How Do We Engage With Activity Trackers?
A Longitudinal Study of Habito

Rúben Gouveia
Madeira Interactive
Technologies Institute, Portugal
rubahfgouveia@gmail.com

Evangelos Karapanos
Madeira Interactive
Technologies Institute, Portugal
e.karapanos@gmail.com

Marc Hassenzahl
Industrial Design, Folkwang
University of Arts, Germany
marc.hassenzahl@folkwang-
uni.de

ABSTRACT

We report on a 10-month in-the-wild study of the adoption, engagement and discontinuation of an activity tracker called *Habito*, by a sample of 256 users who installed the tracker on their own volition. We found 'readiness' to behavior change to be a strong predictor of adoption (which ranged from 56% to 20%). Among adopters, only a third updated their daily goal, which in turn impacted their physical activity levels. The use of the tracker was dominated by *glances* – brief, 5-sec sessions where users called the app to check their current activity levels with no further interaction, while users displayed true lack of interest in historical data. Textual feedback proved highly effective in fueling further engagement with the tracker as well as inducing physical activity. We propose three directions for design: designing for *different levels of 'readiness'*, designing for *multilayered and playful goal setting*, and designing for *sustained engagement*.

Author Keywords

Personal informatics, persuasive technologies, behavior change technologies, physical activity trackers.

ACM Classification Keywords

H5.2. User Interfaces: Evaluation/methodology.

INTRODUCTION

Chronic diseases account for nearly 40% of mortality and 75% of healthcare costs worldwide, with obesity alone being responsible for an estimated 12% of the total health spending growth in the United States [39]. Consequently, policy makers argue for a health care model that stresses patient-driven prevention rather than after the fact cure. This burst of interest in prevention combined with progress in technology has lead to a whole new genre of products:

wearable activity trackers. Their market has rapidly grown to a volume of $1.15 billion worldwide in 2014 [33].

Research on activity trackers so far led to a repertoire of theoretically informed design strategies, such as making use of deliberate goal-setting [6,31], increased self-monitoring [5,24], as well as the exploitation of social influence [31]. Yet, despite promising early results, more recent studies raised concerns about activity trackers' long-term efficacy [19]. Shih et al. [40] studied the adoption of *Fitbit* – a wearable activity tracker – by 26 users. They found that 50% quit using the tool after only two weeks. Similarly, a recent survey [22] revealed that over a third of owners of commercially available trackers discarded them within six months after purchase.

While the quick and profound disengagement with trackers seems disheartening, we do not even know whether this is not actually a positive sign. Trackers as currently designed work primarily as "scaffolding" [12]. They provide structure and motivation to people who feel incapable of implementing their intention of exercising without support. In terms of Deci and Ryan's *Self-Determination Theory* ([9], p. 237), motivation to exercise has to be transformed form external to internal, often through steps of introjection (e.g., the tracker embodies exercising as an activity one should do), identification (e.g., exercising is accepted as necessary) and finally integration (e.g., exercising becomes an intrinsically-motivated activity). Thus, disengagement can signify two opposite outcomes: failure to integrate exercising into daily life or a swift adoption of exercising as an intrinsically motivated practice.

In fact, the majority of the studies have focused primarily on the impact of the tracker on behavior rather than, for example, users' intensity of engagement with the tracker. However, we find user engagement to be an important mediator variable for a number of reasons. First, the most commonly employed strategy for behavior change, self-monitoring, requires engagement. So while we may design features to provide self-relevant feedback, it can only impact behavior if people engage with these features [21]. Second, recent studies have revealed rich qualitative findings on the diversity of motives and behavioral practices that surround the use of physical activity trackers. For instance, prior work has suggested that activity trackers
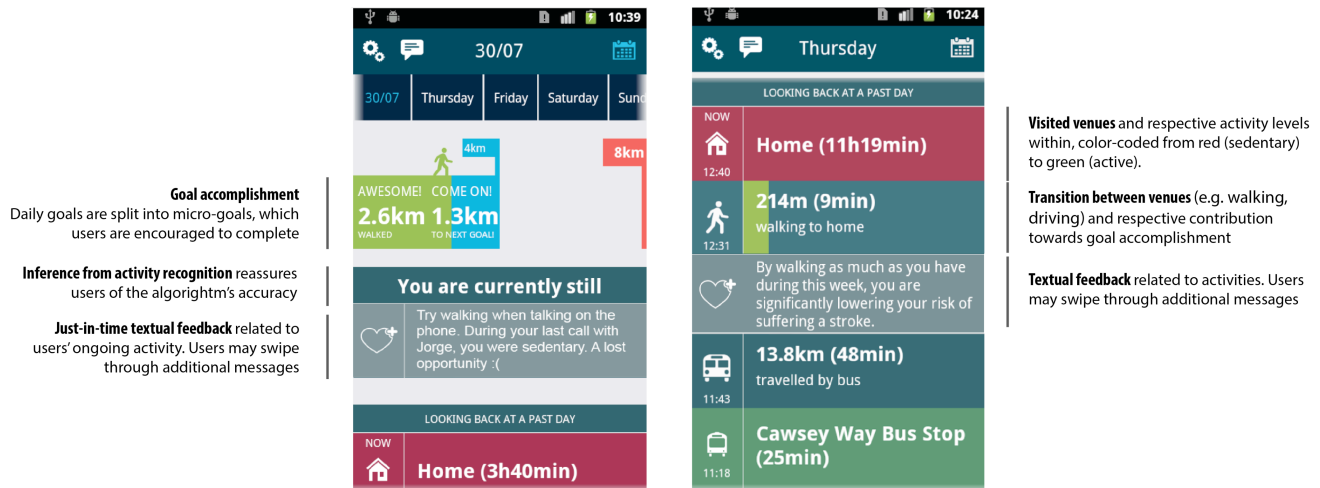
Goal accomplishment
Daily goals are split into micro-goals, which users are encouraged to complete

Inference from activity recognition reassures users of the algoritm's accuracy

Just-in-time textual feedback related to users' ongoing activity. Users may swipe through additional messages

Visited venues and respective activity levels within, color-coded from red (sedentary) to green (active).

Transition between venues (e.g. walking, driving) and respective contribution towards goal accomplishment

Textual feedback related to activities. Users may swipe through additional messages

Figure 1. *Habito* employs three design strategies: *goal setting*, *contextualizing physical activity* and *textual feedback that keeps updating*. The in-the-wild deployment of *Habito* aimed at exploring its adoption, how users engage with feedback, and the impact the design strategies had on users' engagement with the tracker and likelihood to engage with physical activity on the short term.

serve both persuasive and reflective goals [12]. However, we have no knowledge as to when and how users engage with trackers to reflect or to be persuaded. The nature of those different interactions is likely to become apparent through the examination of the frequency and duration of users' engagement with the tracker.

This paper reports an in-the-wild study of 256 users who voluntarily installed *Habito*, a specifically designed physical activity tracker that runs on Android OS. We were interested in acquiring an unbiased estimate of the adoption rate of *Habito* and in understanding how adoption is affected by users' 'readiness', i.e., their motivational stage. Subsequently, we took a closer look at the frequency, duration, and nature of users' engagement with *Habito* itself, the way engagement changes over time, and the impact engagement has on physical activity levels.

In the following section, we describe *Habito* in more detail.

HABITO
Habito (see Figure 1) was specially designed and built to study how users engage with activity trackers. Designing our own application had a number of benefits. First, it enabled us to test different approaches to activity tracking through manipulating the type of feedback given. Second, while commercially available trackers (such as *Fitbit* or the *Moves* app) provide access to people's physical activity data, their APIs have a number of limitations, such as granting access only to the past week's data or providing no tracking of users' actual interaction with the tracker. *Habito* allowed us to not only capture users' physical activity, but also their interactions with the app. *Habito* was developed for Android OS, which allowed us to reach an unbiased sample of users through its deployment on Google Play.

The present version of *Habito* employed three design strategies: *goal setting*, *contextualizing physical activity* with cues relating to location and daily commutes, and *textual feedback that keeps updating* with the goal of sustaining users' interest.

Goal setting
Goal setting is one of the most popular, theoretically informed and empirically grounded approaches to instill behavior change. Research has repeatedly shown that setting concrete goals makes individuals more likely to accomplish them [16,26]. However, goals have to be adopted to become active. The availability of the mere functionality is irrelevant, as long it is not used. Accordingly, knowledge of the extent to which individuals change the defaults goals, of how frequently they update goals, and how those interactions affect engagement and physical activity, would help shed light on individuals goal setting practices and relation to behavioral outcomes.

Upon installing *Habito*, users were prompted to define their daily walking goal. However, a default goal was provided, as prior work has shown that many first-time users are uncertain about how much they walk (or should walk) [20]. While a walking distance of 8 km/day is recommended, we chose to set a default of 1 km/day. We did so for a number of reasons. First, prior work [20] has shown that users tend to underperform compared to medical recommendations. Challenging goals induces initial surprise, experienced as a wake-up call for some, but also induces reactance and higher chance of rejecting the tracker for less motivated individuals. Second, a lower default goal would be achieved easier and, thus, possibly motivate users to reflect (and update) towards their own, appropriate and attainable goal.

Habito provides an awareness of users' current activity levels and goal completion at the top of the screen (see Figure 1). To provide positive reinforcement throughout the day, we split their daily walking goal into four and provided interim milestones. For instance, a goal of 8 km would have three sub-goals: 2, 4 and 6 km. Upon completion of a sub-goal, *Habito* would "reward" the user and motivate her towards achieving the next sub-goal (e.g., "Awesome! 2.6 km walked! Come on, 1.4 km to next goal!").

To track users' physical activity, *Habito* makes use of (1) Google's activity detection API [13], which senses users' state of physical activity (e.g., still, walking, driving) over 30 sec intervals, and (2) an open source step counting algorithm that combines data from the phone's accelerometer and gyroscope [1]. Steps were counted only when 'walking' was detected by the activity detection API. This improved the accuracy of the step counter but also reduced battery drain considerably. Distance walked was further inferred from step count and users' height.

Contextualizing physical activity through location

Habito provides a view of user's daily visited locations and commutes between them (see Figure 1). A new location entry is made if a user spends at least 5 minutes within a 50-meter radius [43]. Locations can further be associated with a name, automatically identified in all subsequent visits. A list of suggested names is provided by *Foursquare*, along two additional places – 'home' and 'work'. Locations are color-labeled according to users' activity levels within this place, ranging from red (sedentary), through orange, to green (physically active). Commutes (as sensed through Google's activity recognition API) and walks outside of places are represented through additional entries.

Contextualizing physical activity through location assists the user in a number of ways. First, presenting additional memory cues (such as places and commutes) supports the recall of episodic memories [8,14], enabling users to identify particular instances of walking that contributed to their daily walking goal. Second, associating physical activity to places supports users in identifying patterns and ill habits, such as places where they are particularly inactive. This should prompt the development of strategies to overcome particularities of places (e.g., when watching TV at home, stand up and move within commercial breaks).

The idea of contextualizing information is not new. Li [23] argued that contextual information may enable users to identify the factors that affect their physical activity levels, eventually "increasing users' awareness of opportunities for physical activity" (p. 53) in the different activities of one's life. In fact, several authors [10, 23] have pointed out that enriching behavioral with contextual information – such as places or people – can reveal factors that affect behavior, and help users to make more informed decisions about how to change their behavior.

However, we have an only limited understanding of how users interact with such contextual information. Epstein et al. [10] explored user preference and perceived values of different visualizations of contextual information (e.g. maps with average time spent in different modes of transportation, and graphs with the total minutes of physical activity for a certain week). However, those insights were based on self-reports while no objective data exists on users' consumption of contextualized data or on their impact towards shaping behaviors (e.g. does a certain representation of contextualized information actually lead users to walk more or eat healthier?).

Textual feedback that keeps updating

The potential of textual feedback in inducing behavior change has been repeatedly highlighted [7]. Rather than "boring users to death with numbers and graphs" [17, p. 48], textual feedback is potentially able to tell a story, is less ambiguous and can help in making sense of the data captured by the tracker. Textual feedback can highlight patterns and draw immediate attention towards important information and instigate action [7] or support reflection over extreme behaviors [30].

Perhaps more importantly, textual feedback can take multiple forms, thus strengthening the tracker's capacity to sustain the novelty of feedback. Prior work has found instant information rewards, such as social media updates and incoming emails on smartphones, to have the capacity to form "checking habits: brief, repetitive inspections of dynamic content quickly accessible on the device" [34]. Consequently, one could wonder whether presenting users with novel textual feedback can lead to checking habits, which sustain engagement with the tracker.

Habito provides users with textual feedback based on their present and past activity levels. Following Munson's classification [30], *Habito's* textual feedback was designed to support either *reflection* or *persuasion*. *Persuasive* messages attempt to instigate behavior change by providing explicit recommendations (*e.g.,* "Try walking when talking on the phone. During your call with Bob, you were sedentary", "Last week, you reached your daily walking goal 2 times, try updating it to 8 km*"). Informational* messages, on the other hand, attempt to assist users in gaining better knowledge about their behaviors, avoiding to employ any form of recommendation or nudging (*e.g.,* "You are the second most active person at work", "You just burned 1560 calories, equivalent to 5 cheeseburgers").

Habito contained a total of 91 different messages, displayed to users over time and in certain conditions. Some of these messages aim to support further inferences about the activities performed. For instance, when registering high physical activity at a given place, messages provided further detail, such as "M-ITI has been your most active location of the week. On average, 400m more than any other location," "In your breaks at M-ITI, you walked an

average of 50 meters. Other messages provided mere facts such as "Only 13% of children walk to school nowadays compared with 66% in 1970" or "Simple movements such as fidgeting, which includes knee shaking or pen tapping can burn up to 800 calories per day." Others provided just in time recommendations such as "You have been sitting for 45 minutes. Try taking a break every 30 minutes," during extended sedentary activities, while others try to create a sense of community, e.g., "M-ITI is the 2nd most physically active community in Madeira. Just 300 meters below the first (University of Madeira)".

STUDY

Habito was posted on Google Play and voluntarily downloaded by users. Over the 10-month period of the study, a total of 256 users downloaded the application. All users had quit *Habito* by the end of the 10-month period.

Contrary to prior work [5,24], we did not sample for users with specific levels of physical activity or increased motivation for becoming fitter, as we wanted to reach out to a representative population of users. We however tried to understand if users' commitment to exercise influenced their adoption of *Habito*. Upon installation of *Habito*, users received an e-mail with the *stage of change questionnaire* [28], which maps people's motivations to change behaviors (i.e. to become more active) to Prochaska's and Velicer's [36] stages of behavior change: *precontemplation* – having no plan to become more active, *contemplation* – not being active but intending to become soon, *preparation* – trying but not yet being regularly active, *action* – being regularly active but for a period less than six months, and *maintenance* – being regularly active for the last six months or more. 54% of users completed the questionnaire.

Users were informed that their data would be stored and analyzed for research purposes. Next to monitoring physical activity and context, application usage was logged, including when the app was launched and quit as well as all interactions within, such as clicking on a specific location, commute or physical activity entry, swiping to a new message, or looking at past days.

Most users (42%) were located in Portugal, followed by United States (25%), United Kingdom (22%), India (9%) and China (2%). All users installed the application on their own volition and no financial incentives were provided.

FINDINGS AND DISCUSSION

"Readiness" for use: motivation and adoption

One hundred and sixty nine users (66%) used *Habito* longer than two days, 97 (38%) longer than a week and only 36 (14%) longer than two weeks. To identify adopters and non-adopters, we ran a k-means cluster analysis on the maximum number of days of use with the number of clusters inferred from the sum of squared error (SEE) curve. This revealed two groups: *adopters*, who used *Habito* for more than a week (97 of 256, 38%), and *non-adopters*, who

quit within the first week (159 of 256, 62%). The former group used the application for a median of 11 days (IQR: 8-16), while the latter used the application for a median of 2 days (IQR: 1-4).

The resulting adoption rate of 38% is clearly below Shih's [40] conservative estimate of 50% for *Fitbit* purchasers. Of course, the present study involved downloading a free mobile app rather than purchasing a wearable device. App acquisition in general is highly exploratory, with only 69% of all apps being kept for longer than two weeks after downloading [37]. For health-related apps this is even worse: Only 1 out of 100 people keep the app, whereas, for example, *Whatsapp*, is kept by every second (50%).

We expected strength of motives to determine whether people adopt *Habito* or not. In fact, previous work has found adoption of interventions for behavior change to be higher in intermediary stages (contemplation, preparation) compared to all other stages (precontemplation, action, maintenance) [24]. 138 users answered the *stage of change questionnaire*. Of these, 43% were found to be in the intermediary stages of behavior change (contemplation: 19%; preparation: 24%). Table 1 shows the adoption rates per stage and in total, along with 95% confidence intervals.

Table 1. Adoption rates of *Habito* per stage of motivation to exercise.

Stages of behavior change	Adopters	%	95% CI (adj. Wald)
Precontemplation	5 of 36	14%	6%-29%
Contemplation	14 of 26	54%	36%-71%
Preparation	19 of 33	58%	41%-73%
Action	7 of 24	29%	15%-49%
Maintenance	4 of 19	21%	8%-44%
Total	49 of 138	36%	28%-44%

The overall adoption rate of 36% (note that this slightly differs from the 38% reported above due to the fact that not all users responded to the stage questionnaire) is clearly a consequence of the stage the person was in. Among the target group (contemplation, preparation), the combined adoption rate was 56% (33 of 59, adj. Wald 95%-CI: 43%-68%), while among the other stages (precontemplation, action, maintenance), the combined adoption rate dropped to 20% (16 of 79, adj. Wald 95%-CI: 13%-30%). A χ^2-test of independence showed adoption not to be independent from the stage a person was in, $\chi^2(1)=18.8$, $p<0.01$.

In sum, given a certain readiness on behalf of users, the adoption rate of *Habito* resembled that found by Shih [40] in the context of *Fitbit*. Obviously, readiness is a strong predictor of adoption, which must be incorporated into studies of the adoption of health-related apps and devices.

In the remainder of the analysis we focus on the adopters' engagement with *Habito*.

Engagement

The 97 adopters had 2737 individual usage sessions (median usage sessions per adopter=28, IQR: 15-45). A session was defined by the moment a user opens the application, until the phone was locked or the application was closed [34]. First we looked at the sessions themselves (and their duration), then we explored patterns across sessions.

Usage sessions

Usage sessions were brief, with 50% of them not longer than 10 sec and 81% not longer than 30 sec. The median session duration was 10 sec (IQR: 4-24). They were thus on average even briefer than in earlier studies of mobile phone use, which found 50% [44], 54% [2] and 61% [11] of usage sessions to last no longer than 30 sec. These differences are unsurprising, as mobile phones are multi-purposive, offering a large range of applications and features, explored sequentially within individual usage sessions (e.g. opening a social media application, switching to a news application and ending up browsing web content) [37]. In the current case, we only focus on using a single application.

Banovic' et al. [2] qualified a usage session as either *glance*, *review* or *engage* session. Glance sessions are brief interactions, in which users check information on the lock screen and then turn the screen off or let the phone timeout [2]. For *Habito*, we define *glance* sessions as sessions in which users open and close *Habito* with no additional actions or inputs (e.g., activating *Habito* to gain awareness of physical activity levels). In Banovic et al. [2], *review* and *engage* sessions involved access to at least one application. These differed however in terms of duration, with *review* sessions lasting up to 60 seconds and *engage* sessions lasting more than 60 seconds. This time split was determined through a head/tail classification [18]. Following this approach, our analysis revealed a natural break point on 22 seconds. *Review* sessions are thus sessions, which last up to 22 seconds, while *engage* sessions last more than 22 seconds, with both sessions involving at least one action within *Habito* (e.g., scrolling through the past day's performance).

Over half (57%) of all usage sessions were *glance* sessions (median duration=5sec, IQR: 2–11), while *review* and *engage* sessions were evenly distributed (review: 22%; median duration=12sec, IQR: 8–18; engage: 21%, median duration=45sec, IQR: 29–67). These results are similar to those of Banovic et. al. [2], who found 47% *glance*, 25% *review* and 22% *engage* sessions with median durations of 14, 23 and 136 sec, respectively.

Review and *engage* sessions were characterized by a high number of interactions with contextual and textual feedback. In fact, 88% and 89% of all *review* and *engage* sessions, respectively, involved exploring an ongoing days' contextual feedback, while exploring textual messages occurred in 84% and 88% of *review* and *engage* sessions.

A χ^2-test of independence revealed a significantly higher frequency of accessing textual messages during *engage* as opposed to *review* sessions, $\chi^2(1)$=4.11, p<0.05. Further, users would access past days with a higher frequency within *engage* (18%) as compared to *review* sessions (13%), $\chi^2(1)$=5.92, p<0.05.

We found the type of session to be linked with goal accomplishment. *Engage* sessions were more frequent when goal accomplishment was low ($\rho(387)$ = -0.58, p<0.05, see Figure 2) while *glance* sessions became more frequent as users progressed towards their set walking goals ($\rho(1084)$ = 0.35, p<0.05). Moreover, the percentage of *glance* sessions would increase over time ($\rho(1053)$ = 0.41, p<0.05), from 45% during first week of use to 68% and 73% during the sixth and twelfth week of use. Additionally, the percentage of *engage* sessions decreased over time ($\rho(1153)$ = -0.38, p<0.05), from 28% during first week of use to 13% and 9% during the sixth and twelfth week of use.
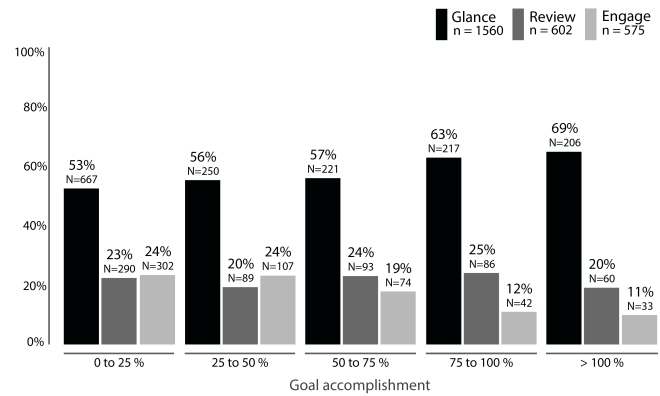

Figure 2. Users' engagement with *Habito* and goal accomplishment (Percentiles and frequencies)

All in all, these results support the notion of activity trackers as "deficit" technologies, to which people turn when they are afraid of failing. During low levels of goal accomplishment individuals exhibit higher dependency on the tracker, as signified by the prominence of engage sessions which focus on feedback. As users progress towards their goal, the prominence of engage sessions decreases and people use the tracker only briefly to acquire an awareness of their current progress towards goal completion (i.e., glance). Over time, when individuals become more self-reliant, use is more and more marked by brief, reassurance-seeking, glance interactions.

Pattern across usage sessions

Approximately a third (29%) of all usage sessions were separated by less than 5 minutes. This resembles the findings of Banovic et al. [2], with 50% of all sessions having been separated by 5 min or less.

We found the time users took to re-engage with the application to increase with their progression towards completion, $\rho(2605)$ = 0.21, p<0.01, see Table 2. Further,

nearly half (46%) of all usage sessions occurred during the first 25% of goal accomplishment, decreasing towards goal accomplishment, $\rho(1067) = -0.32$, p<0.05. These results further support the view of trackers as "deficit" technologies with users taking longer times to re-engage as they become more confident and progress towards their goal.

Table 2. Usage sessions and re-engagement time per goal accomplishment

Goal accomplishment	Usage sessions	Minutes to next session (IQR)
0-25%	46% (N=1259)	10 (1-51)
25-50%	16% (N=446)	27 (2-147)
50-75%	14% (N=388)	82 (1-330)
75-100%	13% (N=345)	133 (9-312)
>100%	11% (N=299)	189 (22-405)

Contrary to what we expected, users would take less time to re-engage with *Habito* after an *engage* session compared to a *glance* session (see Figure 3). In fact, transitions between subsequent *engage* sessions had the lowest re-engagement time (median=4min, IQR: 1-22) than compared to subsequent *glance* sessions (median=27min, IQR: 6-143, Mann-Whitney *U*=54418, p<0.01).

With *engage* sessions displaying short re-engagements and high frequency during lower levels of goal accomplishment, these results hint towards heightened dependency on the feedback and support provided by activity trackers during moments of underperformance. These results may also indicate a break of longer tasks, such as exploring historical information, into closely related micro-tasks, defined by shorter sessions, such as exploring past days over multiple sessions.
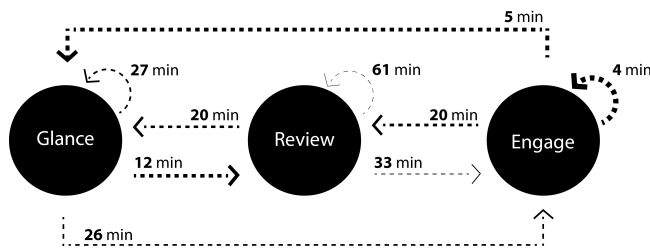

Figure 3. Median transition time between glance, review and engage sessions.

Impact of *Habito's* design strategies

Next we looked at how the three embedded strategies – goal setting, contextualizing physical activity through location and the continuously updating textual feedback – affected users' engagement with the tracker and their levels of physical activity.

Goal-setting

We found only 30 of the adopters (31%) to change their preset goal of 1km per day walking distance. Even more, the large majority of those users (24 of 30, 80%) updated their goal only once: during the first use of *Habito*. During subsequent use, 87% (N=84) of all adopters were recommended, at least once, to update their walking goal (e.g., "last week, you reached your daily walking goal 2 times, try updating it to…"), yet only 5% of them (4 of 84) followed the tracker's recommendation. The median of all updated goals was 8 km (IQR: 5-10).

Figure 4 illustrates the differences among users who updated their goal (self-set goal) and those who didn't (preset goal) in terms of engagement with the tracker as well as physical activity. As one may notice, while both groups had similar patterns of engagement with the tracker, significant differences are found in their levels of physical activity. Users who updated their goal walked more per day (median=6 km, IQR: 3-10) when compared to users that did not update their goal (median=2 km, IQR: 1-4, Mann-Whitney *U*=263, p<0.05).
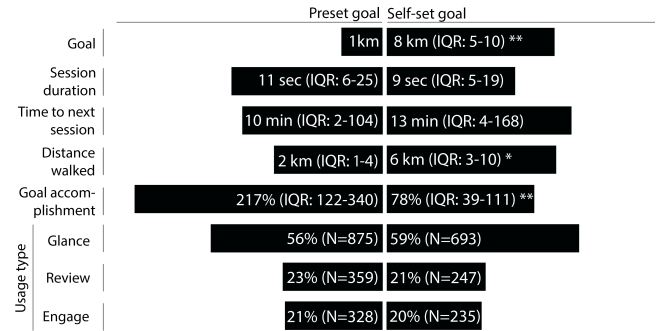
*p<.05, **p<.01

Figure 4. Users' engagement with *Habito* and physical activity levels (Median and IQR values) for those who updated the preset walking goal and those who did not.

This is in line with goal setting theory that argues that setting "difficult goals consistently leads to higher performance than [just] urging people to do their best" [26, p. 706]. However, this did not imply that they were more likely to meet their goal (see Figure 4). In fact, despite a positive correlation between goal and the actual distance walked per day ($\rho(1068)=0.51$, p<0.05), we found a negative correlation between goal and goal accomplishment ($\rho(1068)=-0.67$, p<0.01), which implies that setting a high goal decreases the chance of achievement, but increases physical activity. Supporting users in finding the optimal goal in terms of challenge and achievability is a relevant challenge for activity trackers.

Contextualizing feedback though location and commutes

Users only accessed contextual information in approximately one third (38%) of all usage sessions. This percentage decreased over time, ($\rho(1068)=-0.41$, p<0.05),

from 43% during first week of use to 25% and 18% during the sixth and twelfth week of use.

Interactions with contextual feedback concerned in most cases (89%) the ongoing day, in 7% the past day, and only in 4% a day further in the past. In fact, of all sessions where users looked at a past day's feedback, only 29% involved exploring contextualized feedback. In the majority of cases (71%) it only involved glancing at the overall distance walked during that day.

All in all, these results suggest that contrary to conventional wisdom in personal informatics that portrays behavior change as the result of deep knowledge about one's own behaviors, users lack the interest to reflect on past behaviors. Users' interactions with historical feedback concerned only a small fraction of total use, which further decreased over time, and when users' interacted with past behaviors, these interactions mostly concerned the ongoing, rather than past days. One would expect our added contextual cues (such as location visits and commutes) to strengthen users' capacity to reconstruct past days, which should make past days' history more meaningful and interaction more likely. This was not supported.

One should, however, note that the chosen representation of context may not have been ideal for supporting learning and sustaining users' interest on the long-term. Our representation of context followed the line of the commercial application *Moves*, in which location and physical activity information are presented as unprocessed streams of data in the course of a day. Such low-level representations of context have been found to be less valuable in uncovering the factors that influence behaviors when compared to high-level representations of physical activity (e.g., tables with overall exercise performed during a week at work), as they require "paging days of detailed data to attempt to find trends, correlations, or opportunities for change" [10, p. 2].

Novelty in textual feedback
Users were presented with textual messages (from a pool of 91 different messages) that provided further insights into their physical activity levels. In approximately one third (32%) of all usage sessions, users were presented with a novel message (i.e., one they had not seen before).

When presented with a novel message, users were more likely to swipe to additional messages (median additional messages explored=2, IQR=0-4) as opposed to when presented with a familiar message (median additional messages explored=0, IQR =0-2, Mann-Whitney U=164553, p<0.05). Altogether, after a novel message, users interacted longer with *Habito* (median duration=15sec, IQR: 10-33) than when presented with a familiar message (median duration=7sec, IQR: 2-15, Mann-Whitney U=192711, p<0.05, see Figure 5). A χ^2-test of independence further revealed a significantly higher likelihood of resulting to an *engage* session, when novel

messages were presented (219 of 876) as opposed to familiar messages (354 of 1864, $\chi^2(1)$=13.01, p<0.01), but a significant lower likelihood of resulting to a *glance* session (464 of 876), when novel messages were presented as compared to familiar messages (1100 of 1864, $\chi^2(1)$ = 185.5, p<0.01).

Besides engaging longer, novel message made users return to the application in a shorter period of time (median=13 min, IQR: 2-238), as compared to when a familiar message was presented (median=21 min, IQR: 3-318, Mann-Whitney U=212971, p<0.05).
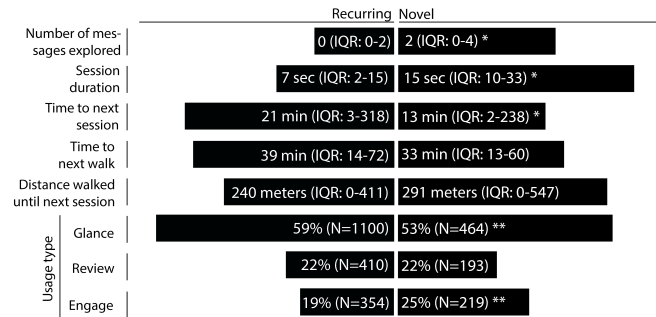
Figure 5. Impact of novel messages on users' engagement with *Habito* and physical activity (Median and IQR values).

Did these bursts of interest that novel content brought inspire users to walk more? Unfortunately not, since no significant differences were found in the time users waited before taking the next walk, or in the distance walked after interacting with a novel or a familiar message (see Figure 5).

All in all, these findings highlight the impact novel content can have on users' engagement with the tracker, both on a single session level (e.g., duration) and in terms of overall patterns of interaction (e.g., time to next usage). However, novelty per se – while intensifying engagement with the tracker – does not translate directly into the target behavior.

Persuasion in textual feedback
We employed two different types of messages in *Habito*: *persuasive* – messages that suggest activities such as "Try walking when talking on the phone. During your call with Jorge, you were sedentary" – and *informational* – messages that provide summative feedback, such as "You just burned 1560 calories, that is equivalent to 5 cheeseburgers"). Prior work has shown that while persuasive messages hold significant motivational power, they can lead to aversion and reactance [3]. Our interest is to understand the impact of both types of messages on engagement with the tracker, and to assess the overall value of persuasive messages with respect to users' level of physical activity.

Approximately two thirds of usage sessions presented exclusively either *persuasive* messages (30%) or

informational messages (35%). *Persuasive* messages led to briefer engagement in the respective session (median=7sec, IQR: 3-11) compared to *informational* messages (median=13sec, IQR: 5-21; Mann-Whitney U=140825, p<0.05). Moreover, users would take significantly more time to re-engage with *Habito* following *persuasive* messages (median=22min, IQR: 5-293) compared to *informational* messages (median=13min, IQR: 2–248, Mann-Whitney U=135533, p<0.05).

However, while persuasive messages led to greater time until re-engagement, users would take less time to start walking and walk for longer distances when presented exclusively with *persuasive* messages (median$_{timewalk}$=29 min, IQR: 16–65, median$_{distancewalk}$=359m, IQR: 0–714) as opposed to *informational* messages (median$_{timewalk}$=38 min, IQR: 21–80, Mann-Whitney U=133181, p<0.05, median$_{distancewalk}$=203m, IQR: 0–493, Mann-Whitney U=113211, p<0.05).
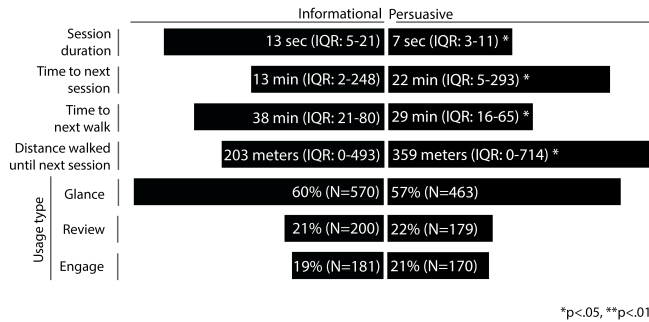

Figure 6. Users' engagement with *Habito* and physical activity (Median and IQR values) when interacting with exclusively informational or persuasive messages.

All in all, our findings seem to support previous research on the dual nature of persuasive messages [3]: while instigating action in the short-term, aversion and reactance also arise, potentially constraining long-term engagements. Further research should employ in-situ methodologies such as Experience Sampling to further inquire into how these effects are mediated through users' subjective experience, such as a momentary decrease in users' perceived autonomy. Next, building upon Munson's [30] guideline for context sensitive messages, research should further estimate the impact context sensing can bring to persuasive messages on increasing the likelihood of opportunistic behavior change and diminishing negative feelings.

IMPLICATIONS FOR DESIGN
All in all, our findings highlight the complexity of the adoption of activity trackers. In the remainder of the text, we discuss some implications for design.

Designing for different levels of 'readiness'
Similar to, but even more than Lin et al. [24], we found 'readiness' for change to be a strong predictor of adoption. Individuals in the contemplation and preparation stages had

an adoption rate of 56%, whereas individuals in precontemplation, action or maintenance stages had an adoption rate of only 20%. This has a number of implications for the design and evaluation of physical activity trackers.

First, it reminds us to take into account people's current motivational stage when evaluating the efficacy of behavior change technologies. Without acknowledging their readiness to behavior change, comparisons of adoption rates and behavior change across studies may not be meaningful.

Second, available findings suggests that current trackers are most likely to work at intermediate stages of behavior change, where individuals have the intention but not yet the means (i.e. motivation, strategies) to change. This leaves out about 57% (in our sample) of the total of potential adopters. Consequently, opportunities to support individuals in the remaining stages is a worthwhile question for the design of activity trackers. For instance, considering the precontemplation stage, an opportunity could be to instill a desire for change rather than merely supporting the process of change. Individuals in the precontemplation stage are often unaware of the extent of their inactivity and are, thus, unwilling to change their behaviors. While existing trackers, if used, just confront them with this "truth" – unblinkingly, in the disguise of a seemingly neutral number – this may turn their initial experiences into something negative, marked by dismay [20]. Engaging users' in the precontemplation stage requires an experiential focus (see [29] for a range of techniques applied by doctors in the precontemplation stage) – one that asks how to increase users' perceptions of self-efficacy and competence.

Designing for multilayered and playful goal setting
Only a third of adopters changed the default daily walking goal. The fact that those who changed their goal walked more, raises the question of how to motivate individuals to reflect upon potential goals and to deliberately set one that is challenging, but achievable. One approach might be enforcing goal-setting. For instance, the commercial tracker 'Basis' asks the user to update their goal once per week. Our results however, seem to indicate this may not be an ideal strategy, as users did not update their goals when recommended to do so, even though achieving their daily goal on a regular basis.

One may attribute this low adherence to more demanding goals as a consequence of low levels of self-efficacy as goal achievement becomes more uncertain. In fact, previous studies have shown that individuals pursue goals more effectively when believing they can be achieved [25]. Activity trackers could allow multiple, simultaneous goals to be set, thus better accounting for the complexity of daily life. This would motivate users to pursue challenging goals in some days, while also guaranteeing a fallback towards more moderate, achievable goals in days that provide limited opportunities for physical activity [25]. Finally,

trackers could build upon persuasive strategies to raise users' perception of the importance of maintaining challenging goals, while also providing reassurance and support regarding their attainability. For instance, opportunities to update goals could be highlighted (e.g. an upcoming weekend or hiking trip), as well as motives towards why goals should be updated (e.g. "setting challenging goals can potentially lead to higher walking distances!").

Further, while setting a challenging, explicit goal, would be beneficial, prior work has shown that users may have no idea how much they should walk and how they compare to others with similar lifestyles [20]. Forcing goal-setting in such a situation is likely to be aversive, as goals may be set above one's real capabilities. One solution would be to devise more playful strategies to goal setting. For example, one's daily walking goal could be set randomly (in sensible bounds of course) like a daily "lottery". It could depend on the weather, drift as users become more skilled with time to make the user more curious about the goal, or it could be imposed by his or her social network. All these would be playful techniques to involve users into a more deliberate reflection and goal-setting.

Designing for sustained engagement

We found 57% of usage sessions to be *glances*, increasing over time to 73% by the twelfth week. *Glances* are brief – with a median duration of 5 sec, spread throughout the day, and mainly provide immediate awareness of one's physical activity. On the contrary, *engage* sessions, where users would spend more time reflecting on the contextual and the textual feedback, were rarer and more frequent during moments of underachievement. As users progress towards meeting their goals, and over time, *engage* sessions become less and less frequent.

On one hand, this supports the dual nature of trackers as "deficit" technologies that "scaffold" behavior during particular problematic moments in time, and as "transformational" technologies that instill and routinize new practices to the point that the technology is no longer necessary [12,17]. On the other hand, it highlights the importance of very brief interactions to maintain engagement.

While glances were frequent and became more frequent with time compared to other types of use, users still became gradually disengaged with the tracker. Note that this decrease in engagement did not lead to reduced physical activity. In fact, user engagement was negatively correlated both with the daily distance walked and the ratio of days in which one's walking goal was met. Similarly to Fritz et al. [12], our findings suggest that users come to disengage with the tracker as they become more likely to meet their daily walking goals. The tracker in its current form is no longer needed. Note however the limitation of our study. We have no knowledge as to what happened after users completely ceased to use and stopped logging their activity with *Habito*. Research has repeatedly found that once the intervention ceases to exist, individuals may relapse to prior stages of behavior change (see [35] for a review). These findings are likely to replicate in the context of activity trackers. Keeping up a minimum of engagement with trackers even in later stages is thus important.

But how could trackers sustain users' engagement? We outline below two plausible directions: *creating checking habits*, and *transitioning glances to moments for reflective engagement*.

Creating checking habits

How can we design behavior change technologies so that they entice users to keep checking their data? Such brief, but frequent interactions (i.e., glances), as we found, can drive much of the usage of the tracker and contribute towards sustained engagement.

Building upon recent work that highlighted smartphones' capacity to create strong *checking habits* [34] – "brief, repetitive inspection[s] of dynamic content" such as social media updates and incoming emails, we asked: *what if the feedback provided by an activity tracker is constantly updating?* In doing so, we would sustain the informational reward people attain from checking feedback, which is assumed to be the primary cause of the formation of checking habits. After all, *novelty* has been a well-established strategy for sustaining engagement in a variety of industries. Consider for instance, the computer gaming industry as well as the airline services, which regularly update their content to sustain user interest.

Our study highlighted that updating the tracker with novel textual messages has the potential to sustain engagement through getting users back to the application faster than when presenting familiar textual messages, possibly hinting towards the formation of checking habits [34]. While the introduction of novel messages per se did not lead to an increase in physical activity, our data suggests that when coupled with persuasive strategies, textual feedback has the potential to lead to an increase in physical activity.

Of course, novelty is not the only means to sustain engagement. Consider for instance the first two concepts illustrated in Figure 7. *TickTock* (left) visualizes when one was physically active, but only over the past half hour. As a result, feedback on one's physical activity is becoming a *scarce* resource [4], and checking habits might be created as a result. As another concept, *Catchup* (middle) enables just-in-time competition with one's past self by contrasting current goal completion (outer ring) to that attained at the same time yesterday (inner ring). Contrary to traditional ways of visualizing goal completion that require projection to establish whether one will meet his or her goal by the end of the day, *Catchup* provides simple, normative feedback (e.g., "you are 200m ahead of the distance you had walked

yesterday by the same time"). As such, the user is likely to frequently reengage with the tracker in order to reassure herself that she is keeping up and ahead of yesterday's performance (see [15] for an elaboration).

Figure 7. *TickTock* (left) visualizes physical activity of only the past half an hour, thus rendering feedback a scarce resource. C*atchup* (middle) enables competition with one's past self through contrasting current goal completion (outer ring) to that attained at the same time yesterday (inner ring). *Predicto* (right) fosters surprise by predicting upcoming day's walking distance from one's sleep pattern, weekday physical activity patterns and recent walking tendencies.

Transitioning glances to moments for reflective engagement
While *engage* sessions are important for reflection and, thus, crucial to sustained behavior change, we found them to represent only a small fraction of use and their frequency to further decrease over time. While this could be understood as the "natural" use cycle of a "scaffolding" technology, one may still attempt to sustain engagement to prevent relapse. Turning glances, which are low on information and reflection, into engage sessions would, thus, be a potential objective. We understand glances as "portals" to deeper engagement with the tracker. One possibility is to avoid providing too flexible, all-embracing and customizable displays, but to provide well-crafted narrative content, a single, tailored and well-crafted story. Or better: The glanceable beginning of a story, which will further unravel when people chose to engage further. For instance, a smart watch may notify a user about his or her high sedentary levels and only through further interaction this story becomes more telling, by for example, providing physical activity levels over the past 30 minutes, creating more opportunities to reflect about reasons for the current lack of physical activity (e.g., place, time, habits).

Such displays could also be sensitive to users' current motivational state, by providing "stories" that match the state. For instance, we found users' informational needs, and consequently their interactions with the tracker, to evolve as they got closer to meeting their goals. While early on, during moments of underachievement, users may need reflection and rich information, later, it may be more about collecting extraordinary accomplishments. When repeated days of inactivity pass by, users may need a wake up call.

Consider for instance, *Predicto* (fig. 7 right), a concept of a tracker that, with its ambiguity, aims at fostering surprise by predicting his or her walking distance from her sleep pattern, general physical activity patterns and recent walking. *Predicto* leverages on the unexpectedness of predictions (overestimation or underestimation) to capture users' attention and make sense of the data.

CONCLUSION
Our study explored the real-world use of an activity tracker. While most studies so far focused on the effectiveness of trackers, that is, their impact on users' levels of physical activity, we sought to better understand the subtleties of how users engage with trackers and how this in turn affects their physical activity.

We found adoption to be strongly influenced by users' 'readiness' concerning the required behavior change. This is certainly one of the reasons why our adoption rate is lower than that of early work [22,40]. The majority of prior studies had systematically biased samples. They selected a specific set of users that already had an appropriate level of 'readiness' for change [24] or provided financial incentives as rewards for participating in the study [5,24,31]. While such studies are useful for tailored efficacy evaluations and greatly advanced our understanding of the effectiveness of different design strategies, they have limited predictive power concerning the adoption and use of a tracker in 'real-life' [21]. Our study revealed the complexities of activity tracking in everyday life, with, for example, users lacking the motivation to set goals, interacting only very briefly with the tracker, and revealing a profound lack of interest in their own historical data. It further showed the dual nature of trackers as a "deficit" technology that "scaffold" change during particular moments in time, and as a "transformational" technology that instills and routinize new practices to the point that the tracker is no longer necessary [12].

We believe activity trackers to be an example for a whole new genre of interactive "transformational technologies" [12], only poorly understood, yet. Detailed and naturalistic inquiries into adoption, use, and disengagement increase our understanding of how current versions of such technologies are actually used. Through this deeper understanding, we may become able to evolve activity trackers from a "deficit" technology, providing the already motivated with some information, into a fully-fledged supportive technology.

REFERENCES

1. Bagi, L. (2013). android-pedometer (Version 1.4). Retrieved from https://github.com/bagilevi/android-pedometer, April 2014.

2. 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.

3. Byrne, S., & Hart, P. S. (2009). The boomerang effect: A synthesis of findings and a preliminary theoretical framework. *Communication yearbook*, *33*, pp. 3-37.

4. Cialdini, R. B., & James, L. (2009). *Influence: Science and practice*, 4. Boston, MA: Pearson education.

5. 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.

6. Consolvo, S., Klasnja, P., McDonald, D. W., & Landay, J. A. (2009). Goal-setting considerations for persuasive technologies that encourage physical activity. In *Proceedings Persuasive'09,* pp. 1-8.

7. 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.

8. Conway, M.A. (2009). Episodic memories. *Neuropsychologia*, *47*(11), pp. 2305-2313.

9. Deci, E. L., & Ryan, R. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), pp. 227–268.

10. 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.

11. Ferreira, D., Goncalves, J., Kostakos, V., Barkhuus, L., & Dey, A. K. (2014). Contextual experience sampling of mobile application micro-usage. In *Proceedings of MobileHCI'14,* pp. 563-572.

12. 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.

13. Google Developers. (2015). ActivityRecognitionApi. Retrieved from https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionApi, April 2014.

14. Gouveia, R. & Karapanos, E. (2013). Footprint tracker: supporting diary studies with lifelogging. In *Proceedings of CHI'13*, pp. 2921-2930.

15. Gouveia, R., Perreira, F., Caraban, A., Munson, S., & Karapanos, E. (2015) You have 5 seconds: Designing Glanceable Feedback for Physical Activity Trackers. In adjunct proceedings of Ubicomp'15.

16. Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta‐analysis of effects and processes. *Advances in experimental social psychology*, *38*, pp. 69-119.

17. Hassenzahl, M., & Laschke, M. (2015). Pleasurable Troublemakers. In S. Walz and S. Deterding, eds., *The Gameful World: Approaches, Issues, Applications*, pp. 167-195.

18. Jiang, B. (2013). Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer*, *65*(3), pp. 482-494.

19. Karapanos, E. (2015) Sustaining User Engagement with Behavior Change Tools, Interactions 22(4), pp. 48-52.

20. Karapanos, E., Gouveia, R., Hassenzahl, M., & Forlizzi, J. (2015) "It's not that hard to walk more": Peoples' experiences with wearable activity trackers. M-ITI Technical Report.

21. Klasnja, P., Consolvo, S., & Pratt, W. (2011). How to evaluate technologies for health behavior change in HCI research. In *Proceedings of CHI'11,* pp. 3063-3072.

22. Ledger, D. and McCaffrey, D. (2014) How the Science of Human Behavior Change Offers the Secret to Long-Term Engagement. Retrieved from http://endeavourpartners.net/white-papers/, December 2014.

23. Li, I. (2012). Personal informatics & context: Using context to reveal factors that affect behavior. *Journal of Ambient Intelligence and Smart Environments*, *4*(1), pp. 71-72.

24. Lin, J. J., Mamykina, L., Lindtner, S., Delajoux, G., & Strub, H. B. (2006). Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *Proceedings of UbiComp'06*, pp. 261-278.

25. Locke, E. A. (1996). Motivation through conscious goal setting. *Applied and Preventive Psychology*, *5*(2), pp. 117-124.

26. Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, *57*(9), pp. 705-715.

27. Lucero, A., Karapanos, E., Arrasvuori, J., & Korhonen, H. (2014). Playful or Gameful?: creating delightful user experiences. *interactions*, *21*(3), pp. 34-39.

28. Marcus, B. H., Selby, V. C., Niaura, R. S., & Rossi, J. S. (1992). Self-efficacy and the stages of exercise behavior change. *Research quarterly for exercise and sport*, *63*(1), pp. 60-66.

29. Miller, W. R., & Rollnick, S. (2012). *Motivational interviewing: Helping people change*. Guilford press.

30. Munson, S. (2012). Mindfulness, reflection, and persuasion in personal informatics. In *Extended Abstracts of CHI'12*.

31. Munson, S. A., & Consolvo, S. (2012). Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *Proceedings of PervasiveHealth'12,* pp. 25-32.

32. Mutsuddi, A. U., & Connelly, K. (2012). Text messages for encouraging physical activity Are they effective after the novelty effect wears off?. In *Proceedings of PervasiveHealth'12,* pp. 33-40.

33. NPD (2013) DigPital Fitness and Wearable Technology. NPD Group's consumer technology report, September 2013

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. 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.

36. Prochaska, J. O., & Velicer, W. F. (1997). The transtheoretical model of health behavior change. *American journal of health promotion*, *12*(1), pp. 38-48.

37. Rahmati, A., Tossell, C., Shepard, C., Kortum, P., & Zhong, L. (2012). Exploring iPhone usage: the influence of socioeconomic differences on smartphone adoption, usage and usability. In *Proceedings of MobileHCI'12,* pp. 11-20.

38. Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. C. (2014). Personal tracking as lived informatics. In *Proceedings of CHI'14,* pp. 1163-1172.

39. RWJF Health Policy Snapshot (2011) "What are the biggest drivers of cost in U.S. health care?"

40. Shih, P. C., Han, K., Poole, E. S., Rosson, M. B., & Carroll, J. M. (2015). Use and Adoption Challenges of Wearable Activity Trackers. In *Proceedings of iConference'15*.

41. Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(2), pp. 411-423.

42. Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological review*, *110*(3), 403.

43. Venkatanathan, J., Ferreira, D., Benisch, M., Lin, J., Karapanos, E., Kostakos, V., Sadeh, N., & Toch, E. (2011). Improving users' consistency when recalling location sharing preferences. In *Proceedings of Interact*, pp. 380-387.

44. Yan, T., Chu, D., Ganesan, D., Kansal, A., & Liu, J. (2012). Fast app launching for mobile devices using predictive user context. In *Proceedings of MobiSys'12,* pp. 113-126.