Exploring Understandable Algorithms to Suggest Fitness Tracker Goals that Foster Commitment

Paweł W. Woźniak
Utrecht University
Utrecht, the Netherlands
p.w.wozniak@uu.nl

Przemysław Piotr Kucharski
Ericsson AB
Łódź, Poland

Maartje M.A. de Graaf
Utrecht University
Utrecht, the Netherlands
m.m.a.degraaf@uu.nl

Jasmin Niess
University of Bremen
Bremen, Germany
niessj@uni-bremen.de

ABSTRACT

While fitness trackers are gaining popularity, they struggle to offer long-term health benefits, largely due to their inability to offer engaging goals. Understanding how trackers can suggest and update fitness goals can lead to building improved systems that support wellbeing. We investigate how to suggest fitness tracker goals to users and ways to help them commit to those goals. We compared algorithms for step goal setting in a pre-study. Next, we conducted two surveys (a vignette study and a survey using the users' Fitbit data) that compared the users' attitudes to suggested goals, with and without disclosing the algorithm to them. We found that explaining how a step goal was computed increased goal commitment and, in one study, contributed to building trust in the goal. Our work shows that explaining how a tracker works can help build engaging fitness tracking experiences. We contribute insights on designing transparent personal informatics systems.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

wellbeing; well-being; health; fitness tracker; goal; transparency

ACM Reference Format:
Paweł W. Woźniak, Przemysław Piotr Kucharski, Maartje M.A. de Graaf, and Jasmin Niess. 2020. Exploring Understandable Algorithms to Suggest Fitness Tracker Goals that Foster Commitment. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society (NordiCHI '20), October 25–29, 2020, Tallinn, Estonia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3419249.3420131

1 INTRODUCTION

Fitness trackers are now commonplace on our wrists, in our pockets or integrated into smartphones. In 2017, 29 percent of the US population tracked their physical activity using a wearable device; most of which hoped to increase their overall wellbeing, health and life satisfaction [19]. Yet, despite the apparent commercial success of fitness trackers, potential health benefits to using a fitness tracker are yet to be confirmed in clinical research [8, 9, 26].

To determine how to build trackers that contribute to the users' wellbeing, research in Human-Computer Interaction (HCI) has been attempting to understand the experience of activity tracking. Notably, Epstein et al. [17] found that most personal informatics experiences were interrupted ‘and users often lacked the motivation necessary to prevent lapsing in their use of the tracker. Note that avoiding lapsing is not an end goal for the design of self-tracking tools. Prior work has emphasised happy abandonment of a tracking tool as a positive tracking experience [17]. However, tracking experiences connected to a history of lapses can significantly affects one's tracker experience as shown by Epstein et al. [16]).

Furthermore, Spiel et al. [60] showed how fitness trackers were designed for a very narrow user group. Niess and Woźniak [47] determined that users often experience a mismatch between their fitness goals and the feedback provided by the tracker. This body of work shows that while trackers are gaining popularity, they still require significant improvement. Consequently, understanding more about how fitness trackers affect users and how data produced by fitness trackers can be used for personal benefit remains a challenge for HCI. A key hurdle is finding ways to provide sufficient motivation and offer a holistic long-term personal informatics experience [17].

In this work, we address this challenge by investigating ways in which fitness trackers can promote sustained engagement through suggesting goals, providing a tailored challenge and supporting an ever-improving physical activity routine. Specifically, we explore how fitness trackers can suggest fitness goals to which users are willing to commit by providing users with suggestions that they understand. This approach is inspired by the Tracker Goal Evolution Model [47], which showed that users expected to know how the tracker worked and why a goal was suggested.

To explore goal suggestions for fitness trackers that are meaningful, i.e. congruent with the user's fitness needs [47], we conducted a pre-study and two additional, consecutive studies in which we analysed different aspects of goal setting. We focused on step goals as these are the most commonly used goals [68]. Since there are no widely recognised ways to compute goal suggestions, we first compared different goal suggestion algorithms sourced from commercial applications and social step campaigns in a pre-study. Next,

NordiCHI '20, October 25–29, 2020, Tallinn, Estonia

we conducted a vignette study (*Study 1*) where we focused on the importance of transparency, i.e. explaining how a suggested goal was computed when suggesting the fitness tracker's goals. Finally, we validated our results in a third study (*Study 2*), which used fitness tracker data from the participants' own trackers to suggest new step goals.

Consequently, this paper contributes the following: (1) a pre-study and two additional, consecutive studies on methods for suggesting new fitness goals for physical activity trackers; (2) empirical proof that transparency in goal suggestions fosters goal commitment in fitness tracker applications; and (3) implications for designing future tracking technologies that support physical activity.

We begin this paper by reporting on past research that inspired our inquiry. We then report on the details of the two studies and the pre-study conducted, and discuss and interpret the obtained results. Finally, we show how the results of our inquiry impact the way fitness tracker experiences should be designed.

2 RELATED WORK

In this section, we contextualise our research within past efforts. First, we review past work in the area of personal informatics. We then discuss physical activity support and goal setting in HCI and beyond, followed by related work focusing on trust and transparency.

2.1 Personal Informatics

Aiming for a holistic understanding of personal informatics is a recognised pursuit in HCI. Notably, Epstein et al. [17] presented the Lived Informatics Model of Personal Informatics. Their work extends the stage-based model of personal informatics systems from Li et al. [37]. The Lived Informatics Model consists of four stages: deciding, selecting, tracking and acting, and lapsing [17]. Consequently, a key design goal for a fitness tracker is to keep the user in the personal informatics loop without lapsing. The Tracker Goal Evolution Model by Niess and Wozniak [47] extends the work from Epstein et al. [17] by addressing goals in more detail. The authors emphasise the importance of goal evolution for sustaining long-term engagement. Specifically, they stress the necessity of trust in the goals and the tracker, and required reflection on these goals in terms of relevance and meaningfulness to users. Inspired by these models, we aim to gain a deeper understanding of how to communicate fitness goals with the aim to make them more transparent and foster user trust, thus fostering long-term engagement.

2.2 Physical Activity Support and Goal Setting

As recent studies have shown that it remains a challenge for fitness trackers to deliver long-term health benefits [8, 9, 20, 26], the HCI field is constantly exploring new methods to keep a user engaged in physical activity. Some of these studies aim for long-term engagement by stimulating social activities. Morrison and Bakayov [45] introduced a social activity tracking system that encouraged face-to-face encounters by triggering discussions regarding physical exercise. Similarly, Rooksby et al. [53] developed a mobile application that supported users in tracking, reflecting on and discussing physical activity with others. Fish'n'Steps [38] encouraged physical activity through creating an environment of cooperation as well as competition, showing how combining different methods to foster engagement with the tracker can be beneficial. RunMerge [31] demonstrated how even complicated metrics can foster an inquisitive attitude in users. RUFUS [67] was a system which showcased how communicating positional data while running to others can lead to an enhanced social experience. These works highlight social interaction as an important aspect to enrich the fitness tracking experience.

Other researchers focus on promoting physical activity through reward systems. EdiPulse [29] created chocolate treats to offer playful reflections on physical activity. Similarly, Khot et al. [30] explored presenting physical activity data as artefacts to prompt reflection, whereas Loop [56] used a moving artefact with a similar purpose. Another strain of research explored how users can be helped in pursuing sustained physical activity by allowing them to cheat. Gal-Oz and Zuckerman [18] conceptualised cheating as a behaviour that could foster engagement in fitness. Similarly, Agapie et al. [2] implemented a system utilising cheat points to support users managing their lapses and found that giving users cheat points could foster motivation. This variety of systems exemplifies multiple ways to address the need for achievement by fitness tracker users. Additonally, prior research aims at integrating lapsing, one of the stages characterising a tracker process [17] into their personal informatics systems [2, 18]. However, as Niess and Woźniak [47] have shown, it is still a challenge to integrate self-tracking that accounts for varying goals into everyday life. These works illustrate how adherence to tracking and reaching goals is a strong theme in personal informatics work. However, less attention was given to the nature of goals per se or how they can be set effectively. Our work explores that gap.

Past research also addressed goal setting techniques as a means of engaging users [10]. Psychologists have determined that specific goals lead to better outcomes than vague goals [39]. Furthermore, studies have shown that difficult goals lead to higher levels of performance than easier ones [35]. Munson and Consolvo [46] found that having both secondary and primary goals were perceived as beneficial to help users be physically active. In contrast, ribbons and trophies have not been perceived as motivating by most users. Our work is interestingly different as it explores ways to explain fitness tracker goals in a way that fosters transparency and trust in the tracker rather than specific goal setting techniques or reward systems.

2.3 Trust and Transparency in Technology

Our study is based on the understanding of trust introduced by the Tracker Goal Evolution Model [47]. We also consider that trust is a key component in continued usage of technologies [52]. Previous work has investigated how different interfaces support transparency and foster trust [32]. Early HCI work showed that if users understand how a system works, they are able to focus on themselves instead of on the system [58]. However, Höök [25] notes that it is not necessarily desirable to have a system explain how it works in full detail because these might be alienating to a layman user. Indeed, Kizilec [32] found that there was a need to balance interface transparency when designing for trust. Too much transparency can be as counterproductive as too little. Other studies found mixed results on the effect of transparency on trust; some showed positive effects while others did not [32]. However, Niess and Wożniak [47] found that trust is one of the key contextual factors to foster meaningful fitness tracker goal engagement. Furthermore, they found that transparency can help users understand how the tracker works and thereby support building trust. Thus, our work investigates how to formulate transparent fitness tracker goals and how this affects attitudes towards and trust in these goals.

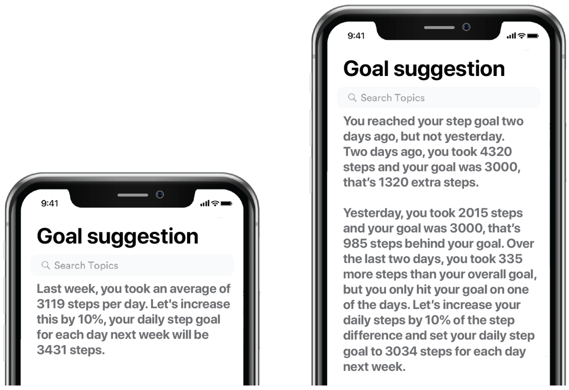
2.4 Explainability

As shown above, transparency alone is often not enough to make a system intelligible for ordinary users. Understanding the value of providing explanations for interactive systems to users is another research topic relevant for our work. When systems provide explanations for their recommendations people will better understand how such systems work [11, 22] which in turn facilitates trust in the system's recommendations [14, 64]. Herlocker et al. [23] showed that, although providing explanations for automated collaborative filtering (ACF) systems increases user acceptance, extracting understandable explanations presented on a usable interface remains challenging. In our work we aim to tackle this challenge in an exploration in a personal informatics context.

Stumpf et al. [61] confronted participants with different explanations of machine learning predictions and explored the willingness of the participants to provide feedback to the learning system. They found that machine learning systems can explain their reasoning and behaviour to users. However, they also found that the willingness to provide feedback is connected to the understandability of the explanations (e.g. rule-based explanations were the most understandable). Pu and Chen [50] outlined the potential of explanation interfaces for recommender agents to foster trust in users. The authors asked participants to evaluate two graphical recommendation interfaces and to determine which interface is more helpful in recommending products to users. The authors showed that trust-inducing interfaces increased the intention of participants to return to the agent and reduced their cognitive effort. Our work explores how insights built in the system explainability can be applied to personal informatics.

Recently, Rader et al. [51] investigated how explaining Facebook News Feed algorithms affects user assessment of the News Feed. Their results show that the explanations increased participants' awareness of how the system works. Binns et al. [6] conducted three consecutive experimental studies exploring how users assess the fairness of algorithm decisions and how explanations affect that perception. They found that there is no simple answer to whether an explanation helps individuals assess the fairness of an algorithmic decision, and stressed that more research was needed to gain deeper insights regarding algorithm explanations in different application areas. Eiband et al. [15] combined designing for transparency and interactive technology for physical activity in a fitness application. They introduced a stage-based design process to support the integration of transparency in real-world scenarios, which is now successfully integrated into the commercial Freeletics Bodyweight Coach[1]. Our work builds on this research and explores

[1]https://help.freeletics.com/hc/en-us/articles/115004675329-The-Freeletics-Bodyweight-Coach-Explained



(a) Goal 1 **(b) Goal 5**

Figure 1: Two goal suggestions from the *Pre-study* presented as mock fitness application screens.

whether explaining how step goals are computed can potentially increase transparency, help users understand how the tracker works and, consequently, build trust. We are inspired by initial results that showed the benefits of transparency for supporting physical activity. We complement past work by specifically investigating fostering transparency and trust in fitness tracker goals in order to stimulate designing more engaging tracking experiences.

3 RESEARCH QUESTIONS

We endeavour to understand how fitness goals suggested by trackers can be perceived as worth committing to and keep the user engaged in the tracking experience. This leads to the following research questions:

- **RQ1**: How can fitness tracker goals be presented to users to foster goal commitment and trust?
- **RQ2**: Does transparency in how a fitness goal was computed result in improved goal commitment?

4 METHOD

To explore how fitness trackers can effectively suggest goals to users, we conducted a pre-study and a series of two studies. The detailed algorithm descriptions and the complete, anonymised study data can be found in supplementary material. In our consecutive study design we move from a theoretical to a more concrete approach. We used step goals in all three studies as they represent the most commonly used goal type [47, 68]. While goal setting was explored in theoretical terms, no studies (to the best of our knowledge) investigated goal suggestion algorithms for fitness trackers. That is why our inquiry started with a pre-study where we compared algorithms sourced from commercial applications. There is a conflict between public health literature (advocating a constant 10000 step goal [62] and psychology (suggesting ever-evolving goals [40]) and, to the best of our knowledge, no scientific source regarding fitness tracker goal calculations exists. Consequently, we take an exploratory, proof-of-concept approach.

We then chose two algorithms from the *Pre-study* and carried out a vignette study (*Study 1*) to explore if transparency in disclosing how fitness goals were computed affected goal commitment and trust in the goal suggestion. Finally, we re-validated our findings in *Study 2*, where we presented transparent and non-transparent goal suggestions to users based on their own current Fitbit data. Before each of the three studies participants were presented with a page informing them that the study was fully anonymous and asking for informed consent. The studies were conducted according to the ethics standards of the conducting institution. According to these rules, the survey was not subject to review by an ethics board. All three studies used participants recruited from Amazon Technical Turk. Theoretical works and empirical studies highlighted potential benefits of Amazon Technical Turk [48]. Paolacci et al. [48] replicated studies from previous judgement and decision making studies and obtained similar results to the original studies. Further, the platform enabled us to source a sample representative of active fitness tracker users worldwide [42].

5 PRE-STUDY

Determining the most beneficial algorithm to suggest a step goal for an individual falls primarily within the scope of the medical and sports sciences. Yet, understanding how to design interactive systems that support effective goal suggestions requires examples of such algorithms. Consequently, we conducted a between-subjects survey comparing five algorithms inspired by those used in commercial devices. As mentioned above, the detailed algorithm descriptions and the complete, anonymised study data can be found in supplementary material.

5.1 Conditions

In order to establish which algorithm could be the most appealing to users and produce most intended goal commitment, we searched for algorithms already used in fitness trackers. We were unable to find any official information from fitness tracker manufacturers or research work that would discuss step goal setting algorithms. Consequently, we turned to internet fora where users tried to determine what the algorithms in their trackers were based on. This resulted in five experimental conditions for our study.

In the survey, users were presented with hypothetical step statistics for two weeks presented as part of a tracker application prototype, as seen in Figure 2. Afterwards, we showed a prototype phone screen with a goal suggestion and explanation computed according to the algorithm in one of the randomly assigned conditions. Figure 1 shows examples of the goal screens.

5.2 Participants

We recruited n=67 participants (44 males, 23 females), aged 19–59 (M = 33.58, SD = 10.57) using Amazon Mechanical Turk (MTurk). The participants resided in the United States or the European Union. We required participants to have completed at least 1,000 HITs with a 95% acceptance rate, in line with past studies of personal informatics [17]. The survey took an average of 3min 35s to complete and the participants received $1.00 as compensation.

Figure 2: A mock fitness app screen with step statistics presented to users in the *Pre-study* and *Study 1*.

(1) I am confident in the recommended step goal from the fitness app.
(2) The recommended step goal from the fitness app is deceptive.
(3) The recommended step goal from the fitness app is reliable.
(4) I trust the recommended step goal from the fitness app.

Table 1: The Trust Scale used in our studies, adapted from work by Cramer et al. [11]. Items were scored on a 7-point Likert scale and item (2) was scored inversely.

5.3 Measures

After the survey introduction, we presented the participants with questions regarding their demographic data. Afterwards, we administered the Goal Commitment Scale (GCS) from Hollenbeck et al. [24]. The GCS has already been applied in various contexts (e.g. [27, 49]. Furthermore, past work [33, 66] demonstrates the connection between goal commitment and behaviour — committing to a goal is a necessary element in behavioural change.

We then conducted a scale measuring trust in the system (Trust Scale, see Table 1). The participants indicated their agreement on a Likert scale from very strongly disagree to very strongly agree. The trust scale was a modified version of the scale used by Cramer et al. [11]. We used the scale to assess if the way the goal was calculated provided potential user benefit. Lastly, we inquired about the participants' propensity to trust, using the faith and trust in general technology scale by McKnight et al. [44]. Propensity to trust has previously been identified as a personality trait independent of a specific trustee as well as independent of the context [55]. Applied to technology this means that trust in technology is given across different technologies and different contexts of use [44]. Thus, we investigated if the trait of propensity to trust technology was correlated to trusting goals suggested by technology, as the literature would suggest.

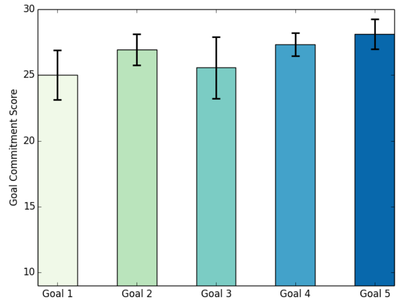


Figure 3: Mean scores on the Goal Commitment Scale for the five goal suggestion algorithms in the *Pre-study*. Error bars show standard errors.

5.4 Results

We conducted a one-way ANOVA to determine the effect of the algorithm used on intended goal commitment, and found no significant difference, $F(4, 58) = 0.94$, $p = 0.45$. Goal commitment scores for the algorithms are shown in Figure 3. Another ANOVA with the aligned rank transform (ART, [65]) applied revealed no significant difference for the effect of the algorithm used on trust in the system, $F(4, 58) = 0.74$, $p = .57$. Additionally, a Pearson's product-moment correlation test was computed to assess the relationship between general faith in technology and intended goal commitment. There was a moderately positive correlation of $r = 0.43$, $p < .01$.

As our study showed no significant differences or even a trend between the goals, we decided to disregard the details of the algorithm in the following studies. We chose Goal 5 as the algorithm to use for further investigation as the highest scoring version. Additionally, we concluded that general faith in technology was a factor affecting trust and intended goal commitment, as suggested by past work [47], and would be included in further analyses.

6 STUDY 1

The second stage of our work (*study 1*) was an experimental vignette study that explored how transparency in step goals affected intended goal commitment and trust in the goal. We decided to use a vignette study (i.e. a study where we ask participants to see the world through the eyes of a hypothetical person in a specific scenario), motivated by past work showing that vignette studies offer the means to balance the benefits of experimental research with high internal validity and the advantages of applied research with high external validity [4]. An additional reason for conducting a vignette study to study tracker goals is the possibility of involving participants who do not own a fitness tracker or would not be willing to contribute their fitness data for the purposes of a study. This is confirmed by privacy research, e.g. [28], which has shown that users have different attitudes towards their own data than towards the data of others. Thus, we decided to first explore our research questions in a more controlled setting with a higher internal validity and a larger sample.

Study 1, our between-subject vignette study used the same step scenario as the *pre-study*. Participants were presented with the same application screen (Figure 2) and then with a goal suggestion calculated according to the algorithm we labelled Goal 5 above. In one of the randomly assigned conditions, Black Box, we only presented the calculated number. In the other condition, Transparent, the full explanation was provided, as shown in Figure 1b. Hence, the case as in the calculated increase of the goal suggestion was the same percentage for all participants. We opted to implement the Black Box condition as it represents the current state of goal suggestions used in commercial systems, where no explanations behind goals are provided[2].

(1) I understand why the fitness app recommended the step goal it did.
(2) I understand what the fitness app bases its recommended step goal on.
(3) I understand how the fitness app calculated the recommended step goal.

Table 2: The Transparency Scale used in our studies, adapted from work by Cramer et al. [11]. Items were scored on a 7-point Likert scale.

6.1 Participants

Using MTurk, we recruited n=105 participants, aged $21 - -68$, $M = 33.71$, $SD = 9.47$, out of whom 65 were male and 40 female. The residence and MTurk performance requirements applied were the same as in the *pre-study*. The participants spent an average of 5min 43s on completing the survey and were remunerated with $1.00.

6.2 Measures and Hypotheses

We used the same measures as used in the *pre-study*. Additionally, we wanted to investigate if the systems were perceived as transparent. To that end, we added a scale adapted from Cramer et al. [11]. The items in the scale are presented in Table 2.

We hypothesised that explaining how the system works would make users perceive it as more transparent and thus build trust and foster intended commitment. Thus, we formulated three research hypotheses:

H1: Presenting a Transparent tracker goal to users will foster significantly more goal commitment than a Black Box tracker goal.

H2: A system using a Transparent tracker goal will be perceived as significantly more transparent than a system using a Black Box tracker goal.

H3: A Transparent tracker goal algorithm fosters significantly more trust in the system than a Black Box tracker goal.

6.3 Results

We conducted three one-way ANCOVAs to determine the effect of disclosing the details of the goal setting algorithm on trust in the system (ART applied), perceived transparency (ART applied) of the system and intended goal commitment, controlling for the users' propensity to trust technology. We found a significant difference for all three measures. Table 3 presents detailed results, shown in Figure 4.

[2]https://www8.garmin.com/manuals/webhelp/vivofit/EN-US/GUID-3C2177B2-5BAE-4324-B709-8148220584D6.html

	Black Box		Transparent		ANCOVA	
	M	*SD*	*M*	*SD*	$F_{1,102}$	*p*
Transparency	14.94	4.97	17.31	2.72	11.70	< .001
Trust	19.04	5.66	21.09	3.78	5.61	< .05
Goal Commitment	25.57	7.08	28.00	4.82	4.93	< .05

Table 3: Means and ANCOVA results for *study 1*. Levene's test and normality checks were performed and the assumptions met. In all cases, the covariate, propensity to trust technology, was significantly related to the measure, p<.001

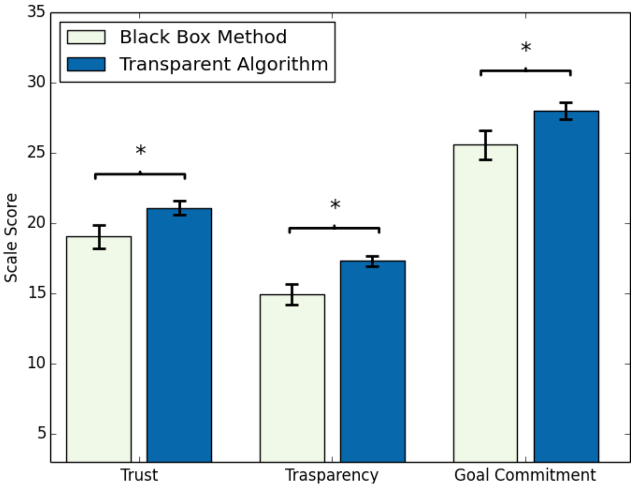


Figure 4: Mean scores on the Trust, Transparency, and Goal Commitment Scale for the two conditions in *study 1*. Error bars show standard errors. Note that the scales have different scoring ranges.

7 STUDY 2

Finally, we conducted a study that used the participants' own data to suggest new goals to them. The conditions and measures used were the same as in the *study 1* with the addition of an open text field where we asked participants if they were willing to pursue the presented goal and explain their decision. Participants were asked to provide at least one full sentence. The suggested goals were computed based on the data collected by the participant using their Fitbit fitness tracker. To that end, we built a custom survey web page that first asked the users to provide two weeks of step data to the study. Participants were asked to log into their Fitbit account and consent to using two weeks of anonymous step data. Through a Fitbit API connection, the data was logged in our custom-designed PHP-based survey system and used to generate the survey.

The request to provide anonymous step data was accompanied by an extensive explanation that the study was fully anonymous and only 14 values of daily steps would be collected. When obtaining the data, we collected step value for the most recent 14-day continued usage period of the tracker. This allowed us to ignore days with very low step values, which were most likely instances of the user forgetting to wear the tracker or the device running out of battery. To mitigate potential bias of the participants (e.g. perceived transparency of the goal suggestion), the data the goal suggestion was based on was displayed to participants in all cases.

Consequently, in the survey, the participants were first presented with a graph of the 14-day step data that was used for calculating a goal suggestion. They were then shown the goal suggestion with a full explanation of how it was computed (Transparent condition) or just given a plain number (Black Box). The conditions were randomly assigned to participants in a between-subjects design. We explored the same research hypotheses as in *study 1*. The case as in the calculated increase of the goal suggestion was the same percentage for all participants. Similar to study 1, the goal suggestion was calculated according to the algorithm we labelled Goal 5 above.

7.1 Qualitative Analysis

In line with established practices in personal informatics, the aim of our additional collection of qualitative data is to understand the qualitative experience of fitness tracker users (e.g. [54]). We used thematic analysis with open coding [7]. Two researchers open coded a representative sample of 15% of the material. Next, a coding tree was established through iterative discussion. The remaining material was split between the two researchers and coded individually. A final discussion session was conducted to finalise the coding tree after the material was coded. The two researchers then identified emerging themes in the material.

7.2 Participants

We recruited $n = 47$ participants (35 males, 12 females) through MTurk and calls on social media. Their ages ranged from 22 to 78, $M = 38.06$, $SD = 11.06$. They received $2.00 as compensation. The participation criteria were the same as in the previous studies with the added requirement of owning a Fitbit tracker.

7.3 Results

The fitness tracker users participating in the study were physically active, taking an average of $M = 13496$, $SD = 413$ steps per day. The step goals that they had set on their trackers ranged from 1500 to 20000 steps, $M = 9408$, $SD = 2895$. Interestingly, 31 out of the 47 participants used the default 10000 step goal.

Similarly to *study 1*, we computed a one-way ANCOVA with ART to investigate the difference between the Transparent and Black Box tracker goal in terms of trust in the system, controlling for the users' propensity to trust technology. We found no significant difference, $F(1, 44) = 0.02$, $p = .88$.

We used another one-way ANCOVA with ART to assess the effect of disclosing the details of the goal setting algorithm on the perceived transparency of the system, controlling for the users' propensity to trust technology. We observed a significant effect,

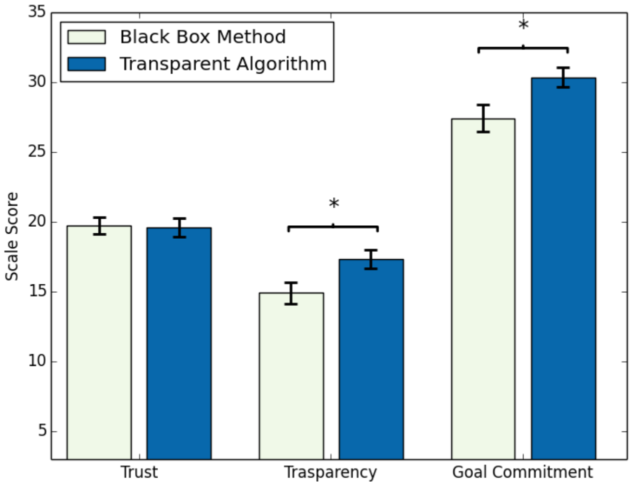


Figure 5: Mean scores on the Trust, Transparency, and Goal Commitment Scale for the two conditions in *study 2*. Error bars show standard errors. Note that the scales have different scoring ranges.

$F(1, 44) = 6.19$, $p < 0.05$. The covariate was significantly related to the perceived transparency of the system, $F(1, 44) = 5.07$, $p < .05$.

A final one-way ANCOVA investigated the effect of disclosing the details of the goal setting algorithm on goal commitment, controlling for the users' propensity to trust technology. A significant effect was found, $F(1, 44) = 7.43$, $p < .01$. The covariate was significantly related to the goal commitment, $F(1, 44) = 8.59$, $p < .01$. We present all results in Figure 5.

7.3.1 Qualitative responses. Here, we present the different views emerging from the data. Four themes were identified: *trustworthy goals*, *meaningful experience*, *contextual factors* and *commitment*.

We found that multiple participants assessed the recommended step goal as *trustworthy*. Even though the participants in the Black Box condition did not receive an explanation about how the fitness tracker goal was computed, they often voiced their confidence in the goal. Interestingly, some participants additionally expressed their uncertainty regarding the computation of the goal at the same time:

I'm not completely sure why it calculated the goal that it did, but I don't think there was anything bad about it. (P22, Black Box)

One user in the Transparent condition emphasised his trust in the fitness tracker goal. The participant directly referred to the algorithm and not the goal:

I trust the algorithm that the fitness app uses. (P41, Transparent)

We further observed that participants reflected on the ways to make the tracking experience more *meaningful*. One user emphasised that it was important to gain increased self-awareness. He recommended making goals personally meaningful:

Why do we need the app to recommend step goals for us? I prefer to recommend my step goals by myself. There was a recommended goal at the beginning when I started to use it [the fitness tracker], but it was difficult to interpret the goals into something meaningful. What does 10,000 steps mean to people? After a few days of tracking, I started to get to know myself, and I was able to adjust the goals. (P9, Transparent)

Similarly, another participant highlighted the importance of the connection between life goals and the numeric tracker goals. She emphasised goal evolution as a means to keep users "inspired" instead of "demoralising" them:

If someone increases their step goal by 10% every X months, there will be a point where they will regularly fall short of that goal, demoralising rather than inspiring them. (...) If the goal is cardiovascular health, then have the goal be to get into a certain heart rate for a certain amount of time every day. (P23, Transparent)

Users mentioned various *contextual factors* that influenced their tracking experience. A statement of a participant in the Black Box condition showed general goal commitment, but also mentioned commitment outside of fitness:

(...) a new video game expansion was released, so I sat at my computer all day instead of being active. (...) Normally, I would try to meet that goal. (P19, Black Box)

Other users struggled to integrate their tracking into their everyday lives, even when they were committed to track. One participant wondered about the fitness tracker taking the interactions with her child into account:

I am not exactly sure what other information it [the fitness tracker] could take into account. It is difficult to count my steps when I am holding my child while walking since I hold him with the arm that is wearing my Fitbit. So that could be taken into consideration, because you are not supposed to wear your Fitbit on your dominant hand/arm (P03, Transparent)

Another theme that we identified in the qualitative data was *commitment* towards the fitness tracker goal. Users eagerly expressed their determination to achieve their goal. One user was striving to meet the goal even when the work conditions hindered it:

I am willing to increase my step goal. I have a job throughout the week that doesn't allow me to wear any kind of watch or jewellery, but I will place my Fitbit on my leg so it won't be noticeable. (P13, Transparent)

In contrast, one participant from the Black Box condition expressed an unwillingness to pursue the recommended goal. The user emphasised the importance of being involved in the decision making process regarding goal recommendations:

I don't want to take on this goal. The app decided without even asking me if I wanted to do it (P12, Black Box)

8 DISCUSSION

In this research, we explored the role of transparency in suggesting goals for fitness tracker users (**RQ1, RQ2**). We found that disclosing the algorithm for computing a fitness goal offered perceived benefits to users. Here, we summarise key findings from our study, then discuss challenges and opportunities for future systems that support physical activity by suggesting fitness tracker goals.

8.1 Transparent tracker goals foster goal commitment

We observed that showing users how a fitness goal was computed increased declared goal commitment both in *study 1* and *study 2* (**RQ1**). Thus, **H1** was confirmed. As the qualitative data shows, users found it easier to relate to the suggested number of steps if accompanied by an explanation. This effect may be caused by a number of reasons. Firstly, showing how a goal was computed illustrates that the system uses the collected information to better understand the user, which Gulotta et al. [21] previously identified as a means to build engagement with personal informatics. Secondly, it appears that transparent goals not only help users understand how the fitness tracker works (which was identified as a currently unmet need for fitness tracker systems [47]), but also contribute to a more personalised experience that results in increased goal commitment. Third, our qualitative results indicate that a lack of transparency of suggested tracker goals caused participants to distrust the goals leading them to not committing to the suggested goal. This link between transparency and trust [11, 22] as well as acceptance [23] has been shown in previous work. We, thus, recommend using transparent goals in future trackers that support evolving personal fitness goals.

8.2 Showing algorithms behind suggested goal increases system's perceived transparency

Study 1 and *study 2* showed that providing the details of the goal calculation algorithm resulted in the fitness tracker application being perceived as more transparent, confirming **H2**. While we did observe a significant effect of providing an algorithm description on perceived transparency, we also note that absolute scores for both conditions were high. This result illustrates that transparency and understanding are connected as the users presented with a more complete explanation of the data reported more transparency, which reflects previous work in other domains, e.g. [11]. On the other hand, we hypothesise that the fact that the collected data was displayed to the participants in all conditions already built a base perception of transparency. Consequently, disclosing available information both about the data collected and the processing of that data could improve the user experience of fitness tracking.

8.3 Disclosing tracker goal calculations fosters trust in some users

We observed a significant increase in trust when the goal suggestion algorithm was disclosed to the users in *study 1*, but no such effect was observed in *study 2* (**RQ1**). This implies that there is currently not enough evidence to fully confirm **H3**. We believe this difference can be primarily attributed to the differences in the participant sample between the two studies. *Study 2* required participants to be active Fitbit users with enough collected data to enable computing a goal suggestion. This implies that these participants trusted the technology enough to purchase and wear the tracker for a certain time. Our work suggests that, while transparency can increase goal commitment in existing fitness tracker users, their level of trust in the tracking system may remain unaffected. However, the fact that we observed an effect on trust in the general sample suggests that transparency is also important for novice tracker users or those who have not yet decided to start tracking fitness in learning to trust their device. This resonates with Epstein et al.'s [17] findings on motivations and hurdles to tracking, but also aligns with research identifying the barriers of technology adoption indicating a lack of understandability and trust as one of the main reason why people resist certain technology [57]. As many users are attracted to tracking through curiosity, transparency may help them enter the tracking experience. Consequently, future tracking systems could use transparency as a means to build trust specifically at the beginning of the fitness tracking experience.

8.4 Transparent fitness tracker design contributes to building commitment

Our results showed that disclosing information about how the tracker works is essential for the creation of engaging fitness tracking experiences and fosters goal commitment (**RQ1, RQ2**). This finding complements the work by Epstein et al. [17] who emphasised that deciding upon the will to change was a key element in the tracking experience. Our results shed light on how tracker systems can make that decision more informed. Based on our results, we hypothesise that trust dynamics with regards to fitness trackers follows a cycle similar to the Lived Informatics Model [17]. Novice fitness tracker users trust their tracker to some extent, enough to begin tracking. This trusting stance, before starting a direct experience with the tracker, can be interpreted as *initial trust* [44] which can diminish quickly. In contrast, *knowledge-based trust* [44], which builds on previous trustor-trustee interactions, can potentially last longer. We, therefore, hypothesise that trust in the tracker evolves over time; as supported by findings in previous work focusing on technology in general [44]. It seemed that we observed primarily initial trust in *study 1* and knowledge-based trust in the *study 2*. However, future work is needed to explore this assumption further.

Thus, explaining how the tracker works is one step towards building trust [47]. We see this as an opportunity to build further engagement with the tracker by capitalising on the knowledge-based trust once established. According to the Tracker Goal Evolution Model, once the users feel they know how their tracker works and trust is established, trackers can help users relate data to their qualitative goals (e.g. becoming fitter). If future trackers can use transparency to effectively build trust and then facilitate the connection to qualitative goals, they will be able to offer a long-term experience that keeps users committed and motivated [39]. Past work in HCI [47] as well as work from psychology shows that static or not meaningful goals are counterproductive for any kind of activity [40]. Consequently, goal setting is essential to sustained engagement.

8.5 Meaningful goals and life context

The results of our qualitative analysis in *study 2* show that users reflected on a suggested goal by immediately contextualising the requirements of fulfilling it. In addition, we found that users took various contextual factors into account. Multiple participants perceived their recommended step goals as reasonable and trustworthy, partly independently of their experimental condition (Black Box, Transparent). This suggests that while a goal recommendation and the reasons behind it are important in committing to a fitness goal, other factors are at play. In *study 2*, we also observed that the attitude towards a goal was influenced by a general commitment towards fitness tracking. These findings illustrate the challenge to integrate current fitness trackers into everyday life in a meaningful way, which is reflected in past research [47, 63].

Users assessed their goals as reasonable. They showed a general commitment towards their goal, but anticipated that they would struggle to pursue it for a variety of reasons. Integrating fitness tracker experiences into everyday life and connecting quantitative tracker goals with qualitative life goals still remains a challenge [47]. The eventual failure of the fitness tracker to support users in achieving their qualitative fitness goals may lead to a loss of confidence in the device and a decrease in trust. Kizilec [32] emphasises that transparency of a system is only relevant when user expectations have not been met. Moreover, our qualitative results show that irregular occurrences disrupts daily use patterns that the system's algorithms did not account for. This creates opportunities for user modelling for fitness goals, which was hinted by recent work [5]. As a consequence, future systems for suggesting fitness goals should enable users to incorporate those contextual factors (e.g. life events, injuries or health issues) into the decision process of committing to a goal and/or include those factors in how the goal is calculated. This can be especially useful if hurdles to physical activity such as injuries or eventual lapses in tracking are present.

8.6 Goals and algorithmic complexity

Next, we reflect on the *pre-study* which showed no differences between the five algorithms that we investigated. As we wanted the study to be close to the current lived practice of fitness tracking, the conditions examined were heavily inspired by solutions from presently available fitness trackers. However, one could easily imagine that trackers would use more complex algorithms, models or even machine learning solutions, especially if contextual factors were considered in the goal suggestion process. Concurrently, recent work showed that automated and crowdsourced exercise plans (and thus exercise goals) are likely to grow more complex and become more useful to users [3]. As suggestions grow more complex, explaining how they are computed to users emerges as a key challenge. We envision that future fitness trackers should incorporate regularly communicating how user data is processed in a transparent manner so as not to jeopardise user engagement, even for the user group which is currently actively tracking fitness.

8.7 Ways forward

Finally, we discuss further research that can enable meaningful goal recommendations for fitness trackers. An issue that needs further investigation is the form in which the explanation of the goal should be provided. While we used plain text explanations, the semantics and form of the explanation require further research. Work in other application domains points to several areas for exploration. Given that algorithms are becoming increasingly complex, users are more likely to require that these systems provide explanations for their decisions [12]. First, analogously to the psychology of everyday human speech, users expect explanations from complex systems in ordinary language [41]. Explanations provided by systems should be compatible with beliefs, desires and other mental states that motivated the decisions [13]. We suggest that future research should further explore how fitness tracker goal explanations can be enhanced by offering reasons, motivations, and justifications [36]. Second, systems should be able to distinct classes of actions (i.e., intentional versus unintentional) and explain each of these classes in the expected way (i.e., unintentional behaviours with (mere) causes, and intentional behaviours with reasons) [1]. Future work should investigate how intentionality can be embedded in goal recommendations through appropriate verbalisations of algorithmic decisions.

Finally, there is a need to balance transparency so that the user is not overwhelmed with the information behind their fitness goal. If goal suggestion algorithms become complex, they cannot simply be communicated with full transparency [12, 32]. Further research is necessary to determine what parts of the goal explanation ordinary users are interested in [43], what the user already knows [59]; and what elements of explanations may build bridges between presumed knowledge and novel information [34].

8.8 Limitations

Our work constitutes a first step towards designing transparent fitness tracker goal recommendations, yet we recognise that the approach used in this paper is prone to certain limitations. We used mainly MTurk to recruit the participants for our three consecutive studies. Even though previous research discussed advantages of that approach [42, 48], we recognise that the target audience of fitness trackers most likely extends beyond the pool of participants available on MTurk. Further, there are sampling and recruitment differences between the participant groups in *study 1* and *Tracker* studies, which may explain the differences in results which we observed. As fitness tracking technologies reach larger audiences, identifying user groups and their specific needs is a challenge that emerges from our work. We studied participants from North America and Western Europe, similar to the majority of past work focused on fitness trackers. However, in line with the work from Spiel et al. [60], we believe that our findings may be applicable only to a subset of the general population. Future research should explore cultural and social factors in the perception of transparency in fitness tracker goals.

We recognise that there are many different ways in which goal-setting can be better supported. Trusting algorithms to offer suggestions is one of them, but not the only one. However, prior work has consistently pointed to goal-setting as a critical challenge, and yet commercial applications today still struggle to help people set appropriate goals. Our works constitutes an exploratory inquiry in this area, aimed to inspire future research. Our choice to use a 'black box' condition was a decision that also impacted our results. Investigating in which way the algorithm should be explained is a viable alternative. However, as commercial systems suggest goals in a arbitrary manner or do not suggest them at all, we opted to use the 'black box' baseline in the pursuit of societal elevance and ecological validity.

Further, as our work primarily investigated transparency and trust in the tracker, the choice of a goal suggestion algorithm was a secondary concern. While we wanted to use goals inspired by commercial solutions, we see that the design of the algorithms might have affected the results of the study. We were unable to find related work that would suggest better algorithms, and research in public health is focused on establishing general step goals rather than evolving challenges. Finding goal suggestion algorithms that are meaningful to users and beneficial from a health (or sports performance) perspective is a challenge for future research. Once these methods are available, our work should be revisited.

Our work explicitly does not adopt a psycho-theoretical approach as research is yet to determine the correct theoretical approach for personal informatics. That is why we take practice-oriented approach which can yield insights for the design of fitness trackers. However, we do recognise that a theory-driven inquiry may lead to other insights.

Finally, most of the participants in *study 2* used the default 10000 step goal, which suggests that they might not have changed or even considered changing their step goal before. There is a possibility that our study primarily promoted them to consider a goal adjustment and thus produced a novelty effect. On the other hand, this would show the benefits of providing goal suggestions and thus the need to better understand how to suggest evolving goals effectively. Given that goal stagnation was previously identified as an issue in personal tracking [17, 47], it appears that the group of users who regularly reflect on their tracker goals is limited.

9 CONCLUSION

This paper investigated the effects of transparency in communicating fitness tracker goals on perceived transparency, trust in the system and goal commitment. We conducted two studies and a pre-study: *Pre-study*, *Study 1*, *Study 2*. We asked users to express their views of proposed fitness goals for both hypothetical scenarios and suggestions based on their own data. We found that disclosing how a suggested tracker goal was computed resulted in significantly increased goal commitment and perceived transparency of the system. We also found limited evidence that a transparent step goal also fostered trust in the system. Our results show that there are complex trust dynamics involved in users contextualising and committing to a suggested step goal. We discussed how future trackers could use transparency to foster increased engagement with fitness goals and offer a more meaningful long-term tracking experience. Further, we recommend that future trackers allow users to explicitly address contextual factors such as life events when setting goals.

Our work sheds new light on the complexity of communicating transparent fitness tracker goals. We hope to stimulate further studies in personal informatics in areas beyond fitness tracking. We believe that building an understanding of how to communicate to users how their tracker works can support building more engaging fitness tracker experiences and assist users on their way to wellbeing. Future work can investigate how our findings can be applied to more complex models of personalising a tracker experience.

ACKNOWLEDGMENTS

This work was supported by Utrecht University's Focus Area Sport and Society. We acknowledge the support of the Leibniz ScienceCampus Bremen Digital Public Health (lsc-diph.de), which is jointly funded by the Leibniz Association (W4/2018), the Federal State of Bremen and the Leibniz Institute for Prevention Research and Epidemiology—BIPS.

REFERENCES

[1] Carole Adam and Benoit Gaudou. 2016. BDI agents in social simulations: a survey. *The Knowledge Engineering Review* 31, 3 (2016), 207–238.

[2] Elena Agapie, Daniel Avrahami, and Jennifer Marlow. 2016. Staying the Course: System-Driven Lapse Management for Supporting Behavior Change. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1072–1083.

[3] Elena Agapie, Bonnie Chinh, Laura R. Pina, Diana Oviedo, Molly C. Welsh, Gary Hsieh, and Sean Munson. 2018. Crowdsourcing Exercise Plans Aligned with Expert Guidelines and Everyday Constraints. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 324, 13 pages. https://doi.org/10.1145/3173574.3173898

[4] Herman Aguinis and Kyle J Bradley. 2014. Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods* 17, 4 (2014), 351–371.

[5] Jakim Berndsen, Barry Smyth, and Aonghus Lawlor. 2019. Pace My Race: Recommendations for Marathon Running. In *Proceedings of the 13th ACM Conference on Recommender Systems* (Copenhagen, Denmark) *(RecSys '19)*. ACM, New York, NY, USA, 246–250. https://doi.org/10.1145/3298689.3346991

[6] Reuben Binns, Max Van Kleek, Michael Veale, Ulrik Lyngs, Jun Zhao, and Nigel Shadbolt. 2018. 'It's Reducing a Human Being to a Percentage': Perceptions of Justice in Algorithmic Decisions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 377.

[7] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. *Qualitative HCI research: Going behind the scenes*. Vol. 9. Morgan & Claypool Publishers. 1–115 pages.

[8] Dena M Bravata, Crystal Smith-Spangler, Vandana Sundaram, Allison L Gienger, Nancy Lin, Robyn Lewis, Christopher D Stave, Ingram Olkin, and John R Sirard. 2007. Using pedometers to increase physical activity and improve health: a systematic review. *Jama* 298, 19 (2007), 2296–2304.

[9] Catherine B Chan, Daniel AJ Ryan, and Catrine Tudor-Locke. 2004. Health benefits of a pedometer-based physical activity intervention in sedentary workers. *Preventive medicine* 39, 6 (2004), 1215–1222.

[10] Sunny Consolvo, Predrag Klasnja, David W McDonald, and James A Landay. 2009. Goal-setting considerations for persuasive technologies that encourage physical activity. In *Proceedings of the 4th international Conference on Persuasive Technology*. ACM, 8.

[11] Henriette Cramer, Vanessa Evers, Satyan Ramlal, Maarten Van Someren, Lloyd Rutledge, Natalia Stash, Lora Aroyo, and Bob Wielinga. 2008. The effects of transparency on trust in and acceptance of a content-based art recommender. *User Modeling and User-Adapted Interaction* 18, 5 (2008), 455.

[12] Maartje MA De Graaf and Bertram F Malle. 2017. How people explain action (and autonomous intelligent systems should too). In *2017 AAAI Fall Symposium Series*.

[13] Daniel Clement Dennett. 1989. *The intentional stance*. MIT press.

[14] Mary T Dzindolet, Scott A Peterson, Regina A Pomranky, Linda G Pierce, and Hall P Beck. 2003. The role of trust in automation reliance. *International journal of human-computer studies* 58, 6 (2003), 697–718.

[15] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In *23rd International Conference on Intelligent User Interfaces*. ACM, 211–223.

[16] Daniel A. Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A. Munson. 2016. Beyond Abandonment to Next Steps: Understanding and Designing for Life After Personal Informatics Tool Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, New York, NY, USA, 1109–1113. https://doi.org/10.1145/2858036.2858045

[17] Daniel A Epstein, An Ping, James Fogarty, and Sean A Munson. 2015. A lived informatics model of personal informatics. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 731–742.

[18] Ayelet Gal-Oz and Oren Zuckerman. 2015. Embracing Cheating in Gamified Fitness Applications. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. ACM, 535–540.

[19] gfk. 2018. Global Studies – Fitness tracking. http://www.gfk.com/global-studies/global-studies-fitness-tracking/

[20] R Gouveia, E Karapanos, and M Hassenzahl. 2015. Activity trackers in the wild: Design strategies and their impact on engagement and physical activity in daily life. *M-ITI Technical Report* (2015).

[21] Rebecca Gulotta, Jodi Forlizzi, Rayoung Yang, and Mark Wah Newman. 2016. Fostering Engagement with Personal Informatics Systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (Brisbane, QLD, Australia) *(DIS '16)*. ACM, New York, NY, USA, 286–300. https://doi.org/10.1145/2901790.

2901803

[22] Steven R Haynes, Mark A Cohen, and Frank E Ritter. 2009. Designs for explaining intelligent agents. *International Journal of Human-Computer Studies* 67, 1 (2009), 90–110.

[23] Jonathan L Herlocker, Joseph A Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM, 241–250.

[24] John R Hollenbeck, Charles R Williams, and Howard J Klein. 1989. An empirical examination of the antecedents of commitment to difficult goals. *Journal of Applied Psychology* 74, 1 (1989), 18.

[25] Kristina Höök. 1998. Evaluating the utility and usability of an adaptive hypermedia system. *Knowledge-Based Systems* 10, 5 (1998), 311–319.

[26] John M Jakicic, Kelliann K Davis, Renee J Rogers, Wendy C King, Marsha D Marcus, Diane Helsel, Amy D Rickman, Abdus S Wahed, and Steven H Belle. 2016. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. *Jama* 316, 11 (2016), 1161–1171.

[27] Richard D Johnson. 2005. An empirical investigation of sources of application-specific computer-self-efficacy and mediators of the efficacy—performance relationship. *International Journal of Human-Computer Studies* 62, 6 (2005), 737–758.

[28] Amy K. Karlson, A.J. Bernheim Brush, and Stuart Schechter. 2009. Can I Borrow Your Phone?: Understanding Concerns when Sharing Mobile Phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. ACM, New York, NY, USA, 1647–1650. https://doi.org/10.1145/1518701.1518953

[29] Rohit Ashok Khot, Deepti Aggarwal, Ryan Pennings, Larissa Hjorth, and Florian'Floyd' Mueller. 2017. Edipulse: investigating a playful approach to self-monitoring through 3D printed chocolate treats. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 6593–6607.

[30] Rohit Ashok Khot, Larissa Hjorth, and Florian'Floyd' Mueller. 2014. Understanding physical activity through 3D printed material artifacts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3835–3844.

[31] Francisco Kiss, Konrad Kucharski, Sven Mayer, Lars Lischke, Pascal Knierim, Andrzej Romanowski, and Paweł W. Wozniak. 2017. RunMerge: Towards Enhanced Proprioception for Advanced Amateur Runners. In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems* (Edinburgh, United Kingdom) *(DIS '17 Companion)*. Association for Computing Machinery, New York, NY, USA, 192–196. https://doi.org/10.1145/3064857.3079144

[32] René F Kizilcec. 2016. How much information?: Effects of transparency on trust in an algorithmic interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 2390–2395.

[33] Howard J Klein, Michael J Wesson, John R Hollenbeck, and Bradley J Alge. 1999. Goal commitment and the goal-setting process: conceptual clarification and empirical synthesis. *Journal of applied psychology* 84, 6 (1999), 885.

[34] Joanna Korman and Bertram F Malle. 2016. Grasping for traits or reasons? How people grapple with puzzling social behaviors. *Personality and Social Psychology Bulletin* 42, 11 (2016), 1451–1465.

[35] Joyca Lacroix, Privender Saini, and Roger Holmes. 2008. The relationship between goal difficulty and performance in the context of a physical activity intervention program. In *Proceedings of the 10th international conference on Human computer interaction with mobile devices and services*. ACM, 415–418.

[36] Pat Langley, Ben Meadows, Mohan Sridharan, and Dongkyu Choi. 2017. Explainable agency for intelligent autonomous systems. In *Twenty-Ninth IAAI Conference*.

[37] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 557–566.

[38] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*. Springer, 261–278.

[39] Edwin A Locke and Gary P Latham. 2002. Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist* 57, 9 (2002), 705.

[40] Edwin A Locke and Gary P Latham. 2013. *New developments in goal setting and task performance*. Routledge.

[41] Bertram F Malle. 1999. How people explain behavior: A new theoretical framework. *Personality and social psychology review* 3, 1 (1999), 23–48.

[42] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods* 44, 1 (2012), 1–23.

[43] John McClure and Denis J Hilton. 1998. Are goals or preconditions better explanations? It depends on the question. *European Journal of Social Psychology* 28, 6 (1998), 897–911.

[44] D Harrison Mcknight, Michelle Carter, Jason Bennett Thatcher, and Paul F Clay. 2011. Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS)* 2, 2 (2011), 12.

[45] Alistair Morrison and Viktor Bakayov. 2017. Stickers for steps: a study of an activity tracking system with face-to-face social engagement. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017).

[46] Sean A Munson and Sunny Consolvo. 2012. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. *PervasiveHealth* 2012 (2012), 2532.

[47] Jasmin Niess and Pawei W Wozniak. 2018. Supporting Meaningful Personal Fitness: the Tracker Goal Evolution Model. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 171.

[48] Gabriele Paolacci, Jesse Chandler, and Panagiotis G Ipeirotis. 2010. Running experiments on amazon mechanical turk. (2010).

[49] Ronald F Piccolo and Jason A Colquitt. 2006. Transformational leadership and job behaviors: The mediating role of core job characteristics. *Academy of Management journal* 49, 2 (2006), 327–340.

[50] Pearl Pu and Li Chen. 2006. Trust building with explanation interfaces. In *Proceedings of the 11th international conference on Intelligent user interfaces*. ACM, 93–100.

[51] Emilee Rader, Kelley Cotter, and Janghee Cho. 2018. Explanations as Mechanisms for Supporting Algorithmic Transparency. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 103.

[52] Jens Riegelsberger, M Angela Sasse, and John D McCarthy. 2005. The mechanics of trust: A framework for research and design. *International Journal of Human-Computer Studies* 62, 3 (2005), 381–422.

[53] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2015. Pass the ball: enforced turn-taking in activity tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 2417–2426.

[54] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2015. Pass the Ball: Enforced Turn-Taking in Activity Tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 2417–2426. https://doi.org/10.1145/2702123.2702577

[55] Julian B Rotter. 1980. Interpersonal trust, trustworthiness, and gullibility. *American psychologist* 35, 1 (1980), 1.

[56] Kim Sauvé, Steven Houben, Nicolai Marquardt, Saskia Bakker, Bart Hengeveld, Sarah Gallacher, and Yvonne Rogers. 2017. LOOP: A physical artifact to facilitate seamless interaction with personal data in everyday life. In *Proceedings of the 2017 ACM Conference Companion Publication on Designing Interactive Systems*. ACM, 285–288.

[57] Neil Selwyn. 2006. Digital division or digital decision? A study of non-users and low-users of computers. *Poetics* 34, 4-5 (2006), 273–292.

[58] Ben Shneiderman and Pattie Maes. 1997. Direct manipulation vs. interface agents. *interactions* 4, 6 (1997), 42–61.

[59] Ben R Slugoski, Mansur Lalljee, Roger Lamb, and Gerald P Ginsburg. 1993. Attribution in conversational context: Effect of mutual knowledge on explanation-giving. *European Journal of Social Psychology* 23, 3 (1993), 219–238.

[60] Katta Spiel, Fares Kayali, Louise Horvath, Michael Penkler, Sabine Harrer, Miguel Sicart, and Jessica Hammer. 2018. Fitter, Happier, More Productive?: The Normative Ontology of Fitness Trackers. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, alt08.

[61] Simone Stumpf, Vidya Rajaram, Lida Li, Margaret Burnett, Thomas Dietterich, Erin Sullivan, Russell Drummond, and Jonathan Herlocker. 2007. Toward harnessing user feedback for machine learning. In *Proceedings of the 12th international conference on Intelligent user interfaces*. ACM, 82–91.

[62] Ann M Swartz, Scott J Strath, David R Bassett Jr, J Brian Moore, Beth A Redwine, Maureen Groër, and Dixie L Thompson. 2003. Increasing daily walking improves glucose tolerance in overweight women. *Preventive medicine* 37, 4 (2003), 356–362.

[63] Lie Ming Tang, Jochen Meyer, Daniel A. Epstein, Kevin Bragg, Lina Engelen, Adrian Bauman, and Judy Kay. 2018. Defining Adherence: Making Sense of Physical Activity Tracker Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 37 (March 2018), 22 pages. https://doi.org/10.1145/3191769

[64] Nava Tintarev and Judith Masthoff. 2015. Explaining recommendations: Design and evaluation. In *Recommender systems handbook*. Springer, 353–382.

[65] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 143–146. https://doi.org/10.1145/1978942.1978963

[66] Jerry C Wofford, Vicki L Goodwin, and Steven Premack. 1992. Meta-analysis of the antecedents of personal goal level and of the antecedents and consequences of goal commitment. *Journal of Management* 18, 3 (1992), 595–615.

[67] Paweł Woźniak, Kristina Knaving, Staffan Björk, and Morten Fjeld. 2015. RUFUS: Remote Supporter Feedback for Long-Distance Runners. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Copenhagen, Denmark) *(MobileHCI '15)*. Association for Computing Machinery, New York, NY, USA, 115–124. https://doi.org/10.1145/2785830.2785893

[68] Rayoung Yang, Eunice Shin, Mark W. Newman, and Mark S. Ackerman. 2015. When Fitness Trackers Don'T 'Fit': End-user Difficulties in the Assessment of Personal Tracking Device Accuracy. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Osaka, Japan) *(UbiComp '15)*. ACM, New York, NY, USA, 623–634. https://doi.org/10.1145/2750858.2804269