivate her to achieve a fitness goal. Who do people see as *comparable* to them, and thus a potential source of guidance, when making decisions about physical activities?

Integrating these two approaches has the potential to better support users by helping them refine their fitness goals and to address pragmatic barriers to achieving their goals. In this paper, we focus on users of fitness trackers who are committed to physical activity though they report challenges in reliably meeting physical activity goals. We explore their reactions to designs that integrate tracked behavioral data (at the individual and the population level) with individual routines. We hope to re-engage these users – that is, to enhance the usefulness of their tracking data and hence increase their interest in interacting with these data. Through interviews and feedback on design prototypes we provide a preliminary understanding of user needs and preferences for interacting with data collected by the Fitbit fitness trackers.

This is an important step before performing long-term evaluation [1]. Our goal is to answer the following research questions:

- How can systems use population- and self-tracked data to motivate and engage users in fitness and tracking?
- How can systems support goal refinement and adaptation to address challenges resulting from changing individual routines?

We interviewed 26 IT company employees who have used one of the current Fitbit devices for between three months and two years. Although most of the sample of users has positive attitudes toward increasing or maintaining physical activities, they reported decreasing interest in their activity data once they had achieved awareness of their baseline levels. Based on the interview data, we built four use cases and two digital prototypes aimed at helping to increase or maintain positive intention to exercise and perceived behavior control over pragmatic challenges. In a second interview we presented these prototypes to probe participant reactions. We found differences among employees as to which of three different personalized reference groups (self, social groups, or similarity groups) they find relevant and potentially motivating of their fitness activity. They also expressed a need to maintain a consistent activity level in response to daily fluctuations in work and family demands, which is at odds with their expressed desire for goal adaptation. Based on the study results, we discuss design opportunities to support beyond readiness and awareness and the use of varied reference groups to support goal attainment.

2. BACKGROUND

In this section, we review some of the most relevant literature in promoting physical activity and self-tracking and some of these efforts in an organizational setting that motivate our study.

2.1 Strategies for Increasing Physical Activity

Self-tracking and reflection have proven effective in improving physical activity levels [23] as well as other health-related behaviors [3][8]. Many applications have applied various persuasive strategies to keep their users motivated for improving their health and wellness status.

One common strategy is using goal setting. Specific and realistic goals are more effective than general goals, for example in weight loss [26]. Goals are also more effective when users can track their progress [22]. Giving users options of pursuing a primary and/or a secondary goal also helps sustain their motivation when the primary goal is difficult to achieve [7][25]. Our study aims to address practical challenges of goal attainment by adapting goals based on a user's behavior pattern and routine.

Another way to motivate people to improve their physical activity is through social comparison. Sharing progress with users of the same application [6], within the same work environment [14], and with a chosen role model [31] have been shown to increase user step count. The second aim of our study is to understand trade-offs in sharing and comparing fitness data with different reference groups, such as one based on an anonymous population dataset, a social group or an individual's personal history data.

2.2 Engagement with Fitness Tracking

To engage users with physical activity and self-tracking, researchers have explored different ways of visualization, such as goal-oriented "visual cuts" [10] or visualized performance progress using various persuasive techniques[7][18][21]. Since most user interactions with a fitness tracker are "glances", that is, brief views less than five seconds in duration [17], strategies such as instantiating competition with a user's past behavior have been suggested to encourage users check and reflect on the tracking information more often [17].

These research findings have led employers to develop programs to promote walking for health among their employees [4][16][32]. In the short-term, these studies showed promising results, finding that social relationships in the organization affect employee inclination to participate [4][16]. However, industry reports show that 30-50% of employees stop using their fitness tracker in less than six months [9], despite the compensation for the device and additional incentives to encourage exercise and tracking. To probe design opportunities to reengage these users with fitness tracking, we focus on employees who are committed to continued physical activity and already gained awareness of their physical activity behavior through prior tracking experience.

3. METHOD

3.1 Study Site and Participants

We conducted the study in a 1500-employee corporate facility of a major IT company in the United States. For the past two years, the company, through its wellness program, has been subsidizing employee purchase of Fitbit devices (up to USD $100) to encourage physical activity. The company also rewarded step achievements with virtual points redeemable for merchandise.

We recruited 26 employees who were current users of any of the models of Fitbit devices or had used one for more than three months in the past two years. Recruitment was through the "snowball" sampling method and through flyers posted in the cafeteria at the company site. The sample includes 7 females; all participants achieved at least a master's level of graduate education. The majority (11) of the participants were in the 35-44 age group, with 11 in older groups (six in the 45-54 age group, five in the over-55 group) and 5 in younger age groups (four in the 25-34 group, one in the 18-25 group). Seven participants used a Fitbit device for more than one year, eleven for 6-12 months, and six used it for less than 6 months (Table 1). We compensated each participant at the completion of the second interview with a lunch voucher (USD $8) redeemable in the company cafeteria.

3.2 Study Design

The study was approved by the company's internal IRB process. We conducted two interviews with each participant. The first took 20-40 minutes. We asked participants when and why they started using the Fitbit device and their experiences with the device. In addition, we asked about their typical work and physical activity routines. At the end we asked them for permission to access their Fitbit data via the Fitbit API. We analyzed the data — including steps, sleep, and weight — to ascertain their patterns.

Using data from a total population base of approximately 30,000 users of fitness devices, with step, weight and sleep records ranging over 1-1.5 years, we created a comparison reference group for each user based on their age range, weight, gender, average activity level (classified as sedentary, low active, somewhat active, active, highly active [30]), and weekly activity pattern. Based on these reference groups and the interview data, we created four storyboards (Figure 1) and two digital prototypes (Figure 2) to use as design probes [15] in the second interview. We describe the storyboards and prototypes in section 5.

We carried out a second interview with participants 2-3 weeks after the first. We first followed up on the previous interview to understand whether the participant use of or thoughts about Fitbit changed since the first interview. We then used the storyboards and prototypes to probe their reactions, expectations, and concerns about our designs that integrated a population data-informed view with their routine. The second interview took 30-45 minutes.

All interviews were audio-recorded and transcribed for analysis. We analyzed them using a combination of an inductive and a deductive approach. Both authors first coded one transcript from each interview round for emergent themes and then met to discuss discrepancies and consolidate codes. Then each coded another transcript from each round of interviews to verify the combined codes and to capture additional themes. Then they went on to use the final codes to code the rest of the transcripts.

4. INTERVIEW 1: FITBIT EXPERIENCE

4.1 Goals in Using the Fitbit Device and App

Less than 25% of the reasons users gave for initiating use of the device and associated app included specific behavioral goals, such as losing weight or managing one's health to better meet family caretaking obligations. Another roughly 25% of reasons could be classified as "external" to the user, such as the availability of the device from the employer or a spouse's urging. The majority (56% of total) referred to a general interest in gaining knowledge about oneself, including wanting awareness of activity levels or wanting to track one's behavior over time.

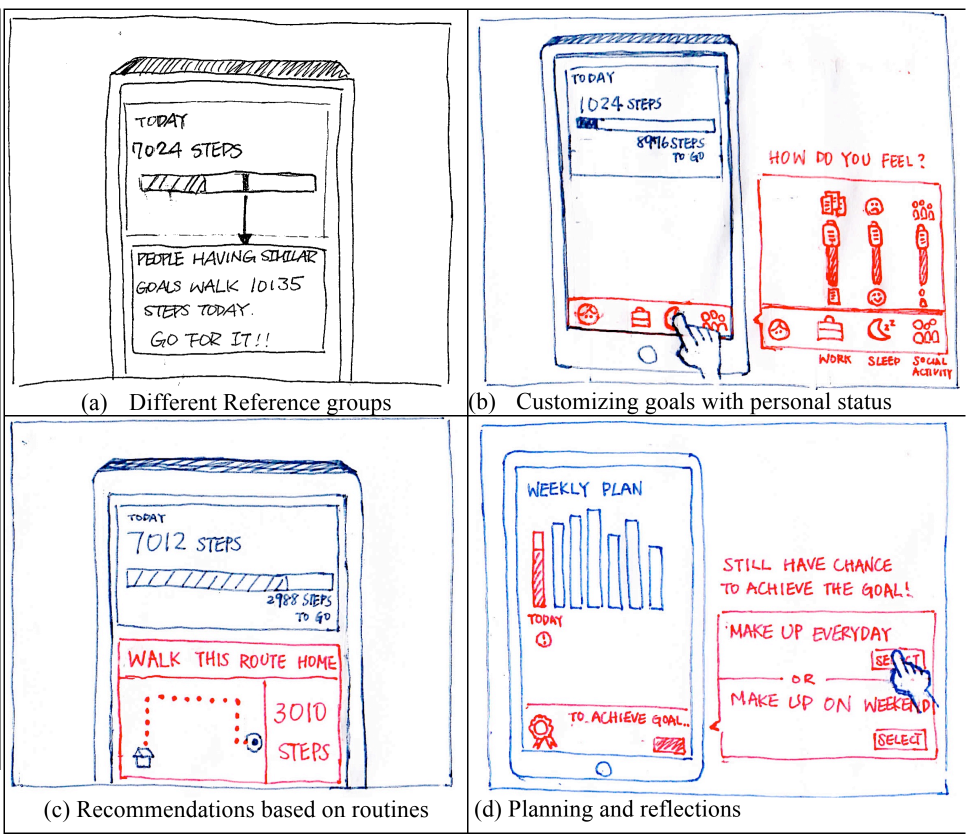
Once they started using the device, individuals found it difficult to regularly achieve the 10K step-count most commonly identified in step programs as the desirable target [19]. Only 35% of our sample had an average level at, or above, this target when computed across their full period of device usage. On average, the participants produced 8,608 steps per day, with a range of just fewer than 4,400 to almost 18,000. The activity level for each participant (using the nomenclature in [30]) is listed in Table 1. The majority (n=13) falls into the "somewhat active" group (7,500-10,000 steps/day). Many participants also reported participating in strength or cardio activities not fully quantified by pedometer-style devices.

Even those who did achieve the 10K level on at least some days reported significant fluctuations in step counts over the course of a week. Weekend patterns for most participants were less predictable than weekday patterns. Many noted that family and social obligations were the major determinants of their activities on the weekends and thus they were unable to make their physical activity as high a priority as during the week. Weekends could translate into a large number of steps, for example due to taking a hike with one's spouse (e.g., P26) or accompanying a child on her weekend routine in the city (e.g., P23), or could produce very few steps if one spent the weekend doing home-based activities. As one participant noted, *"The weekdays are very structured. The weekends are a different story. There's no structure. Well there's structure, but it's usually not around exercise."* (P12)

Therefore, instead of a fixed, daily step-count goal, participants often mentioned striving to achieve consistency in their activity levels, where consistency could be defined over a longer interval, such as a week. For such participants, consistency was defined as a weekly goal where low-activity days were supplemented by planned high activity days. Some did try to maintain consistent levels of activity on a daily basis and mentioned needing to monitor themselves in order to do so: *"When I was behind I*

	gender	age	activity level	duration (months)
P01	female	25-34	low active	3
P02	female	45-54	somewhat active	3
P03	male	45-54	somewhat active	8
P04	male	35-44	sedentary	*5
P05	male	25-34	somewhat active	7
P06	female	35-44	somewhat active	7
P07	male	25-34	somewhat active	11
P08	male	45-54	highly active	23
P09	male	35-44	low active	11
P10	male	35-44	low active	6
P11	male	35-44	somewhat active	9
P12	male	35-44	active	7
P13	female	45-54	active	12
P14	male	18-24	somewhat active	*3
P15	male	35-44	active	14
P16	male	45-54	low active	6
P17	female	>55	active	22
P18	male	45-54	low active	*12
P19	male	>55	somewhat active	11
P20	male	>55	active	20
P21	male	>55	low active	10
P22	female	35-44	somewhat active	6
P23	male	35-44	somewhat active	5
P24	male	35-44	somewhat active	5
P25	female	35-44	somewhat active	10
P26	male	>55	somewhat active	24

Table 1 Study participants showing gender, age, activity level and duration of use.
(* Users who stopped using Fitbit before participating in the study.)



(a) Different Reference groups
(b) Customizing goals with personal status
(c) Recommendations based on routines
(d) Planning and reflections

Figure 1 Four selected screens showing scenarios incorporating population data to address problems with the Fitbit application revealed in interview 1. (Full storyboard is available at: http://doi.org/10.5281/zenodo.47428)

usually took a walk in the middle of my work when I had been sitting for too long and I walked in the aisles here." (P11)

4.2 Perceived Benefits of the Fitbit App

Most study participants reported that the application was adequate for helping them develop an awareness of their activity level and for tracking their performance over time. Part of their developing awareness included becoming familiar with the variability in their activity routines across the week. People reported that they were able to quickly learn to quantify their daily routines in terms of steps and could easily predict how many steps various activities or even various type of days (e.g., "in the office" work days) would produce. Unusual events, such as attending a conference and exploring its host city (e.g., P05) or completing a new hike (e.g., P26), continued to produce novel step counts. Beyond simple awareness of their step patterns, participants also found it valuable to explore the resultant historical patterns at various granularities (weekly, monthly). Many of the participants noted that they wanted to monitor their performance to ensure that they were maintaining their activity level across time *"So if I'm missing particular days or certain weeks have started falling off of my habit I try to regain it in subsequent weeks."* (P05).

Study participants who checked their data more frequently than at the once-a-day norm reported additional benefits from doing so. Those who checked the step count or distance display immediately after completing an activity not only gained a more refined awareness of their activities but also experienced some satisfaction from the validation of their accomplishment (e.g., P12). Others who monitored their progress over the course of the day reported being able to make adjustments to their activities to increase step counts at various times. Some appeared to use the awareness they had of their average baseline performance as an *implicit* goal to meet or exceed in subsequent days. Several people talked about "make-up strategies" that they activated when they realized they had been mostly inactive in a day: *"If I saw around 2 pm while at work that I had only 5,000 steps or 7,000 steps, I'd get up and start – if I had to go for a meeting, I'd make sure I take the long way around the office – or walk around a little more to try to get that step."* (P14)

Checking data more frequently enabled some users to engage in social activities that they found motivating for driving further activity. Whether through organized "challenges" or through more informal commenting on the step levels of colleagues or friends, social competition was often described as a motivator for increased activity: *"It has been a very good motivational tool especially as a social aspect. We are teasing each other and cheering each other on. I think that's been probably the most helpful aspect of Fitbit for me."* (P08)

However, others preferred focusing on their own historical patterns for motivating their step achievements. Some explained that while it was fun to interact with friends or family, differences in activity levels precluded meaningful comparisons for motivating their activity. Some preferred to also use historical records of their own step activity as the basis for broader reflection on their behavior. For example, people were able to understand their routines and thus better able to predict the kinds of facilitators and inhibitors to activity they might encounter and plan around them. Some found it useful to reflect on the impact of a particular activity on their overall health: *"I found that if I'm doing very intense exercise I don't have the stamina to keep going for a long time versus if I'm doing something more moderate like walking which I can do much longer."* (P03).

4.3 Dissatisfaction with the Fitbit App

While only three participants stopped using the device, the majority reported checking their data less frequently over time. Many switched from a routine of at least once daily to once or twice a week. Several reported that they reviewed their data only in response to the weekly summary email from the Fitbit application. They felt that once they gained awareness of their activity patterns, there was no other information they could gain from the application. *"I do look at the data once in a while, but I know how much I run a week now, so I'm not clear what additional information I'm getting from the device."* (P05)

Some also described other ways that the device and application did not satisfy their needs, including finding that using the device did make them more active (e.g., P06, P10, P11, P12). Others had hoped to better understand what their step activity meant for their health (e.g., P01, P03, P22). For example, those who did aerobic activity were perplexed that high intensity aerobic activity may be, in the Fitbit scheme, valued less than a leisurely walk.

5. APP REDESIGN FOR DYNAMIC & PLANNING USE

The goal of the application redesign was to explore ways to address some of the shortcomings of the current application that

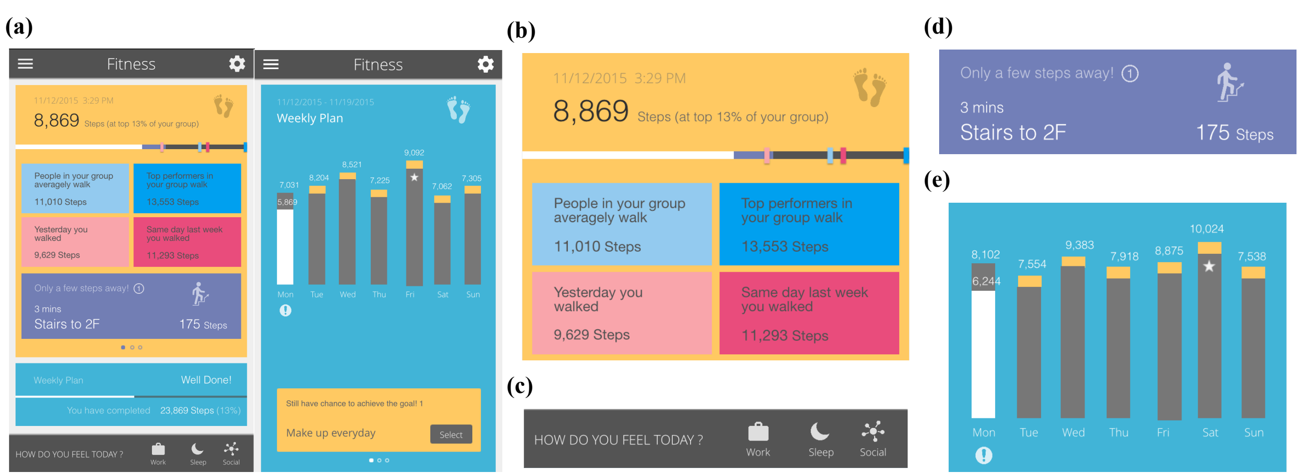


Figure 2 Interactive prototypes to support (a) dynamic use (left) and planning use (right), (b) different reference group (c) customizing goals with personal status, (c) recommendation based on routines, (d) weekly planning and reflections.

we discovered through the first interview. The first interview revealed two different modes of using the Fitbit application:

- Some participants wished to understand their fitness status and progress toward the goal over the course of a day. They also expected to adjust their goals and progress based on their work, life, and social routine.
- Others hoped to plan their physical activities ahead, with most hoping to make weekly plans. These participants realized they have different work or life events each day, and therefore thought that planning ahead might help them better achieve their weekly goal.

We also wanted to address differences in the reference points people mentioned when tracking changes in their performance over time. Recall that some liked to engage in social comparisons, whereas preferred to compare themselves to their own previous performance. Based on these aspirational uses we developed four use cases (Figure 1) and two prototype interfaces (Figure 2(a)) that integrated population data with personal routine. We discuss these next.

5.1 Four Use Cases

5.1.1 Using similarity group and personal history as daily performance reference

Some participants had a specific health or fitness goal, such as training for a marathon (e.g., P02). For them, similarity analysis can place them among a population group who have a similar goal and have similar physical capability, which might be inferred by age, gender, or body mass index (BMI). Knowing how they are doing among similar people (their similarity group) might give them a better measure of their performance and motivate them to achieve more (Figure 1(a)). On the other hand, other participants mentioned goals that they referred to as "*monitoring their status*" (e.g., P03, P22). These participants were usually more interested in comparing themselves with their own previous history.

We illustrated these two different goals by marking two types of comparison in the prototype. For comparison with a *similarity group*, we indicated the average number of steps from the top 10% of performers (dark blue tick mark) and from the whole group (light blue) that we computed from the population data analysis. For comparison with *personal history*, we used the number of steps from the participant's previous day and from the same day in the previous week (e.g., last Wednesday); we used two different shades of pink for the markers. The color of the tick marks is coded to the color of the panels below that provide step count details. (Figure 2(b)).

5.1.2 Customizing daily goals in response to personal demands

Many participants felt that a fixed goal was not appropriate in circumstances where they had unusual demands on their time. For example, one participant felt frustrated about not being able to achieve the default 10K-step goal when she was working toward a tight project deadline (P25). Another mentioned that he always over-achieved the goals when he traveled (P26). By learning the person's routine and anomalies to her routine, the system could adjust a goal to more appropriate levels and thereby encourage users not to give up or to even strive harder (Figure 1(b)).

Therefore, we designed a case to support goal adaptation based on daily routines. We chose three data points – how busy they are at work, how well they slept the night before, and how many social events they have after work – as example parameters for goal adaptation. (Figure 2(c))

5.1.3 Recommending activities based on behavior pattern and routine

Employees described various strategies for planning and making up steps when they were lagging behind their implicit goals. During workdays, they might take a walk around the building after lunch or during a break. They might also choose to go to the gym or walk around their house after work to boost the step-count.

Systems could recommend strategies to supplement step counts to meet goals based on an understanding of an individual's behavior routine. (Figure 1(c)) In our prototype, we presented a sample feature that recommend strategies over the course of a day. For example, during a coffee break, systems could suggest users to walk to the coffee shop thereby adding some steps. (Figure 2 (d))

5.1.4 Making a weekly plan through analysis of similarity group and personal history

Many participants reported having a habit of planning weekly activities ahead of time to anticipate their busy schedules or to ensure that they would maintain a consistent level of long-term performance. They may also plan a hike or other physical activities during the weekend to compensate for sedentary weekdays. However, they reported having difficulties creating a day-to-day plan to maintain the long-term consistency. By analyzing activity patterns from an individual's history and from the similarity group, systems might be able to help users set up daily goals for a week at a time. (Figure 1(d))

We designed the prototype to display the daily goals for the week in the same screen so that users could glance at the plan at once. We also designed it to enable users to reflect on their goal attainment in the same view. When users review their performance and find they cannot meet the goal, they can choose to distribute their missing steps to the rest of the week, their most active days, or weekends. They could also reflect on the reasons why they did not meet the goal – work, sleep, social events, or indicate that they had done other exercise that Fitbit did not recognize, with the idea that systems could adjust the plan in the long run based on the feedback (Figure 2(e)).

6. INTERVIEW 2: REACTIONS TO PROTOTYPES AND USE CASES

6.1 Who should be in the Reference Group?

Employees expressed interest in seeing their real-time performance data along with reference group information. A frequent update of their step count, contrasted with the target counts from a reference group, might enable them to adjust their activity amounts over the course of the day and reach their daily goals, often through relatively small, individual efforts. Participants mentioned using reference group comparisons to increase their motivation and to help them plan their daily or weekly activities. Further probing revealed variation in the nature of the reference group that they felt would be most helpful in motivating their activity.

6.1.1 Similarity group

Some participants mentioned that understanding the performance of people similar to them could help keep them motivated. For these people, activity level is an important indicator for establishing similarity and for enabling fair comparison: "*You want to play against people or play with people that are close to your level, people in your league*" (P11). Besides basic demographic information, such as age, gender, and weight, many employees also mentioned work or life routine as an important indicator to determine a similarity group. For example, they would not want to compare themselves against those who regularly perform physical activity as a part of their job. A few participants also mentioned that having a newborn baby dramatically changed their lifestyle and argued for the importance of incorporating a major change in status in specifying a similarity group (P15, P23). Other participants wanted to use the activity patterns of people who live their aspirational lifestyle as motivators. For example, one employee mentioned hoping to learn about the exercise pattern of the senior vice presidents in the company because "*they must be extremely organized. They're probably very consistent because they have to schedule it*." (P12)

These participants hoped to obtain guidance from similar individuals to help them develop an exercise plan. For example, "*Because your time gets super constrained when you have kids, what are those people able to do?*" (P15) They expected that examples from people with similar work or daily life experiences could help them formulate more realistic personal fitness plans.

6.1.2 Social group

Some participants prefer to compare themselves to people they normally interact with, such as friends, colleagues, or family members. One participant mentioned that he is comfortable competing only with his family because he knows winning or losing will not hurt the other's feelings (P19). Other participants have "*walking buddies*" in their teams or in the company with whom they normally take after-lunch walks or have walking meetings (e.g., P08, P26). They also "friend" each other on the Fitbit app and participate in "challenges" with each other (e.g., P11, P13). These participants also imagined a more in-depth comparison with the social group with whom they have face-to-face interaction. Instead of only looking at their total number of steps on the current Fitbit app, they would like to know how many steps they walk together or separately (P26), whether their families or buddies achieve their goals (P09), and how they plan their physical activities across the day (P08).

6.1.3 Self

Some participants, especially those who have only general fitness goals, like monitoring their activities, are not interested in interacting or competing with other people. These participants, however, were motivated by comparisons with their own past daily performance (e.g., P04). Some of these people habitually skim through the historical data summary view in the Fitbit app, most often in the weekly view. However, many of them stated that they were not able to identify patterns in their historical data. Exploring the prototypes, they thought that simply contrasting current and previous data (a single data point or an average over a period of time) might give them a clearer idea of their performance (e.g., Figure 2(a)). In turn, this could help them to achieve goals of maintaining consistent activity levels (e.g., P11, P23) or enable them to monitor their overall activity pattern, such as a running regime on certain days of the week (e.g., P05).

6.2 Should Systems Adapt Daily and/or Weekly Goals?

Most participants expressed an interest in functionality to help them adapt their goals according to changes in their routines. Besides the varying levels of work, sleep, and social commitments, some participants also suggested health status and other personal obligations as possible inputs for adjusting goals.

However, several noted practical limitations to making adaptations automatically. For example, one employee mentioned that whether he is busy might not be a reliable predictor of his step counts: "*I could be busy meeting or thinking or having a lot of work that makes it so that I can't move*" (P04). Another described a scenario in which she might be busy with her kids, but it could include walking around with her kids and thus result in many steps, or playing together at home and thus result in a small number of steps (P06). In these cases, simply indicating the "busy level" or importing calendar to determine blocked events might not provide enough context for deciding how to scale the goal.

Privacy considerations might also limit the accuracy of automatic adjustments. For example, even though study participants are used to making their calendar information public, indicating blocked events to allow meeting scheduling, most participants reported being reticent to indicate how busy they are off work. Although participants agreed that this information might influence their ability to exercise, nevertheless, they were reluctant to share the information with systems: "*I wouldn't want to record that. I'm kind of sensitive of getting that information out*." (P01). People reported feeling discomfort if the information were to be automatically determined by or shared with systems.

However, some participants preferred maintaining a consistent activity level rather than adapting goals based on daily activity level fluctuations. One participant described the adaptation as "*finding an excuse for my failure*" (P08). Other participants mentioned the over-arching importance of goals such as "*run as much as last summer*" (P05) or "*to be active every day*." (P24)

6.3 How should Systems Support Planning?

Some participants mentioned having developed a habit of checking the device or app at certain times of the day to know how far away they are from their daily goal (e.g., P11, P26). This provides an estimate of whether they could achieve their goal and constrain strategies they might take to reach their goal. Making this salient might help users be aware of their status throughout the day and thus remind them to continue moving. Participants provided various suggestions for extra activities, for example, walking reminders when they had been sitting for a long time (P10), suggestions for new routes when traveling (P07), and an event prompt to encourage them to leave their office for company events (P12).

Participants also reported needing support for planning at a granularity of a week. For example, it is sometimes difficult for participants with long-term goals to track how daily performance contributes to weekly or longer-term goals (P06). Participants with a particular lifestyle or routines that might include children or frequent business trips often need to plan ahead in order to include physical activities in their busy schedule (e.g., P15, P21). Even people who do not have specific goals other than simply being active often do not realize how many days they failed to exercise and therefore misjudge their activity level (e.g., P14).

People preferred to make weekly activity plans in advance and usually to plan their make-up strategies as well. A majority stated wanting to make up for low activity days on the days when they have more time to exercise, such as on what are already their most active days or during the weekend (e.g., P23). Participants who want to maintain a certain level of consistency across all days, on the contrary, reported wanting to make up a missing goal on their least active days to maintain consistency (e.g., P20). Still others, who desired a long-term improvement, preferred to evenly distribute the missing steps across days (e.g., P26).

7. DISCUSSION

7.1 Design Beyond Readiness/Awareness

Most of our participants demonstrated that they were already committed to exercising and as such they were in need of support to motivate them to continue exercising, and moreover, to do so in the face of significant time challenges. The Fitbit product app satisfied their interest in awareness of their activity level but did not support them in ways that would continue to engage them with their data and perhaps enable them to further understand their activity patterns. As a result, they became less and less interested in checking their step count information despite the fact that they kept wearing the device. This is consistent with the argument that most current fitness trackers and apps are designed for people who are in preliminary stages of exercise readiness, for example those who are identified as being in the contemplation or preparation stages according to the Transtheoretical model [17]. We believe that the feedback we received to the prototypes provides some evidence that reference groups – behavioral data from similarity analysis, family or friends, or personal historical data – combined with dynamic support for planning at a daily and weekly level have the potential to continue to engage people with their data stream and help guide them towards their goals.

7.2 Reference Groups & Goals

Goal setting has been widely studied and adopted in many personal informatics tools to keep users motivated towards their goal [7][25]. However, translating a long-term goal into sustained action is challenging, as was reported by our participants. Involving users in ongoing comparisons with some target might be one way to drive more frequent interactions with one's data. Participants' expressed preferences for one or another of the three reference groups in our prototype reveal that there is variation in the comparisons that people find of interest for evaluating their performance. Providing alternatives in an app enables individuals to utilize the comparisons they find most motivating.

Comparing oneself to a similarity group helped some people to feel confidence that they would be able to be as active as the targets indicated on the status bar because it was people like themselves who achieved the target step counts. However, our data suggests that the population datasets we utilized as the basis for computing the similarity groups may not include all of the features that are relevant for correctly finding *someone like me.* Simple demographic information (e.g., age, gender, or weight) or behavioral matrix (e.g., activity level) might not be enough to create a meaningful match. One dimension that should be incorporated is the personal goal. For example, people wanting to lose weight might have different walking routines than people using walking for rehabilitation from knee surgery. People also have different constraints in their life, such as family (e.g., with or without kids) or work style (e.g., sedentary vs. walking around in the laboratory), and such constraints need to be incorporated for the recommendations to be relevant to an individual.

Thus, population datasets with a richer set of features would be needed to provide more engaging and relevant similarity comparisons. While some of the feature information might be inferable by analyzing individual behavioral pattern, it needs to be complemented by the presence of those features in the population datasets. Going forward, it is worth considering whether organizational wellness plans could design incentive plans around these additional segmentation factors and thus get people to share more nuanced information for generating population datasets capable of supporting more personalized recommendations.

7.3 Leveraging Social & Self Comparisons

However, our study found that not everyone is motivated by comparing their performance with anonymous others. In our study we also found some people have preferences toward comparison with known-social group and themselves.

Similar to studies working with children or teenagers, sharing physical activity data with family or close friends might support motivation or affection needs [24][29]. In our study, we found that family and friends often influence people's decision of choosing a fitness tracker and apps. They also have similar lifestyles or understand each other's goal and constraints well. Participants also use these data as an indicator of other's health status and to build relationships. Systems to support this type of social sharing should provide the sharing information at various granularities. For example, as Epstein et al. found in [11], a detailed, granular activity might enable family members or close friends to identify behavior patterns during a specific timeframe (e.g., using commute time for more walking). On the other hand, a more general view might be sufficient to determine whether the other is active.

We found another group of participants who are interested in using themselves as the comparison. They closely monitor and "*benchmark*" themselves using their previous performance. Although many current systems support a history view, most people only *glance* at their apps on an average of 5 seconds and do not *interact* with these apps often [17]. Similar to the *Catchup* concept in [17] and the intrapersonal retrospective recommendation mechanism in [12], our design attempted to address this phenomenon by bringing the history information to the front page to allow users to directly compare their real-time performance against their past data. Future deployment studies should investigate whether and how people glance at this view. Future systems should also examine other constraints on self-comparisons. For example, systems could allow users to compare themselves during a particular time frame, such as when they are healthy and are able to do a full range of activities or when they are under tight schedules and might only be able to squeeze in workout time during a lunch break.

8. CONCLUSION

In this study, we explored user perceptions of integrating self- and population- tracked data with individual routine to motivate users to improve their fitness behavior and to address their challenges. Our findings suggest that participants – committed to and ready for fitness and wellness improvement – perceived benefits in comparing their performance with different personalized reference group and in getting planning guidance through these groups.

Our study represents only a first step in the exploration of these ideas and will require replication with other samples, including with individuals who did not obtain their device in the context of a workplace wellness program. It is likely that we were not able to identify issues or expectations that would emerge through a long-term field deployment. In addition, data may hold an inherently greater interest for our participants who are by and large a highly technical group than for members of the general public. Our participants are also potentially more motivated to maintain physical activity and the associated tracking behavior. However, we recruited participants with various activity level, goals, and length of use to introduce as much variability as possible into our sample. As such we believe our findings provide some initial understandings in engaging users with wearable fitness devices that may form the basis for future research and system design.

9. ACKNOWLEDGMENTS
We thank the participants in this study for sharing their experience and insights with us. We also thank our colleagues for helpful discussions regarding our study, especially Henry Chang and Ching-Hua Chen. We thank Sean Munson, Daniel Epstein, and reviewers for thoughtful feedback on earlier versions of the paper.

10. REFERENCES
[1] Abowd GD, Mynatt ED. 2000. Charting past, present, and future research in ubiquitous computing. *Acm T Comput-Hum Int (TOCHI)*; *7*(1): 29-58.

[2] Agapie E, Colusso L, Munson SA, Hsieh G. 2016. PlanSourcing: Generating Behavior Change Plans with Friends and Crowds. *CSCW 2016*. 119-133.

[3] Bandura A. 1991. Social cognitive theory of self-regulation. *Organ Behav Hum Dec*; 50(2): 248-287.

[4] Brown DK, Barton JL, Pretty J, Gladwell VF. 2012. Walks4work: Rationale and study design to investigate walking at lunchtime in the workplace setting. *BMC public health*; *12*(1): 550.

[5] Clawson J, Pater JA, Miller AD, Mynatt ED, Mamykina L 2015. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. *UbiComp 2015*. 647-658.

[6] Consolvo S, Everitt K, Smith I, Landay JA. 2006. Design requirements for technologies that encourage physical activity. *CHI 2006*. 457-466.

[7] Consolvo S, Klasnja P, McDonald DW, Avrahami D, Froehlich JE, LeGrand L, Libby R, Mosher K, Landay JA. 2008. Flowers or a Robot Army? Encouraging Awareness & Activity with Personal, Mobile Displays. *UbiComp 2008*. 54-63.

[8] Eakin EG, Lawler SP, Vandelanotte C, Owen N. 2007.Telephone Interventions for Physical Activity & Dietary Behavior Change: A systematic review, Am J Prev Med; 32(5): 419-434. PMID: 17478269.

[9] Endeavour Partners, Inside Wearables, http://endeavourpartners.net/assets/Endeavour-Partners-Inside-Wearables-Part-2-July-2014.pdf

[10] Epstein DA, Cordeiro F, Bales E, Fogarty J, Munson SA. 2014. Taming Data Complexity in Lifelogs: Exploring Visual Cuts of Personal Informatics Data. *DIS 2014*. 667-676

[11] Epstein DA, Borning A, Fogarty J. 2013. Fine-grained sharing of sensed physical activity: a value sensitive approach. *UbiComp 2013*. 489-498.

[12] Farrell RG, Danis CM, Ramakrishnan S, Kellogg WA. 2012. Intrapersonal retrospective recommendation: lifestyle change recommendations using stable patterns of personal behavior. *LIFESTYLE 2012*. 24.

[13] Forbes, http://fortune.com/2014/04/15/in-corporate-wellness-programs-wearables-take-a-step-forward/

[14] Foster D, Linehan C, Kirman B, Lawson S, James G. 2010. Motivating physical activity at work: using persuasive social media for competitive step counting. *MindTrek 2014: Envisioning Future Media Environments*. 111-116.

[15] Gaver B, Dunne T, Pacenti E. 1999. Design: cultural probes. *interactions*, *6*(1), 21-29.

[16] Gorm N, Shklovski I. Steps, Choices and Moral Accounting: Observations from a Step-Counting Campaign in the Workplace. *CSCW 2016*. 148-159.

[17] Gouveia R, Karapanos E, Hassenzahl M. 2015. How do we engage with activity trackers?: a longitudinal study of habito. *UbiComp 2015*. 1305-1316.

[18] Halko S, Kientz JA. 2010. Personality and persuasive technology: An exploratory study on health-promoting mobile applications. *In Persuasive technology*. 150-161.

[19] Hatano Y. Use of the pedometer for promoting daily walking exercise. 1993. *ICHPER*; 29: 4-8

[20] Lazar A, Koehler C, Tanenbaum J, Nguyen DH. 2015. Why we use and abandon smart devices. *UbiComp 2015*. 635-646.

[21] Lin JJ, Mamykina L, Lindtner S, Delajoux G, Strub HB. 2006. Fish'n'Steps: Encouraging physical activity with an interactive computer game. *UbiComp 2006*. 261-278.

[22] Locke EA, Latham GP. Goal Setting Theory. In Smith KG, Hitt MA (eds.). *Great minds in management: the process of theory development*. 2005.

[23] Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective Techniques in Healthy Eating and Physical Activity Interventions: A Meta-Regression. *Health Psychol* 2009; 28(6). PMID: 19916637.

[24] Miller AD, Mynatt ED. StepStream: A Schoolbased Pervasive Social Fitness System for Everyday Adolescent Health. *CHI 2014*, 2823–2832

[25] Munson SA, Consolvo S. 2012. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. *PervasiveHealth 2012*. 25-32.

[26] Nothwehr F, Yang J. Goal setting frequency and the use of behavioral strategies related to diet and physical activity, *Health Education Research* 2006; 22(4). PMID: 17032703.

[27] Rabbi M, Aung MH, Zhang M, Choudhury T. 2015. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. *UbiComp 2015*. 707-718.

[28] Shih PC, Han K, Poole ES, Rosson MB, Carroll JM. 2015. Use and adoption challenges of wearable activity trackers. *iConference 2015*.

[29] Toscos T, Faber A, Connelly K, Upoma AM. Encouraging Physical Activity in Teens: Can Technology Help Reduce Barriers to Physical Activity in Adolescent Girls? *PervasiveHealth 2008*, 218–221

[30] Tudor-Locke C, Bassett DJ. 2004. How many steps/day are enough? Preliminary pedometer indices for public health. *Sports Med*;34(1):1-8.

[31] Turchaninova A, Khatri A, Uyanik I, Pavlidis I. 2015. Role model in human physical activity. *WH 2015*. (21): 6.

[32] Vyas D, Fitz-Walter Z, Mealy E, Soro A, Zhang J, Brereton, M. 2015. Exploring Physical Activities in an Employer-Sponsored Health Program. *CHI 2015 EA*. 1421-1426.

[33] Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, Bradley R, Heywood J. 2010. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res; 12(2). PMID: 20542858