The Falsified Self: Complexities in Personal Data Collection

Alessandro Marcengo[1], Amon Rapp[2(✉)], Federica Cena[2], and Marina Geymonat[1]

[1] Telecom Italia - Research and Prototyping, Via Reis Romoli 274, 10148 Turin, Italy
alessandro.marcengo@telecomitalia.it
[2] Computer Science Department, University of Torino, C.so Svizzera 185, 10149 Turin, Italy
amon.rapp@gmail.com, cena@di.unito.it

Abstract. Personal Informatics systems collect personal information in order to trigger self-reflection and improve self-knowledge. Users can now choose among different wearable devices for collecting these data according to their needs and desires. These tools exploit not only different shapes and physical forms, but also diverse technologies and algorithms, which may impact the effectiveness of data gathering. In this paper we explored whether there are significant differences in their reported measures and how these can impact the user experience, along with the perceived accuracy of the gathered data and the perceived reliability of the device. To this aim, we carried out an autoethnography which lasted 4 weeks, monitoring the number of steps and the distance covered during the day and the sleep period through different wearables. The results showed that there are wide differences among diverse tools and these differences greatly influence how data collected and devices used are perceived.

Keywords: Quantified self · Personal informatics · Self-tracking · Wearable devices · Autoethnography

1 Introduction

Technological advances in wearable and ubiquitous technologies have recently opened new opportunities for Personal Informatics (PI). These systems aim to leverage sensors and mobile devices for collecting personal information in order to trigger self-reflection and enhance self-knowledge [1].

While first PI systems were employed mainly in clinical purposes for supporting patients in self-tracking dysfunctional behaviors or problematic medical conditions, they were then adopted by researchers, technical fanatics, and members of the Quantified Self community. Quantified Selfers use them to discover factors that may influence their behaviors, and are engaged in self-experimentation, i.e. the practice of systematically changing aspects of daily lives in order to discover variables that affect physical parameters, psychological states, and, by and large, aspects of daily life [2, 3]. However, thanks to the recent diffusion of wearable devices on the market, we are assisting to the commercialization of a plethora of tools that track a variety of personal information, from steps to sleep, from posture to arousal levels, from heartbeat to blood pressure.

© Springer International Publishing Switzerland 2016
M. Antona and C. Stephanidis (Eds.): UAHCI 2016, Part I, LNCS 9737, pp. 351–358, 2016.
DOI: 10.1007/978-3-319-40250-5_34

These instruments take mainly the form of wearables that users can choose according to their needs: for example, physical activity can be traced by necklaces (Misfit Shine), bracelets (Jawbone Up), watches (Apple Watch), or mobile apps that run on the user's smartphone (Moves). All these technologies measure steps by leveraging not only different forms, that can be differently integrated in people's daily lives and personal styles, but also different technologies and algorithms, which can affect the data accuracy.

Moving from these considerations, we aimed at understanding whether there were significant differences in the reported measures and their possible causes (e.g., characteristics of the device, the position in which they were worn, etc). Moreover, we aim at studying the possible effects of these differences on the perceived accuracy of the gathered data and on the consequent perceived reliability of the device. To this aim, we carried out a four-week autoethnography, monitoring the number of steps, the distance covered during the day and the sleep period with different devices. The results of the study are somehow surprising: (i) the gathered data for the same target parameter were very different depending on the device used, and the difference depended mainly on where the devices were positioned and on the user's habits; (ii) this affected the perceived reliability of the devices fostering the research of alternative strategies for accounting the data collected.

The paper is structured as follow. Section 2 provides the most relevant related work in relation of technologies for self-tracking and their reliability. Section 3 provides a picture of the practice of autoethnography both in anthropology and in Human-Computer Interaction. Section 4 describes the setting of our research while Sect. 5 describes its results. Finally, Sect. 6 concludes the paper providing the future directions of the work.

2 Related Work

Different works have explored how users perceive accuracy and reliability of wearable devices and ubiquitous technologies.

Kay et al. [4], for example, investigated how users perceive accuracy, finding how they react negatively when perceive inaccuracies and which kind of unrealistic expectations they have about their weight. Lazar et al. [7], in their study on why and how users abandon their smart devices, noted how they place a great deal of importance on accuracy, impacting the kind of devices they choose and keep using, and being one of the main cause of the abandonment of these self-tracking tools.

Consolvo et al. [5] found differences in participants' reactions to diverse kinds of errors in detecting data. They found seven types of errors made by the fitness device they were evaluating: (i) errors in the start time; (ii) errors in the duration; (iii) errors about confusing an activity it was trained to infer with another it was trained to infer; (iv) errors about confusing an activity it was not trained to infer with one it was trained to infer; (v) failures to detect an activity it was trained to infer; (vi) failures to detect an activity it was not trained to infer; and (vii) errors in detecting an activity when none occurred. Participants were particularly frustrated when the device failed to detect any activity when they performed an activity that the tool was trained to infer, or when the

device detected an activity when none occurred: these two kinds of error questioned the overall credibility of the device. The other type of perceived error that had a great impact on the device's credibility was when the device detected an activity when none occurred.

Mackinlay [8], instead, focused on how users test the accuracy of a device's measurements, reporting barriers in evaluating and bettering the accuracy of their data due to the limited visibility of the system's status that undermined the users' endeavors in calibrating and testing the device.

Yang et al. [6] analyzed 600 Amazon reviews and interviewed 24 participants describing the different methods that users employ to assess accuracy of their self-tracking devices identifying the issues they encountered. They found that differences in users' expectations, physical characteristics, types of activities and lifestyle made them to have different perceptions of the devices' accuracy. The authors conclude how it is essential to focus on how users perceive and assess accuracy of their data in order to determine the reliability of the self-tracking devices. They further suggest (i) to support testability, (ii) to increase transparency of what these instruments can and cannot recognize and (iii) to allow for ways to calibrate the device to personal movement patterns and purposes, enabling users "to record unique movements into a device to let the device know which movements to record and which to ignore" [6].

3 The Autoethnographic Method

We employed an autoethnographic method in order to detect differences among different self-tracking devices and the impact they may have on the user experience. Autoethnography is an ethnographic method in which the fieldworker's experience is investigated together with the experience of the other social actors observed. It is considered valuable on its own and it is reported in the ethnographic recounting [9]. This makes autoethnography close to the autobiographical genres of narration, tying the personal to the social and the cultural in a multi-level form of description of reality [10]. The autoethnographer uses her self-observation and the episodes happened to her as a starting point to make reflections on cultural and social accounts, for returning then to her self and her interpretations of what she observed.

"Autoethnography requires that we observe ourselves observing, that we interrogate what we think and believe, and that we challenge our own assumptions, asking over and over if we have penetrated as many layers of our own defenses, fears, and insecurities as our project requires" [11]. Goodall [13] stresses that good autoethnography "completely dissolves any idea of distance, doesn't produce 'findings,' isn't generalizable, and only has credibility when self-reflexive, and authority when richly vulnerable… When it is done well, we can learn previously unspoken, unknown things about culture and communication from it"

Autoethnography has been employed in HCI for evaluating technologies and gaining empathy with users of various types of devices [16]. It has been used in autobiographical design as a design research method that "drawing on extensive, genuine usage by those

creating or building the system" allows designers "to uncover detailed, subtle understandings that they likely wouldn't have found with other user-centered design techniques because they might seem unremarkable" [14].

The recent popularity of this kind of self-study has to be retraced to the need of finding less-demanding techniques than traditional ethnographic methods, which are very expensive in terms of time and costs [15]. "Typically, ethnography will take place over a period of several months with at least the same amount of time spent in analysis and interpretations of the observations" [Bentley]. So they can be inscribed in those approaches called as "rapid ethnography", which aims to understand users and their environments in a shortened timeframe [18].

O'Kane et al. [16], for example, used autoethnography for evaluating a wrist blood pressure monitor used by people with conditions of hypertension. They found that this method enables researchers "to understand and empathize with the experiences mobile device users can face in difficult to access contexts", allowing them "to better understand user experiences with mobile devices, including mobile medical technology, especially during non-routine times that can be difficult to study in-situ with traditional user studies" [16]. By using this method we tried to overcome the difficulties in observing users in private setting, such as during sleep, gathering a variety of data that would have been impossible to collect otherwise.

4 Ethnographic Setting

In the light of the aspects identified above and the chosen autoethnography methodology, we choose for a four week session of self-observation wearing different kinds of wearable devices.

The parameters that we decided to compare were the *steps* and the estimated *distance* covered during the day and the total amount of *sleep* (also segmented into *light/heavy sleep* and *awake time)* for each sleep cycle.

The wearable instruments have been differentiated depending on the model and the position on the body; it was also used an application running background on the phone in order to collect data during the day (*steps* and *distance*).

The purpose of the experiment was to understand, in an explorative way, whether there were significant differences in the resulting measures and whether they could be attributed to the characteristics of the devices or to the position in which they were worn. The objects chosen in particular were the following, each one placed in a different position on the ethnographer's body (the first author):

- Withings Activité on the right wrist: the Withings Activité is primarily a classical watch that also measured *steps* and *sleep*, it does not require to be recharged and it is waterproof, this will let you keep it continuously with no need to ever separate from it.
- Shine Misfits necklace: the Misfits Shine is a waterproof unit that can be worn in various positions, the one that offers the least friction is the use in combination with the necklace accessory. It does not require to be recharged.

- Sony SWR30 on the left wrist: the Sony SWR30 is a hybrid between a smartwatch and a wristband. It has indeed both telephony and logging features. It is waterproof and requires to be briefly recharged every 5/6 days.
- GoogleFit application running background on a Sony Xperia Z3. This application uses the accelerometer of the phone for the *steps* estimation and the GPS signal to calculate the *distance* covered.

The starting hypothesis has been that the recorded data would not suffer the influence of the body positioning, recording approximately the same values regardless of the device used. The app on the phone has been used as a method of comparison between wearable and non-wearable paradigm. Actually the result has been quite surprising. In the four weeks period the recorded data resulted completely different especially as a function of the body positioning of the device and the distortions in the accuracy caused by the peculiarities of the personal lifestyle.

5 Results

Analyzing *sleep* related data, there were several interesting findings, particularly related to the aspects that follow.

Regarding the *sleep* total amount logs the data results reliable with negligible deviations in the order of minutes. However, it has emerged as the *sleep* total amount recorded by the Misfit Shine worn as a necklace is always higher of about thirty minutes. This point, in relation to the personal experience appears due to the fact that the necklace considers the horizontal still position as "sleeping" not taking into account that in fact the user might be lying reading a book before falling asleep. So the *sleep* total amount will always be increased by the reading time in the bed. Furthermore, by comparing the data with the personal observation it is clear that the device with higher accuracy results the one worn on the right wrist (Withings Activité). This is because the right wrist makes possible to discriminate the browsing of pages in a horizontal still position as "non sleep" (obviously in the case of a left-handed user the most discriminating device would be the one on the left wrist).

What emerged in a rather surprising way is the total discrepancy between the different devices about the light sleep and deep sleep data. It was not possible to discriminate any specific reason related to the positioning of the device for the differences in the data collected, so we have to hypothesize that the cause relies in the poor quality of the algorithms that discriminate against the two types of *sleep*. About the *steps* logging the following evidences have emerged. The total *steps* amount is strongly affected by the location on the body on which the device is worn and the wearer's activities dictated by his specific lifestyle. Indeed in relation to the data collected by Withings Activité on the right wrist it has been observed that on days in which the tester has performed much talk in public (meetings, presentations, etc.) *steps* were very biased towards high figures due to the gesticulation of the speech. On the other hand the opposite effect was found in some other lifestyle variables. In particular, the data has been surprisingly distorted toward low figures for the device worn on both wrists for two conditions. The first one is about walking while pushing a stroller. In this case probably the algorithm does not register the dangling of the hands and

does not log the activity as *steps*. The second one occurs walking while carrying a moderately heavy bag (e.g. a small suitcase) depending which hand holds the bag. Lastly we recorded a highly distorted *steps* data toward low figures for the phone app due to the fact that the device failed to record in all the occasions when the phone was placed outside of the user's pockets (e.g. weekend, sports, home, etc.).

About the covered distance the data appear completely unreliable: this does not seem due to the position on the body but mainly in relation to the calculation algorithms embedded by the device manufacturer. The conversion from *steps* to *distance* covered appears totally arbitrary, making this the more improbable data among the information collected. The coupling of geo positioning does not seem to improve the accuracy since most of the *steps* are made indoor on the same spot (gps geo positioning seems to work better for outdoor sport situations, like running, hiking, etc.).

In general, on the weekend all the data appear distorted by incomplete or peculiar usage of the devices due to different life activities (i.e. working in the garden, playing with kids, etc.).

Because all the distortions in the measures noted by the ethnographer, however unexpected, he developed a quite disturbing feeling. In particular, the ethnographer had the impression that the devices were drawing, despite the limited nature of data collected, a false self, an image in which the ethnographer could not identify himself. The impression was that the measures were counterfeiting his self-perception.

In the light of these aspects that have become evident at the beginning of the autoethnography the ethnographer developed some strategies targeted to consider only the data streams deserving good accuracy in relation with his personal lifestyle. For example, the total *sleep* duration data stream he considered closer to reality was the one coming from the right wrist (Withings Activitè). Instead, the accuracy related to heavy/light sleep patterns remained completely unknown for the above reasons.

In terms of total *steps* instead the more accurate data stream was considered the one provided by the Misfits Shine necklace as its position on the body was not affected by the oscillation (or not) of the arms. Even in this case, however, the measures of the distance covered and the calories burned appeared totally unreliable and obscure.

It appears that even if all the devices could be able to record several measures (e.g. *Sleep* + *Steps*) no one resulted enough accurate in all the measures: this led to the necessity of splitting on two different devices, in two different parts of the body the recording of the different data streams.

From these considerations some design insights can be derived such as the manufacturer's need to consider different designs for different usage styles induced by different types of users with different habits (e.g. reading in the bed, pushing a stroller, carrying a folder, gesturing or drawing a lot during work time). These could be condensed into a few personas that can lead to different models of the same device or different tracking algorithms on the same device. This customization may be transferred directly into the experience of the user by collecting certain aspects of her habits that may impact on the accuracy of the device, possibly also advising her on the best body location in which to wear the device in relation to her personal lifestyle.

6 Conclusion and Future Work

The goal of this paper is to study the differences in the measures reported by different PI tools and to investigate possible causes, as well as to discover if such differences have effects on the perceived data accuracy and device reliability. To this aim, we carried out a four-week autoethnography, monitoring the number of steps, the distance covered in a day and the sleep period through different tools. The results of the study showed that (i) there are wide differences due to the device position and the user lifestyle, and (ii) this lack of reliability requires the user to search for personal strategies for making the data accountable.

The next step will be to study different devices with more subjects, in order to confirm our initial findings. Moreover, we want to investigate in a deeper way the subjective users' perceptions with respect to the reliability of the device in relation to the accuracy of the data gathered. We want also to study whether different visualizations may affect the user's perception of the reliability of the data collected. A further interesting experiment would be a comparison with data gathered through a specialized medical device in order to evaluate which commercial device is actually more effective in terms of accuracy.

References

1. Li, I., Dey, A.K., Forlizzi, J.: A stage-based model of personal informatics systems. In: 28th International Conference on Human Factors in Computing Systems, pp. 557–566. ACM Press (2010)
2. Marcengo, A., Rapp, A.: Visualization of human behavior data: the quantified self. In: Huang, L.H., Huang, W. (eds.) Innovative Approaches of Data Visualization and Visual Analytics, pp. 236–265. IGI Global, Hershey (2013)
3. Rapp, A., Cena, F.: Self-monitoring and technology: challenges and open issues in personal informatics. In: Stephanidis, C., Antona, M. (eds.) UAHCI 2014, Part IV. LNCS, vol. 8516, pp. 613–622. Springer, Heidelberg (2014)
4. Kay, M., Morris, D., Schraefel, M.C., 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 2013), pp. 401–410 (2013)
5. 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 2008), pp. 1797–1806 (2008)
6. 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 2015), pp. 623–634. ACM, New York (2015)
7. 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 2015), pp. 635–646. ACM, New York (2015)
8. 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)

9. Tedlock, B.: From participant observation to the observation of participation: the emergence of narrative ethnography. J. Anthropol. Res. **47**(1), 69–94 (1991)
10. Ellis, C., Bochner, A.: Autoethnography, personal narrative, and personal reflexivity. In: Denzinand, N.K., Lincoln, Y.S. (eds.) Handbook of Qualitative Research, 2nd edn, pp. 733–768. Sage, Thousand Oaks (2000)
11. Jones, S.H., Adams, T.E., Ellis, C. (eds.): Handbook of autoethnography. Left Coast Press, Inc, Walnut Creek (2013)
12. Tami, S.: Performing autoethnography: an embodied methodological praxis. Qual. Inq. **7**(6), 706–732 (2001)
13. Goodall, Jr., H.L.: Notes for the autoethnography and autobiography panel NCA. National Communication Association Convention in New York City (1998)
14. Neustaedter, C., Sengers, P.: Autobiographical design: what you can learn from designing for yourself. Interactions **19**(6), 28–33 (2012)
15. Cunningham, S.J., Jones, M.: Autoethnography: a tool for practice and education. In: the 6th ACM SIGCHI New Zealand Chapter's International Conference on Computer-Human Interaction: Making CHI Natural (CHINZ 2005), pp. 1–8. ACM, New York (2005)
16. O'Kane, A.A., Rogers, Y., Blandford, A.E.: Gaining empathy for non-routine mobile device use through autoethnography. In: the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI 2014), pp. 987–990. ACM, New York (2014)
17. Bentley, R., Hughes, J.A., Randall, D., Rodden, T., Sawyer, P., Shapiro, D., Sommerville, I.: Ethnographically informed systems design for air traffic control. In: Conference on Computer Supported Cooperative Work, pp. 123–129 (1992)
18. Millen, D.R.: Rapid ethnography: time deepening strategies for HCI field research. In: Boyarski, D., Kellogg, W.A. (eds.) the 3rd Conference on Designing Interactive Systems: Processes Practices Methods and Techniques (DIS 2000). ACM, New York (2000)