Exploring the Design Space of Glanceable Feedback for Physical Activity Trackers

Rúben Gouveia[1], Fábio Pereira[1], Evangelos Karapanos[1,2], Sean A. Munson[3], Marc Hassenzahl[4]
[1]Madeira-ITI, [2]Cyprus University of Technology, [3]University of Washington, [4]Folkwang University of Arts
{rubahfgouveia, fpereirauma, e.karapanos}@gmail.com, smunson@uw.edu, marc.hassenzahl@folkwang-uni.de

ABSTRACT
Recent research reveals over 70% of the usage of physical activity trackers to be driven by *glances* – brief, 5-second sessions where individuals check ongoing activity levels with no further interaction. This raises a question as to how to best design glanceable behavioral feedback. We first set out to explore the design space of glanceable feedback in physical activity trackers, which resulted in 21 unique concepts and 6 design qualities: being *abstract*, *integrating with existing activities*, *supporting comparisons to targets and norms*, being *actionable*, having the capacity to lead to *checking habits* and to act as a *proxy to further engagement*. Second, we prototyped four of the concepts and deployed them in the wild to better understand how different types of glanceable behavioral feedback affect user engagement and physical activity. We found significant differences among the prototypes, all in all, highlighting the surprisingly strong effect glanceable feedback has on individuals' behaviors.

Author Keywords
Physical activity tracking; glanceable displays; behavioral feedback interfaces; personal informatics.

ACM Classification Keywords
H.5.2. User Interfaces: Evaluation/Methodology.

INTRODUCTION
People increasingly adopt technologies to track their everyday behavior [39]. Personal informatics tools rest on the assumption that people develop a better understanding of their habits through self-monitoring, which in turn promotes self-knowledge, reflection and ultimately change upon undesirable habits [28]. Examples are counting steps to increase levels of physical activity [11] or measuring water spent in the shower to reduce waste [17].

Since knowledge of existing behavioral patterns seems at the heart of self-tracking, according tools focus on the rich visualization and the deep exploration of personal data [14, 7]. This implies a certain way of using such tools. First people collect data, then explore and review summaries of longer periods in retrospect (i.e., days, weeks) to identify patterns and plan alternative future courses of action [16]. For example, some people use tools provided by their phone companies to analyze their monthly costs to pick the "best" tariff or to optimize own future usage behavior.

UbiComp '16, September 12-16, 2016, Heidelberg, Germany

DOI: http://dx.doi.org/10.1145/2971648.2971754

Figure 1 – *TickTock* (left) and *Normly* (right), two of the concepts developed as watchfaces. *TickTock* portrays periods in which one was physically active over the past hour. *Normly* compares one's goal completion to that of others having a similar walking goal.

In addition to this rather analytical approach, people use self-tracking to monitor and regulate immediate behavior [8]. For example, somebody may have told Ruben that paced walking (e.g., 6 km/h) is a valuable opportunity to get a little more exercise throughout the day. Unfortunately, Ruben is a slow walker. To get into the habit, he measures his walking pace *while* walking home from work to keep up the speed. This scenario requires frequent feedback while actually being engaged in the activity of walking [8].

In the case of physical activity trackers, brief and frequent monitoring may in fact be the dominant mode of interaction. In a prior study [19], 70% of all interactions with an activity tracker were *glances* – brief, 5-second sessions where users checked their current activity levels with no further exploration or interaction.

While researchers have noted the value of glanceable feedback as a complement to the deeper and reflective analysis [11], research focusing specifically on glanceable activity feedback displays has been scarcer than research on deep, reflective feedback displays. In particular, literature lacks detailed inquiries into the design space of glanceable behavioral feedback, guidelines for what makes feedback glanceable, and an understanding of the effects of different glanceable feedback displays.

To provide a better understanding of glanceable behavioral feedback, we first explored the design space of glanceable feedback in the context of physical activity trackers. We created a total of 21 concepts and a total of 6 design qualities through an iterative ideation and reflection process. We argue that glanceable feedback for behavior change should *be abstract*, *integrate with existing activities*, *support comparisons to targets and norms*, *be actionable*, and have the capacity to lead to *checking habits* and act as a *proxy to further engagement*. Second, we prototyped four of the concepts and deployed them in the wild to better understand how different types of glanceable feedback affect users' engagement and physical activity.

RELATED WORK

So far, the importance of glanceable feedback in behavior change tools has been noted by a number of researchers. Ham and Midden [20] emphasized the persuasiveness of glanceable feedback since it requires minimal attention to be perceived and processed. Consolvo et al. [11] found individuals to increase long-term commitment to physical exercise when presented with glanceable feedback. Mullet and Sano [32] further argue that the frequent monitoring of behavior can lead to early correction of slips and relapse.

But what makes feedback especially glanceable? Consolvo et al. [12] define "glanceability" in terms of how *quickly* and *easily* feedback is able to convey information after one pays attention. To accomplish high glanceability, feedback should be "reduced to the essence through a process of simplification and abstraction" [32]. Feedback should provide "just enough" to be perceived and processed [30]. A further quality of glanceable feedback is its ability to be perceived at the periphery of one's attention [5]. Feedback should be "*working in the background while we attend to foreground activities … [enabling people] to get the essence of the information with a quick visual glance"* [29].

Empirical studies have provided support for the effectiveness of glanceable feedback. Jafarinaimi et al. [23] developed *Breakaway*, a small human sculpture aimed at encouraging regular breaks from work. *Breakaway* mimicked its user's posture throughout the day. It was placed on the office desk, offering persistent, yet unobtrusive and quickly consumable feedback. A case study with a single participant showed the likelihood of taking a break from work to increase when the sculpture slouched. In addition, the participant commented on how easily *Breakaway* could be ignored, when busy. In this case, healthy sitting is a secondary task to be monitored and regulated throughout the day while actually completing primary, work-related tasks.

Another example is Consolvo et al.'s [11] *UbiFit Garden*, a mobile application designed to support overall physical activity by tracking users' physical activity, and presenting feedback on the background screen of mobile phones. In a comparative study, participants using *UbiFit Garden* had higher activity levels than participants without persistent feedback on behaviors. The always-available information on activity levels acted as a reminder to stay engaged and committed to the goal of increasing physical activity. Fortmann et al. [15] created *WaterJewel*, a wearable wrist bracelet to motivate users to maintain adequate hydration levels throughout the day. *WaterJewel* has eight LEDs, which light up when users progress towards their daily goal of water intake. Participants using *WaterJewel* were more likely to accomplish their goals for water intake than participants who received the information on their phones.

All in all, research suggests that presenting abstract, easily consumable information, at locations where the individual is likely to gaze frequently positively affects self-regulation of particular behaviors.

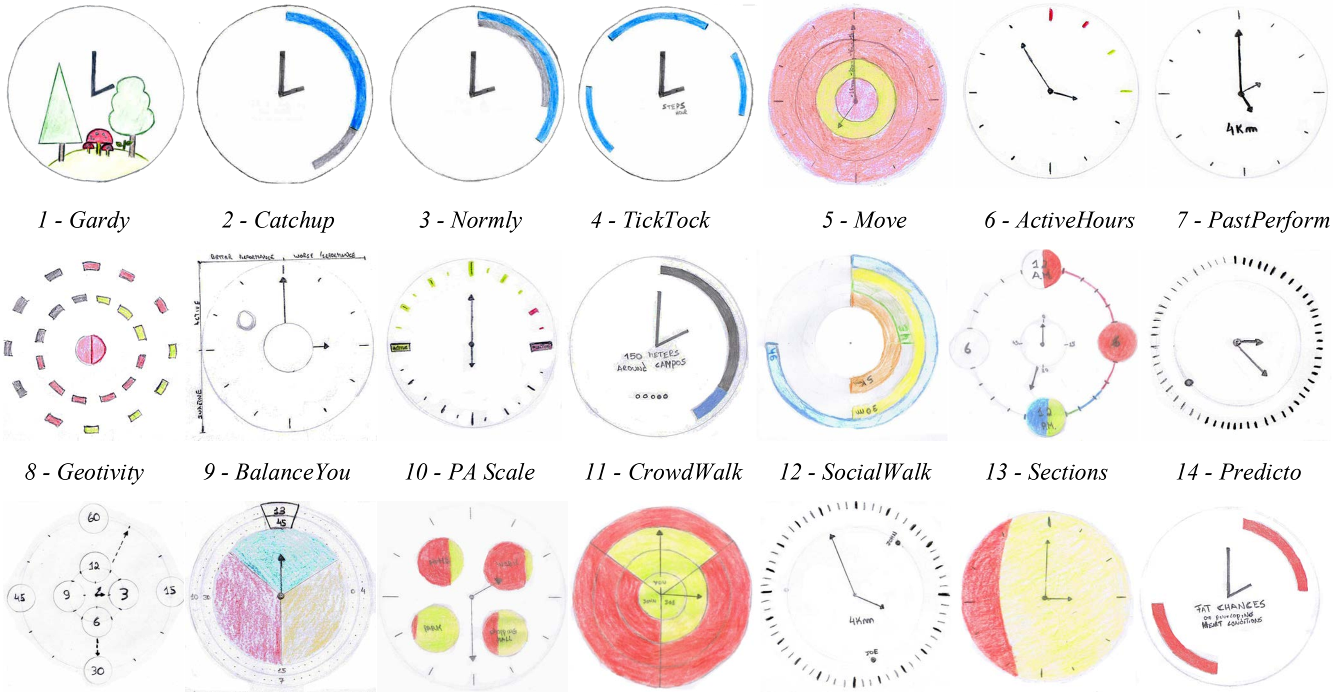
Yet, while the strengths of glanceable feedback have been recognized, previous literature has highlighted the need to explore the efficacy of different forms of glanceable feedback. In Consolvo et al.'s study [11], for example, men were more skeptical of the garden display than women, raising questions about the effectiveness of different stories told through feedback. Are some forms of glanceable feedback more effective compared to others? [12].

In the remainder of the paper, we present our design space exploration, which led to 21 concepts and 6 design qualities important for glanceable feedback, followed by an empirical exploration of the four prototyped concepts.

DESIGN SPACE EXPLORATION: CONCEPTS AND QUALITIES

Our first goal was to explore the design space of glanceable feedback for activity trackers. Since wrist-worn devices (e.g., smartwatches, wristbands) are the most glanced mobile feedback displays available [31, 35], we focused our exploration on smartwatch interfaces. As a technology, smartwatches allow for the widest variety of ways to present feedback in glanceable ways.

The design space exploration was performed by the first three authors. Starting with a design brief of 'glanceable watchfaces reflecting physical activity', we followed an iterative process of synthesis and analysis, whereby new ideas were compared to each other to reveal the underlying differences and qualities of glanceability, followed by new rounds of ideation aimed at further deepening the understanding of each emerging quality. Existing research prototypes (e.g., *UbiFit*) or commercial products (e.g., *Fitbit*) were often used as reference points during the analysis, while theoretical frameworks and constructs (e.g., Cialdini's [9] scarcity principle) often helped us elaborate on the design qualities. This process led to a total of twenty-



1 - Gardy, 2 - Catchup, 3 - Normly, 4 - TickTock, 5 - Move, 6 - ActiveHours, 7 - PastPerform

8 - Geotivity, 9 - BalanceYou, 10 - PA Scale, 11 - CrowdWalk, 12 - SocialWalk, 13 - Sections, 14 - Predicto

15 - JustaWatch, 16 - Rotations, 17 - Locals, 18 - Scarcition, 19 - DistantYou, 20 - DayBalance, 21 - Meanfull

Figure 2. The 21 concepts of glanceable physical activity feedback

one concepts (see Figure 2) and six design qualities. We briefly summarize each of the six design qualities and illustrate them either with one of our 21 concepts or prototypes already existing in the literature (see Table 1).

Abstract

Abstraction of data is perhaps the most prevalent quality of glanceable displays [30,38]. A number of existing prototypes and products apply this principle. Abstracting data, as opposed to displaying raw data, allows users to process and perceive information with minimal consciousness [20], enabling quick awareness and reflection on one's behaviors [11].

To support abstraction, all 21 concepts convey step count through abstract forms, such as circles (e.g., Fig.2.12), or stylized representations (e.g., Fig.2.1). *Gardy* (Fig.2.1), for instance, uses the metaphor of a blossoming garden to highlight one's progress towards goal completion – a simplified variant of UbiFit Garden's abstraction of user's activity levels [11]. Similarly, *Geotivity* and *SocialWalk* (Fig. 2.8 and Fig. 2.12), use shapes to represent different facets of one's physical activity – *Geotivity* displays the moments in which one was active and sedentary (green and red rectangles) over the course of a day, while *SocialWalk* displays different aspects of one's physical activity, such as the total distance walked or time sedentary, through circles.

Integrates with existing activities

Another principle that often came out in our analysis of the emerging concepts was that of *integration with existing activities*. Embedding feedback into frequently occurring activities makes the feedback more likely to be glanced. In fact, glanceable displays have been commonly placed in frequently accessed locations - such the background of one's mobile phone [11] or the periphery of one's vision [5]. Prior work has found that users check their smartwatch 60-80 [35] and 95 [31] times in a day, with more than half of the usage being fueled by checking the time, or triggered by an incoming notification. Following upon this, we decided to integrate all 21 concepts with the practice of checking the time; feedback was placed on the periphery or the background of the primary screen of the smartwatch, whose main function was to tell the time.

Support Comparisons to Targets and Norms

Activity trackers commonly provide descriptive feedback – they tell us how much we have walked but not whether this is enough [33]. Feedback that presents *progress in comparison to a target* can be easier for the user to process, helping the user evaluate their behavior relative to a certain goal rather than presenting raw data requiring further inferences. Consider, for instance, *Fitbit Flex's* glanceable feedback. The wristband features five LEDs that illuminate for each 20% of a daily walking goal achieved. However, even this seemingly simple display requires some quite difficult projections, if one wants to use it for immediate self-regulation. Since for an office worker physical activity is not a constant background task, users need to estimate how likely it is to meet the daily goal based on the distance walked so far and opportunities to walk in the future.

Table 1. We identified 6 underlying design qualities in our 21 concepts

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Being abstract		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Integrating with activities		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Comparison to targets and norms			■	■		■	■	■	■	■	■		■	■				■	■	■	■	
Being Actionable						■						■										
Leading to checking habits	Novelty	■										■	■					■				
	Scarcity				■	■	■												■			
Proxy to further engagement															■							■

Normative comparisons can reduce this burden of projection. For instance, *PastPerform* (see Fig 2.7) and C*atchup* (see Fig 2.2) compare the distance walked so far to the distance walked at the same time yesterday, or at a day when one's goal was barely met, respectively. Following the same logic, *Normly* (see Fig 2.3) employs a large database of other people's walking on different days and compares at every glance, the distance one has walked so far to that of other users, who usually are equally active. *DistantYou* (see Fig 2.19) follows the same approach as *Normly*, but highlights the specific time in which other people met their goal. *ActiveHours* (Fig 2.6) and *Sections* (Fig 2.13) further attempt to project norms by highlighting how balanced a user has been (i.e. active vs inactive) over the course of an hour, while *PA Scale* (Fig 2.10), *BalanceYou* (Fig 2.9) and *DayBalance* (Fig 2.20) highlight how balanced a user has been over the course of a full day.

All the previously described interfaces provide normative, directly interpretable feedback that helps users maintain an awareness of their performance at a glance.

Actionable

Another quality that often surfaced in our exploration was that of *actionable* feedback. Effective glanceable feedback interfaces should not only inform but also instigate short, goal related actions [25]. An example is *CrowdWalk* (Fig 2.11), which presents in a brief text walking challenges one may perform from the current location, and visualizes the contribution these would make towards meeting the daily goal. For instance, as users enter a building, *CrowdWalk* may suggest taking the stairs; when entering a supermarket, users may be challenged to leave their shopping cart behind while walking back and forth to gather items. As another example, *Move* (Fig 2.5) suggest moments, every 15 minutes, where a user should try to fit in exercise over the course of a day. *Move* takes into account users' calendar, and levels of past activity to make such recommendations.

Leads to checking habits

While glancing is the dominant form of interaction with smartwatches [35] and physical activity trackers [19], prior work has shown the frequency of glances as well as the overall engagement with feedback to decrease over time [19]. This drop in engagement may have detrimental effects on behavior change as individuals quickly relapse once self-monitoring stops [36], while the frequent monitoring of one's behaviors can help prevent relapse. We thus argue that glanceable feedback should be able to sustain the frequency of glancing over the long run, or in other words to *instigate checking habits* [41]. Prior work has suggested this to be feasible. For instance, Oulasvirta et al. [34] linked the information gratification users derive from social media updates and incoming emails on their smartphones to the creation of "checking habits: brief, repetitive inspections of dynamic content quickly accessible on the device".

Our ideation process resulted to two approaches for the creation of checking habits: *novelty*, and *scarcity*.

Novelty asks: what if the feedback provided by an activity tracker constantly presents new information? This is a well-employed strategy in the computer gaming and airline industries, which regularly update content to sustain interest in games or safety instructions. According to Oulasvirta et al. [34], the gratification people derive from encountering novel content as they check their smartwatch would reinforce the habit of checking for new information. Gouveia et al. [19] employed this strategy in the design of the *Habito* mobile app, which, among other features, presented users with textual messages providing feedback about their physical activity. They found that when users read a novel message, they would take less time to come back to the app than when encountering a message they had read before. In the case of glanceable displays, feedback should be short and quickly apprehensible. For instance, *Locals* (Fig. 2.17) portrays random places where a user has walked over the course of the day, indicating his activity (and inactivity) levels within. *CrowdWalk* (Fig 2.11) further leverages on novelty by constantly updating the walking activities suggested to the user. *SocialWalk* (Fig 2.12) compares a user's progress towards goal completion to the progress of random friends. *Locals, CrowdWalk and SocialWalk* leverage on the idea of novelty by updating the places, activities and friends, respectively, multiple times per day. *Gardy* (Fig 2.1) further supports novelty by introducing new elements into users' garden as they progress towards their walking goal.

Scarcity suggests that checking habits may be created if feedback is turned into a scarce resource [13]. Scarcity is a powerful persuasion strategy – individuals are, for instance, more likely to subscribe to a workshop if they know seats are limited [9]. Existing media already apply this principle. For instance, individuals often endure TV commercials to assure they do not miss parts of an interrupted show. Likewise, social media users, such as those on Facebook, frequently reengage to ensure that they do not miss major content among many updates. Overall, people often build their revisit patterns around the update patterns of content to be viewed [4]. Building upon this principle, behavioral feedback could be displayed for a limited amount of time, thus reinforcing re-engagement habits and the frequent monitoring of behaviors. As an example, *TickTock* and *Scarition* (Fig 2.4 and 2.18) portray moments in which a user was active over the past hour and, respectively, the same information but in comparison to his friends.

Acts as a proxy to further engagement

Prior work has found that individuals quickly lose interest in deep data exploration [24]. We argue that glanceable feedback can be designed with the goal of creating "aha" moments, thus acting as cues for further engagement with the feedback. One strategy could be to present information that raises questions rather than provides answers. For instance, *Meanfull* (Fig 2.21) highlights patterns in user data through textual messages (e.g., "Lazy Tuesdays..."), while offering users the opportunity to further explore the underlying data. Another strategy could be to present insights that surprise the user. For instance, *Predicto* (Fig 2.14) analyzes parameters such as past night's sleep quality, the weather over the upcoming day and existing patterns in physical activity to predict the activity levels of the upcoming day. When predictions challenge a user's expectations, the user may become interested to explore the grounds for this surprising prediction.

FIELD DEPLOYMENT OF 4 GLANCEABLE INTERFACES

Next, we wanted to evaluate some of the assumptions that were generated during the ideation phase, in the real world. We selected and prototyped four concepts and deployed them over 28 days with twelve participants. The goal of the study was to compare concepts in terms of their adoption, how participants engaged with them, and what impact they had on their physical activity. We did not design this study to evaluate each concept's efficacy towards behavior change, given the limited sample and short, seven-day exposure participants had to each of the interfaces. Rather, we wanted to inquire into participants' experiences with the four interfaces that go beyond their initial reactions.

Interfaces

We selected four of the twenty-one concepts based on two criteria: *diversity* and *feasibility*. First, we excluded certain concepts, as they were infeasible to prototype to a mature stage within our available resources. Next, we selected two concepts (goal completion and stylized representation) due to their similarity of existing work (*Fitbit Flex's* wristband LED feedback and *UbiFit Garden*, respectively). Finally, we selected two additional, diverse concepts that we deemed represented interesting design claims. We do not argue that these concepts represent the entire design space of glanceable interfaces. We also do not assume the interfaces as a direct representation the theories that motivated them. Their performance during the field study depended on their implementation as much as the design claims they encapsulate.

Figure 3. *Gardy* (left) and *Goal Completion* (right), two of the concepts developed as watchfaces

All interfaces were developed as watch faces for Android Wear. Each comprised the primary screen of the smartwatch. Their only interactive feature was to allow users to set a daily goal for physical activity. We developed, debugged, and field-tested all interfaces on the LG G Watch R to control for variations in interfaces across hardware or other confounders related to hardware variation [1].

TickTock

TickTock (see Fig. 1) portrays, in the periphery of the smartwatch, the periods in which one was physically active over the past hour. We expected *TickTock* to present two main advantages over the other interfaces. First, through turning the feedback into a scarce resource [9] – by constraining it to only the past hour – we expected to build "checking habits", i.e., frequent monitoring of the smartwatch to make sure that no feedback goes unnoticed. This increased frequency of self-monitoring may, in turn, lead to increases in individuals' physical activity. Secondly, we expected that presenting physical activity of only the past hour would inherently lead participants to strive for keeping a balance of physical activity throughout their days. For instance, if they notice that they have been inactive for the past hour, they may try to have a short walk. As a result, contrary to the remaining three interfaces which aim at assisting individuals in achieving a daily goal, *TickTock* may be pushing individuals to avoid prolong periods of sedentarism, which has been found to be a health risk factor independently of the amount of physical activity one performs over the course of a day [37].

Normly

Normly (see Fig. 1) compares at each glance one's daily progress to that of others having the same goal. To establish normative data, we leveraged a database of the daily walking progress of 25 individuals, on a total of approximately 20000 days. We split the database in 10 groups, reflecting the distance walked at the end of the day (i.e., 7km, 8km etc). We then split the data in 1-min intervals, averaging the values within each group. As a result, if a user defines a goal of 8 km/day, *Normly* will compare, at a resolution of 1-min, his daily progress to the average progress of people who walked 8 km by the end of the day. We expected this normative feedback would lead to more frequent action and increases in overall physical activity at the end of the day, for instance in comparison to *Goal Completion*, which simply presents but does not evaluate one's daily progress.

Gardy

Gardy (see Fig. 3) abstracts physical activity levels through a garden, blossoming as individuals' progress towards their daily walking goal. At the start of each day, the garden is bare, with elements such as leaves, mushrooms and trees appearing as they reach their goal. Such abstract, stylized representations have been previously found to sustain users' engagement, through fostering curiosity on users as they anticipate the unfolding of the story, while individuals tend to appreciate the attractiveness and variety of metrics conveyed in such displays [11]. Yet, little is known as to how individuals engage with such representations and the impact they have on users' behaviors.

Goal Completion

Goal Completion (see Fig. 3) presents one's progress towards their daily goal. Participants were presented with a preset goal of 10K steps [44] and were allowed to modify it. Ample evidence exists on the efficacy of goal setting [33] - individuals that set specific goals (e.g., walk 10K steps per day) to be more likely to enhance self-regulation and activate self-evaluations than those which set abstract goals as "do my best" or "try hard" [27]. We decided to set a challenging default goal that reflects medical practitioners' recommendations (i.e., 10K steps) as previous studies on activity tracking have found individuals to have limited understanding of their daily physical activity and to go with the preset goal, even when this is unrealistically low [19], while setting a challenging goal is strongly linked to greater performance [27]. *Goal setting* is no different to commercial prototypes (e.g., Fitbit's feedback on band). We included *Goal Setting* as a baseline, against which we could compare the remaining glanceable interfaces.

Participants

We recruited participants through the reddit community, via the *lggwatchr* subreddit. To qualify, participants had to own an *LG G Watch R* and be willing to commit to use the four interfaces for a total of 28 days. A total of 12 participants successfully completed the study (median age = 25, all male). Seven participants were located in the U.S., two in Canada (25%), and one in Italy and Sweden respectively. They all had prior experience with physical activity tracking. Participants were rewarded with a 40€ voucher upon successful completion.

Readiness to change

We did not limit our sample to participants of certain 'readiness' to change as we wanted to have a diverse sample. However, we measured the stage of behavior change individuals were in using a five-item questionnaire [26]. Our population was biased towards physically active people: no participant was in the precontemplation stage, three in the contemplation stage, two in preparation, two in action and five in maintenance. Prior work has shown activity trackers to work best for people in the intermediary stages of behavior change (contemplation and preparation); in other means, individuals that have the will but not yet the means to change their behaviors [19]. This has to be taken into consideration when interpreting our results. We chose not to use participants' readiness as a variable in our analysis process due to our limited sample size.

Procedure

We debriefed participants and assisted them in installing our application. They used each interface for seven days, followed by a Skype interview, which introduced the upcoming interface and inquired into their usage and experience with the past one. The order of interfaces was counterbalanced across participants. Each interview lasted up to 15 minutes. All interviews were audio-recorded and transcribed by two independent researchers.

Participants were asked to keep the interfaces for the full duration of the study. They were informed that their physical activity and smartwatch usage would be tracked. During our final interview, we asked participants to rank the interfaces in terms of general preference and motivation to exercise, from most to least preferred, and we allowed them to continue using all interfaces after the study elapsed.

We logged participants' physical activity and smartwatch use in order to compare our concepts in terms of adoption, engagement and impact on physical activity. To track participants' physical activity, we made use of Android's step counter, tracking the start and end time of walking activities as well as the number of steps taken while walking. Regarding smartwatch usage, we tracked the time and duration of individual usage sessions, as well as interactions within a session, such as swiping to settings or launching additional applications. A usage session was defined by the time the smartwatch screen was turned on (i.e. interactive mode), until the screen was turned off or timed out (i.e. ambient mode). We also tracked incoming notifications, in an attempt to distinguish smartwatch use motivated by checking notifications versus our interfaces.

Findings

We first summarize overall participant engagement with all interfaces and their physical activity over the course of the 28 days. Next we delve into engagement, experience and impact on physical activity of each of the four interfaces.

Overall engagement and physical activity

All in all, participants checked their smartwatch on average 107 times per day (SD=80), which is slightly higher than in previous studies [31]. Over 80% of all usage sessions were *glances*: sessions in which a participant briefly checks his smartwatch and lets the screen timeout, with no further interaction [6]. Such sessions were short, with a median duration of 7 seconds (SD= 10).

Participants primarily used their smartwatch to check the time or incoming notifications: interactions following a notification (up to one minute), accounted for 41% of all usage. Participants often commented that while they did not engage with the watch in order to check their physical activity, they often paid attention to physical activity feedback, which became a constant reminder to move:

> *I would actually look at the time, but I would also happen to look at the steps*. [P3]

> *I've always expected to see this information privately, such as on a website or my mobile. But, I feel it's a little bit more motivating to have it always 'in my face'.* [P8]

Overall, participants engaged fewer times per day and walked less per day while using Gardy than with any of the other interfaces (see Table 2). Pairwise comparisons with Bonferroni correction revealed significant differences in participant engagement between *Gardy* and *Normly* ($p<0.05$), *Gardy* and *Goal Completion* ($p<0.05$), *Gardy* and *TickTock* ($p<0.05$) and marginal differences in terms of physical activity ($p<0.10$) between *Gardy* and all of the remaining interfaces.

These findings were consistent with participant preferences. Normly was the most preferred prototype (for 9 of 12 participants), followed by TickTock (2 of 12) and Goal Completion (1 of 12). *Ticktock* was also the most controversial as 3 participants considered it their least preferred. The least preferred prototype was Gardy, with 8 participants considering it their least preferred.

Table 2. Mean daily usage sessions and step count per interface

	Normly	*TickTock*	*Goal Comp.*	*Gardy*
Usage sessions	122 (SD: 99)	110 (SD: 81)	108 (SD: 69)	86 (SD: 60)
Step count	5460 (SD: 4528)	5150 (SD: 4543)	5340 (SD: 4528)	3760 (SD: 3511)

Participant experiences with Normly

We expected that providing participants with normative feedback on their performance would lead to more frequent action and higher overall levels of physical activity, as compared to *Goal Completion*. This was not confirmed at an overall analysis, as an independent samples t-test showed no significant differences among the daily number of steps walked across both interfaces ($mean_{Normly}$ = 5460 steps, $mean_{\text{Goal Compl.}}$ = 5340 steps, $t(165) = -0.18$, $p=0.86$).

However, we noticed differences among participant behaviors based on how far ahead or behind others they were at each given moment. More specifically, we looked at participant physical activity upon interacting with the watch. Participants interacted with *Normly* a total of 9472 times. In 1855 of those (20%), they were up to 500 steps behind or ahead of others. In 5764 of the times (61%) they trailed behind others by over 500 steps, while in 1799 of the times (19%) were more than 500 steps ahead of others.

We found that, when close to others, participants would take a mean of 5 min after the interaction to start a new walk, and they would walk on average 394 steps. In contrast, when lagging behind by over 500 steps, participants would take significantly more time to start a new walk (mean=19 minutes, ($t(7614) = -10$, $p<0.01$) and walk significantly less steps (*mean*=156 steps, $t(7614) = 19.3$, $p<0.01$), as confirmed by independent samples t-tests. The same happened when participants were far ahead of others, where they would take 10min on average to start a new walk, $t(3649) = -13.1$, $p<0.01$, and walk for an average of 248 steps, $t(3649) = 9.94$, $p<0.01$.

Participants felt motivated to walk when sensing they could easily catch-up or stay ahead of others. This effect would disappear, though, once differences grew bigger in either direction:

> *If I was way far ahead, I wouldn't do much. If I was just a little ahead, I would try to walk and keep ahead.* [P3]

In certain ways, these findings are not surprising. More than providing normative feedback, *Normly* engaged participants in a competition with others, even though they had no relationship to or understanding of who these others were. Participants accepted these others as *similar to themselves* – knowing they shared they same walking goals, and competed with them on a daily basis.

> *... I mean, we have the same goal so we should be walking about the same* [P6]

> *I liked being able to see how good or bad I did against others at a glance (...) even though I didn't know them, It made me want to keep up with other people.* [P12]

From a social comparison perspective, individual motivation and performance is expected to be heightened when outperforming others is attainable but not certain [43]. However, in over 60% of the times individuals checked their watch, they trailed behind others considerably. In fact, participants achieved their daily goal on average only once over the seven days, and as a result were compared to others who consistently performed better, which had a toll on their motivation:

> *It was tough seeing others always ahead of me and knowing I couldn't catch up to them (because I was having a busy week). I just ignored how much others had walked and tried to focus only on mine* [P9]

We must note that participant's underachievement was emphasized as they were being compared to people which met that goal by the end of the day. This was not the case of participants, as they were trying to achieve it - either successfully or not.

A question raised is: if participants witnessed that they consistently underperformed compared to others and that this has a toll on their motivation, why didn't they decrease their daily walking goal? Our analysis suggests that participants wanted to retain their initial target, as they felt the reward of reaching a more demanding goal was more enticing than outperforming a less-competitive group, *e.g.*:

> *I was mostly behind [others], but I didn't really think about changing [my goal] (...) I know I can achieve 8000 steps, so why change it to 5000?(...) It's pretty sweet when I hit my goal before them* [P12]

These insights have implications for the design of normative glanceable feedback, suggesting a need for more dynamic systems that maintain comparative levels of performance for a higher percentage of the time. This might be achieved through deception (e.g., artificially lowering the performance of others to provide an opportunity for the participant to catch up).

Further, many felt frustrated with the flexibility of the interface, as they had to keep putting in steps throughout the day to keep up with others:

> *it's not easy to keep ahead (…) an hour ago I was 90 steps behind so I walked a bit until I was 100 steps ahead. But now I am already 80 steps behind! It is frustrating, but if I don't keep up they will get a lot of steps ahead* [P16]

Participant Experiences with TickTock

By displaying behavioral feedback for a limited amount of time, we expected *TickTock* to reinforce re-engagement habits. This was true as participants re-engaged with *TickTock* more frequently – on average every 9 min - as compared to *Goal Completion* (every 15 min, $t(16675) = 6.59$, $p<0.01$). As one participant noted:

> *It only shows me how active I've been over the last hour, so I need to come back to it ever now and then to see how I'd been.* [P7]

Not only did *TickTock* lead to more frequent interactions, it also triggered more frequent walking activities. When using *TickTock*, participants would make on average 61 walking activities per day. An independent samples t-test revealed a significantly higher number of daily walking activities when participants used *TickTock* as compared to *Goal Completion* (mean=50, $t(162)=-2.5$, $p<0.05$), *Normly* (mean=51, $t(166)=-2.3$, $p<0.05$), and *Gardy* (mean=50, $t(166)=-2.77$, $p<0.01$). They would, however, walk for an average of 77 steps in each walking activity, which was significantly shorter than in *Goal Completion* (mean=106, $t(6910)=4.8$, $p<0.01$), *Normly* (mean=107, $t(8313)=5.8$, $p<0.01$), and *Gardy* (mean=99, $t(8678)=4.6$, $p<0.01$).

Our qualitative data suggest two main reasons for the effectiveness of *TickTock* on triggering short, frequent action from individuals. First, it strengthened individual *accountability* for maintaining minimum levels of physical exercise every hour by making this information explicit and easy to glance upon. Second, it rewarded short breaks from sedentary behaviors by making their impact *visible in an instant*:

> *It rewards my sporadic movements since I can see the colors change when I start moving.* [P3]

We further found that the feedback provided by *TickTock* was a significant predictor of later behavior. We performed a linear regression analysis to predict the time participants took until the next walk after interacting with *TickTock*, based on the feedback they received, namely their active time (min) over the past hour. The number of minutes from a participant looking at *TickTock's* feedback until their next walk can be predicted as 1.06 + 0.95 * active-time; $F(1,8465) = 26734$, $p<0.001$, with a R^2 of 0.76. In other words, for every additional 10 min of physical activity that the participants saw they performed over the past hour, they would take an extra 9.5 min till their next walk (Fig. 4). Participants who saw that they walked 10 or less min over the past hour had a 77% probability of starting a new walk in the next 5 min. As one participant noted:

> *... every time I was at work and saw 0 steps in the last hour, it was a signal to get up and walk around.* [P17]

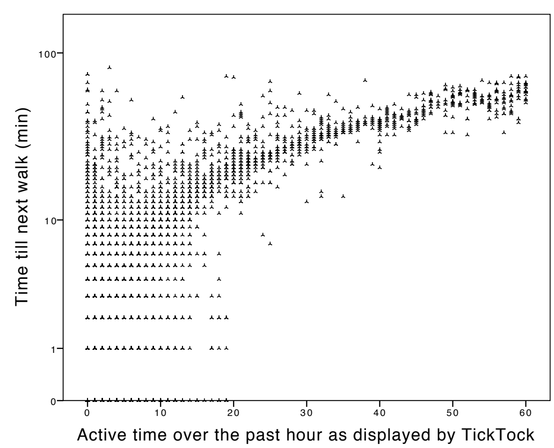


Figure 4. Witnessing that one was sedentary over the past hour would trigger physical activity in shorter period of time.

This push for frequent engagement, however, took a toll on participants' motivation, with some experiencing *reactance* and most reporting that they often felt a lack of credit for physical activity that took place earlier in the day:

> *When I looked and it said I had 0 steps over the last hour, I felt that I hadn't walked for the whole day, which was not the case, so I would think to myself: it's simply not showing the total steps from the whole day (…) I also had no clue how much I had walked over the day.* [P3]

In fact, *TickTock* was the most controversial interface, being the most preferred by two participants and the least preferred by four participants. In addition, three participants ranked it as the most motivating to exercise, while five ranked it as the least motivating. We found that these two groups of participants differed primarily on their fitness goal: participants who rated *TickTock* as motivating had already adopted the goal of breaking sedentary activity throughout the day as their primary motivation.

> *I'm not trying to hit a target so I don't really care about the total [steps] (...) I care more about seeing the steps in the last hour and keeping balanced during my day.* [P11]

In contrast, participants that rated *TickTock* as the least motivating were driven towards larger, daily goals. They found TickTock inflexible and unforgiving on days where their schedule did not allow for frequent physical activity:

> *It's less flexible depending on the day I'll have ahead of me. If I had a goal, I could adjust it on busier days, but in this one I can't really do that. If I stop for 60 min I've gone sedentary.* [P17]

As expected, these two groups differed in terms of their behaviors. An independent samples t-test of found the 'anti-sedentary' group to engage with *TickTock* more frequently (N=165) than the 'daily goal' group (N=75, $t(54) = 5.36$, $p<0.01$), perform more physical activities in the course of their days (mean=82) than they 'daily goal' group (N=51, $t(53) = 5.02$, $p<0.01$), and have marginally higher step count (N=6529) than the 'daily goal' group (N=4176, $t(54) = 1.83$, $p<0.10$) of participants which considered it the least motivating.

Participant Experiences with Gardy

Contrary to our expectations, *Gardy* was the least preferred interface and least motivating to exercise (for 8 and 7 participants, respectively), with participants engaging and performing significantly less physical activity with this interface as compared to the remaining (see section 'Overall engagement and physical activity').

Moreover, single linear regressions revealed that participant engagement decreased over the course of the seven days, by an average of 11 sessions per day ($N_{Engage}=86–11*day$, $F(1,82) = 11.93$, $p<0.01$, $R^2 = 0.13$). The number of steps would also decrease by an average of 442 steps per day ($N_{steps}=3760–442*day$, $F(1,82) = 5.62$, $p<0.05$, $R^2 = 0.06$).

Participants displayed an initial interest in the interface to see how the garden fills up. Some participants would even lower their goal to explore all the stages of the story, e.g., *"[P9]: To be honest, I lowered my goal to get to the last screen faster"*. However, after encountering all stages of the story, their engagement with *Gardy* would be halved.

> *I feel my interest wore off after time (...) probably after I figured out the cycle (...) it's fun to figure out what is going to show up next, but after you get the hang of it, it kind of loses a bit of interest* [P9]

Participants further reported difficulties in estimating exactly their progress over the course of a day, as *Gardy* did not provide numerical feedback on one's step count. In fact, many participants complemented *Gardy* with an external numerical step count (e.g., Google Fit).

> *I knew it changed at every 20% of my goal but I couldn't know how much I walked, precisely. I'm sure I could figure that out, but not by just glancing at it* [P12]

Finally, the public nature of the watch, combined with *Gardy's* simple graphical representation, had a significant impact on participants' attitudes towards the interface. For some, being public was a benefit as it spurred discussion, especially in the presence of children:

> *The garden is definitely the one that attracts more attention (...) I work at a dining and some kids came up with their parents and asked me what it is. I feel good having it full when I explain, it's double rewarding… having them see I've reached my goal.* [P15]

For others, however, it was demotivating as they felt the design of *Gardy* was inconsistent with their self-identity. This would have an impact on its adoption, as participants often reported avoiding checking their watch in public:

> *I would avoid looking at the garden with other people around (...) I would hide it beneath my jacket (...) my own watchface is much simpler and not childish (...).* [P7]

Participant Experiences with Goal Completion

Contrary to *TickTock*, *Goal Completion* seemed to work best for people who preferred defined daily walking goals.

> *I like having a hard goal to hit. It motivates me more than just seeing numbers*. [P4]

Participants appreciated it's minimalistic graphical representation, at which they glanced frequently to maintain an awareness of physical activity and to reassure themselves that they had adequate progress:

> *I feel I was glancing quite often to see where I was (...) by quickly looking at the circle I could tell if I was around 15%, 30% or 50% of my goal.* [P4]

They often developed shortcuts in their decision-making, such as the following one, who developed the strategy of having a short walk if goal completion was less than 50%:

> *I would try to walk when the circle was only half full* [P4]

Interestingly, when interacting with Goal Completion, participants performed the fewest updates of their daily walk goal (N=13) among all interfaces ($N_{Normly}=35$, $N_{Gardy}=66$, $N_{TickTock}=20$). A plausible explanation for this was the lack of novelty of Goal Completion, as all participants had prior exposure to similar feedback through their own activity trackers.

> *I can't say it took me by surprise, I already track my progress (...) It just makes it a bit more glanceable (...) I don't feel it gives me the extra push like the rest do.* [P10]

DISCUSSION AND CONCLUSION

Our design space exploration of glanceable feedback for physical activity trackers resulted in 21 concepts and six overall design qualities. Based on this we prototyped and deployed four concepts "in the wild", representing different elements of the design space. We found that, as expected, integrating feedback with frequently performed activities, such as checking the time, provides a promising path for self-monitoring tools. Participants engaged with their watches about 100 times per day, which is substantially higher than the number of times people engage with an activity tracker app on a smartphone [19]. While checking activity levels was most of the time not their primary intention, they would still glance at it, which impacted their subsequent behavior.

In our analysis of how people responded to the different prototypes, their use to support self-regulation was striking. When using *TickTock*, people who saw that they had a sufficient number of active minutes in the preceding hour were less likely to initiate a new walk, while an individual who had not been active was more likely to initiate a new walk soon.

In people's reactions to *Normly*, however, we see how some presentations of a lack of activity can rather be demotivating. Users took less time to start a new walk, and walked for longer distances, when they were closely behind or even ahead of others. If the difference was large in either direction, however, the feedback did not inspire new walks: the user was either comfortably ahead or too far behind to catch up. These findings corroborate social comparison research – motivation increases when outperforming others is attainable but not certain [43]. These demotivating examples are common in social comparison. In over 60% of the glances at *Normly*, participants saw themselves substantially underperforming. Rather than presenting demotivating feedback in these instances, feedback should maximize its effect on behaviors. One approach, as we discussed earlier, could be the use of *benevolent deception* – for instance, artificially lowering the performance of others, or changing how it is portrayed, to communicate an opportunity for the user to catch up [3,10].

Our study further showed how the different interfaces support self-regulation of different targets, and thus lead to different behavioral patterns. For instance, displaying behavioral feedback for a limited amount of time, as in the case of *TickTock*, led participants to *re-engage* and *walk more frequently*. In contrast, feedback about completion of traditional step goals best supported reaching one's target step count. These are quite subtle effects designers have to consider. Aligning measures and feedback with the desired behavior is key.

Previous research led us to expect *Gardy* to be a popular interface. Participant's responses, however, did not support this. First, this serves as an important reminder that interfaces for smartwatches are more public than smartphones. They must be evaluated not just for their efficacy, but also for their fit with user self-identify [15] and even fashion [42]. Second, while *Gardy*'s stylized representation created some interest in the beginning, it could not sustain interest. After observing one full cycle, participants got bored. More variation, as offered by *UbiFit Garden*, would be important here. How to fit this on a watch interface, however, remains a challenge. Third, many participants experienced difficulties in evaluating their exact progress and planning actions over the course of a day *Gardy's* vague representation. Participants seem to mainly associate exact measurements with a tracker and expect according feedback. This may be a consequence of the all-male, already physically active sample of participants, who in fact already owned smart watches. However, this does not imply that vague feedback is wrong. It can be a way to motivate other people (e.g., novices), who do not respond favorably to a framing of activity in terms of numbers and performance. Fourth, the semantic link between a garden and physical activity is rather weak. Because of this, the garden does not offer the most meaningful story. Letting a garden grow through activity appears slightly arbitrary. Other "stories", such as tending to a Tamagotchi-like dog, which wants and needs to be walked, might be more acceptable and interesting over a longer period of time.

All in all, our study shows that glanecable feedback has a positive effect in general, through its increased availability. More importantly, we showed how subtle differences in interaction emerge, depending on the exact concept (i.e., form) chosen. While some may argue that the how doesn't matter as long as people become more active, we believe that especially for a more sustained use the exact mechanism invoked matter. While the garden may not have been the wisest choice, a vague, varying, more story-like concept could be more motivating than, for example, social comparison in the long run. The story unfolds, while social comparison simply becomes demotivating the moment one realizes that there is no chance of getting ahead of others. This hints a noteworthy limitation of our study. While it was in the wild, it still featured only seven days of using each interface. This prevents drawing any conclusions about long term behaviors from the present results [21].

While future research is needed to assess long-term use and effects, as well as differences in more diverse populations, our study outlines a rich design space for these further explorations. The results of the field study show the importance of aligning feedback with the desired behavior, and highlights opportunities to present more motivating feedback and in ways that are have greater potential to sustain user interest.

ACKNOWLEDGMENTS

This research was supported by the Portuguese Foundation of Science and Technology grants CMUP-EPB/TIC/0063/2013 and SFRH/BD/89796/2012.

REFERENCES

1. Adam Noah, J., Spierer, D. K., Gu, J., & Bronner, S. (2013). Comparison of steps and energy expenditure assessment in adults of Fitbit Tracker and Ultra to the Actical and indirect calorimetry. Journal of medical engineering & technology, 37(7), pp. 456-462.
2. Adams, A. T., Costa, J., Jung, M. F., & Choudhury, T. (2015). Mindless computing: designing technologies to subtly influence behavior. In Proceedings of Ubicomp'15, pp. 719-730.
3. Adar, E., Tan, D. S., & Teevan, J. (2013). Benevolent deception in human computer interaction. In Proceedings of CHI'13, pp. 1863-1872.
4. Adar, E., Teevan, J., & Dumais, S. T. (2008). Large scale analysis of web revisitation patterns. In Proceedings of CHI'08, pp. 1197-1206.
5. Bakker, S., van den Hoven, E., & Eggen, B. (2015). Peripheral interaction: characteristics and considerations. Personal and Ubiquitous Computing, 19(1), pp. 239-254.
6. Banovic, N., Brant, C., Mankoff, J., & Dey, A. (2014). ProactiveTasks: the short of mobile device use sessions. In Proceedings of MobileHCI'14, pp. 243-252.
7. Bentley, F., Tollmar, K., Stephenson, P., Levy, L., Jones, B., Robertson, S., Price, E, Catrambone, R., & Wilson, J. (2013). Health Mashups: Presenting statistical patterns between wellbeing data and context in natural language to promote behavior change. Transactions on Computer-Human Interaction, 20(5), 30.
8. Choe, E. K., Lee, B., Kay, M., Pratt, W., & Kientz, J. A. (2015). SleepTight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In Proceedings of Ubicomp'15, pp. 121-132.
9. Cialdini, R. B., & James, L. (2009). Influence: Science and practice (Vol. 4). Boston, MA: Pearson education.
10. Colusso, L., Hsieh, G., & Munson, S. A. (2016). Designing Closeness to Increase Gamers' Performance. In Proceedings CHI'16, pp. 3020-3024.
11. Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., Libby, R., Mosher, K., & Landay, J. A. (2008). Flowers or a robot army?:encouraging awareness & activity with personal, mobile displays. In Proceedings UbiComp'08, pp. 54-63.
12. Consolvo, S., Klasnja, P., McDonald, D. W., & Landay, J. A. (2014). Designing for Healthy Lifestyles: Design Considerations for Mobile Technologies to Encourage Consumer Health and Wellness. Human–Computer Interaction, 6(3-4), pp. 167-315.
13. Döring, T., Sylvester, A., & Schmidt, A. (2013). A design space for ephemeral user interfaces. In *Proceedings of* TEI'13*,* pp. 75-82.
14. Epstein, D., Cordeiro, F., Bales, E., Fogarty, J., & Munson, S. (2014). Taming data complexity in lifelogs: exploring visual cuts of personal informatics data. In Proceedings of Ubicomp'14, pp. 667-676.
15. Fortmann, J., Cobus, V., Heuten, W., & Boll, S. (2014, November). Waterjewel: Design and evaluation of a bracelet to promote a better drinking behaviour. In Proceedings of Mobile and Ubiquitous Multimedia, pp. 58-67.
16. Fritz, T., Huang, E. M., Murphy, G. C., & Zimmermann, T. (2014). Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In Proceedings of CHI'14, pp. 487-496.
17. Froehlich, J. E., Larson, E., Campbell, T., Haggerty, C., Fogarty, J., & Patel, S. N. (2009). HydroSense: infrastructure-mediated single-point sensing of whole-home water activity. In Proceedings of Ubicomp'14, pp. 235-244.
18. Goffman, E. (1963). Stigma: Notes on the management of spoiled identity. Englewood Clifls, NJ: Prentice-Hall
19. Gouveia, R., Karapanos, E., & Hassenzahl, M. (2015) How Do We Engage with Activity Trackers? A Longitudinal Study of Habito. In Proceedings of Ubicomp'15, pp. 1305-1316.
20. Ham, J., & Midden, C. (2010). Ambient persuasive technology needs little cognitive effort: the differential effects of cognitive load on lighting feedback versus factual feedback. In Persuasive Technology, pp. 132-142.
21. Hazlewood, W. R., Stolterman, E., & Connelly, K. (2011). Issues in evaluating ambient displays in the wild: two case studies In Proceedings of CHI'11, pp. 877-886.
22. Houben, S., Golsteijn, C., Gallacher, S., Johnson, R., Bakker, S., Marquardt, N., Capra, L., & Rogers, Y. (2016). Physikit: Data Engagement Through Physical Ambient Visualizations in the Home. In Proceedings of CHI'16, pp. 1608-1619.
23. Jafarinaimi, N., Forlizzi, J., Hurst, A., & Zimmerman, J. (2005). Breakaway: an ambient display designed to change human behavior. In Proceedings of CHI'05, pp. 1945-1948.
24. Karapanos, E., Gouveia, R., Hassenzahl, M., & Forlizzi, J. (2016). Wellbeing in the Making: Peoples' Experiences with Wearable Activity Trackers. *Psychology of Well-Being*, *6*(1), pp. 1-17.
25. Katz, D., Dalton, N., Holland, S., O'Kane, A., & Price, B. A. (2016). Questioning the Reflection Paradigm for

Diabetes Mobile Apps. In EAI International Conference on Wearables in Healthcare.

26. Kearney, J. M., de Graaf, C., Damkjaer, S., & Engstrom, L. M. (1999). Stages of change towards physical activity in a nationally representative sample in the European Union. Public health nutrition, 2(1a), pp. 115-124.
27. Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. Organizational behavior and human decision processes, 50(2), pp. 212-247.
28. Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. In Proceedings of CHI'10, pp. 557-566.
29. Mankoff, J., Dey, A. K., Hsieh, G., Kientz, J., Lederer, S., & Ames, M. (2003, April). Heuristic evaluation of ambient displays. In Proceedings of CHI'2003, pp. 169-176.
30. Matthews, T., Blais, D., Shick, A., Mankoff, J., Forlizzi, J., Rohrbach, S., & Klatzky, R. (2006). Evaluating glanceable visuals for multitasking. Technical Report EECS-2006-173. UC Berkeley.
31. Min, C., Kang, S., Yoo, C., Cha, J., Choi, S., Oh, Y., & Song, J. (2015). Exploring current practices for battery use and management of smartwatches. In Proceedings of ISWC'15, pp. 11-18.
32. Mullet, K. & Sano, D. (1995) Designing visual interfaces: Communication oriented techniques. Sunsoft Press
33. Munson, S. A., & Consolvo, S. (2012). Exploring goalsetting, rewards, self-monitoring, and sharing to motivate physical activity. In Proceedings of PervasiveHealth'12, pp. 25-32.
34. Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. Personal and Ubiquitous Computing, 16(1), pp. 105-114.
35. Pizza, S., Brown, B., McMillan, D., & Lampinen, A. (2016). Smartwatch in vivo. In Proceedings of CHI'16, pp. 5456-5469.
36. Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: applications to addictive behaviors. American psychologist, 47(9), pp. 1102-1114.
37. Proper, K. I., Singh, A. S., Van Mechelen, W., & Chinapaw, M. J. (2011). Sedentary behaviors and health outcomes among adults: a systematic review of prospective studies. American journal of preventive medicine, 40(2), pp. 174-182.
38. Rogers, Y., Hazlewood, W. R., Marshall, P., Dalton, N., & Hertrich, S. (2010, September). Ambient influence: Can twinkly lights lure and abstract representations trigger behavioral change?. In Proceedings of Ubicomp'10, pp. 261-270.
39. Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. C.(2014). Personal tracking as lived informatics. In Proceedings of CHI'14, pp. 1163-1172.
40. Schneider, H., Moser, K., Butz, A., & Alt, F. (2016). Understanding the Mechanics of Persuasive System Design: A Mixed-Method Theory-driven Analysis of Freeletics. In Proceedings of CHI'16, pp. 309-320.
41. Stawarz, K., Cox, A. L., & Blandford, A. (2014). Don't forget your pill!: designing effective medication reminder apps that support users' daily routines. In Proceedings of CHI'14, pp. 2269-2278.
42. Stusak, S., Tabard, A., Sauka, F., Khot, R. A., & Butz, A. (2014). Activity sculptures: Exploring the impact of physical visualizations on running activity. Transactions on visualization and computer graphic*s*, 20(12), pp. 2201-2210.
43. Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect?. Current directions in psychological science, 11(5), pp. 159-163.
44. Tudor-Locke, C., & Bassett Jr, D. R. (2004). How many steps/day are enough?. Sports medicine, 34(1), pp.1-8.