Would I Lie to You - Would You Notice?

Felix Huppert(✉), Matthias Kranz, and Gerold Hoelzl

University of Passau, Passau, Germany
{felix.huppert,matthias.kranz,gerold.hoelzl}@uni-passau.de

Abstract. The quantified self-paradigm is well established. Its main purpose is to use numbers from sensors to derive self-knowledge. The massive availability of persuasive technology to monitor physiological parameters of humans made the paradigm available to a tremendous number of people. A multitude of different hard- and software platforms are available at the market. They all have different properties at different levels of quality. All in common is their promise to provide accurate and precise data about the humans' physiological condition and performed activities. Basically, they all provide a tool to make people aware of formerly hidden, non-observable, body signals. The gained awareness can then be used by people to e.g. improve their health or fitness level. In this work, we emphasize the perception of the gathered sensory data by the people. We focus on the question of how the trustworthiness of the recorded and presented data is perceived by people. As a fact, non-credible data can be understood by the user as being trustworthy and can have a negative impact on users' behavior. This can be especially critical for human's health in the fitness and medical application domain. It is of high importance to understand how people perceive and correlate their intrinsic body feelings with the data collected and presented by a mobile smart device like a smart watch or a fitness tracker.

Keywords: Fitness tracking · User perceived credibility · Quantified self · Trust in data

1 Introduction

With the rise of consumer-targeted ubiquitous computing technology over the last decade, significant advances in self-tracking and self-monitoring of physical activities and hidden body parameters (e.g. heart-rate and step count) to optimize personal health behaviors have been achieved. Sensing solutions for essential tracking parameters have been implemented into a wide range of affordable everyday pervasive devices such as fitness trackers, smart watches or the omnipresent smart phones. The fundamental concept of any kind of activity tracker device or application can be paraphrased by three essential steps: (I) collect activity related physical sensor measurements, (II) process and analyze gathered measurements to gain semantically abstracted data and (III) provide comprehensible feedback to the user about the tracked activity. In line with the personal informatics and quantified self context, the available feedback from tracking devices and applications

Y. Yin et al. (Eds.): MobiCASE 2019, LNICST 290, pp. 230–243, 2019.
https://doi.org/10.1007/978-3-030-28468-8_17

is intended to be used to reflect on current activity patterns, monitor the progress towards a pursued long-term behavior adaptation or goal and provide motivational support throughout a change process.

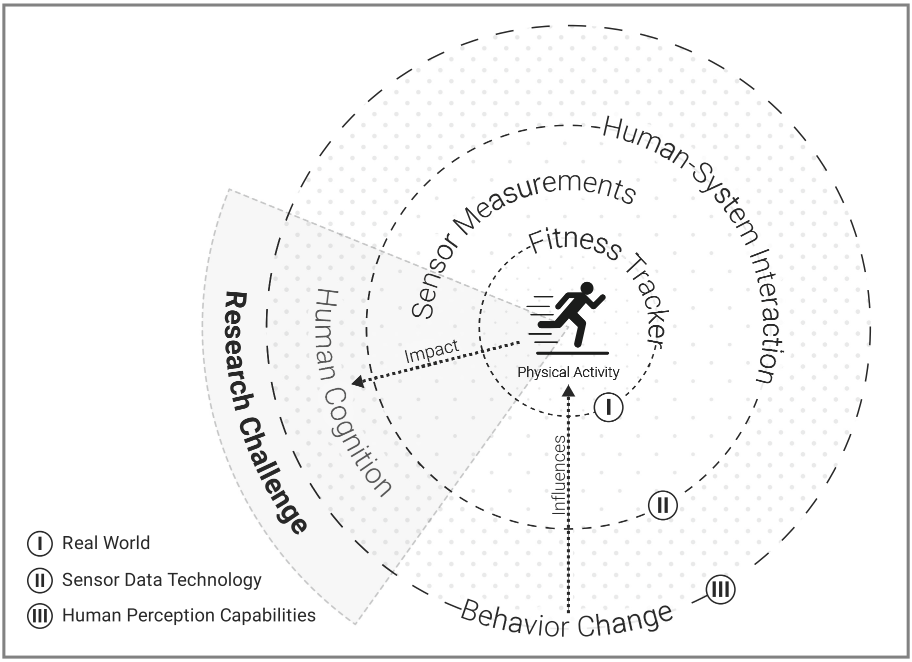

Fig. 1. Layered model, visualizing the abstract activity tracking principle, its dependencies and connectedness between the layers and the related research areas.

To provide a better overview of the essential fitness tracker systems structure, we conceptualized a layered model (cf. Fig. 1). The model visualizes the transitions and relations between the physical activity, sensor data collection and processing, human-computer interaction, psychological influences and impact on the human behavior in the context of activity tracking. Altogether, the model represents a self-regulating circle, where the measured physical activity is followed by an eventual behavior change induced by the feedback from the fitness tracker device, which will have an influence on future activities. The layers and transitions are abstract representations of the related research areas that are omnipresent in the activity tracking context. Research tends to be focused around problems and questions within one of these areas of interest (e.g. data science, psychology or human-computer interaction). Limited multidisciplinary research has been done, where possible transboundary effects - tracker data quality and its impact on human behavior - were evaluated.

As highlighted in the layer model (cf. Fig. 1), we focus on incorporating a wider range of research topics into one combined research effort to investigate the perceived trustworthiness and user expertise in the context of fitness tracking systems and the possible impact of data quality on human behavior. Based on these research intentions, we outlined the research question as follows: *Does a significant variation of pedometer feedback data accuracy from a fitness tracker have an influence on the tracking system credibility perceived by the user?* The research challenge is highlighted in the layer model (cf. Fig. 1) and can be outlined as followed: The physical activity in the research challenge represents the focus on pedometer walking tracking. The sensor measurement and data progressing includes not only the data logging, but also the systematical variation of the data validity in order to assess the research question. The human-system interaction interface presents the system feedback to the user in an understandable way. On the most outer layer of the model, the human cognition represents the research on the perceived system credibility and influences on the users.

The aim of our research is to provide a first insight and impression of the correlation between accuracy of fitness tracker measurements and user-perceived system credibility. In this paper, we present a study setup for the evaluation of the expected influences of data validity on system credibility. We conduct an exploratory, out-of-the-lab study to evaluate and discuss the system and its perceived impact on the user-perceived credibility.

2 Related Work

The fundamental idea to use pedometers to quantify the human physical activity has been established in the 1980s. The intention behind these first generation devices was similar to todays product goals, where step count data would be presented to the users to allow them to reflect on their daily activity and function as a motivational source to be more active [9,17,18]. Due to the primitive implementation and limited range of functions, the general public quickly lost interest and trust in the rudimentary mechanical tracker devices [6,27]. With the rise of smart-devices in the late 2000s, a new era of activity tracking devices was started. The increase in processing power and availability of better sensor technology allowed for complex algorithms and real time evaluation of the movement data, which resulted in more accurate and precise pedometer tracking [27]. These advancements in persuasive technology systems have lead to an increased interest in the related research areas [3,9].

Alongside persuasive fitness tracker devices emerged the personal informatics systems terminology. The description was first brought up by Li et al. in 2010 and was defined as "[Systems that] ... help people collect personally relevant information for the purpose of self-reflection and gaining self-knowledge" [17]. This concept references back to the basics of the quantified self concept. Wolf et al. defined the quantified self movement as the integration and acceptance of continuous data collection technology in the everyday lifestyle [31]. Wrist-worn fitness trackers are a one prime example out of many for a quantified self- and personal informatics systems.

Pedometer measurements represent an integral part of human activity tracking approaches. This common feature integration is justified with the importance of steps in the fundamental human activity [3,29]. The most dominant physical actions typically carried out throughout a day can be associated with taking steps. Steps are objective, intuitive and comprehensible in the context of understanding personal activeness, which makes this measure ideal for humans to reflect on their own physical activity [3]. Most other common fitness tracker measures (e.g. flight of stairs, active minutes, calories burned, etc.) are derived from the pedometer count [8,12,13]. In cross-sectional studies, the negative correlation between steps taken per day and common health issues has been proven. Active individuals, who achieve a higher than average step count per day, were identified to be less likely to have health related issues in the future [3,5,15,30].

Research studies confirm that both accuracy and precision of activity tracking devices have been steadily improved over the last decade [3,8,9,14,16,25,26]. A trustworthy tracking system should offer accurate and precise data transformation under all circumstances to avoid misleading customers [28]. Controlled lab studies where participants walk, jog or run on a treadmill with different testing setups are used by researchers to evaluate the performance of consumer fitness tracker devices [16,25]. The results reveal an optimal accuracy and precision for a typical walking pace of 2.5 mph for most trackers. Faster or slower walking speeds would lead to devices consistently under- or overestimating the step count by 3.5% to 10% [16]. Inaccuracies and precision of the tested devices are consistent and independent of the total steps taken during the study [8]. A high correlation between the results of the examined controlled and field studies was shown, indicating that the findings of lab studies are valid in a real-world setting [9].

Motivation is a key factor especially for physical activity, since unmotivated humans tend to lose interest in their goal [17]. Fitness tracker represent an extrinsic motivation source [1]. Further, motivation is an essential part of the biopsychosocial transtheoretical model [22], which outlines the general process of intentional behavior change. The five stages of change (I Precontemplation, II Contemplation, III Preparation, IV Action and V Maintenance), the influence of decisional balance and the connection to Bandura's self-efficacy theory were first described by Prochaska et al. [23]. The decisional balance describes the user-perceived balance between benefits and drawbacks of the behavior change throughout the stages. Fitness tracker and the common related motivational methods (e.g. gamification, goal setting or social integration) can help shift and maintain a positive decisional balance and support the user to gain higher self-efficacy [10,11,13,23].

The general acceptance and cease of usage motives of the tracker devices has been extensively covered in survey studies [7,19,20,32]. Besides technical difficulties, lack of motivation and support is one of the most common reasons, why users tend to abandon their fitness tracker devices [19]. Furthermore, the perceived trustworthiness of fitness trackers seems to be linked to the long-term user acceptance [24,28]. Presenting the user knowingly false data samples increases the mistrust and decrease the motivation, thus leading users to stop using their fitness tracker. Both motivation and trust in the technology have a significant impact on users pursuing to use fitness tracker devices. The quality of all the gathered activity data has to be interpreted with caution [8,28]. Long-term adaption of fitness trackers and the related behavior change has been assessed in a wide range of studies [20,32]. Researcher focused on HCI or human behavior influences, pay little to no tribute to the data quality of the underlying measurement trackers. The impact of false data through inaccuracies on the data reflection process has been untended in this research field.

3 User Study Concept

Our hypotheses for the user study was subject to following wording: *An influence on the pedometer data feedback accuracy has no significant effect on the system credibility perceived by the user when compared to a neutral, unaltered control group.* The chosen study concept was inspired by related work studies and was a combination of survey and long-term field experiments. The controllability and high accuracy means of lab studies were judged to be disadvantageous in the given context, since the unfamiliar surrounding conditions might influence the perception of the study participants (e.g. suspect tracker accuracy is tested). A field study shifts the focus point away from tracker and offers a more realistic application case, which results in higher external validity for gathered data [2].

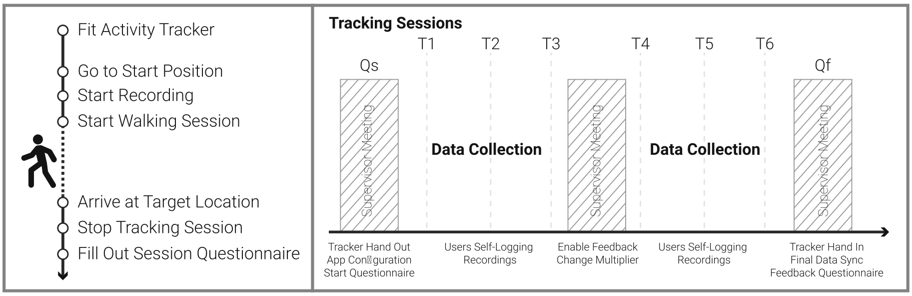

Fig. 2. The left side illustration depicts the study procedure for one data recording session. The right side graph gives an overview of the complete study schedule that spans over a couple of weeks.

For the two-phase, self-guided, longitudinal study design, the participants were instructed to collect regular pedometer data on their walking commute route. An overview of the study schedule is shown in Fig. 2. Initially, a meeting with all willing participants was held to collect pre-study data with a short questionnaire, setup the fitness trackers and provide instructions to the participants. In the first data collection phase, baseline pedometer data sets ($3-4$ per participant) without any provided feedback were recorded. These baseline measurements were recorded to verify the validity and reliably of the used pedometer tracker. After an intermediate meeting with the study supervisor, where additional guidance and an update of the tracking system was provided, another set of commute recordings was collected. In these following sessions, altered feedback values were shown to the participants at the end of each recorded walk. Dependent on the study group, a negative (−15%), neutral (0%) or positive (+15%) alteration multiplier was used. These manipulation values were chosen to provide a significant difference between the baseline and shown measurements without exaggerating the effect, therefore making the changes too obvious. After the second phase had been completed, in a final meeting, the subjective perceived data and tracker credibility was evaluated with a questionnaire.

The procedure for one run of the data collection is illustrated in Fig. 2. After the tracker has been fitted to the wrist of the non-dominant hand, the study participant walks to the starting point on their daily commute route. At this defined location, the recording process is started with our Android control application and the regular commute walk can begin. Once arrived at the target location, the data collection is stopped and the session questionnaire is filled out from inside the app. Both the start and finish locations (e.g. front door, street sign or building) were self-selected by the study participants to provide fixed points of reference for the data collection.

4 Activity Tracking System

Today, the most common consumer-grade fitness tracker are wrist-worn bracelets. Since these rudimentary activity trackers on their own have limited functionality and typically come with a additional smart phone application, we decided to pair the selected tracker with our own Android app. A large variety of purpose-built bracelets are available and most common devices (e.g. Fitbit, Garmin or Polar) have been reviewed or used in some related research [8,16,25,26].

For our study intentions, the Mi Band 2 (cf. Fig. 3) bracelet by Xiaomi was ideal, since it is a relatively popular, low-cost and reliable fitness tracker bracelet. It features an accelerometer based pedometer, photoelectric heart rate sensor, small display and one touch button for user interaction in a small robust wrist-worn package. In a comparative study of 17 different activity trackers in 2015, the first generation Mi Band tracker scored well [8]. The Mi Band achieved an average pedometer accuracy of 96.56% and a variation coefficient of 5.81% across the three study setups (200, 500 and 1000 step trials). It was ranked among the top 5 of the compared devices, which included trackers from many renowned brands. The Mi Band tracker was recommended: "[The] ... Xiaomi Mi Band showed the best package compared to its price." [8].

Access to the pedometer tracker data can be gained with the bluetooth API, which has been reverse engineered by the open source community for Android smart phones. The custom designed application (cf. Fig. 3) was used to (I) start

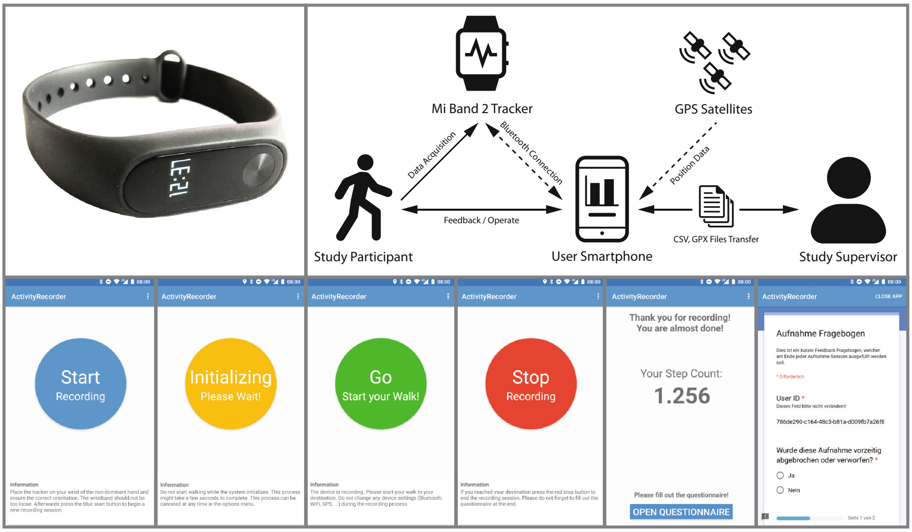

Fig. 3. Illustration of the architecture to record and transfer the collected data from the Mi Band 2 to a central repository of the study supervisor, and to present the user the calculated step count. In addition to the data transfer, a questionnaire to evaluate the run and the perceived step count was implemented.

and stop the recordings sessions, (II) backup the Mi Band data, (III) track GPS location, (IV) show the manipulated feedback score to the user and (V) to fill out the session questionnaires. Both the pedometer and GPS tracking data is stored locally and the latest recordings are transferred over the internet to the study supervisor after a session has been completed. The goal of this implemented system was to have the study participants do the recordings on a self-reliant basis, whilst the remote study supervisor can maintain full control over the study conduction and settings [21].

5 Evaluation

For the user study, eleven student participants from non-technical areas of study were recruited from a selected pool. The mean age was 22.27 years and a median of 23 years with 18.18% being male and 81.81% female. Most participants (90.90%) were right-handed and 81.10% regularly wore a watch-like device (e.g. (smart) watch, fitness tracker). The pre-test questionnaire indicated a predominant interest (81.8%) in technology and 54.5% already use some sort of tracking application or device. In total, six participants completed the all data collection sessions over a duration of four weeks. These six individuals were split evenly into the three test groups (negative, neutral and positive feedback alteration) for the second phase of the user study.

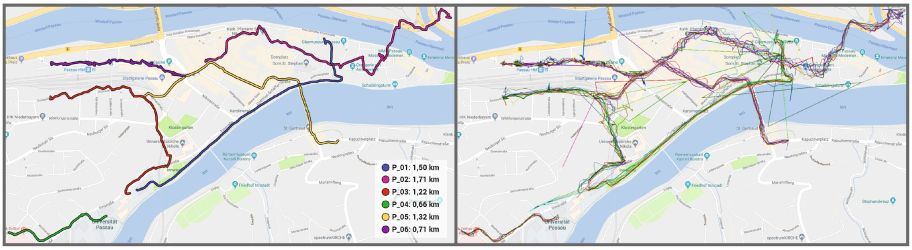

Fig. 4. These maps show the GPS tracks from the pathways the participants recorded during the study. The location data in the left map has been cleaned from outliers and slightly smoothed, while in the maps on the right the raw location data is illustrated.

Due to the self-reliant study design, the GPS position was recorded besides the pedometer data to ensure comparable data sets, free of larger deviations from the typical commute route. All location based data was slight smoothed and larger outliers and measurement glitches removed. The raw and processed GPS data is illustrated in Fig. 4. Over the course of four weeks, a total of 40 usable recordings were collected by the six active participants. No significant deviations were detected in the data set. A descriptive evaluation of the GPS data is listed in Table 1.

Table 1. Descriptive statistics of the recorded GPS data from the commute walks.

Participant	P_01	P_02	P_03	P_04	P_05	P_06
Distance	1.50 km	1.71 km	1.22 km	0.66 km	1.32 km	0.71 km
Time	16:18 min.	19:14 min.	13:11 min.	6:40 min.	14:01 min.	11:23 min.
Speed	5.52 km/h	5.30 km/h	5.55 km/h	5.94 km/h	5.65 km/h	3.74 km/h

The key point of interest is the manipulated pedometer data and the user-perceived system credibility score. The averaged pedometer measures for all six participants are listed in Table 2. The base step count represents the averaged raw step count. The calculated modified data is based on the study group (negative, neutral or positive) for the between subject user study test. The intention of the baseline data collection was to show that the Mi Band 2 has a high measurement reliability. With a repeatability error, which ranges between 2.38% and 3.58% (cf. Fig. 5), the Mi Band 2 produces precise pedometer data comparable with other trackers tested in related work lab studies [8].

The positive and negative 15% step count variation should have presented the study participants with a value that is significantly lower or higher than the baseline step count, but not too large of a deviation to give away the research intention. A paired sample t-test was conducted on the pedometer data measurements to confirm the statistical significant difference. The results of the t-test

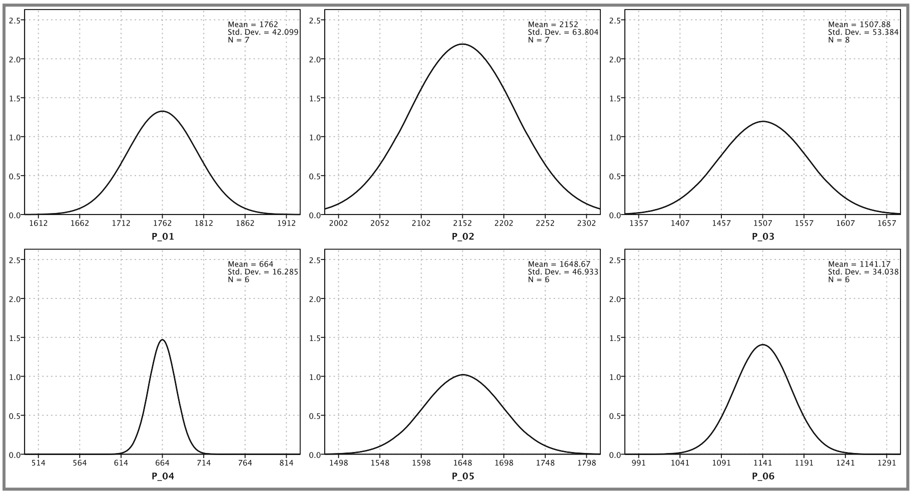

Fig. 5. Deviation of steps counts per study participant during the baseline data collection. Knowing the deviation from the users path, it was possible to deduce the +/− 15% manipulation threshold for the real value that was presented to the user.

(cf. Table 2) indicated a significant difference between the baseline and manipulated pedometer feedback for the positive and negative study groups. For the neutral control group no significant distinction was identified, which was to be expected since the data was not altered.

Table 2. The significance of the difference between the averaged baseline and manipulated pedometer data is evaluated in a t-test and the precision of the baseline data set is shown in this table.

Participant	P_01	P_02	P_03	P_04	P_05	P_06
Manipulation	Positive	Negative	Neutral	Neutral	Negative	Positive
Base step count	1.762	2.152	1.507	664	1.648	1.141
Mod. step count	1.992	1.846	-	-	1.427	1.317
Base Precision	2.38%	2.96%	3.54%	2.84%	2.45%	2.98%
Sig. test	1.9%	1.8%	80.5%	92.5%	1.2%	4.1%

A post-study questionnaire and short interview session was carried out to evaluate the credibility and perceived accuracy of the used fitness tracker. The likert scale questions were derived from related survey research work. We used a one to five scale, where a one indicates a negative denotation and a rating of five presents a positive statement towards the asked question. This answer scale is reflected in the result representation in Fig. 6. Negative answers are color coded with red color nuances and positive statements are denoted with green nuances. The amount of fill of each circle represents the percentage split of people that gave that respective answer. Figure 6 clearly presents the overall consenting appraisal of the survey questions.

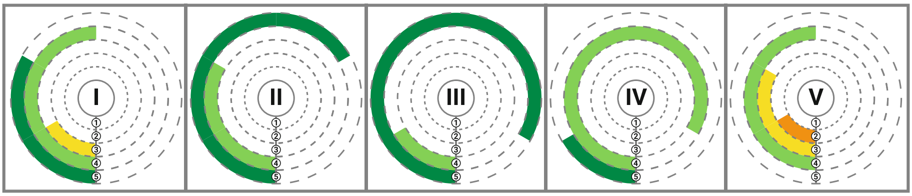

Fig. 6. Visualization of the results from the post-study likert scale questionnaire, where the participants had to evaluate (I) relevance of the feedback, (II) expected value match, (III) trust in the fitness tracker, (IV) perceived accuracy of the measurements and (V) possibility of influence. (Color figure online)

The overall personal relevance (I) of the step feedback was rated above average (median 4), which is an indication that the participants interpret the shown feedback as valid. The impact (V) of the pedometer feedback on behavior change was judged as possible. Furthermore, some participants mentioned in the interview that a lower feedback value would increase their personal interest to be more active. These statements indicate that the fitness tracker would be used as intended to monitor the daily activity and provide progress feedback towards being more active.In regard to the expected step count value (II), 83% mentioned an exact match between the presented feedback value and their expectations. One participant pointed out slightly higher pedometer feedback values, even though the individual was in the negatively influenced study group.

Five out of the six participants evaluated the presented pedometer score as accurate (IV) and one rated it to be very accurate. One participant mentioned in the interview that the used wrist-worn tracker provided "better" accuracy than other prior tested methods. This positive statement was also emphasized in two other interviews, where the shown feedback of the Mi Band 2 was judged to be more accurate than tracking methods already used by some participants.

The credibility (III) of the fitness tracker used during the user study was assessed as very credible by all six participants, indicating a high trust in the Mi Band and its measurement capabilities.

Overall, the trust and credibility evaluation provides coherent results that indicate a high trust in the data validity regardless of the manipulated feedback.

6 Result Discussion

Relating back the Fogg et al. definition of credibility in the context of computer systems [10], where credibility was defined as trustworthiness and expertise of a system. The user study participants indicated an overall knowledgeable system, since their confidence in the data measurements and validity of the presented feedback values was high. Trustworthiness was defined by Fogg et al. as the perceived goodness of system. Regarding the trust for our given tracking system, the participants judged the used fitness tracker as very reliable and trustworthy in the questionnaire and interview. Based on the general computer credibility definition and the clear trends in the study data, the Mi Band 2, despite the added inaccuracies, was rated as credible. As a conclusion to the user study, the outlined research question is answered: A significant variation of pedometer feedback data accuracy does not seem to have a significant influence on the perceived system credibility.

The Mi Band 2 tracker provided repeatable step count figures with an overall precision of 2.85%. The intentional manipulated pedometer feedback values were significantly higher or lower by 15%, depended on the study group. From the evaluation of the post-study questionnaire and the interview session, it is clear that the participants had little to no awareness of the intentional induced data inaccuracies. This was further underlined by the statements that the provided wrist-worn fitness tracker was more accurate and trustworthy than other already used smartphone applications, since the Mi Band 2 is a "purpose-build" device. The difference between the independent user measurements with a smartphone and the shown feedback from our Mi Band setup were not further questioned and seemingly had no influence on the credibility rating. The statistical assessment indicated no significant difference in the perceived system trust and accuracy between the neutral and groups with manipulated feedback.

In related work, authors [8,9,28] expect users to think critical about tracking technology and the data quality. One of the key factors for long-term behavior change through the usage of fitness tracker systems is the data reflection process [17]. Users reflect on their hourly, daily or weekly achievements and try to adapt their behavior. Our presented study shows that especially inexperienced fitness tracker users can not associate well between their real-life activity and the presented feedback values, even for short-term tracking. They did not reminisce precisely about details of the logged activity. The verification of the data validity and reliability was assumingly based more on the overall user experience with the tracker rather than the perceived walking activity during the recorded sessions. The users seem to lack a understanding of the correlation between these data values. Inexperienced, first time tracker users seem to take the system credibility for granted, until some major inaccuracies raises concerns [4,9,11]. This almost careless attitude about fitness tracker technology and the accompanying possibility of false feedback might have a significant impact on the intended long-term behavior change.

7 Conclusion

Would I Lie to You - Would you Notice? - A question we can definitely answer with No, you wouldn't notice.

Based on the fact that activity trackers only provide very abstract data (e.g. step count or minutes active) it is nearly impossible, even for the informed users to judge if the values presented by the fitness trackers are trustworthy or not. End users have next to no possibilities to verify the validity of their recorded fitness- or health data. The tracker always has more information and dependent on the quality of the algorithms, or the intent of the App developers, can present even statistically significantly changed values, without the user being able to identify them as being not trustworthy. This problem domain can be seen as closely related to the market of lemons paradigm. Users have to see their smart devices as a black box, thus have no control over the data processing and feedback generation.

In a first explorative study, we introduced a 15% offset in step count value for daily commute walks over the span of four weeks. Three study groups received either negative, neutral or positively influenced feedback values. The participants had to reflect on the shown pedometer feedback and rate the credibility of the shown values. The results indicated that the variation of the step count value had no significant impact on the user-perceived credibility and awareness, thus the users didn't notice the change. This is especially interesting as the overall walking distance was short and recallable by the users. We argue that although the walking distance was short, users were not able to judge the values correctly. This means in fact, the less recallable the activities they perform are, the more variation can be in the abstract data without them being able to judge if the the data is trustworthy or not.

Participants blindly trusted the wrist-worn fitness tracker devices and did not question the presented mismatched information at all, which was also confirmed in the post-study interview when the research intention was revealed. With this paper, we provide a first thought-provoking impulse to question the impact of fitness feedback data quality on the user-perceived trustworthiness. It is of highly importance to understand that people have a high trust in technology and no way to proof if the data is correct or not. Especially in the fitness and health domain, it can expose risks and health issues to the users or influence them to change their behaviour based on wrong assumptions. Up until now, researchers, manufacturers and users assumed the abstract feedback data to be valid and easily usable for studies without prior detailed verification. Even users were adjudged the capability to spot inaccuracies and compensate for misleading information. In the conduced user study, we have demonstratively shown major flaws in the widespread fundamental belief that user can judge general data quality and fitness tracker credibility.

References

1. Asimakopoulos, S., Asimakopoulos, G., Spillers, F.: Motivation and user engagement in fitness tracking: heuristics for mobile healthcare wearables. In: Informatics, vol. 4, p. 5. Multidisciplinary Digital Publishing Institute (2017)
2. Attig, C., Franke, T.: I track, therefore i walk-exploring the motivational costs of wearing activity trackers in actual users. Int. J. Hum. Comput. Stud. **127**, 211–224 (2018)

3. Bassett, D.R., Toth, L.P., LaMunion, S.R., Crouter, S.E.: Step counting: a review of measurement considerations and health-related applications. Sports Med. **47**(7), 1303–1315 (2017)
4. Benedetto, S., Caldato, C., Bazzan, E., Greenwood, D.C., Pensabene, V., Actis, P.: Assessment of the Fitbit charge 2 for monitoring heart rate. PloS one **13**, e0192691 (2018)
5. Bravata, D.M., et al.: Using pedometers to increase physical activity and improve health: a systematic review. Jama **298**(19), 2296–2304 (2007)
6. Crouter, S.E., Schneider, P.L., Bassett, J.D.: Spring-levered versus piezo-electric pedometer accuracy in overweight and obese adults. Med. Sci. Sports Exerc. **37**(10), 1673–1679 (2005)
7. Day, S.: Self-tracking over time: the fitbit® phenomenon. In: The 7th Annual Conference of Computing and Information Technology Research and Education New Zealand (CITRENZ2016) and the 29th Annual Conference of the National Advisory Committee on Computing Qualifications, Wellington, New Zealand, pp. 1–6 (2016)
8. El-Amrawy, F., Nounou, M.I.: Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthc. Inform. Res. **21**(4), 315–320 (2015)
9. Evenson, K.R., Goto, M.M., Furberg, R.D.: Systematic review of the validity and reliability of consumer-wearable activity trackers. Int. J. Behav. Nutr. Phys. Act. **12**(1), 159 (2015)
10. Fogg, B.J.: Persuasive technology: using computers to change what we think and do. Ubiquity **2002**, 5 (2002)
11. Fritz, T., Huang, E.M., Murphy, G.C., Zimmermann, T.: Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 487–496. ACM (2014)
12. Gouveia, R., Pereira, F., Karapanos, E., Munson, S.A., Hassenzahl, M.: Exploring the design space of glanceable feedback for physical activity trackers. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 144–155. ACM (2016)
13. Harjumaa, M., Segerståhl, K., Oinas-Kukkonen, H.: Understanding persuasive software functionality in practice: a field trial of polar FT60. In: proceedings of the 4th International Conference on Persuasive Technology, p. 2. ACM (2009)
14. Hoelzl, G., Kranz, M., Schmid, A., Halbmayer, P., Ferscha, A.: Size does matter - positioning on the wrist a comparative study: smartwatch vs. smartphone. In: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 703–708, March 2017
15. Katzmarzyk, P.T., et al.: The international study of childhood obesity, lifestyle and the environment (ISCOLE): design and methods. BMC Public Health **13**(1), 900 (2013)
16. Leth, S., Hansen, J., Nielsen, O.W., Dinesen, B.: Evaluation of commercial self-monitoring devices for clinical purposes: results from the future patient trial, phase I. Sensors **17**(1), 211 (2017)
17. Li, I., Dey, A., Forlizzi, J.: A stage-based model of personal informatics systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 557–566. ACM (2010)
18. Liu, X., et al.: Characterizing smartwatch usage in the wild. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pp. 385–398. ACM (2017)

19. Maher, C., Ryan, J., Ambrosi, C., Edney, S.: Users' experiences of wearable activity trackers: a cross-sectional study. BMC Public Health **17**(1), 880 (2017)
20. Mercer, K., Li, M., Giangregorio, L., Burns, C., Grindrod, K.: Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR mHealth and uHealth **4**(2), e40 (2016)
21. Pilgram, N., Mohamed, A., Kranz, M., Hoelzl, G.: Biofeedback in the wild - a smartwatch approach. In: IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), pp. 312–317, March 2018
22. Prochaska, J.O., DiClemente, C.C.: Transtheoretical therapy: toward a more integrative model of change. Psychother. Theor. Res. Pract. **19**(3), 276 (1982)
23. Prochaska, J.O., DiClemente, C.C.: Toward a comprehensive model of change. In: Miller, W.R., Heather, N. (eds.) Treating addictive behaviors. ABBI, vol. 13, pp. 3–27. Springer, Boston (1986). https://doi.org/10.1007/978-1-4613-2191-0_1
24. Rupp, M.A., Michaelis, J.R., McConnell, D.S., Smither, J.A.: The impact of technological trust and self-determined motivation on intentions to use wearable fitness technology. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 60, pp. 1434–1438. SAGE Publications, Los Angeles (2016)
25. Sears, T., Alvalos, E., Lawson, S., McAlister, I., Eschbach, L.C., Bunn, J.: Wrist-worn physical activity trackers tend to underestimate steps during walking. Int. J. Exerc. Sci. **10**(5), 764–773 (2017)
26. Shcherbina, A., et al.: Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J. Pers. Med. **7**(2), 3 (2017)
27. Troiano, R.P., McClain, J.J., Brychta, R.J., Chen, K.Y.: Evolution of accelerometer methods for physical activity research. Br. J. Sports Med. **48**(13), 1019–1023 (2014)
28. Trommler, D., Attig, C., Franke, T.: Trust in activity tracker measurement and its link to user acceptance. Mensch und Computer 2018-Tagungsband (2018)
29. Tryon, W.W.: Activity Measurement in Psychology and Medicine. Springer Science & Business Media, Boston (2013)
30. Wattanapisit, A., Thanamee, S.: Evidence behind 10,000 steps walking. J. Health Res. **31**(3), 241–248 (2017)
31. Wolf, G.: Know thyself: Tracking every facet of life, from sleep to mood to pain (2009). https://www.wired.com/2009/06/lbnp-knowthyself/
32. Yang, R., Shin, E., Newman, M.W., Ackerman, M.S.: 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, pp. 623–634. ACM (2015)