Accuracy and Reliability of Personal Data Collection: An Autoethnographic Study

Amon Rapp
University of Torino
Computer Science Department
C.so Svizzera, 185, Torino Italy
amon.rapp@gmail.com

Alessandro Marcengo
Telecom Italia
Via Reiss Romoli, 274
Torino Italy
alessandro.marcengo@telecomitalia.it

Federica Cena
University of Torino
Computer Science Department
C.so Svizzera, 185, Torino Italy
cena@di.unito.it

ABSTRACT
Accuracy of self-tracking devices is a key problem when dealing with personal data. Different devices may result in different reported measure, and this may impact on the users' perceived reliability of the devices they used. We conducted an autoethnography to investigate how different devices collect data on specific parameter in order to highlight discrepancies in the measures reported. Results highlight that designers should account for the variability of activities that users may face during their daily practices, as each of them may impact on the device's capability of collecting accurate data.

CCS Concepts
• **Human-centered computing→Human computer interaction.**

Keywords
Personal informatics; Quantified Self; Personalization; Autoethnography.

1. INTRODUCTION
Personal Informatics systems are currently appealing a large number of users, spreading beyond the traditional user group of Quantified Selfers [6]. Quantified selfers have a deep knowledge of tracking technologies, finding solutions for the possible barriers that they may encounter during the data collection and management. However, this is not true for all those people that are interested and curious toward Personal Informatics, and may try this kind of technologies for the first time [8].

One of the issue that this new user base may encounter is related to the bewilderment induced by the different possibilities of tracking the same parameter. Thanks to the spreading of multiple wearable devices for personal data collection, in fact users can now rely on different instruments to measure the same parameter. Each of them has its own physical structure, uses specific recognition algorithms and is addressed to be worn on certain part of the body: and all these elements may affect the reported measures and thus the data collected by the device. The differences in the data collected that may result from such a diversity might impact on the user's perceived accuracy of the gathered data and on the consequent perceived reliability of the instrument used.

We carried out a four-week autoethnographic study to investigate how different self-tracking tools may lead to different results in terms of the values of the collected data. The results of the study reveal that: i) the data collected for a specific target parameter were different depending on the tools used, and such difference was primarily due to the position in which these instruments were worn and the activities performed during the day by the ethnographer; ii) the discrepancies among the measures reported by the different tools impacted on their perceived reliability, pushing the ethnographer to seek strategies to account for the data collected.

2. RELATED WORK
Various research has studied how users perceive reliability and accuracy of self-tracking instruments. Kay et al. [3] found that users react negatively to the inaccuracies of their devices, while Lazar et al. [4] emphasized that they do care about the accuracy of the data collected, so that failing to produce accurate information is one of the main reason for abandoning a specific device.

Consolvo et al. [1] listed seven different types of errors that a fitness tracker device may produce during its daily use, such as exchanging one activity or another one, completely failing to detect an activity, or detecting an activity that was not occurred: this kind of errors produces frustration in users, directly impacting on the instrument's credibility. While Mackinlay [5] highlighted that users put to test their devices' accuracy, but often find difficulties in calibrating them due to the scarce visibility of their status. Finally, Yang et al. [9] outlined the various techniques that users use to evaluate trackers' accuracy, emphasizing the different perceptions that they may have of accuracy and reliability.

3. METHOD
We used autoethnography to individuate discrepancies among diverse trackers and analyze how they may affect the user's experience. This method considers the ethnographer's subjective experience worth to be analyzed and reported, valuable as that of the other individuals. The autoethnographer continuously observes herself to account for the reality she is interested to explain [2].

The second author self-examined the use of four different wearable devices to compare the data collected and eventually individuate criticalities due to discrepancies in their accuracy and/or reliability. The devices were chosen by taking into account the position in which they are worn, with the goal of exploring the differences in the gathered measures by them.

We selected: Withings Activité on the right wrist; Shine Misfits necklace; Sony SWR30 on the left wrist; GoogleFit application running background on a Sony Xperia Z3.

The hypothesis was that the recorded data would not be affected by the influence of the body positioning, all recording approximately the same data. The self-observation session was carried out for four weeks. We provide here a brief summary of the study findings pointing to Marcengo et al. [7] for a more detailed description.

4. RESULTS AND DISCUSSION

Sleep data analysis showed interesting problems related with the personal style of "going to sleep" in relation with the used device. For instance, the *sleep* total amount recorded by the Misfit Shine (necklace) is always higher of about thirty minutes. This point is due to the fact that the Shine considers the lying position as the user is already sleeping even if she's reading a book or watching her tablet in the bed. So the *sleep* total amount will always be increased by the activity performed before falling asleep. The device with the best accuracy results the one worn on the right wrist. This makes possible to distinguish the activities performed with the right hand while lying in the bed as something different from sleeping (for left-handed user the same principle will work for the left wrist).

Also *steps* showed interesting evidences and relations through life style and devices. The total *steps* amount is very biased by the interaction between the location on the body (if wearable) and the activities performed by the user. Indeed, considering the data collected by Withings Activité (on the right wrist) it is clear that if the user performed a lot of public talking on a specific day (meetings, showing slides, etc) *steps* becomes inclined towards high figures due to the gestures involved. Opposite results become evident according to different life circumstances. In particular data became surprisingly low for two conditions. The first one is when the user walk pushing a stroller. In this case the device does not log the alternate hanging of the hands and does not see the activity as walking. The second one occurs if the user carry a moderately heavy bag (e.g. a small suitcase) depending which hand holds the bag.

If the steps are collected by a phone app even more life situation distortions become evident toward low figures because of all the occasions when the phone is not on the body (e.g. weekend, sports, home, etc.). This, in a minor evident manner, is also true also for wearable devices. On the weekend all data appears distorted by incomplete or peculiar usage of the device due to different life activities (i.e. working in the garden, playing with kids, etc.).

From these evidences some needs of personalization in the design of logging devices and apps emerge. Manufacturers need to consider different designs for different life styles brought by different types of users with different life patterns (e.g. watching videos in the bed, walking with a stroller, carrying a bag, gesturing a lot, etc.). These patterns could be compressed into a few personas that can lead to different declinations of the same device or slightly different tracking algorithms on the same device. This personalization may be transferred directly into the user experience by collecting specific aspects and habits that impact on the accuracy of the logging system. In certain case should be possible also to advise the user about the best body location to wear the device in relation to her personal lifestyle.

5. CONCLUSION

Our study emphasizes the need of considering the idiosyncratic activities that users carry out during their daily practices in order to produce more accurate and thus reliable trackers. Activity recognition algorithms should be tailored to the specific habits of the single individual as these may be the main culprit for the inaccurate reporting of the target parameters. Personalization, thus, should be not only a matter of the services provided by the new personal informatics technologies, but also a key requirement for the design and implementation of the modalities for collecting the data.

6. REFERENCES

1. Consolvo, S., McDonald, D.W., Toscos, T., Chen, M. Y., Froehlich, J., Harrison, B., Klasnja, P., LaMarca, A., LeGrand, L., Libby, R., Smith, I., Landay, J. A.: Activity sensing in the wild: a field trial of ubifit garden. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08) 1797–1806 (2008)
2. Ellis, C., Bochner, A.,: Autoethnography, personal narrative, and personal reflexivity. In Handbook of qualitative research (2nd. ed.), Norman K. Denzinand Yvonna S. Lincoln (eds.). Sage, Thousand Oaks, CA, 733-768 (2000)
3. Kay, M., Morris, D., schraefel, mc, Kientz, J. A.: There's no such thing as gaining a pound: reconsidering the bathroom scale user interface. In: ACM international joint conference on Pervasive and ubiquitous computing (UbiComp '13), 401–410 (2013)
4. Lazar, A., Koehler, C., Tanenbaum, J., Nguyen, D.H.;Why we use and abandon smart devices. In: the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15). ACM, New York, NY, USA, 635-646 (2015)
5. Mackinlay, M.: Phases of Accuracy Diagnosis:(In) visibility of System Status in the Fitbit. Intersect: The Stanford Journal of Science, Technology and Society 6, 2 (2013)
6. Marcengo, A., Rapp, A.: Visualization of Human Behavior Data: The Quantified Self, in Huang L. H. and Huang, W. (Eds.) *Innovative approaches of data visualization and visual analytics*. IGI Global, Hershey, PA, 236-265 (2013)
7. Marcengo, A., Rapp, A., Cena, F. The Falsified Self: Complexities in Personal Data Collection. To appear in Proceedings of HCI International '16, Springer, 2016.
8. Rapp, A., Cena, F.: Self-monitoring and Technology: Challenges and Open Issues in Personal Informatics. In: HCI International. Universal Access in Human-Computer Interaction. Design for All and Accessibility Practice. Lecture Notes in Computer Science, Volume 8516, 2014, 613-622 (2014)
9. Yang, R., Shin, E., Newman, M. N., Ackerman, M. S.: When fitness trackers don't 'fit': end-user difficulties in the assessment of personal tracking device accuracy. In: the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15). ACM, New York, NY, USA, 623-634 (2015)