Activity Trackers in the Wild: Design Strategies and their Impact on Engagement and Physical Activity in Daily Life

1st Author Name	2nd Author Name	3rd Author Name
Affiliation	Affiliation	Affiliation
Address	Address	Address
e-mail address	e-mail address	e-mail address
Optional phone number	Optional phone number	Optional phone number

ABSTRACT
We report on an in-the-wild study of *Habito*, a physical activity tracker that employs three design strategies: *goal setting*, *contextualizing physical activity* and *continuously updating textual feedback*. We find 'readiness' to behavior change to be a strong predictor of the adoption (which ranges from 50% to 7%). Among adopters, only a third updated their daily goal, which in turn impacted their engagement and 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 a 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 reflect on the findings and propose three directions for design: designing for different levels of 'readiness', designing for playful goal setting, and designing for a glance-dominated world.

Author Keywords
Personal informatics, persuasive technology, physical activity trackers.

ACM Classification Keywords
H5.2. User Interfaces: Evaluation/methodology.

INTRODUCTION
Chronic diseases account for nearly 40% of mortality cases and 75% of healthcare costs worldwide, while obesity alone is responsible for an estimated 12% of the total health spending growth in the United States [31]. 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 and progress in technology has lead to a whole new genre of products: wearable activity trackers. Their market has grown to a volume of $1.15 billion worldwide in 2014 [25].

Accompanying research on activity trackers resulted in a wide repertoire theoretically informed design strategies, such as making use of the beneficial psychological effects of deliberate goal-setting, increased self-monitoring, as well as the exploitation of social influence [3,4,5,17,23].

Yet, despite promising early results, more recent studies painted a less positive picture and researchers have raised concerns over trackers' long-term efficacy [9,12,14,32]. Shih et al. [32] studied the adoption of *Fitbit* – a wearable activity tracker – by 26 users. They found 50% to quit using the tool after only two weeks. A survey [15] 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 in fact a rather positive sign. Trackers as currently designed work primarily as "scaffolding". They provide structure and motivation to people, who feel incapable of implementing their intention of exercising more without support. In terms of Deci and Ryan's *Self-Determination Theory* ([6], p. 237) the motivation for exercise has to be transformed form external to internal, often through the 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, a part of the Self). Of course, only short engagement can signify two opposite outcomes: a general 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. Second, recent studies have revealed rich qualitative findings on the diversity of motives and behavioral

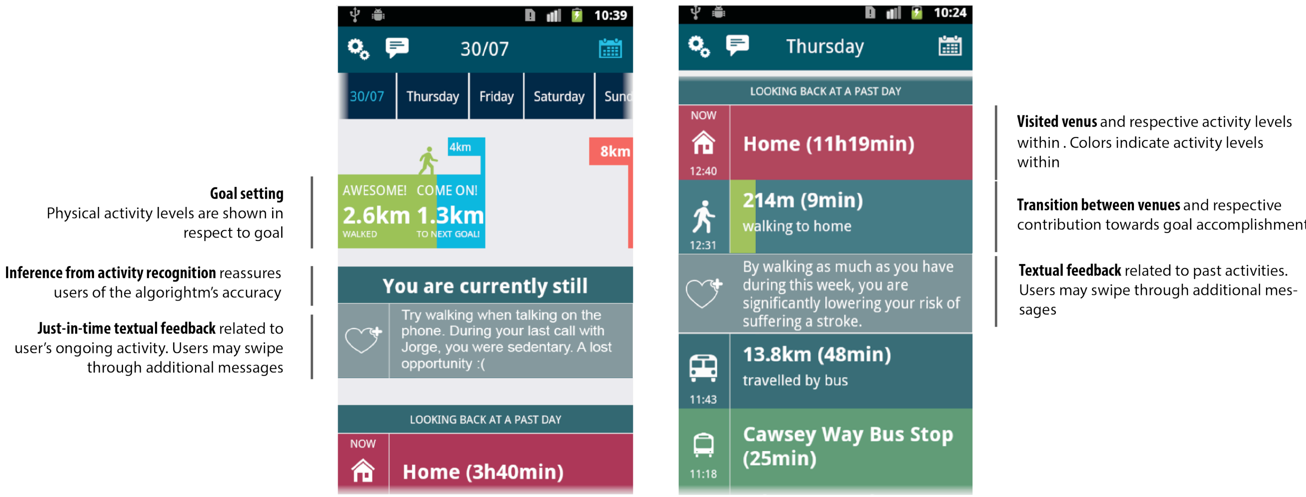

Figure 1. *Habito* employs three design strategies: *goal setting*, *contextualizing physical activity* and *textual feedback that keeps updating*. The in-the-wild deployment of *Habito* aimed at exploring its adoption, how users engage with feedback, and the impact the design strategies had on users'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

practices that surround the use of physical activity trackers [9,22,30]. For instance, prior work has suggested that activity trackers serve both persuasive and reflective goals [9]. 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 86 users, who voluntarily installed *Habito*, a physical activity tracker that runs on Android OS. *Habito* employs three design strategies: *goal setting*, *contextualizing physical activity* and *continuously updating textual feedback*. Specifically, we were interested in acquiring an unbiased estimate of the adoption rate of *Habito* and understand how this 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 their physical activity levels. We further explored the impact of the three design strategies employed by *Habito* (i.e., *goal setting, contextualizing, textual feedback)*.

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 (and disengage from) 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*, *Jawbone* or the *Moves* app) provide access to people's physical activity data, their APIs do 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 (e.g., in terms of number or duration of usage sessions). *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 participants through its deployment on Google Play.

The present version of *Habito* employed three design strategies: *goal setting*, *contextualizing* and different *types of textual feedback*.

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 [10,18]. However, to be supportive, goals need to be set by the individuals themselves. The mere availability of the functionality is irrelevant, if 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 be interesting.

In *Habito*, upon installation, 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) in terms of a concrete distance [4]. While recommended is walking in the range of 8 km per day, we chose to set a low default (i.e., 1 km/day). We did so because of a number of reasons. First, prior work [12] has shown that users tend to significantly underperform compared to medical recommendations. This induces an initial surprise, which is experienced as a wake-up call for some, but also induces reactance and higher chance of rejecting the tracker for less motivated individuals. Second, such a low goal would be achieved easily and, thus, would motivate users to update it with their own, thereby reflecting on what goal would be appropriate and attainable.

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 parts and provided interim milestones. For instance, assuming a goal of 8 km, the user would have three additional sub-goals: 2, 4 and 6 km. Upon the completion of a sub-goal, *Habito* would "reward" the participant and motivate her towards achieving the next sub-goal (e.g., "Awesome! 2.6 km walked", "Come on! 1.4 km to next goal", see Figure 1).

To track users' physical activity, *Habito* makes use of (1) Google's activity detection API, which senses users' state of physical activity (e.g., still, walking, commuting by car) over 30 sec intervals, and (2) an open source step counting algorithm that combines data from the phone's accelerometer and gyroscope. 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

Habito contextualizes users' physical activity through associating it with locations and the commutes between locations (see Figure 1). A new location entry is made, if the user spends at least 5 minutes within a 50-meter radius [35]. By tapping on this entry, the user is presented with the location on a map and a list of nearby places that are retrieved from *Foursquare*. Two additional places – 'home' and 'work' – are presented at the top of the list. Once the user associates a location with a name (i.e., make it into a meaningful place), the place is used in all further displays. When individuals walk within a place, the physical activity is automatically associated with this place. The location entry is then color-labeled to represent how sedentary the user was at this place, with three levels 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 assists the user in a number of ways. First, through presenting additional memory cues (such as places and commutes between places) it supports the recall of episodic memories [5], enabling users to identify the particular instances of walking which contributed to their daily walking goal. Second, associating physical activity with places supports users in identifying patterns and ill habits. For instance, tying physical activity to particular locations may enable users to identify the places, where they are particularly inactive. This should prompt the development of strategies to overcome the particularities of the place (e.g., walking while making telephone calls in the office).

The idea of contextualizing information is not new. Li [16, p. 53] argued that contextual information may enable users to identify how contextual factors affect their physical activity levels, eventually "increasing users' awareness of opportunities for physical activity" in the different activities of one's life. In fact, several authors have pointed out that enriching behavioral with contextual information – such as places or people – can reveal the factors that affect behavior, and help users to make more informed decisions about how to change their behavior [16]. However, we have an only limited understanding of how effective this contextualization is in cueing episodic memories and providing novel insights, and how users interact with such contextual information. Prior work on the adoption of activity trackers has revealed that while users keep wearing the tracker and checking their activity levels through a glanceable wrist-worn display, they stop reflecting over historical data [12]. However, those insights were based on self-reports while no objective data exists on users' consumption of contextualized historical data.

Textual feedback that keeps updating

The potential of textual feedback in inducing behavior change has been repeatedly highlighted [4], yet only rarely employed [4,24]. Rather than "boring users to death with numbers and graphs" [11 2015, 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 [4] or support reflection over outlying behaviors [23].

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 such instant information rewards, found most commonly in 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" [26]. Consequently, one could wonder whether presenting users with novel textual feedback can lead to checking habits, and thus sustain their engagement with the tracker?

Habito provides users with textual feedback based on their present and past activity levels. Following Munson's classification [22], *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 the user in gaining better knowledge about her 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, that's equivalent to 5 cheeseburgers").*

Habito contained a total of 91 different messages, which were displayed to users over time and given certain conditions were satisfied. Some of these messages aim to support further inferences about the activities performed. For instance, when *Habito* had registered high physical activity at a given place, it colored this place in green, and the text below provided further detail, such as "XX has been your most active location of the week. On average, 400m more than any other location," "In your breaks at XX, you walked an average of 50 meters. Others messages provided mere facts such as "Only 13% of children walk to school nowadays compared with 66% in 1970" or "Keep active. Simple movements such as fidgeting, which includes knee shaking or pen tapping can burn up to 800 calories per day." Others provide just in time recommendations such as "You have been sitting for 45 minutes. Try taking a break every 30 minutes," when the system has sensed extended sedentary activity or "If you have time, park your car further away and walk the remaining distance!" when the system has sensed commuting by car, while others try to create a sense of community, e.g., "XX is the 2nd most physically active community in XX. Just 300 meters below the first (XX)".

STUDY

Habito was posted on Google Play and downloaded by users on their own will. For this paper, we selected a sample of 86 participants who had all installed the application by a minimum of 7 weeks before the sample date. This happened 10 weeks after the application deployment, so any users that installed the app after the first three weeks from deployment were discarded from the analysis.

Contrary to prior work [3,17], we did not sample for participants 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* [Marcus, 1992], which maps peoples motivations to change behaviors (i.e. to become more active) to Prochaska's and Velicer's [28] 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. Forty-nine (of 86, 57%) completed the questionnaire.

Participants were informed that their data would be stored and analysed 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 participants (39 of 86, 45%) were located in the [Anonymized country], followed by [Anonymized] (20 of 86, 23%), [Anonymized] (18 of 86, 21%), [Anonymized] (7 of 86, 9%) and [Anonymized] (2 of 86, 2%). All participants installed the application on their own volition and were provided no financial incentives.

FINDINGS AND DISCUSSION

"Readiness" for use: motivation and adoption

Fifty-two users (of 86, 60%) used *Habito* longer than two days, 34 (39%) longer than a week and only 18 (21%) 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 [Tishb]. This revealed two groups: *adopters*, who used *Habito* for more than five days (22 of 86, 26%), and *non-adopters*, who quit within the first five days (64 of 86, 74%). The former group used the application for a median of 14 days (IQR: 10-21), with none quitting before the first week of use. The latter group used the application for a median of 2 days (IQR: 2-4).

The resulting adoption rate of 26% is clearly below Shih's [32] most 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 [29]. 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, *Habito* is targeted a particular user group: People who contemplate to exercise more, but haven't fully fletched routines, yet. Consequently, we expected adoption to be higher in intermediary stages (contemplation, preparation) compared to all other stages (precontemplation, action, maintenance) [17,28]. Forty-nine participants answered the *stage of change questionnaire.* Table 1 shows the adoption rates per stage and in total, complete with 95% confidence intervals.

The overall adoption rate of 27% (note that this slightly differs from the 26% reported above due to the fact that not all participants 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 50% (11 of 22, adj. Wald 95%-CI: 31%-69%), while among the other stages (precontemplation, action, maintenance), the combined adoption rate dropped to 7% (2 of 27, adj. Wald 95%-CI: 1%-24%). A χ^2-test of independence showed adoption not to be independent from the stage a person was in, $\chi^2(1) = 11.28$, $p<0.01$.

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

Stages of behavior change	Adopters	%	95% CI (adj. Wald)
Precontemplation	1 of 13	8%	0%-35%
Contemplation	5 of 9	56%	27%-81%
Preparation	6 of 13	46%	23%-71%
Action	1 of 8	13%	0%-49%
Maintenance	0 of 6	0%	0%-36%
Total	13 of 49	27%	16%-40%

In sum, given a certain readiness on behalf of users, the adoption rate of *Habito* resembled that found by Shih [X] 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 22 adopters had 503 individual usage sessions (median usage sessions per adopter=23, IQR: 12-35). A session was defined by the moment a user opens the application, running until the phone was either idle, locked or the application was closed [26]. First we looked at the sessions themselves (and their duration), then we explored patterns across sessions.

Usage sessions

Usage sessions were brief, with 46% of them not longer than 10 sec and 78% not longer than 30 sec. The median session duration was 12 sec (IQR: 5-28). They were thus on average even briefer than in earlier studies, which found 50% [Yan, 2012], 54% [1] and 61% [7] of mobile usage sessions to last no longer than 30 sec.

Banovic' et al. [1] further qualified usage sessions as either *glance*, *review* or *engage* sessions. 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 [Banovic, 2014]. 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. [1], *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 [13]. Following this approach, our analysis revealed a natural break point on 24 seconds. *Review* sessions are thus sessions, which last up to 24 seconds, while *engage* sessions last more than 24 seconds, with both sessions involving at least one action within *Habito* (e.g., scrolling through the past day's performance).

Approximately half (258 of 503, 51%) of all usage sessions were *glance* sessions (median duration=5sec, IQR: 2–12), while *review* and *engage* sessions were evenly distributed (review: 124 of 503, 25%; median duration=14sec, IQR: 9–21; engage: 121 of 503, 24%, median duration=53sec, IQR:34–78). These results are similar to those of Banovic et. al. [1], which found 47% *glance*, 25% *review* and 22% *engage* sessions with median durations of 14, 23 and 136 sec, respectively.

We found the type of session to be linked with goal accomplishment. *Engage* sessions were more frequent when goal accomplishment was low ($\rho(501) = -0.43$, $p<0.05$, see Figure 2) while *glance* sessions became more frequent as users progressed towards their set walking goals. Moreover, the percentage of *glance* sessions would increase over time ($\rho(251) = 0.41$, $p<0.05$), from 50% during first week of use to 71% and 78% during the third and sixth week of use. Additionally, the percentage of *engage* sessions decreased over time ($\rho(251) = -0.52$, $p<0.05$), from 26% during first week of use to 12% and 11% during the third and sixth week of use.

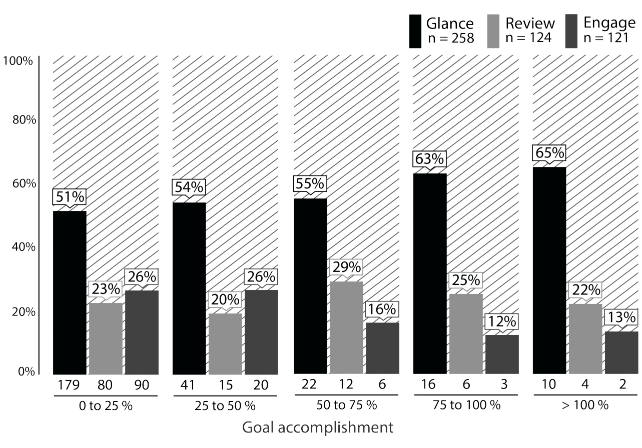

Figure 2. Users' engagement with Habito and goal accomplishment (Percentiles and frequencies)

All in all, these results support the view 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 exhibited by the prominence of engage sessions which signify stronger engagement with the 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 sessions). Over time, as individuals become more self-reliant the use of the tool becomes more strongly centered on such brief, reassurance-seeking interactions.

Pattern across usage sessions

Over a third of usage sessions (171 of 503, 34%) were separated by less than 5 minutes. This resembles the findings of Banovic et al. [1], 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 ($\rho(251) = 0.18$, $p<0.01$), taking a median of 12 min (IQR: 1- 173) to re-engage during the first 25% of goal accomplishment, 32 min (IQR=3-249) during 25-50%, 86 min (IQR=1-330) during 50-75%, 125 min (IQR=7-455) during 75-100%, and 200 min (IQR=9-675) when users had exceeded their goal.

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: 1-36).

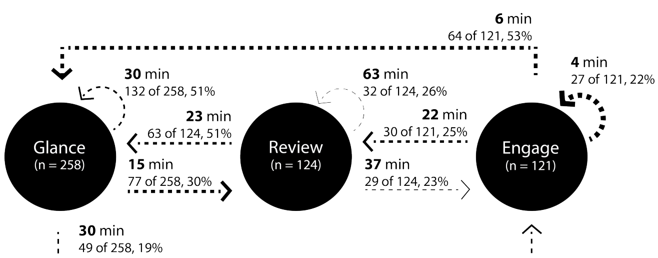

Figure 3. Frequency and time to re-engage among glance, review and engage sessions.

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 completion. Engage sessions serve to empower users during moments of underperformance and are concentrated in time, while glance sessions are spread throughout the day and serve to provide an awareness of physical activity levels.

Impact of *Habito's* design strategies

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

Goal-setting

We found only eight (of 22, 36%) participants to change the preset goal of a 1km per day walking distance. The median goal of the eight participants at the end of the study was 5 km (IQR: 2-8). This is still only half of the daily distance recommended by experts. Almost all participants (7 of 8, 90%) updated their goal only once, namely during the first use of *Habito*.

We found significant differences in users' engagement with the tracker as well as their levels of physical activity, depending on whether they updated their daily goal or not. Participants who updated their goal had briefer usage sessions (median=11 sec, IQR: 4-24, N=221) than those that didn't (median=14 sec, IQR: 5-34, N=282, Mann-Whitney U=26845, $p<0.05$), further reflected in the percentage of *glance* and *engage* sessions (see Figure 4). Furthermore, participants who updated their goal took more time to re-engage with *Habito* (median=37 min, IQR: 4-288, Mann-Whitney U=24325, $p<0.05$, see Figure 4).

Despite their lower levels of engagement with the tracker, participants who updated their goal, on average, walked more per day (median=2 km, IQR: 1-3) when compared to users that didn't update their goal (median=1 km, IQR: 0.7-2, Mann-Whitney U=5205.0, $p<0.01$).

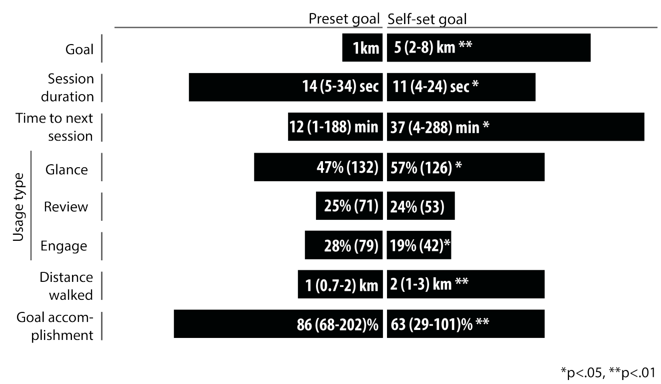

Figure 4. Users' engagement with *Habito* and physical activity levels (Median and IQR values) for those who updated the preset walking goal and those who didn't.

This is line with goal setting theory that argues that setting "difficult goals consistently leads to higher performance than [just] urging people to do their best" [18, p706]. 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 ($\rho(251)=0.38$, $p<0.01$) we found a negative correlation between goal and goal accomplishment ($\rho(251)=-0.41$, $p<0.01$), which implies that setting one's goal high decreases the chance of achievement, but increases physical activity. Supporting users in finding the optimal walking goal in terms of challenge and achievability is a relevant challenge for activity trackers (see [4,23]).

Contextualizing feedback though location

Users only accessed contextual information in approximately half (225 of 503, 45%) of all usage sessions. This percentage decreased over time, ($\rho(251)=-0.45$, $p<0.05$), from 47% during first week of use to 29% and 22% during the third and sixth week of use.

Interactions with contextual feedback concerned in most cases (207 of 225, 92%) the ongoing day, in 5% the past day, and only in 3% any day further in the past. In fact, of all 77 sessions where users looked at a past day's feedback, only 18 (23%) involved exploring the contextualized feedback.

This provides a number of insights. First, there is a lack of interest from users' in contextual feedback, with 45% and 15% of sessions respectively involving an exploration of contextual feedback about the ongoing or a past day. Second, users interest in contextual feedback decreases even further over time, which replicates the findings of [PUC] with overt behavioral data next to their self-reported data. Third, we find users' interactions with contextual feedback to center around the ongoing, rather than past days (92%). 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 confirmed. One explanation is that the chosen representation of context did not work. Future work should explore richer cues (*e.g.,* visual cues from one's perspective such as photos obtained from the sensecam prototype [5,8]) or richer representations of time within the historical feedback (e.g., visualizing the duration of events and allowing for event concurrency). The other explanation is that users truly lack interest in exploring historical data. Browsing historical data could be a strategy tied to particular, but rare occasions, such performing projective analyses (e.g., 'Am I likely to meet my goal?') or "account balancing" (e.g., 'Can I afford walking less today, because I overshot my goal yesterday?').

All in all, next to an overall lack of interest in historical data, users seem to be more interested in the present and the future rather than the past. When users display an interest towards their past, it is mostly in the form of comparing past to ongoing performance.

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 (179 of 503, 36%) of usage sessions, users were presented with a novel message (i.e., one they had not seen before).

Presented with a novel message, participants were more likely to scroll through additional messages (median messages explored=2, IQR=1-4) as opposed to when presented with a familiar message (median messages explored=1, IQR =1-2, Mann-Whitney U=874, $p<0.01$). Altogether, after a novel message, participants interacted longer with the tracker (median duration=19sec, IQR: 8-42) than when presented with a familiar message (median duration=10sec, IQR: 3-23, Mann-Whitney U=12983, $p<0.01$, see Figure 5). A χ^2-test of independence further revealed a significantly higher frequency of *engage* sessions when novel messages were presented (60 of 179, 34%) as opposed to familiar messages (61 of 324, 19%, $\chi^2(1)=13.6$, $p<0.01$), but a significant lower frequency of *glance* sessions occurring when novel messages were presented (80 of 179, 45%) as compared to familiar messages (178 of 324, 55%, $\chi^2(1) = 4.9$, $p<0.05$).

Next to engaging longer, presenting a novel message would make participants to return to the application in a shorter period of time (median=14 min, IQR: 1-150), as compared to when a recurring message was presented (median=26 min, IQR: 4-252, Mann-Whitney U=15512, $p<0.05$).

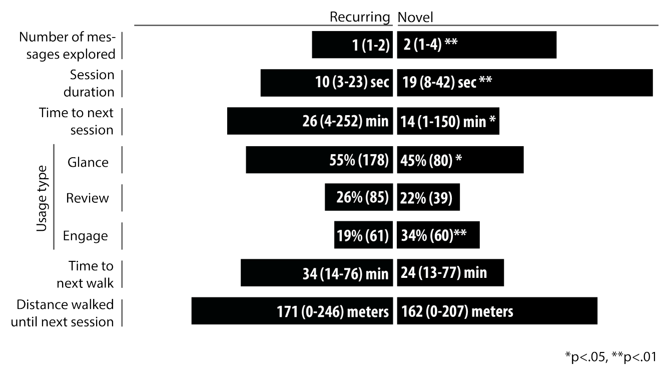

Figure 5. Impact of novel messages on users' engagement with Habito and physical activity (Median and IQR values).

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

All in all, these findings highlight the role 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 activity 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's equivalent to 5 cheeseburgers"). Prior work has shown that while persuasive messages hold significant motivational power, they can lead to aversion and reactance (e.g., recommending users to walk during an important meeting) [22]. 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' likelihood of physical activity.

Approximately two thirds of usage sessions presented exclusively either *persuasive* messages (143 of 503, 28%) or *informational* messages (157 of 503, 31%). *Persuasive* messages led to briefer engagement in the respective session (median=7sec, IQR: 2-18) compared to *informational* messages (median=15sec, IQR: 8-29; Mann-Whitney U=9616, $p<0.05$). Moreover, users would take significantly more time to re-engage with *Habito* following *persuasive* messages (median=27min, IQR: 4-238) compared to *informational* messages (median=17min, IQR: 1–255, Mann-Whitney U=9585, $p<0.05$).

However, while persuasive messages led to greater time till re-engagement, users would take less time to start walking and walk for longer distances when presented exclusively with *persuasive* messages ($median_{timewalk}$=25min, IQR: 13–97, $median_{distancewalk}$=205m, IQR: 0–324) as opposed to *informational* messages ($median_{timewalk}$=39 min, IQR: 13–97, Mann-Whitney U=9313, $p<0.05$, $median_{distancewalk}$=143m, IQR: 0–183, Mann-Whitney U=6692, $p<0.05$).

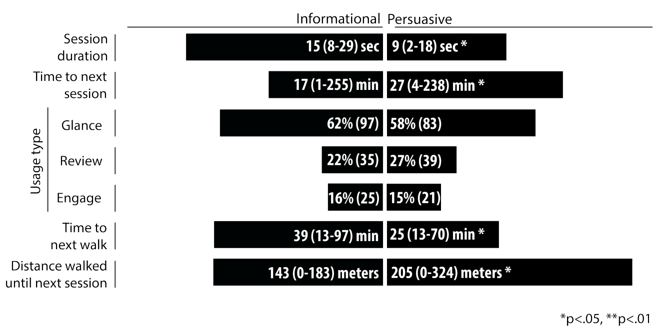

Figure 6. Users' engagement with Habito and physical activity (Median and IQR values) when interacting with exclusively informational or persuasive messages.

All in all, our findings seem to support the dual nature of persuasive messages: while aversion and reactance are possible, they are likely to instigate action in the short-term. Further research should employ in-situ methodologies such as the Experience Sampling Method 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 [22] 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.

DIRECTIONS FOR DESIGN

All in all, our findings highlight the complexity of the adoption of activity trackers. In the remainder of the text, we attempt to reflect on some directions for design.

Designing for different levels of 'readiness'

Similar to, but even more than Lin et al. [17], we found 'readiness' for change to be a strong predictor of adoption. Individuals in the contemplation and preparation stages had an adoption rate of 50%, whereas individuals in precontemplation, action or maintenance stages had an adoption rate of only 7%. This has a number of implications for the design and evaluation of physical activity trackers.

First, it reminds us that when evaluating the efficacy of behavior change technologies, people's motives and readiness for change should be taken into account. Without accounting for such external factors, comparisons of adoption rates and behavior change across studies may not be meaningful.

Second, it suggests that current trackers are most likely to work at the 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 55% (in our sample) of the total pool of potential adopters. Consequently, how to support individuals in the remaining stages is a pressing question for activity trackers. For instance, considering the precontemplation stage, a goal could be to instill a stronger desire for change rather than supporting merely the process of change. Individuals in the precontemplation stage are often unaware of their extent of inactivity and are unwilling to change their behaviors. While existing trackers just confront them with this "truth" – unblinkingly, in the disguise of a seemingly neutral number – this often turns their initial experiences negative, marked by dismay [12]. Engaging users' in the precontemplation stage requires an experiential focus – one that asks how to heighten users' perceptions of self-efficacy and competence. At the same time, trackers may not have enough agency and authority to convey the importance of behavior change (see [21] for a range of techniques applied by doctors in the precontemplation stage).

Designing for playful goal setting

We found only a third of adopters to update the default daily walking goal to a self-set value. This raises the question of how to motivate individuals to reflect upon and set their daily walking goal with activity trackers. One approach might be enforcing the choice. For instance, the commercial tracker 'Basis' asks the user to update their goal once per week. We chose to follow a slightly different approach. When users would repeatedly meet their goal in a week, we would propose an increase (i.e., "last week, you reached your daily walking goal 2 times. Perhaps updating it to 8 km?" This, however, didn't lead to satisfying results.

Prior work has highlighted that users are often clueless about how much they should walk and how they compare to others that have similar lifestyles [9,12]. Traditional approaches to goal setting are thus likely to impose an uncertain and uncomfortable judgment to users in the absence of an established record of their walking activity.

We recommend the exploration of playful strategies to goal setting as a means to providing initial motivation for reflecting upon and setting one's goal. For instance, leveraging upon the Playful Experiences framework [19], which prescribes 22 potential sources of playfulness in users' interactions with technology, one could ask how to design goal-setting strategies that invoke the feelings of completion, fantasy, submission (i.e. being part of a structure), or subversion (e.g. breaking social rules and norms). For example, one's daily walking goal could drift with time to capture his or her curiosity, or be imposed by his or her social network to support reflection through nudging and related social practices.

Designing for a glance-dominated world

We found 50% of the tracker's use to be characterized by *glances*, which further increased to 70% by the third week. Such sessions were brief – with a median duration of 5 sec, spread throughout the day, and served to provide 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 concentrated in time, in particular during moments of underachievement. As users would progress towards meeting their goals, and over time, engage sessions would become less frequent.

On one hand, this supports the dual nature of trackers as learning technologies that scaffold behavior during particular moments in time, and as 'gateway' technologies that routinize new practices to the point that the tracker is no longer necessary [9]. On the other hand, it highlights the potential of glanceable interactions as proxies to further engagement. Based on these findings, we discuss the following two directions for design.

Increasing the frequency and impact of glances

While glances fueled much of the usage of the tool, their frequency would decrease over time and users would gradually come to disengage with the tracker. Note that this decrease in engagement didn't come at the price of 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. [9], our findings suggest that users come to disengage with the tracker as they become more likely to meet their daily walking goals. One has to note however the limited timeframe of our study, while Fritz' et al. [9] insights were only qualitative in nature. Research has repeatedly found that once the intervention ceases to exist, most individuals relapse to prior stages of behavior stage (see [27] for a review). These findings are likely to replicate in the context of activity trackers. Supporting engagement with trackers should thus be important on the long term.

The question raised is how can trackers boost and sustain users' engagement? Our study highlighted that updating the tracker with novel (textual) feedback has the potential to increase engagement through getting users back to the application in faster rates, possibly hinting towards the formation of checking habits [26]. While the introduction of novel messages per se did not lead to an increase in physical activity, our data seem to suggest that when coupled with persuasive strategies, textual feedback has the potential to lead to an increase in physical activity.

The potential of glanceable feedback to lead to opportunistic engagement with behavior change has been previously noted [3, 12]. However, with the emergence of smart watches, glanceable feedback becomes more prominent as their bandwidth of information increases compared to current activity trackers. Fitbit's wearable tool, for instance, includes five LEDs, each lighting up when another 20% of the user's daily goal has been reached. While goal setting is a widely proven technique for behavior change, sustaining one's awareness of goal completion throughout the day may not necessarily be the most effective glanceable feedback, as this requires a projection of one's likelihood to meet his or her daily goal based on the distance walked at the time. The ability and willingness to perform this judgment may vary in the course of a day, as individuals get closer to meeting their goals (see [34] for a review on temporal distancing). An approach to circumventing this issue would be comparing one's walked distance to the distance walked during the previous day at the same time. Similarly, by expanding to the wide spectrum of behavior change theories and techniques, such as social persuasion and just-in-time recommendations, we can create a breadth of new approaches to glanceable feedback in activity trackers.

Moments of learning: transitioning glances to engages

While engage sessions provide the premise of behavior change through reflection, we found them to represent only a small fraction of use and their frequency to further decrease over time. We believe this to be one of the causes of users' gradual disengagement from the tracker, as glance sessions contribute limited new knowledge to users, and thus the perceived value in sustaining the use of the tool would be expected to decrease over time.

We argue however that glance sessions, due to their frequent occurrence, can be leveraged to act as proxies for deeper engagement with the tracker. Designing for these transitions requires a thinking of the nature of such moments of micro-learning. Rather than providing flexible, all-embracing displays that enable users to ask their questions to the data, we argue that such displays should be narrative in their own means (i.e., they should contain a single, tailored and well-crafted story). For instance, a smart watch may notify a user about his or her high sedentary levels and upon further interaction provide her physical activity levels over the past 30 min.

Such displays should also be sensitive to users' current motivational state. For instance, we found users' informational needs, and consequently their interactions with the tracker, to evolve as they progressed to meet their goals. Feedback should thus be dynamic to address the changing informational needs of the individual. While early on, during moments of underachievement, users seem to have heightened need for reflection and rich information, trackers should transition to awareness-enabling devices as users get closer to meeting their daily walking goals.

CONCLUSION

Our study aimed at exploring the real-world use of activity trackers. While most studies have focused on the final outcome of trackers, their impact on users' levels of physical activity, we sought to understand how users engage with trackers and how this in turn affects their physical activity.

In line with recent studies, we found a less positive picture of the adoption of trackers than that painted by early studies. We found the adoption rate, however, to be strongly influenced by users' 'readiness' for behavior change. This likely constitutes one of the reasons for this contrast to early work in activity tracking. The majority of early studies have had systematic biases on their samples, through for instance, selecting a specific set of participants that have the 'readiness' for change [Lin], or providing financial incentives to participants as rewards for participating in the study [2,3,17,23]. While such studies are useful for tailored efficacy evaluations and have greatly advanced our understanding of the effectiveness of different design strategies, they have limited predictive power over the adoption of such tools in 'real-life' [14].

All in all, our study revealed the complexities of activity tracking in real life, with two thirds of users lacking the motivation to set their daily walking goal, the use of the tracker being dominated by brief inspections, and with users displaying a true lack of interest in historical data. Last, the supported the dual nature of trackers as learning technologies that scaffold behavior during particular moments in time, and as 'gateway' technologies that routinize new practices to the point that the tracker is no longer necessary [9].

REFERENCES

1. Banovic, N., Brant, C., Mankoff, J., & Dey, A. (2014, September). ProactiveTasks: the short of mobile device use sessions. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (pp. 243-252). ACM.
2. Consolvo, S., Everitt, K., Smith, I., & Landay, J. A. (2006). Design requirements for technologies that encourage physical activity. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (pp. 457-466). ACM.
3. Consolvo, S., Klasnja, P., McDonald, D. W., Avrahami, D., Froehlich, J., LeGrand, L., ... & Landay, J. A. (2008). Flowers or a robot army?: encouraging awareness & activity with personal, mobile displays. In*Proceedings of the 10th international conference on Ubiquitous computing* (pp. 54-63). ACM.
4. 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), 167-315.
5. Conway, M.A. (2009). Episodic memories. Neuropsychologia, 47(11), 2305-2313.
6. 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), 227–268. doi:10.1207/S15327965PLI1104_01
7. Ferreira, D., Goncalves, J., Kostakos, V., Barkhuus, L., & Dey, A. K. (2014). Contextual experience sampling of mobile application micro-usage. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services* (pp. 91-100). ACM.
8. Gouveia, R., & Karapanos, E. (2013, April). Footprint tracker: supporting diary studies with lifelogging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2921-2930). ACM.
9. 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 the 32nd annual* ACM.
10. Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta‐analysis of effects and processes. Advances in experimental social psychology, 38, 69-119.
11. Hassenzahl, M., & Laschke, M. (2015). Pleasurable Troublemakers. In S. Walz and S. Deterding, eds., *The Gameful World: Approaches, Issues, Applications*, 167.
12. Karapanos, E., Gouveia, R., Hassenzahl, M., & Forlizzi, J. (2015) "It's not that hard to walk more": Peoples' experiences with wearable activity trackers. Submitted for Initial Review in *Personal and Ubiquitous Computing.*
13. Jiang, B. (2013). Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer*, *65*(3), 482-494.

14. Klasnja, P., Consolvo, S., & Pratt, W. (2011). How to evaluate technologies for health behavior change in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3063-3072). ACM.

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

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

17. 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 *UbiComp 2006: Ubiquitous Computing* (pp. 261-278).

18. 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), 705.

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

20. 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), 60-66.

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

22. Munson, S. Mindfulness, reflection, and persuasion in personal informatics. Ext. Abstract CHI'12.

23. Munson, S. A., & Consolvo, S. (2012). Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on* (pp. 25-32). IEEE.

24. Mutsuddi, A. U., & Connelly, K. (2012). Text messages for encouraging physical activity Are they effective after the novelty effect wears off?. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2012 6th International Conference on* (pp. 33-40). IEEE.

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

26. Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing*, *16*(1), 105-114.

27. Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: applications to addictive behaviors. *American psychologist*,*47*(9), 1102.

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

29. 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 the 14th international conference on Human-computer interaction with mobile devices and services* (pp. 11-20). ACM.

30. Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. C. (2014). Personal tracking as lived informatics. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1163-1172). ACM.

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

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

33. 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), 411-423.

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

35. Venkatanathan, J., Ferreira, D., Benisch, M., Lin, J., Karapanos, E., Kostakos, V., ... & Toch, E. (2011). Improving users' consistency when recalling location sharing preferences. In *Human-Computer Interaction–INTERACT 2011* (pp. 380-387). Springer Berlin Heidelberg.

36. Yan, T., Chu, D., Ganesan, D., Kansal, A., & Liu, J. (2012). Fast app launching for mobile devices using predictive user context. In *Proceedings of the 10th international conference on Mobile systems, applications, and services* (pp. 113-126). ACM.