Quantified Factory Worker - Expert Evaluation and Ethical Considerations of Wearable Self-tracking Devices

Päivi Heikkilä[†]
VTT Technical Research Centre of Finland Ltd.
Tampere, Finland
paivi.heikkila@vtt.fi

Anita Honka
VTT Technical Research Centre of Finland Ltd.
Tampere, Finland
anita.honka@vtt.fi

Sebastian Mach
Chemnitz University of Technology
Chemnitz, Germany
sebastian.mach@psychologie.tu-chemnitz.de

Franziska Schmalfuß
Chemnitz University of Technology
Chemnitz, Germany
franziska.schmalfuss@psychologie.tu-chemnitz.de

Eija Kaasinen
VTT Technical Research Centre of Finland Ltd.
Tampere, Finland
eija.kaasinen@vtt.fi

Kaisa Väänänen
Tampere University of Technology
Tampere, Finland
kaisa.vaananen@tut.fi

ABSTRACT

Following the Quantified Self trend, everyday self-tracking practices have become common. Still, self-monitoring of people at work is a rather new research topic. Self-tracking of employees' activities, mental state and emotions enables data-based feedback, which could improve the employees' awareness of issues influencing their well-being and performance. We contribute to this topic from two perspectives. First, we explored the potential of wearable self-tracking devices for providing personal feedback to machine operators working in a factory. We used the expert evaluation method to lay ground to the user perspective of self-tracking at work. User experience experts evaluated five tracking devices for their user experience, perceived accuracy and fit to factory workers. Second, we conducted a workshop with the experts to systematically assess the ethical considerations that may arise when adopting self-tracking at work. The results provide insights into the potential of the use of self-tracking devices in a factory context.

CCS CONCEPTS

• Human-centered computing → Human Computer Interaction (HCI) • Ubiquitous and mobile computing

KEYWORDS

Quantified Self, User experience, Ethics, Wearable devices, Expert evaluation, Factory workers

ACM Reference format:

Päivi Heikkilä, Anita Honka, Sebastian Mach, Franziska Schmalfuß, Eija Kaasinen and Kaisa Väänänen. 2018. Quantified Factory Worker: Expert Evaluation and Ethical Considerations of Wearable Self-tracking Devices. In *Proceedings of Academic Mindtrek conference (Mindtrek 2018). ACM, Tampere, Finland, 10 pages.* https://doi.org/10.1145/3275116.3275119

1 Introduction

Wearable self-tracking devices are becoming increasingly common in everyday use by consumers and employees, especially the wristbands and smartwatches for tracking activity, sleep and heart rate [see e.g. 22]. Recently, the range of wearable self-monitoring devices, on the market, has extended to tracking one's emotions or mental state.

While the number and variety of wearable self-tracking devices have increased, the Quantified Self trend has raised the awareness of ways to practice self-tracking and gain meaningful data-based insights to foster positive behavioural changes. Enthusiastic early adopters have found self-tracking interesting and useful [4,14,23], but less is known about the user experience and usage practices of less technology-savvy users or the use in work contexts.

Extending self-monitoring practices to workplaces or integrating them to corporate wellness programs is a relatively new, but rapidly increasing phenomenon [18]. However, not much scholarly literature exists on the consequences of incorporation of self-tracking at the workplace [2,18]. The potential for true benefits for the worker and the employer depends on a range of issues: ways to adopt self-tracking practices, awareness and consideration of ethical issues and the technical possibilities of the available wearable devices. Furthermore, user experience issues are an important factor in determining whether use feels effortless and pleasant and whether employees accept the new technologies. The boundaries of voluntariness, autonomy of practicing self-monitoring and the purposes of self-tracking raise important ethical issues, when self-monitoring opportunities are introduced in a work context.

In this paper, we focus on the potential of wearable self-tracking devices for factory workers. Even though factory work may be regarded as rather monotonous and predictable work, the modern factory floor work is assumed to become closer to knowledge work, emphasizing decision-making and problem-solving [8]. When operating a highly automated manufacturing machine, the work already requires process management and problem-solving skills. Compared to typical knowledge work, however, solving problems quickly is more crucial, to avoid or minimize delays in the manufacturing process. A single factory operator may be responsible for a whole machine, and should handle the problem situations without finding it too stressful. In this kind of factory work, self-tracking data may give insights on how one responds to challenging situations physiologically and motivate finding ways to stay calm. Self-tracking may also help to notice the achievements and efforts made during the workday, and thus, create job satisfaction.

The presented study is a part of a wider research project (Factory2Fit), which aims at engaging and empowering factory workers by different means, with one being the opportunity to receive self-tracking feedback, which could improve the workers' awareness of the issues having an impact on their well-being and performance. The gathered self-tracking data will be intended only for the factory workers themselves, not for the employers. The aim of this study was to evaluate the potential of different types of commercially available wearable self-tracking devices, in the context of factory work, and to identify ethical considerations related to self-tracking of factory workers. Before self-tracking devices should be implemented in actual pilots, we wanted to assure that the devices introduced are easy to use and fit the factory context.

Thus, the research questions of this study are:

1) What kind of self-tracking devices have potential to be introduced to factory workers?
2) What ethical considerations need to be taken into account when quantifying the factory worker?

The first research question is studied through an expert evaluation. The expert evaluation method was chosen because it has been proven to be effective in early assessment of potential alternative solutions [27] – experts can systematically evaluate several devices before taking the most appropriate device(s) to the actual users. User experience experts used the selected wearable self-tracking devices as part of their daily life, and evaluated the potential of each device in terms of user experience and perceived accuracy. After this, they evaluated whether the selected devices seem applicable for use by factory workers. As a response to the second research question, ethical considerations were identified in a workshop with the experts, focusing on ethical issues related to the theme of letting the factory workers quantify themselves.

The paper is structured as follows: First, we present related work on wearable devices for tracking well-being and on the potential ethical issues of self-monitoring at work. Second, we describe the study methods and the evaluated devices. Then, the results are presented, first related to the research question 1 and then, to the research question 2. Finally, we discuss the implications and limitations of the results.

2 Related Work

In this section, we first provide background on the expanding range of wearable devices for tracking well-being. Then, we shed light on the potential of self-tracking, especially in work context, and the ethical issues related to adopting self-tracking practices at work.

2.1 Wearable Devices for Tracking Well-being

Commercial wearable devices, for tracking well-being and behaviour, have become popular amongst consumers, with around 25% of Americans owning a wearable wrist device [22]. The most prominent devices include wristbands and smartwatches used to track physical activity and sleep based on accelerometer and heart rate measurements [22], e.g. Fitbit, Samsung, Apple, Polar and Nokia Health, to mention a few of the manufacturers. The advantages of these devices are that they can track objective data round-the-clock, without requiring continuous charging, and they are non-invasive and compact in size. In addition, the accuracy of these devices is sufficient for everyday life contexts for step count [3] and heart rate measurements [24].

Devices monitoring mental well-being, such as negative stress or emotions, are less common in the consumer market, but their popularity is anticipated to increase [25] as the maturity of these devices improves, regarding their user experience and accuracy. Measurements for detecting emotional states that already can, or are foreseen in the future to be recorded with good quality, with off-the-shelf devices, include heart rate variability (HRV), heart rate (HR) and electrodermal activity (EDA) for sensing the functions of the autonomous nervous system, and electroencephalography (EEG) for measuring brain waves [6]. In general, the data tracked with wearables have the potential for providing personalized, immediate and objective feedback [21].

2.2 Quantified Self and Self-tracking at Work

The trend of self-tracking and quantification of oneself has emerged during the last ten years. Quantification of oneself refers to self-tracking of biological, physical, behavioural or environmental data [26]. The term Quantified Self has been used for the practice of self-monitoring, as well as the movement and international community of people practicing self-tracking. The term was coined by Gary Wolf and Kevin Kelly, who created the website (http://quantifiedself.com) for sharing self-tracking practices in 2007. Also, other terms, such as personal informatics [14], personal analytics [17] and lifelogging [see more in 8] have been used to describe practices for gaining self-knowledge, through recording personal data. Most of the research on self-tracking practices has focused on exploring the active members of the Quantified Self community [4,20] or early adopters, for example using several trackers [23] or attending web forums related to self-tracking [14].

An ultimate goal of the Quantified Self movement is to gain meaningful insights and make positive behavioural changes, based on personal data [4]. Even though this goal is commonly shared by people quantifying themselves and there is optimism towards adopting self-tracking at a workplace [18], the long-lasting impacts have not yet been systematically examined [2,18]. However, the number of trials of self-tracking at work and the inclusion of this to corporate wellness initiatives have expanded during the last years [18].

Although applying the Quantified Self approach to the workplace is a relatively new topic and there is not much research conducted yet, several potential benefits, also applicable to work contexts, have been identified. The potential benefits include reducing the physical and cognitive burden at work [13], earlier detection of health problems [15] and fostering healthy behaviour [1]. Adopting wearable self-tracking technologies may also enable personal recommendations on the sequence of the pending work tasks or practices to reduce stress at work [25]. If the worker can be localized through wearable technologies, the impacts extend to improved safety and the request of immediate help to one's location [13]. Besides all the potential benefits, wearable devices first need to be suitable and provide a solid user experience for the working context. This will be the starting point in the present paper.

2.3 Ethical Issues related to Self-tracking at Work

Adopting self-tracking at the workplace raises a range of ethical issues, as self-monitoring may not be truly voluntary. Lupton [16] has categorized self-tracking at work into five different modes: private, communal, pushed, imposed and exploited. While private self-tracking is initiated for purely personal reasons, pushed self-tracking is externally encouraged or advocated. Inviting workers to voluntarily attend to a corporate wellness program is one example of pushed self-tracking.

The development of ambient intelligent applications, in general, poses a number of ethical concerns. Application design can be guided by different ways of identifying and considering ethical questions, for example, by assessing ethical impacts of information technology [28], identifying the values of the target users and responding to them [6], or following ethical guidelines or principles. The framework of Ikonen et al. [9,12] includes six ethical principles for designing ambient intelligent applications. The identified principles are privacy (ability to control access to one's personal information and protect one's own space), autonomy (right to decide the ways and the purposes for technology use), integrity and dignity (users of technology shall be respected), reliability (sufficient reliability of the technology for its purpose of use), e-inclusion (accessibility for all user groups) and benefit to the society (increases the quality of life and causes no harm).

Regarding the ubiquitous computing at work, Nihan [19] lists values and principles that are frequently used to explain, justify or challenge the development of Ubicomp in the workplace. The listed ten aspects have different meanings and emphases, depending on the stakeholders concerned (e.g. employees vs management). The identified values are privacy, autonomy, health, safety, security, control, responsibility, justice, performance, as well as social interactions and integrations.

Moore and Piwek [18] raise ethical questions specifically related to the context of the emergence of wearable devices and self-tracking technologies in the workplace. They discuss the ethics through four themes: work intensification, the challenges in measuring productivity, stigma and shame for opt-outs and the lack of legal regulation. Fundamental issues, such as "dehumanizing employees" and inadequacy caused by attempts to achieve complete well-being, are discussed. As this is a new area for scholarly research, they call for critical research related to the consequences of incorporating self-tracking technologies in the workplace.

3 Study Design

The study consisted of an expert evaluation of selected wearable devices and an ethics assessment workshop. This section describes the methods and the study procedures.

3.1 Initial User Studies

Before the expert evaluation and the ethics assessment workshop, the experts either participated in or familiarised themselves with the user studies conducted earlier in the research project [10]. Prior starting this study, initial user studies were carried out in the three pilot factories of the research project to understand the factory context of use and the workers' first impressions and thoughts of self-tracking at work. The studies revealed that the workers might welcome positive feedback, enabled by self-tracking, that many are not used to receiving at work. However, the idea of monitoring of oneself at work raised scepticism, as it was regarded as a source of stress or negatively highlighting the differences between workers. This raises the importance of paying attention to ethical issues and the appropriate selection of the self-tracking devices to be used. The results of the initial user studies were used as a basis for understanding factory work and creating scenarios for ethics assessment workshop.

3.2 Expert Evaluation

The first part of the study was conducted by using an expert evaluation method, which has been proven to be efficient in finding user experience issues in a systematic manner [see e.g. 27]. The expert evaluation method was selected, as the project team wanted to gain expertise of the possibilities of the commercially available wearable self-tracking devices and evaluate their potential, to be able to select appropriate devices for the pilots with factory workers. Researchers that had solid experience in user experience design, acceptability of products and/or digital health served as experts. As the devices, especially for tracking mental states, are still new in the market, experts should evaluate their user experience, perceived accuracy and suitability first, so that factory workers would only use devices that are suitable for real use.

Altogether, nine experts (7 females, 2 males) with a broad age range (average age 38, from 21 to 60 years) tested the selected five devices in their daily life. The usage period of the devices ranged from 9 days to 45 days. The experts used the devices continuously at work and on free time as well as during night, except a headset device that could be used only for short periods, due to practical problems with it. The usage period depended on the schedule of the research project, as all the evaluations needed to be conducted in the early phase of the project. Each device was used and evaluated by four of the experts, one device at a time. An exception to this was a headset device that was immediately found to be unsuitable for factory workers and thus, was only used by two persons. Each expert evaluated 1-3 devices.

3.2.1 Evaluated devices. The devices for the evaluation were chosen based on the following selection criteria: 1) The devices and the related smartphone apps had a consumer-friendly design, regarding their appearance and the expected usability; 2) The self-tracking feedback provided via the apps included advanced data interpretation, instead of presenting the raw data; 3) It seemed possible to transfer time-series data collected by the devices to 3rd party apps or data sheets for research purposes; 4) For tracking mental states, the attempt was to find a variety of different measurement technologies for testing.

The selected devices included a smartwatch and a wristband for self-tracking heart rate, activity and sleep, as well as three device options for mental state detection. The selected devices vary from each other in terms of their form factor and way of wearing them, as well as in terms of the sensors and technologies utilized in the measurements. Three devices for emotion and stress detection were included as less obvious options. The devices are described below.

Figure 1: Devices from left to right, up to down: Samsung Gear S3, Fitbit Charge HR, Spire, Moodmetric, and Emotiv Insight.

The Samsung Gear S3 (www.samsung.com) is a smartwatch with various supporting functionalities, such as managing health and well-being, making phone calls, and messaging. The technical features include an optical heart rate sensor, a 3-axis accelerometer and gyrometer, GPS, an ambient light sensor, a barometer and altimeter as well as a microphone. A large variability of apps is available for the device.

The Fitbit Charge HR (www.fitbit.com) is a wristband for tracking activity, heart rate and sleep. The device includes an optical heart rate sensor, a 3-axis accelerometer and an altimeter sensor. The device monitors several metrics, such as resting heart rate, steps, minutes of activity and stationary hours, as well as duration and quality of sleep. Call, message and calendar notifications can be displayed on the screen of the device. The provided data can be explored through a dashboard app.

The Spire device (www.spire.io) analyses the breathing patterns of the user. Based on the personal average breathing rate, it identifies the mental states of tense, focus and calm. It also monitors sedentary minutes and counts steps, as well as stores location information related to the states. The device notifies different mental states and long periods of inactivity by vibrating. The mobile app displays the current state of the user and gives a detailed summary of the identified states, during each day. The app also includes calming exercises.

The Moodmetric ring (www.moodmetric.com) supports stress management by estimating the intensity of emotions, based on measurements of electrodermal activity (EDA) of the skin. The ring notifies the user of the emotional intensity by a led indicator (high alerts and calm notifications). The mobile app (Moodmetric) displays the current emotional intensity and a summary view of the emotional intensity, during each day and night. The app also shows the step count and offers a feature for practicing calming down.

The Emotiv Insight (www.emotiv.com) is a headset, for monitoring cognitive performance and well-being, based on brainwave (EEG) signals. The device can monitor six cognitive and emotion metrics: focus, stress, excitement, relaxation, interest, and engagement. The mobile app shows the cognitive state of the user in terms of the metrics and an image of the currently activated brain areas.

3.2.2 Data Gathering and Analysis. After each testing period, the experts filled in an evaluation questionnaire. The questionnaire included open questions regarding: 1) the user experience of the device and the accompanying app (e.g. how did you feel using the device?), 2) the perceived accuracy of the self-tracking feedback provided by the apps (e.g. what is your feeling of the accuracy of each parameter monitored?), as well as, 3) the devices' applicability to a factory environment (e.g. do you recommend using the device during work in a factory?). Experts evaluated the first two aspects also quantitatively using a scale from 1 (very poor) to 7 (very good).

The experts' user experience observations were coded and analysed by two researchers, to identify the factors that had an impact on the user experience and perceived accuracy of the evaluated devices. The main categories of the content analysis were: 1) the general user experience, 2) the perceived accuracy of the device, and 3) the applicability for factory workers. Sub-categories for the general user experience and accuracy were retrieved by using guiding questions, based on the dimensions of the Technology Acceptance

Model for Mobile Services (TAMM) [11], i.e. the *ease of adoption, ease of use, trust* and *value.* These aspects were considered as critical factors determining the potential of the devices.

3.3 Ethics Assessment Workshop

As a second part of the study, we organized a workshop for identifying ethical considerations related to the topic. Four of the experts who evaluated the devices participated in the two-hour ethics assessment workshop. All participants had expertise in considering ethical issues, for example when designing digital services or when evaluating e-health solutions.

3.3.1 Scenarios. Before the workshop, two researchers created two scenarios, based on the initial user studies, to illustrate the idea of a quantified factory worker. The scenarios varied in terms of the worker (age, gender, working experience, and mood) and the situation (day vs night work shift, problem solving vs no problems). The imaginary factory was the same in both scenarios, representing the pilot sites of the project and the work contexts that we had observed during the initial user studies. Receiving of both real-time feedback and summarized feedback of the work shift were included in the scenarios. The first scenario concentrates on the work context only, while the second scenario extends the use of the self-tracking device to one's free time (tracking sleep).

The **first scenario** focused on a problem-solving situation:

Matthew, 24, has worked as a machine operator in a sheet metal factory, for three years. He has recently started to use a new multipurpose machine, and finds the work mentally loading, as the machine frequently stops running, which he needs to handle quickly. The problems vary, and thus, there is no clear solution how to get the machine running again.

Matthew starts his work shift, wearing a wearable measurement device on his wrist. Everything goes smoothly until noon, when the machine suddenly stops. He is working alone and feels panicked, as he does not know what to do. His wrist device vibrates to notify him of his increased heartbeat. The device guides him to calm down, breathe deeply and think clearly again. He starts systematically trying options that have helped with problem situations, in the past. The wearable device indicates that he is concentrated, which gives him encouragement to carry on solving the problem. Finally, he gets the machine running again.

After his work shift, he looks over the day's statistics on his phone's app. He sees that he had handled the problem situation quicker than before and realizes that the machine stoppage was only a short one. In addition, he sees that he has already achieved more than half of the recommended step count for the day. He feels satisfied.

The **second scenario** focused on an experienced older worker, who is more sceptical towards wearing a measurement device during a work shift and in free time:

Brenda, 48, has worked in the same factory for 20 years. In general, she finds most renewals and corporate campaigns welcome, but is a bit sceptical towards a new opportunity to wear a measurement device during a work shift and even on free time. She does not want to show others and management that she is reluctant to change, and thus, starts to use the new device.

Brenda starts her work shift, a third night in a row. She finds the second night shift the hardest, while the third one feels easier as she starts to get used to the night rhythm. Still, she feels tired. When starting the shift, she checks the statistics of her wearable device on the operators' computer, especially how she has slept during the past days. She also checks her resting heartbeat and some work performance metrics. During a one-month's use time, she has actually found it interesting to see measures related to herself and her work, and the changes in them, but she does not want to share these results with others. As she feels tired and the measures reinforce her feeling, she starts the work shift with routine tasks that she knows well and that are not likely to result in problems. After sitting for one hour, the wearable device indicates that it is time to walk a bit. This helps Brenda to become more energetic.

After the work shift, Brenda is still a bit tired, but feels better when she checks the statistics of the work shift. She has been able to carry out a decent number of tasks and the machine has been working the whole time. She feels good when heading back home and takes a daytime nap.

In the workshop, participants walked through both scenarios, one at a time, and identified ethical issues. Each participant wrote down relevant ethical issues to post-its, and placed them on the analysis template, including ethical aspects selected for the analysis (listed in the Analysis section). This was done separately for both scenarios. After this, the experts discussed the remarks and added issues raised in the discussion into the template.

3.3.2 Analysis. As a basis for the ethical assessment, we utilised two analysis frameworks: the six ethical guidelines for ambient intelligent applications by Ikonen et al. [9], and a list of ethical values related to ubiquitous computing in the workplace, compiled by Nihan [19]. We utilized the former framework as such, and complemented it with the additional values, listed by Nihan, to include values of particular relevance in the work context. Altogether, our resulting template included ten ethical aspects: *privacy, autonomy, integrity and dignity, reliability, e-inclusion, benefit to society, health and safety, responsibility, justice* and *social interactions.*

4 Findings of Expert Evaluation of the Devices

Table 1 summarizes the results of the expert evaluation as ratings given to the evaluated devices. After that, we describe the results in more detail one by one in terms of the user experience, perceived accuracy, and the fit of each device to factory workers, based on the understanding of the user studies in the project.

Device (what data is tracked)	**User experience** (Rating 1-7 median, range, N=4)	**Perceived accuracy** (Rating 1-7 median, range, N=4)	**Suitability for factory workers**
Samsung Gear S3 smartwatch (activity, heart rate, sleep)	5.5 (4-7)	6.0 (5-6)	Good, except for the big size of the watch.
Fitbit Charge HR wristband (activity, heart rate, sleep)	6.0 (6-7)	5.5 (5-6)	Good. Unobtrusive & fairly good perceived accuracy.
Spire clip (mental states and activity)	6.0 (5-7)	4.5 (3-5)	Maybe. Unobtrusive, but unconvincing accuracy.
Moodmetric ring (emotional intensity)	5.0 (4-5)	6.0 (5-6)	Only for some tasks, cannot be worn for all tasks.
Emotiv Insight, EEG headset (emotions)	1.5 (1-2)	Could not be evaluated	Not suitable.

Table 1. Comparison of the devices by the experts.

4.1 Samsung Gear S3

4.1.1 User Experience and Perceived Accuracy. The experts evaluated the Samsung Gear S3 as easy to adopt and easy to use. The feedback of the smartwatch was evaluated as clear and intuitive. The device appeared to be modest and solid. However, the big size caused problems when pulling one's sleeve over the device and felt disturbing, while running or sleeping. The experts perceived the mobile apps as being intuitive and well structured. The only downsides were irritation of a repeated default reminder to go for a walk and concern caused by the apps' permission requests. However, the content of the apps and the recorded data were considered interesting.

The perceived accuracy of the sensors and the displayed data was rated as good for GPS, heart rate, step count, distances, barometer and sleep tracking, when the device was tightly connected to the wrist. The accuracy of the floor count was estimated as moderate. When compared with another smartwatch, no differences in accuracy were noticed.

4.1.2 Applicability to Factory Workers. The experts regarded the Samsung Gear S3 as suitable for use in a factory, due to its large variety of functions and sensors, and the additional messaging possibilities, which could increase its acceptability. The vibrating alarms can make the workers aware of new messages or notifications in areas with a high noise level. In other contexts, the device is able to provide audible information via speakers. The connection via Wi-Fi allows greater distances than via Bluetooth, and the NFC data connection enables exchanging sensitive data, in close ranges. The wearing position of the device was regarded as practical for doing work tasks with both hands. However, the big size and the resulting potential danger, for example while manually interacting with machines, must be considered.

4.2 Fitbit Charge HR

4.2.1 User Experience and Perceived Accuracy. The experts perceived the device as easy to adopt and connect with the app. It was easy to use, and interpreting the information on the device display was straightforward. Wearing the wristband felt unobtrusive as it was slim, light and fitted well on the wrist. In addition, the experts appreciated the reasonably long battery life, of 4 to 5 days. The dashboard app was perceived as easy to use and the presented metrics were regarded as interesting. The positive indications of achievements and the vibration indication of an arriving call, with the caller's name visible, were liked. However, one of the experts reported the synchronizing of the device data with the dashboard as being rather slow. One of the experts decided to buy a Fitbit device for herself, based on her positive experiences during the testing period.

The daily step count, and the automatic detection of walking, running and cycling were considered as accurate. Two of the experts highlighted that the resting heart rate and sleep quality reflected their perceived stress and tiredness levels well. The experiences regarding the accuracy of sleep detection were mixed. Three of the experts perceived the automatic sleep time detection as being fairly correct, but one expert felt that the device frequently interpreted lying in bed awake as being asleep.

4.2.2 Applicability to Factory Workers. The experts found the Fitbit Charge HR suitable for factory workers. Due to the slim and plain design of the device it was not foreseen to disturb manual work or draw undue attention to itself. The material of the band did not seem particularly sensitive to dirt or dents. However, possible problems in usage with gloves were noted. The accuracy of the monitored metrics seemed to be at a good level.

4.3 Spire

4.3.1 User Experience and Perceived Accuracy. The Spire device was perceived as easy to adopt and easy to use. It was unobtrusive and comfortable to wear, could be hidden under clothes and easily forgotten. Charging the device and connecting it to the phone was easy, but two experts found synchronization with the phone to be slow. In addition, the experts experienced loss of data due to not opening the Spire app frequently enough (as the device can store only 6 hours of data). The experts found the vibration indicator of the mental states and sedentary periods useful, but the vibration patterns were difficult to differentiate from each other. The information provided by the Spire app was easy to understand, and the daily statistics were found interesting. The experts considered the app views pleasant, especially the main view, with a beautiful background scenery giving a calming feeling.

The detected states were not always in line with the personal feelings. For example, focused moments were occasionally identified as calm, some feelings of tenseness were not detected (possibly because the state needs to last at least 2 minutes to be detected) and light physical activity (e.g. house chores) were often identified as a tense mental state.

4.3.2 Applicability to Factory Workers. As wearing the device was perceived comfortable and unobtrusive, it would fit well in a factory environment. It would not be in the way when performing manufacturing work, and the vibrating feedback could be beneficial for the workers. However, as factory work typically includes physical activity, it could provide wrong results (light physical activity may be identified as mental tenseness). The synchronization of the data might be an issue, as the device can only store 6 hours of data. It should also be noted, that Spire does not provide continuous data, but detects the user state only 30% to 50% of the time it is worn. However, as the feedback from Spire is based on breathing, the acceptability could be higher than with devices that may be perceived as more ambiguous. With Spire, the factor to be measured (breathing) is easy to understand and it is also rather easy to influence (e.g. for calming oneself).

4.4 Moodmetric

4.4.1 User Experience and Perceived Accuracy. Adopting the Moodmetric device was easy, except two experts had problems in finding the most suitable finger for the ring. Use of the ring was mainly easy and effortless, but some problems occurred. One evaluation could not be completed, because the ring stopped measuring. This may have been due to the user's dry hands, during the cold weather, when using the device (around -20 Celsius). Other problems included the difficulty to notice whether the ring was switched on and the low battery indication of it, and problems in everyday use, when washing hands, wearing gloves or doing physical tasks. The ring was felt to be in the way when changing clothes, and holding or lifting items. However, the ring was found durable, since it continued to work regardless of water splashes or accidentally hitting it on items. The experts evaluated the appearance of the ring either as neutral or too bulky. One expert commented that the ring should either look more stylish or clearly indicate that it is a measurement device. The feedback provided by the app was perceived as understandable and interesting, but real-time feedback and more detailed event-related feedback were wished for. The data visualization was liked, but one user found it confusing to follow the daily data, as the day (from 6 am to 6 pm) and night views (from 6 pm to 6 am) did not reflect the real day and night time. One expert experienced connection problems with the phone, but in general, synchronizing the ring with the app was fast.

The experts noticed that most of their emotional reactions were correctly detected and that the device detected both low and high intensity (active/nervous/alert state vs relaxed/calm state). In addition, they reported that the device could detect a rushed feeling, annoyance, stress, uncertainty and multi-tasking. However, the measures also included some false positives, at least partly due to a high body temperature, which seemed to have a clear impact on the measurements. High emotional intensity was shown, for example, after having a sauna or during physical activity. One expert noticed that her tiredness and longer-term stress were reflected in the measurements, during a stressful week.

4.4.2 Applicability to Factory Workers. The experts did not consider the ring as suitable for factory workers. Wearing it can be difficult when doing physical tasks, and it is impossible when wearing gloves. Also taking the ring on and off, when washing hands, could be problematic. Physical tasks and gripping items tightly could also lead to false measurements. Use of the ring could even be prevented by safety regulations. However, as the accuracy of the data seems promising, the data stream is continuous, and the device seems to be reasonably durable, the device might be suitable for some, non-physical tasks in a factory.

4.5 Emotiv Insight

4.5.1 User Experience and Perceived Accuracy. The experts evaluated the user experience for Emotiv Insight device as poor. A proper sensor contact, between the device and the sculpt, was impossible to achieve, and the device itself was uncomfortable to wear. One of the experts reported that after wearing it for an hour, her head began to ache.

The accuracy of the detected metrics could not be evaluated, due to the low usability of the device.

4.5.2 Applicability to Factory Workers. For the above reasons, the device is not suitable for real-life settings.

4.6 Summary of the Expert Evaluation

According to the results, Samsung Gear S3 and Fitbit Charge seemed to be the most suitable devices for factory workers for measuring activity, heart rate, and sleep. Moreover, these devices have additional strengths: Samsung Gear S3 can be also used for communication, while Fitbit Charge HR is compact in size (unobtrusive) and has a reasonably long battery duration, with a continuous heart rate measurement. However, none of the reviewed trackers for emotional state seemed completely suitable for manufacturing work. The Moodmetric ring and the Spire breathing device were the best options, amongst the evaluated emotion trackers, but both have their downsides. The accuracy of the Moodmetric ring is good, but wearing the rather large ring during manufacturing work is foreseen as being impossible or at least inconvenient. The Spire device would fit well in factory environments, in terms of its ease of use, but its accuracy was not perceived as convincing.

5 Ethical Considerations

This section presents the results of the ethical assessment workshop. In the following, the considerations related to each ethical aspect are summarized. Altogether, 47 expert comments were collected in the workshop. The aspects of privacy (7 comments), autonomy (8 comments), as well as health and safety (8 comments) received the largest number of ethical remarks, which highlights the importance of their role in adopting wearable self-monitoring devices at work.

Privacy is a crucial issue, when adopting this technology at work, due to the personal and sensitive nature of the collected data. The privacy of the user should be guaranteed by ensuring data security and by appropriate design of the wearable devices used. Only data, which is needed for analysis and feedback should be collected. The user's state should not be visible to others, and others should not notice the device's feedback, for example, when indicating a fast heartbeat or calming the user. If the data is used for making anonymous summaries, then the privacy of work teams and the possible cues to identify workers may pose ethical issues.

To support *autonomy*, the employer should guarantee true voluntariness of the adoption of this technology. The line between voluntariness and pressure to start using the device may be thin; the user may feel social pressure or be afraid of negative consequences, such as being stigmatized as a difficult person resistant to changes. The user's autonomy should also be respected, when using the device, by offering the user the possibility to decide what data is collected and how the feedback is given (what device, vibration/sound etc). The feedback should be in the form of recommendations or encouragement, suggesting actions rather than instructing the user.

Any technology should respect the *integrity and dignity* of its users. In this case, dignity needs to be considered when using the wearable device and in the form of the feedback. Is it possible that the device will create a negative image of the user, to herself or others? Could the worker start to consider herself as a nervous person because the device frequently indicates high stress? These possibilities can partly be reduced by giving the feedback in a discrete and appropriate format, by recognizing positive moments and trends, and by giving gentle guidance, and not only indicating the physiological signs. Still, the feedback should be truthful to be valuable to the user.

Reliability of this technology can be approached on several levels. First, it is important to achieve the user's trust in the ways the data is utilized. The user should not have concerns of the employer seeing her personal data, if the data is only collected for the employee's personal use. In addition, the accuracy of the self-monitoring data should be reliable.

E-inclusion was not identified as an issue that would be particularly relevant in the context of factory work. When adopting self-monitoring technologies, however, it would be important to include also older employees in the pilot groups, if possible.

Encouraging self-monitoring at the workplaces is a significant societal issue, which should be implemented to provide *benefit to society and* the workers. Is it possible to truly empower workers and increase their well-being or is it possible that self-tracking creates new pressure, concerns and mental load? Can it create awareness that guides the user towards a healthier life or does it encourage excessive self-monitoring, increasing workers' stress?

Supporting *health and safety* is a crucial issue, as it is one reason for introducing this technology. The wearable devices should be able to meet the individual needs of different users. For example, encouraging the user to collect as many steps as possible may not be a good idea for everyone; sometimes rest would be more essential. It is important to consider whether the device may strengthen the current feeling, such as a feeling of stress, panic or tiredness, when one sees the physiological data related to it. The device should not only remind the user of one's state, but it should also give discreet guidance to change the undesired feeling, for example, by guiding the user in calming down.

The issues of *responsibility* become critical in cases of misuse of the employees' data or accidents related to manufacturing work. Wearing the device and using the accompanied app should not distract the factory worker's attention from the work tasks. Even though use of the device would include gamification features, the user should be able to concentrate on work.

The possibilities to start using self-tracking devices at work or opt out of using them should convey a feeling of *justice.* If the employer offers the opportunity to use the new devices to only some of the employees, others may feel excluded. Feelings of jealousy or inequality may hamper the workplace spirit. On the other hand, some employees may find using the new devices troublesome; - as work becomes increasingly digitalized, they may add to the technological load, by introducing another new technology to learn.

Adoption of self-monitoring devices is likely to have an impact also on *social interactions* at the workplace and raise discussion among workers. Even if the data was not officially shared with others, one may feel social pressure to share it with workmates in a situation where others ask about it or share their data with others. Using self-monitoring devices may also decrease interaction with others in situations where the device is sufficient in assisting the user.

The ethics assessment workshop highlighted several ethical issues to be considered when factory workers are offered an opportunity for self-tracking at work. Some of the ethical issues form ethical dilemmas that need special consideration. For example, due to privacy reasons, the collection of unnecessary data should be avoided. However, to support the user's reflection and thus, to ensure the value of self-tracking, a rich variety of data can prove to be useful.

6 Discussion

Quantified Self research has mainly been focused on self-tracking practices of consumers and early adopters [4,14,20,23], but it has potential also in work context to increase workers' awareness of their well-being and work performance. However, work context raises different requirements and also ethical considerations. This study investigated the potential five different wearable self-tracking devices in a factory context, and ethical issues related to adopting self-tracking on the factory floor.

According to the expert evaluation, the two wrist devices, Fitbit Charge HR and Samsung Gear S3, were evaluated as being the most potential for factory workers. The experts considered them easy to adopt, easy to use, reliable (especially in terms of data accuracy) and valuable for receiving interesting data. The experts considered the data provided by the emotion trackers interesting. Unfortunately, they did not perceive any of the three evaluated emotion trackers suitable for factory work, due to their restricted usability or perceived inaccuracy.

To be applicable in factory settings, a wearable device needs to be **unobtrusive** during wear, in order to not distract from work tasks, to draw unnecessary attention or to disturb wearing personal safety equipment. **Discrete feedback**, such as vibration feedback, is desirable, due to its practicality in noisy environments and its invisibility for others. The device needs to provide **reliable data**, for example not interpreting physical activity as mental arousal. Both the **immediate feedback and the longer-term trend data** are expected to be valuable. The instant feedback may help, for instance, in staying calm or decreasing long sedentary periods, when needed, while trend data enables personal reflection, which may increase awareness of issues influencing performance and well-being, and thus, lead to positive behavioural changes. To support positive changes and positive user experience, the feedback should not be given as instructions, but rather as notifications or **encouraging suggestions.**

Regarding ethical issues, the experts emphasized the importance of privacy and autonomy, as well as health and safety. The purpose of self-monitoring needs to be clearly defined and the process of collecting and using the data kept as transparent as possible. Special attention should be paid to ensuring the true voluntariness of self-monitoring. As Lupton [16] states, the line between pushed and imposed self-tracking is thin. If self-tracking at work is proposed by the employer, it becomes pushed self-tracking, which may be difficult to refuse. Also, Moore and Piwek have emphasized the same issue; opting out includes the risk of being excluded and stigmatized [18].

Taking ethical considerations into account may require balancing ethical dilemmas. When the purpose for self-monitoring is to increase the opportunities for self-reflection, it is difficult to know, beforehand, which data is relevant to the employees, which may lead to gathering a vast amount of personal data. However, to protect the privacy of the employees, unnecessary information should not be collected. The same kind of balancing is needed with the degree of autonomy provided to the employees. Even though the employees should be provided a possibility to define what information is collected from them, having too many options to decide on may increase their stress and make the adoption of self-monitoring too complicated.

In this study, the scenarios used in ethics assessment workshop naturally had an impact on the findings of the workshop. On the other hand, they helped concretising the topic from different workers' perspective and focusing on aspects identified relevant for the research project. Additional ethical issues may emerge, when involving the actual workers of the pilot sites. In this paper, we focused on the worker's perspective. Expectations of employers or managers should also be studied in the future. If accepted by the workers, anonymous quantification data could provide insights to management or employers.

The expert evaluations gave insight about the potential of different wearable self-tracking devices in factory work. Thus, they created a good basis to continue studies with actual factory workers. In the next phase of the project, we will pilot wearable self-tracking devices at the project pilot sites. The pilots will be carried out as co-designed activities, involving factory workers, as well factory management. In addition to considering user experience and ethical issues, the implementation needs to be adjusted according to the safety regulations of the factories.

7 Conclusion

This article presents an expert evaluation of selected wearable self-tracking devices, which we studied to identify the most suitable device(s) for factory workers. Based on the evaluation, the wearable wrist devices measuring activity, heart rate and sleep were evaluated as being the most suitable devices, while none of the devices measuring the emotional state was considered as suitable, due to the impracticality of the devices in the factory context or the perceived shortcomings in the accuracy of measurements. In addition, the arranged ethics assessment workshop highlighted a variety of ethical issues, especially related to privacy and autonomy of users, as well as their health and safety.

This work contributes to the field of HCI by providing insights into the potential of the use of self-tracking devices in a factory context. Unobtrusiveness of use, as well as reliable and encouraging feedback that is discretely conveyed are expected to be key attributes for the potential of self-tracking devices. As quantifying the worker is a sensitive topic, the paper highlights ethical issues, to be considered when adopting such devices at workplaces.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 723277. We wish to thank the experts who participated in the reviews of the wearable devices and the ethics workshop.

REFERENCES

[1] Stavros Asimakopoulos, Grigorios Asimakopoulos, and Frank Spillers. 2017. Motivation and user engagement in fitness tracking: Heuristics for mobile healthcare wearables. *Informatics* 4(1), 5.

[2] Kathryn Buchanan and Dan Lockton. 2015. Understanding human connectivity and the Quantified Self. *Working Papers of the Sustainable Society Network+*.

[3] Meredith A. Case, Holland A. Burwick, Kevin G. Volpp, and Mitesh S. Patel. 2015. Accuracy of smartphone applications and wearable devices for tracking physical activity data. *Jama* 313, 6: 625-626.

[4] Eun K. Choe, Nicole B. Lee, Bongshin Lee, Wanda Pratt, and Julie A. Kientz. 2014. Understanding quantified-selfers' practices in collecting and exploring personal data. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 1143-1152. DOI: http://dx.doi.org/10.1145/2556288.2557372

Benjamin Cowley, Marco Filetti, Kristian Lukander, Jari Torniainen, Andreas Henelius, Lauri Ahonen, Oswald Barral et al. 2016. The Psychophysiology Primer: A Guide to Methods and a Broad Review with a Focus on Human-

Computer Interaction. *Foundations and Trends® Human–Computer Interaction* 9, 3-4: 151-308.

[5] Batya Friedman, Peter H. Kahn Jr, Alan Borning, and Alina Huldtgren. 2013. Value sensitive design and information systems. In *Early engagement and new technologies: Opening up the laboratory*, Springer, Netherlands, 55-95.

[6] Dominic Gorecky, Mathias Schmitt, Matthias Loskyll, and Detlef Zühlke. 2014. Human-machine-interaction in the industry 4.0 era. In *Industrial Informatics (INDIN), 2014 12th IEEE International Conference on Industrial Informatics (INDIN)*, IEEE, 289-294.

[7] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. Lifelogging: Personal big data. *Foundations and Trends® in Information Retrieval* 8, 1: 1-125.

[8] Veikko Ikonen, Eija Kaasinen, and Marketta Niemelä. 2009. Defining Ethical Guidelines for Ambient Intelligence Applications on a Mobile Phone. In *Intelligent Environments (Workshops)*, 261-268.

[9] Eija Kaasinen, Franziska Schmalfuß, Cemalettin Özturk, Susanna Aromaa, Menouer Boubekeuer, Juhani Heilala, Päivi Heikkilä, Timo Kuula et al. 2018. Empowering industrial workers with engaging factory adaptation solutions. Submitted to *Computers & Industrial Engineering.*

[10] Eija Kaasinen. 2009. User Acceptance of Mobile Services. *International Journal of Mobile Human Computer Interaction* 1, 1: 79–97.

[11] Eleni Kosta, Olli Pitkänen, Marketta Niemelä, and Eija Kaasinen. 2010. Mobile-centric ambient intelligence in health- and homecare-anticipating ethical and legal challenges. *Science and Engineering Ethics* 16, 2: 303-323.

[12] Martin Lavallière, Arielle A. Burstein, Pedro Arezes, and Joseph F. Coughlin. 2016. Tackling the challenges of an aging workforce with the use of wearable technologies and the quantified-self. *Dyna* 83, 197: 38-43.

[13] 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*, 557-566.

[14] Xiao Li, Jessilyn Dunn, Denis Salins, Gao Zhou, Wenyu Zhou, Sophia M. Schüssler-Fiorenza Rose, Dalia Perelman et al. 2017. Digital Health: Tracking Physiomes and Activity Using Wearable Biosensors Reveals Useful Health-Related Information. *PLoS Biology* 15, 1:1-30.

[15] Deborah Lupton. 2016. The diverse domains of quantified selves: self-tracking modes and dataveillance. *Economy and Society* 45, 1: 101-122.

[16] MIT Technology Review: Stephen Wolfram on Personal Analytics. Retrieved June 19, 2017 from https://www.technologyreview.com/s/514356/stephen-wolfram-on-personal-analytics/

[17] Veronica P. Moore, and Lukasz Piwek. 2016. Regulating wellbeing in the brave new quantified workplace. *Employee Relations* 39, 3: 308-316.

[18] Céline E. Nihan. 2015. Ubiquitous computing in the workplace: ethical issues identified by the interdisciplinary IWE and HRM Research Group. In *Ubiquitous Computing in the Workplace*, 75-93. Springer International Publishing, 2015.

[19] Jeungmin Oh and Uichin Lee. 2015. Exploring UX issues in Quantified Self technologies. In *Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*, 53-59. IEEE.

[20] Lukasz Piwek, David A. Ellis, Sally Andrews, and Adam Joinson. 2016. The rise of consumer health wearables: promises and barriers. *PLoS Medicine* 13, 2: 1-9.

[21] Rock Health. 2016. Digital Health Consumer Adoption: Report, December, 2016. https://rockhealth.com/reports/digital-health-consumer-adoption-2016/

[22] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 1163-1172. DOI: http://dx.doi.org/10.1145/2556288.2557039

[23] Anna C. Shcherbina, Mikael Mattsson, Daryl Waggott, Heidi Salisbury, Jeffrey W. Christle, Trevor Hastie, Matthew T. Wheeler, and Euan A. Ashley. 2017.Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. *Journal of personalized medicine* 7, 2: 3.

[24] Melanie Swan. 2012. Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. *Journal of Sensor and Actuator Networks* 1, 3: 217-253.

[25] Melanie Swan. 2013. The quantified self: fundamental disruption in big data science and biological discovery. *Big Data* 1, 2: 85-99.

[26] Kaisa Väänänen-Vainio-Mattila, and Minna Waljas. 2010. Evaluating user experience of cross-platform web services with a heuristic evaluation method. *International Journal of Arts and Technology* 3, 4: 402-421.

[27] David Wright. 2011. A framework for the ethical impact assessment of information technology. *Ethics and information technology* 13, 3: 199-226.