How Does Sleep Tracking Influence Your Life? Experiences from a Longitudinal Field Study with a Wearable Ring

ELINA KUOSMANEN, University of Oulu, Finland
AKU VISURI, University of Oulu, Finland
SABA KHEIRINEJAD, University of Oulu, Finland NIELS
VAN BERKEL, Aalborg University, Denmark HELI
KOSKIMÄKI, Oura Health, Finland
DENZIL FERREIRA, University of Oulu, Finland
SIMO HOSIO, University of Oulu, Finland

A new generation of wearable devices now enable end-users to keep track of their sleep patterns. This paper reports on a longitudinal study of 82 participants who used a state-of-the-art sleep tracking ring for an average of 65 days. We conducted interviews and questionnaires to understand changes to their lifestyle, their perceptions of the tracked information and sleep, and the overall experience of using an unobtrusive sleep tracking device. Our results indicate that such a device is suitable for long-term sleep tracking and helpful in identifying detrimental lifestyle elements that hinder sleep quality. However, tracking one's sleep can also introduce stress or physical discomfort, potentially leading to adverse outcomes. We discuss these findings in light of related work and highlight the near-term research directions that the rapid commoditisation of sleep tracking technology enables.

1 INTRODUCTION

Increasing technological advancements in sensor technologies, algorithms, and electronic components have enabled new forms of self-tracking. These developments have led to new standalone devices, *e.g.* Fitbit and Polar watches, as well as the myriad of applications that run on today's widely available sensor-rich smartphones. One of the more recent advances in self-tracking focuses on sleep quality. Novel devices, such as the Oura Ring[1] and the Withings sleep tracking mat[2], offer users the opportunity to obtain detailed insights into their sleep patterns and subsequently to

[1]https://ouraring.com/
[2]https://www.withings.com/mx/en/sleep

increase their overall well-being. Earlier research has, however, also raised concerns regarding the use of self-tracking technology, including difficulties in maintaining interests [51], the normative nature of these devices and related applications [45], and challenges in converting initial insights into long-lasting habits [47]. Identifying the needs and habits of end-users is critical to inform the design of self-tracking applications and technologies.

Obtaining sufficient sleep is essential in terms of both physical and mental health [34, 50]. Supporting people in obtaining sufficient sleep is therefore a critical goal towards increasing overall well-being. Prior work on sleep and self-tracking has investigated the practices of data collection and analysis [9, 25, 39, 53]. New consumer-targeted tracking tools offer automatic and unobtrusive data collection, visualisations, and recommendations towards achieving sleeping goals, raising new questions in relation to the interaction with collected sleep tracking data.

We present a longitudinal field study ($N = 82$) on the experience of using a wearable sleep tracking ring. We focus on user perceptions of the self-tracking data and how it affects their daily life decisions. Through an analysis of interview and questionnaire data, we answer the following research questions:

RQ1 How does the use of a sleep tracking wearable influence end-user's sleep-related behaviour?
RQ2 How do end-users reflect on the reliability of the collected sleep data?

We contribute to a better understanding of the daily use of sleep tracking technology and highlight implications for future self-tracking technologies. We replicate and expand on prior work on self-tracking technology in terms of user categorisation [22, 25, 39, 53] and data perception [7, 9, 27].

2 RELATED WORK

2.1 Measuring Sleep Quality

Sleep quality and quantity have a many-faceted impact on both our physical and mental well-being [36]. For example, sleep affects our immune system and blood sugar level, as well as our ability to learn, memorize and make logical decisions [52]. Sleep can be measured in different dimensions, such as quantity, continuity, and timing. Sleep's physiological level is typically described in sleep stages [5]. The measurement of these sleep stages has traditionally been performed through polysomnography [3, 6], primarily in clinical settings [14].

While the accurate physiological tracking of sleep, including transitions between different stages of sleep, requires clinical-grade measurement devices, the rise of affordable smartphones and wearable consumer electronics has enabled a new class of applications for measuring and reflecting on one's sleep. Ko et al. identified five primary delivery platforms for consumer sleep technologies: mobile device, wearable, embedded, desktop or website and accessory appliance platforms [23]. Mobile device solutions may use inertial sensors for sleep detection, such as accelerometer, microphone and ambient light sensor, or rely on traditional diaries [10, 23, 40]. For non-intrusive sleep tracking, embedded platforms, including sensors that can be placed into the bed or in the bedroom, are developed [23, 40]. Activities carried out during the daytime, such as exercise or diet, affect sleep [44]. Hence, wearable sensors tracking body movements or biometric information providing continuous data throughout day and night are relevant to sleep tracking [23]. Such wearables, including the Oura ring, usually utilize accelerometer in sleep tracking (such as Fitbit Flex, Jawbone UP) [21]. They may also include heart rate, blood flow, respiratory rate, and movement sensors, with prior work highlighting their ability to accurately offer detailed information on the wearer's sleep stages [12, 13]. In addition to the accelerometer, the Oura ring measures body temperature and utilizes PPG for heart rate and respiration measurements.

Sleep hygiene is an often-used term for behavioural and environmental factors affecting sleep quality; for example, drinking coffee and exercise are behavioural factors, and bedroom lighting or temperature relate to the sleep environment [31, 38]. Sleep tracking can increase the awareness of sleep quality and help to identify the factors affecting one's sleep, but not necessarily help to improve the sleep quality [26]. Sleep quality consists of several contributors. In addition to sleep duration, modern sleep trackers provide the duration and timing of different sleep stages. However, it is not known how to directly affect the sleep stages, *i.e.* amount of REM sleep, via behaviour changes [38].

Prior work has investigated the validity of sleep tracking devices [12, 13], including validation of the Oura ring against polysomnography [13] and actigraphy [32]. This paper focuses on how tracking activity impacts the user's behaviour and how users reflect on their personal sleep data.

2.2 Data Perception in Sleep Tracking

Due to a lack of standardisation in sleep tracking technologies, most consumer devices base their sleep analysis on their proprietary algorithms, with many of them missing validation against clinical-grade measurement protocols [41]. Most consumer sleep tracking devices use data generated from in-built accelerometers to define sleep parameters [21]. The growing literature contrasting these devices against actigraphy or polysomnography illustrates that this approach tends to underestimate sleep disorders and overestimate sleep efficiency and total sleep times in typical sleepers [24], and should not be considered as a substitute for PSG [19]. The benefit of consumer sleep tracking devices is that they measure sleep quality in the natural environment [38]. In addition, the adherence towards tracking, i.e. how regular and continuous the data collection is, may affect the perceived accuracy of the device [48].

Nevertheless, the user perception of the device accuracy is often not based on technical validation. Yang et al. [54] identified two different methods for assessing the accuracy of their tracking devices. In 'folk-testing', users compare the measurement data with a reliable source or crosscheck readings with another device. In 'ad-hoc testing', the user compares a measurement result against one's intuition. Related, Liu et al. [29] analyzed data from five online forums on the use of sleep-tracking technologies and identified the users' concerns towards data trustworthiness. Similar to the 'folk-testing' in [54], users reported the data from different devices might differ. The users had doubts about the sleep recognition, for example, in which they suspected sensors to detect the movement or voice from a sleep partner, which maps to 'ad-hoc testing' [54]. Sometimes users detected a clear error in sleep detection, such as the sleep tracker detecting sleep on the way to work. According to [37], if metrics do not reflect the user's beliefs, the user may generate doubts about other collected data, causing a mistrust towards the whole system.

Perceived quality of sleep is highly subjective [28], *e.g.* how much sleep is enough to feel rested? In addition to our daytime actions, our biochemistry affects our sleep [1]. Hence, also the perception of sleep tracking metrics is subjective and might lead to a mismatch between the metrics and the user's perception of sleep quality [27, 38]. Lack of trust in the data that the wearable captured reduces user acceptance [49] and might challenge the long-term adoption of wearable devices [11]. According to Shih et al. [42], unrealistic expectations of the device's technical capabilities can cause a perception of inaccuracy.

Liang et al. [26] studied the barriers to improving sleep through the use of sleep tracking. They interviewed 12 participants who had used Fitbit to track their sleep for at least 1.5 months. They found that self-tracking raised participants' awareness of sleep patterns, but only three participants actively changed their habits to improve their sleep. The authors identified reasons for participants not changing their (sleep) habits. The users did not know what 'normal' sleep is, as they were missing a reference to the population data. The accuracy of sleep detection was also a problem. The users were unable to identify reasons for sleep problems and did not know how to improve their sleep. Based on their findings, Liang et al. provide three design considerations related to data accuracy, visualisation, providing guidance and personalised recommendations.

The perceived credibility of sleep trackers was studied in [27]. A total of 22 students (with an average age of 25) tracked their sleep using three different devices simultaneously: Fitbit Charge 2, Neuroon EEG, and SleepScope. The participants studied their sleep metrics from one night across all three devices and were subsequently interviewed about their experiences and the perceived credibility of each device. The authors describe that, in addition to the sleep data provided by the device, the functionality of the device, as well as the interface and physical appearance affected the perception of credibility. The conflicting data between three devices forced the user to decide which data source they believed to be the most accurate. The strategies to relate their own experience of sleep quality with the provided data were categorized into three groups: 1) favouring subjective experience, 2) combining subjective experience with the device data, and 3) favouring the device data.

Related to data perception, Choe et al. [7] analyzed participants' descriptions of self-reflection on sleep tracking and categorized the self-reflections according to the certainty level, dividing the self-reflection into Conclusive findings and Hypothesises. Conclusive findings were further refined to neutral statements, confirmation of existing knowledge, and disproof of existing knowledge. In another paper, Choe et al. [8] studied how users of self-trackers discover insights from the data during the self-reflection. They defined types of visualization insights as Recall, Detail, Comparison, Trend, Value judgement, Distribution, Correlation, Outlier, Data summary, and Prediction. The most common insight type in their study was Recall, with further subtypes describing the perception towards the data, External context (explaining a past scenario seen in the data), Confirmation (confirming existing knowledge) and Contradiction (contradicting existing knowledge). They focused on the conclusions and interpretation of data, not the people's actions as a result of these obtained insights. We will discuss how data perception in [7] and [8] relates to our findings in Discussion.

2.3 Types of Self-Tracking Users

Following the initial introduction of pedometers (also known as step counters) as early as the 18th century [30], technological developments have provided consumers access to a wide range of devices for activity tracking. Karapanos et al. [22] recruited users of activity-tracking wristbands (Fitbit, Jawbone UP, and Nike+ Fuelband) from Amazon Mechanical Turk for an online survey. They divided the users into 'purposive' and 'explorative' groups based on their initial purchase decision. The 'purposive' group had a clear use case for the device, such as achieving a healthier lifestyle, whereas the 'explorative' bought the device intuitively, or received it as a gift.

Rooksby et al. [39] interviewed 22 people using or wanting to use a pedometer or activity tracker. They identified five distinct styles of tracking personal data. 'Directive tracking' is goal-driven, aiming to *e.g.* lose weight, and set a suitable goal, or adopt a goal set by the tracker, such as daily step count. Partly overlapping, some participants tracked activity to collect rewards, either competing against others, participating in a lottery, or achieving some (self-set) goals. In 'documentary tracking', people focused on documenting activities instead of changing their behaviour. In 'diagnostic tracking', the user is looking for connections between different life elements, such as finding a reason for occasional stomach problems. 'Fetished tracking' was done due to enthusiasm for technology. In addition to these, Rapp et al. [37] found yet another category, 'playful tracking': New users try out the various features of the device in a playful manner to assess the potential benefit of using it.

Li et al. interviewed 15 users of self-tracking tools identified six themes of questions related to the data collected from themselves: 'Status', 'History', 'Goals', 'Discrepancies', 'Context', and

Fig. 1. The Oura Ring and the Oura app, showing user's health data summarised into three scores: readiness, sleep, and activity.

'Factors' [25]. They noticed people asking different questions at different times and defined two phases of data reflection: 'Maintenance' and 'Discovery'. In the 'Maintenance' phase, the user has already identified a goal to be achieved and factors affecting their behaviour. In the 'Discovery' phase, the user has not (yet) identified a goal or the factors influencing their behaviour. The user can move between the phases; a user in the 'Discovery' phase can identify a connection between actions and data that is of interest to them, and through this connection discover a goal and move to the 'Maintenance' phase, as the user in 'Maintenance' phase might reach the goal and move to the 'Discovery' phase.

Whooley et al. [53] analysed 51 videos from a Quantified Self website. They found self-tracking to be driven by two intentions, 'self-improvement' and 'curiosity'. Similar 'Maintenance' phase in [25], the 'self-improvement' group had a specific goal to achieve, and they used tracking to quantifying the factors contributing to the goal and outcomes. The 'curiosity' group corresponds the 'Discovery' phase, in which tracking was not goal-oriented and also relates to the 'playful tracking' by [37]. In the discussion, we compare these user categorisations from prior work with our findings.

3 STUDY PROTOCOL

The goal of this study is to investigate whether and how sleep tracking impacts end-user's habits, in particular in relation to their sleep-related behaviour. Further, we are interested in learning how end-users reflect on the collected self-tracking data provided through sleep trackers.

3.1 Apparatus

The participants used the Oura Ring and the accompanying Oura mobile application (Fig. 1). Further, the participants installed a lightweight logging application, the goal of which was limited to tracking the use of the Oura mobile application. This logging application required no user interaction. We selected the Oura Ring for our study due to its advanced sleep tracking capabilities [13, 32].

The ring measures the duration and timing of different sleep stages as well as how quickly the user falls asleep. PPG is used for heart rate related measurements and tracking respiration during the night. Based on the metrics shown in Fig. 1, a daily sleep score describing an overall rating for sleep quality is calculated. The application provides insights on sleep stages (REM, light, deep), total sleep time, and sleep latency. Oura also provides data related to activity and recovery. Fig. 1 shows the activity and readiness score contributors. Based on individual historical metrics, the application sets a target for each day's activity and detects whether it is achieved. The daily readiness score describes the balance between sleep, activity, and recovery and is personalised according to the user's data. The ring data is stored locally and synced to the participant's smartphone when the Oura application is active. The ring data also synchronises to a cloud service. Each participant was asked to share their data with the researchers through the cloud service. The ring has a battery life of 4–5 days, can be worn during all activities and in all normal outside temperatures, and is waterproof.

3.2 Participants and Protocol

We recruited participants via student and employee email lists from the local university, as well as people from other communities through word of mouth. Initially, 237 potential participants volunteered to take part in the study. We randomly invited 100 participants among the volunteers, while ensuring equal distribution between men and women and three predetermined age groups (-24, 25-34, 35+). To reduce the potential for participation bias, we did not provide any monetary rewards for participating. The self-tracking device was loaned to the participants by the Oura company for the duration of the study. Participants returned the devices following the completion of the study. Our study uses version two of the Oura ring, for which no monthly subscriptions are offered. Following a number of participants that failed to show up or dropped out throughout the study period, a total of 82 completed the study (39 men, 43 women). The age range of the participants was 18–61, with an average age of 27.3. The age of our sample provides a typical representation of health tracker users, with a recent survey indicating that most users are within the age groups 20–29 and 30–39 [46].

Participants attended three meetings over the study period. In the first meeting, the study was introduced and participant consent was collected. We installed the logging application and took measurements for correct ring size. In the second meeting, approximately two weeks from the initial meeting, participants received their rings and were given written instructions for installing the Oura mobile application. The instructions did not introduce the content of the app or the collected data, the participants were guided to follow the Oura mobile application for those details. The participants were asked to wear the ring whenever appropriate according to their daily activity, *i.e.* in a natural way and not enforced by the researchers. The total study period from the first meeting to the end of the study was approximately two months. In practice, the actual data collection period varied from 25 to 105 days (M = 65, SD = 22.5).

At the end of the study period, we organised a face-to-face interview in which we discussed participants' experiences in sleep tracking with the Oura ring. Due to the COVID-19 pandemic, we conducted the final interview as an online questionnaire for approximately 40% of our participants, following the same set of questions and themes that were used in the earlier interview sessions. This paper focuses on the results of the interviews and overall use of the sleep tracking ring. This combined interview/questionnaire data collection focused on the following main categories:

- How frequently the participant paid attention to the data and insights offered by the ring and the Oura application.
- Which particular sleep-related metrics and details participants paid attention to.
- Did wearing the ring influence their sleep habits, both in terms of planning or sleep performance.

- Did they notice large differences in sleep habits between workdays and holidays or holiday periods.
- Did wearing the ring reveal any new insights about the participant's sleep habits or learn anything new about their behaviour or physical well-being.

Participants were able to voice their experiences in any category throughout the interview. We made extensive written notes during the interview sessions, no audio or video material was recorded. The online questionnaire contained multiple-choice questions as well as open-ended questions and free commentary sections. Our data collection procedure is in line with both our own institution's as well as the Finnish National Advisory Board on Research Ethics guidance, which does not prescribe a formal ethical review for the given study context [35]. We ensured that participants were informed of the potential risks in participating, followed best-practices in relation to handling data storage, and informed participants of their right to withdraw their consent at any point throughout the study prior to the data collection phase.

4 DATA ANALYSIS

4.1 Methodology

Based on the interviews and questionnaires, we report the results in terms of how often the participants investigated their data, perceived reliability of the provided sleep data, and, most importantly, the participants' own ideas of their sleep quality as well as the lifestyle changes that were introduced as a direct result of introducing sleep tracking. For each of these overarching themes, two of the paper's authors conducted a thematic analysis [4] to identify different underlying behaviours and attitudes. In classifying the different user archetypes, the coders first individually read through the interview data, after which they agreed upon a set of labels. Participants were subsequently categorised as based on their overall comments through a collaborative labelling process. We invited a third author to dispute any of the themes of the individual responses, ending up with constant comparison [17] based coding for the participant responses. The coded categories and the number of participants in each category are presented in Table 1. This analysis also led us to derive a distinct set of user archetypes that emerged among all the analysed material.

4.2 Self-Reported Frequency of Sleep Tracking

As one of the goals of this study we wanted to establish an idea of how frequently people engaged in sleep tracking, i.e. intentionally observed the logged and aggregated sleep statistics. First, based on the participants' own descriptions in the interview, we categorised participants into two groups: 'daily' and 'rare' users. Participants who use the app (almost) every day were denoted as 'daily' users (65/82 participants). Some of these participants described that they use the app several times throughout the day; "*Yes, daily. Often I was interested to check the sleep score and readiness right after waking up. I also checked the activity during the day. I paid attention if some value was red.*" (P20). The 'rare' users (17/82 participants) told they checked the app a few times a week, or after a couple of days. P129 shared "*Because I was really sceptic about the data provided by the ring during the sleep, I didn't follow the data often.*" (P129)

To cross-validate participants' self-perceptions on their sleep-tracking behaviour against the ground truth data, we analysed the available log data from both the logging application that was installed to the users' devices as well as the data provided by Oura's mobile application. Validating the interview-based classification, the majority of the participants in our field study indeed tracked their sleep rather extensively throughout the study period. We utilised the ring's data to count the number of nights the ring was worn (with enough battery). The rate of nights the ring was worn varied between 49% and 100%. For 'rare' users, the ring was on for 85% of nights, whereas for 'daily users’ this was true for 96% of nights – with 27 ‘daily’ users wearing the ring every night during the study period. Based on the application usage logs, the Oura mobile application was launched on average 4.4 (SD = 7.61) times per day per participant.

For both groups we measured how the daily usage frequency of the application changes over time. Using a linear regression model we can observe a decrease in usage time per day over the course of the study period ($p < .005$, R2 = .12) and a less of an impact on daily usage frequency ($p <$ .005, R2 = .01). Novelty effect typically declines after initial use [33, 43] and using a threshold of two weeks we can observe a clear decline in usage frequency (7.6 > 2.7 times per day, $p < .0005$ using Student’s t-test, for during and after first two weeks respectively). Usage time over a full day remains similar during the whole study period, with an average daily application usage time of 55.9s (SD = 88.2s). Additionally, one-way ANOVA shows that self-reported ‘daily’ participants used the application for longer periods on average (M = 71.5s) and their usage frequency declined less than the average during the study period ($p < .005$, F = 5.92). The ‘rare’ users show similar significant differences with an average application usage time of 36.6s. Lastly, using Student’s t-test, we find a significant difference in daily application usage between ‘daily’ (average of 5.53 times per day) and ‘rare’ (average of 4.25 times per day), with $p < .005$, although the distinction is not as clear as the participants’ self-perception of their usage frequency. Despite the extensive study period (ranging from 1–3 months for participants), device usage is likely to shift when considering usage extending beyond the study period.

4.3 Perceived Reliability of Sleep Data

We divided the participants to three categories based on their own perception of the data or, in other words, how well they felt the data matched to their own interpretation and understanding of their sleep quality. Approximately half of the participants (52%) were ‘neutral’, as they never mentioned anything specific about the validity of the data in the interviews/questionnaires, and did not express agreement or disagreement with the data provided.

The ‘trusting’ participants (24%) felt the ring metrics ensured their own perception of sleep quality, and that the metrics correlated with own physical sensations. Participants reflected on the observations provided by the application in relation to their own experiences. For example; “*[...] it correlates with my actual alertness level. When I was hiking for four days, the resting heart rate was higher, I think the strain of the travel affected to that.*” (P05).

The ‘doubtful’ participants (23%) often described a mismatch between their own experiences and the ring’s metrics. “*All the time the ring seems to complain about the quality of sleep, even though I felt I slept well. Sometimes the ring requires unnecessarily me to rest for a day or to calm down after hard exercise.*” (P54). Similarly; “*Often I felt good and spry, but the ring indexes showed something totally different. I have learned to be very critical about data provided by this kind of devices, since your own feeling is the only valid measure.*” (P113).

4.4 User Archetypes

Having a clear definition of a products’ users helps in product creation cycles and is largely considered as a key design strategy when implementing new products [18]. As a key outcome of the detailed participant response analysis, we derived five distinct user archetypes; ‘passive’, ‘aware’, ‘reactive’, ‘motivated’, and ‘critical’ (Table 1). Figure 2 illustrates the shares of each user archetype and how they perceived the data from the Oura device.

18 percent of participants were classified as ‘passive’ users. They did follow the sleep metrics, but did not observe a strong connection with their daily life. This category contains the relatively largest share of doubtful users, 50% (9/18). Even among the participants which considered the data to be reliable, the self tracking of sleep failed to make an impact on their lives; “*I sleep less than I*

Data perception	Distribution	Description
Neutral	52% (43/82)	Not mentioning anything about validity of the data, not agree or disagree with the results.
Trusting	24% (20/82)	Data ensured their perception of the sleep, and correlated with own feeling.
Doubtful	23% (19/82)	Data mismatch with own feeling.
User archetype	**Distribution**	**Description**
Aware	37% (30/82)	Drew conclusions of own behaviour based on the sleep metrics, and noticed a connection between actions and sleep. May recognise the need of change.
Passive	22 % (18/82)	Followed the sleep metrics, but perceived them as observations,they did not make any conclusions or connect the metrics to own their own behaviour.
Reactive	20% (16/82)	Sometimes took action to improve the own sleep quality, either actively based on the data, or afterwards noticed that they had sometimes changed routines.
Motivated	16% (13/82)	Actively tries to improve the sleep quality based on the metrics.
Critical	6% (5/82)	Experienced the sleep tracking had a negative effect to sleep.

Frequency	Distribution	Other device	Distribution
Daily	79% (65/82)	**No**	74% (61/82)
Rare	21% (17/82)	**Yes**	26% (21/82)

Table 1. Overview of participant categorisation following the analysis of interview and application usage data. The table presents distribution of participants according to their perception towards data, user archetypes, how often they engage with the app, and if they have another tracking device in use.

thought. The ring did not affect my sleeping habits or sleep quality. I consider the data mainly reliable. It somewhat reflected my feelings." (P132).

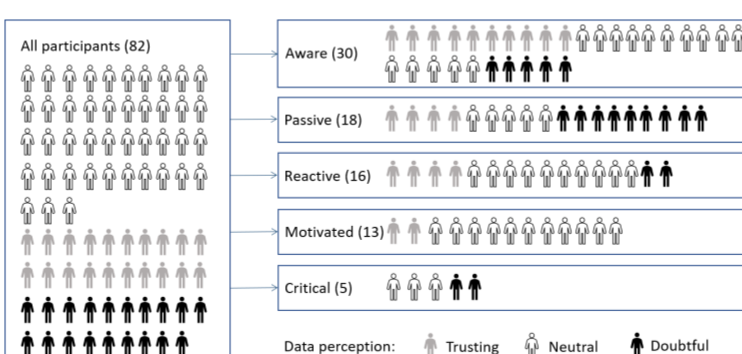

Fig. 2. Data perception across user archetypes. Numbers in brackets indicates the amount of participants. In 'all participants' section each full row adds up to 10 participants.

Over third of the users (37%) were categorised as 'aware' users. They acknowledged that their habits and behaviour effects their sleep quality. This awareness did, however, not necessarily result in changes in participant behaviour; "*I learned how much low amount of sleep affects to the readiness on following day, and how much even a small amount of alcohol affects to the sleep quality. The ring didn't affect to my sleep routines or sleep quality, even though it increased my awareness.*" (P109).

20% of users were considered as 'reactive'. These users undertook some actions to improve their sleep quality. They might have tried something, or afterwards noticed that in practice they changed their habit. As an example, P17 told "*Using the ring did not affect to my sleeping routines much, but I started to think about sleeping and why I keep waking up in the night. I have thought about going to sleep and what I do before it. [...] I have tried to affect the amount of REM and deep sleep. Often I sleep very light, too short periods of REM and deep sleep.*" (P17). The data provided by the ring allowed these participants to make concrete steps to improve their sleep quality: "*I don't get enough REM-sleep, and I might wake up during a night surprisingly often. I thought earlier that the quality of my sleep is quite good, but seems there is room for improvement.[...] It made the maintaining the sleep rhythm easier, since the app told when is good to go to sleep. Usually if I obeyed it, the next days, study requirements and challenges were easier.*" (P140).

16% percent of participants were classified as 'motivated' users, who actively target to improve the quality of their sleep. One of the motivated participants described: "*Yes it changed my sleeping routines so that if the sleep score was bad, I tried to go bed earlier. It motivated to sleep better after a bad night, hence go to bed earlier.*" (P70). Several participants showed signs of being highly target oriented, expanding beyond the scope of the sleep metrics: "*I tried to go to sleep regularly at the same time, and also wake up always at the same time. I added exercise, since I wanted to achieve the activity targets. I noticed I sleep better and more than before I had the ring. I wanted to reduce alcohol consumption, since I noticed it has an radical affect to the quality of sleep.*" (P138).

Finally, we note that a small portion of participants (6%) had a negative experience from wearing the ring and dubbed this archetype as 'critical'. As stated by these participants, the additional data caused additional stress or resulted in a competition with themselves; "*Wearing the ring affected negatively, because when staring at the numbers you think a lot about number of waking up. I tried to get good results.*" (P33); "*The ring stressed me and it was difficult to sleep with it.*" (P30).

4.5 Perceiving and Improving Sleep Quality

The interviews provided a rich array of perspectives on the quality of sleep as well as the use of technology in informing people about (changes to) their sleep quality. We identified three core themes in the participants' responses: factors that affect sleep quality, metrics used by the participants, and the changes that resulted in people's lives as a result of tracking sleep.

4.5.1 Factors Affecting Sleep Quality. During the interviews we noticed several factors that the participants expressed to have had an influence their sleep quality. The most frequently mentioned factor was time of going to bed (22 participants), which participants often outlined as the most critical factor; "*The time of going to bed matters to having all sleep stages.*" (P71).

Participants mentioned the wake-up time (three participants), amount of sleep (ten participants), and sleep rhythm (three participants) to affect sleep quality. Fifteen participants state that the amount, intensity, and/or timing of exercise to affect their sleep quality. Seven participants noted that exercising late causes a higher resting heart rate during the night. For example; "*As a person exercising a lot, I learned to combine different sensations in my body to the data provided by the app, such as high resting heart rate.*" (P139).

Participants report to actively explore the effect of external factors on their sleep quality. In addition to sports, this for example also included food consumption (eleven participants) and alcohol (fourteen participants). "*It was fun to follow the resting heart rate, especially when drinking alcohol. It caused restless sleep and high heart rate.*" (P74).

Mental and physical discomfort and pain were also discussed in relation to participant sleep quality. Sickness or pain was mentioned as decreasing the sleep quality by twelve participants. The effect of stress was mentioned by eight participants, with some describing how stressful periods could be identified through the sleep data. Other decreasing factors mentioned by participants were noise (six participants), the sleep environment (four participants), and medication (one participant). Five participants described the ring itself negatively affected their sleep. Three of these participants mentioned that the ring caused stress about their sleep quality.

4.5.2 Relevance of Self-Tracking Metrics. Based on the interview data, we calculated the frequency with which in-app metrics were discussed by the participants. Participants either discussed that they took note of the metrics, or described otherwise observations or conclusions related to a metric. Here, we report the most frequently discussed metrics.

60 out of the total of 82 participants mentioned heart rate, with 55 specifying that they were interested in the resting heart rate. Heart rate variation was mentioned by 22 participants. Furthermore, we note that 64 of the participants mentioned sleep stages as a metric of interest in monitoring their sleep quality. A total of 38 participants said that they assess their REM sleep, and an identical number of participants said they assess their deep sleep duration. Twenty-one participants mentioned light sleep, and 20 mentioning the night-time awake time metric as relevant.

The Oura application also keeps track of the user's physical activity, with 28 participants in our sample stating that they keep track of their activity levels through the app. Participants are interested in a variety of physical metrics, including temperature (seven participants), readiness score (four participants), breathing (two participants), step count (one participant), and recovery (one participant).

4.5.3 Changes following Self-Tracking of Sleep. In the interviews, the 'motivated' and 'reactive' participants described the actions they took for improving the sleep quality, as well as one of the participants classified as 'critical'. Seventeen participants revealed the type of changes they have made to their lives as based on the data provided by the Oura ring. Ten told they follow the recommendations provided by the Oura app.

In addition to the application-based recommendations, many participants described self-initiated actions they took for improving their sleep. Most of these actions were related to the timing of their sleep (twelve participants). Participants described how the information obtained through the wearable provided an initial incentive for action: "*The ring inspired to fixing the sleep rhythm. I even paid more attention to the sleep quality and amount, and balancing the sleep rhythm. Using the ring is a good incentive for good actions in terms of sleep, when in the morning you were able to check the results and compare it to own feelings.*" (P137).

Six participants mention a change in their actions in relation to eating or drinking, and five participants state that they now exercise more. One participant mentioned that he tries to achieve the activity target as set by the Oura ring, providing a concrete incentive for daily exercise goals; "*There was an increase in my activity on many days, when I was able to check the step count from the Oura-app.*" (P121).

4.5.4 Challenges in Self-Tracking of Sleep. The interviews also revealed that participants experienced various challenges in relation to the collection and interpretation of the self-tracking data. Nine participants said that although the data is motivating or encouraging them to be more active, it is sometimes challenging to find a balance between looking at the numbers and respecting their own felt experience; "*It might start to control your life too much, if you follow it very closely. Then* *you might forget to listen to your body, even though that is the most important thing. Hence, it would be important to learn how to interpret your body utilising the device, not just read the device. For example, if the app says that your readiness is bad, you might believe it even though it wouldn't be correct.*" (P113).

The constant availability of data could result in compulsive behaviour among users or demand too much of their time and focus. Participants commented that the way data is presented could result in unintended consequences related to their perception of rest and sleep. "*I noticed I am addicted to the data given by the ring. For example, the first thing I do after waking up, is to check the sleep data.* (P133), and "*It made me a bit neurotic related to sleep. Sleeping should not be considered as a performance.*" (P131).

5 DISCUSSION

In this paper, we present participants' experiences with a wearable sleep tracker, the Oura ring. We now reflect on the archetypes we identified in our study and contrast our findings with earlier work in the self-tracking domain. Subsequently, we discuss the meaning of trust towards data in sleep tracking. We conclude our discussion with an overview of future directions for sleep tracking research.

5.1 Reflections on Wearable User Archetypes

As wearable sleep tracking technology is relatively new, the future development of these devices can benefit from identifying different classes of users [18]. Through our thematic coding process, we identified five different user archetypes, distinguishing the behaviour and reactions of the participants towards their sleep data as 'motivated', 'reactive', 'aware', 'passive', or 'critical'.

As described in the Related Work section, prior research has identified different types of tracker users based on their motivation or actions [22, 25, 37, 39, 53]. Through a contrasting of these user categories (see Table 2), we can identify a clear overlap between the different categorisations presented in prior work. All papers identify the target-driven users, trying to achieve some goal (either self-set or provided by the tracking device) through self-tracking. We label these as 'motivated' users. In our study, these participants use the data for achieving better sleep quality, are willing to change their habits, and have a precise aim to improve their sleep quality. Many participants also mention they aim to achieve the activity goals as set by the app or to improve their sleep or a specific sleep subcategory (*e.g.*, amount of deep sleep). Next, the user group deriving conclusions from their data, being curious, and exploring ways to utilise the data is also identified in prior studies. Through our interviews, we further refine this category based on whether or not the user

Study	Target driven	Conclusion driven	Data driven	Experience driven	Addiction driven
[22]	Purposive	Explorative	-	-	-
[25]	Maintenance	Discovery	-	-	-
[39]	Directive, Collecting rewards	Diagnostic Playful*	Documentary	-	Fetished
[53]	Self-improvement	Curiosity	-	-	-
Our study	Motivated	Reactive & Aware	Passive	Critical	-

Table 2. The user categories identified across different studies, highlighting which categories broadly overlap or contain similar characteristics. The star denotes the extension by [37] to the categories of [39].

adjusts their behaviour based on the sensor readings. The ‘reactive’ and ‘aware’ groups all made conclusions between actions and sleep quality. The ‘reactive’ group, in contrast to the ‘aware’ group, also makes life changes based on the data to achieve better sleep. In addition to these two obvious user archetypes, Rooksby et al. named the users focusing merely on documenting activities as ‘documentary tracking’ [39]. In our study, the ‘passive’ group merely collected data; they might check it daily but did not draw any conclusions based on this data. Rooksby et al. found some people collect data due to interest in gadgets or the data itself [39], referring to the group as ‘fetished tracking’. In our study, this interest in gadgets was not brought up in the interviews.

The ‘critical’ user archetype which we identified was not discussed in earlier studies. Sometimes the activity tracking might reduce joy in the activity itself [16], hurting sleep quality. As in earlier studies [2], we also noticed that increased awareness and that not following recommendations might cause feelings of guilt. Related to this, Epstein et al. have studied the user experiences of those who abandon their tracking device [15]. Among other reasons, they find ‘discomfort with information revealed’ and ‘data quality concerns’ to be reasons for abandoning the device. These experiences are seen in our ‘critical’ group. Hence the ‘critical’ users will probably not continue using the device. Self-selection bias may explain why this archetype has not appeared as widely in prior work. Karapanos et al. collected data through MTurk, asking the users of tracking devices to answer a survey. The ‘critical’ users might have already stopped tracking or are not interested in taking such tasks due to negative experiences [22]. Whooley et al. [53] used videos of Quantified Self websites, biasing data collection towards enthusiastic trackers. Li et al. [25] recruited 15 current users of self-tracking or personal informatics tools. They first identified six kinds of question themes related to the self-tracking data. Subsequently, they defined the phases of ‘Maintenance’ and ‘Discovery’ by identifying the different questions users have towards the self-tracking data in these phases. The ‘critical’ user might not be identified through this procedure as they do not have such questions towards the data to begin with. Rooksby et al. [39] invited people who use or have previously used a pedometer or activity tracker to join their study. In addition, they provided an option to borrow the device during the study period. Hence, among the participants borrowing the device, the ‘critical’ user archetype could have appeared. However, only four out of 22 participants in their sample were new to self-tracking and opted to borrow a device from the researchers. In our study, only five of the 82 participants were ‘critical’, highlighting their relatively small number and the subsequent difficulty to be identified in studies with smaller sample size.

Rapp et al. [37] studied how beginning self-trackers differed from experienced trackers and expanded the categorisation of [39] with a category describing especially non-experienced trackers, ‘playful tracking.’ This user category experiments with how their actions affect the data and try to identify the benefit of using the device. In our study, these users are labelled as ‘reactive’ or ‘aware’ users, depending on how the users react on the data provided. Rapp et al. pointed out that beginning self-trackers are often less goal-driven than experienced users [37]. In our study, there were also very target-oriented users, the ‘motivated’ users, even though not many had prior experience on activity or sleep tracking. Users faced problems in tracking in [37], with some similar observations as in our study; some felt physical discomfort, and some felt mistrust towards data accuracy or even towards the whole system. However, Rapp et al. did not categorise these users as a separate group [37].

The different user archetypes (Table 1) reacted differently towards the presented data. The three user archetypes who identified the connection between their actions and sleep quality (‘aware’ users) and even took steps towards better sleep (‘motivated’ and reactive’ users), mainly experienced sleep tracking as something positive. The users merely read the numbers without interpreting them in relation to their body or actions (‘passive’ users), did not see benefits in using the tracker, but no significant disadvantages either. We want to stress the experiences of the ‘critical’ user archetype.

This user group is often overlooked – with the notable exception of the work by Spiel et al. who highlight the normative nature of self-tracking applications [45] – and reports feelings of stress due to the self-tracking data. These users often find a strong contradiction between the tracking data and their own experience.

5.2 Meaning of Trustworthiness in Sleep Tracking Data

Our analysis raised several questions regarding the participants' trust towards the self-tracking data. Half of the participants were found to have a 'neutral' perspective towards the data, and 19 (23.1%) participants were doubtful about the data presented to them. Whereas any difference between the reported metrics and the user's assessment was experienced merely as a surprise to the 'trusting' users, the 'doubtful' participants quickly considered the collected metrics unreliable. Similar to 'ad-hoc testing' in [54], the 'doubtful' participants often perceived a discrepancy between their own intuition of sleep quality and the provided metrics.

Choe et al. categorized sleep tracking self-reflections according to certainty levels [7], dividing self-reflection into Conclusive findings and Hypotheses. Conclusive findings were further refined to neutral statements, confirmation of existing knowledge, and disproof of existing knowledge. Another study categorized insights based on the content of the insight [8] (*e.g.*, identifying details in the data, or comparing different time segmentations). Recall was most common, in which the user connects the phenomenon in the data to his behaviour. Recall is divided into three subtypes: External content (insight where an occurrence explains the phenomenon in the data), Confirmation (data confirms existing knowledge), and Contradiction (conflicts with existing knowledge).

These findings [7, 8] relate to our data perception categories. Users with 'neutral' data perceptions often described the effect of past scenarios, such as noticing stress or sickness affecting data. This maps to the neutral statements in Conclusive findings [7] and the External context subtype [8]. The 'trusting' users felt the metrics confirmed their feelings, following confirmation in Conclusive findings [7] and the Confirmation subtype [8]. The 'doubtful' participants described the discrepancy between the metrics and their sensations, overlapping with contradiction in Conclusive findings [7] and the Contradiction subtype [8]. Liang et al. [27] identified three categories in users' strategies for combining their own experiences with the data provided by the sleep trackers. The category of participants favouring subjective experience would be labelled as 'doubtful' in our study, participants combining subjective experience with the sleep metrics as 'neutral', and the ones favouring the device data as 'trusting' users.

While we find similarities in our results, [7] and [8] consider data perception as a feature of a certain type of insight (Conclusive findings [7], Recall [8]). Instead, we posit data perception as a feature of the user, covering the full use of tracking devices, *e.g.* if users doubt the device's accuracy they will also doubt the collected data and are less likely to act based on collected findings. This approach is supported by [37], who found that occurrences of experienced mismatch in data may rapidly cause mistrust for the whole system.

We note that accurately measuring sleep is not straightforward, with some participants noticing that *e.g.* watching TV was erroneously categorised as sleeping. This resulted in lower overall trustworthiness towards the metrics. Literature shows that the perceived trustworthiness is connected to user acceptance in activity tracking [49]. As sleep classifications are unlikely to become 100% reliable shortly, we align ourselves with prior research [27] that suggests that sleep tracking applications should make their detection algorithm more transparent and allow the user to understand why misinterpretations happen. For example, when watching TV, the stillness and decreasing heart rate might lead to sleep detection. In addition, the users should be able to make corrections to the collected data (*e.g.*, remove, edit duration) – as currently many applications already do. This could help to overcome some of the frustrations and mistrust towards the collected and reported data.

5.3 Future Directions and Design Implications

Sleep tracking provides a tool for users to increase self-awareness of their sleep. Sleep duration, which has been considered the most critical aspect of sleep in several questionnaire-based studies, has recently shifted into the background. Sleep quality, including dimensions of quantity, continuity, and timing, is now considered the new norm. The possibility to understand night-by-night variation in these aspects allows us to obtain a better understanding of human well-being. In addition, sleep, as a stable and undisturbed measurement time for several biosignals and values, will enable a broader perspective for human physiology and recovery. Yet, presenting these new types of information to end-users in a valuable and informative way is not straightforward. As stated by Jensen et al., "*providing the 'right' information within the framing of 'smart' technology adds complexity to how such information is understood and experienced in the context of everyday life*" [20].

Although the duration of our study was substantial, how participants change between the different archetypes is outside of our paper's scope and remains an area for future work. Yet, it is likely that participants shift between different roles in their use of self-tracking technology. Identifying these moments can help application designers to better support the users. For example, how can we incentivise 'aware' users to use the information they already consume to inform a change in their behaviour? Similarly, when users reduce their interest in self-tracking sleep or managing their sleep quality, the messages and support offered by the application may no longer be appropriate.

The experience of sleep quality is individual, and people differ in which sleep features they pay attention to [28]. A general statistical scale can determine how much sleep a person needs, considering user demographics, but there are personal differences to consider. Companion applications often show personalised scoring mechanisms, yet many of our participants reported not being able to affect these metrics. Most of the participants (78%) paid attention to metrics related to the sleep stages when monitoring sleep quality, with sleep stages as the most often mentioned sleep contributor. Considering there are no direct behavioural actions known to affect the duration or timing of the sleep stages [38], following sleep stages when making actions towards better sleep might be misleading. Hence, there is a clear need for sleep self-tracking technology to be personalised to the user's needs and focus on metrics that the user can affect through their behaviour. In addition, the personality of the user, or the user type, affects how they experience the data and recommendations provided by the app, and the application should consider this in the level and tone of the feedback. Although the companion app in our study does not allow users to 'compete' with others in terms of sleep (quality), our interview data revealed that users often 'compete' with their historical data – sometimes causing participants additional stress and even an inability to fall asleep. How to best manage this tradeoff between user insight and the potentially undesired side-effect of 'self-competition' remains an open challenge for future work focusing on sleep tracking.

5.4 Limitations

Our study contains several limitations which may have affected the results presented in our paper. First, as we did not provide any monetary reward for joining the study, our participant pool likely contained participants with a pre-existing interest in sleep tracking technology. Therefore, we expect that the distribution of the archetypes as shown in Table 1 would differ if the study was run on a random sample of participants. Second, we experienced technical problems with the logging application that caused a subset of the log data to be lost. We have adjusted the analysis on this part by only considering data from days where the logs contain entries from a minimum of eight hours. The application usage logs were missing from a total of twelve participants. However, this did not have a major impact in our study, as our focus in this paper is in the interview data. Restrictions due to the COVID-19 pandemic affected the second half of the study. A few participants indeed noted having changed sleeping habits due to restrictions in our country. Since we did not analyze participants' actual sleep metrics but instead focused on the experience of using the device, the pandemic's effect on our work is minimal.

6 CONCLUSION

In this paper, we present user experiences of users of a novel ring-form sleep tracking device. Through a longitudinal field study with 82 participants, we collect insights into their self-tracking experiences, focusing on which data was most relevant to the participants and how they reacted to these data. Analyzing the interview and self-tracking data, we identify five distinct user archetypes; 'motivated', 'reactive', 'aware', 'passive', and 'critical'. These archetypes highlight differences in the user interpretation and use of self-tracking data and their actions in response to the data. In addition to these user archetypes, our results highlight differences between user perceptions and the data as reported by the application. A majority of participants found the sleep tracking to support maintaining their sleep rhythm and offering concrete decision support towards improving their sleep quality. However, some participants struggled to find a balance between the tracking device and their perceptions of their sleep. Some participants even experienced additional stress from the tracking device, reporting pressure or guilt about their 'sleep performance'. These results highlight the importance of carefully designing self-tracking technology for a variety of end-users. Future work should explore in more detail how these differences in participant profiles can be considered in the design of sleep-focused self-tracking technologies.

REFERENCES

[1] Saeed Abdullah, Mark Matthews, Elizabeth L Murnane, Geri Gay, and Tanzeem Choudhury. 2014. Towards circadian computing: "early to bed and early to rise" makes some of us unhealthy and sleep deprived. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 673–684.

[2] Jared S Bauer, Sunny Consolvo, Benjamin Greenstein, Jonathan Schooler, Eric Wu, Nathaniel F Watson, and Julie Kientz. 2012. ShutEye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1401–1410.

[3] Konrad E Bloch. 1997. Polysomnography: a systematic review. *Technology and health care* 5, 4 (1997), 285–305.

[4] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[5] Daniel J Buysse. 2014. Sleep health: can we define it? Does it matter? *Sleep* 37, 1 (2014), 9–17.

[6] Andrew L Chesson Jr, Richard A Ferber, June M Fry, Madeleine Grigg-Damberger, Kristyna M Hartse, Thomas D Hurwitz, Stephen Johnson, Gihan A Kader, Michael Littner, Gerald Rosen, et al. 1997. The indications for polysomnography and related procedures. *Sleep* 20, 6 (1997), 423–487.

[7] Eun Kyoung Choe, Bongshin Lee, Matthew Kay, Wanda Pratt, and Julie A Kientz. 2015. SleepTight: low-burden, self-monitoring technology for capturing and reflecting on sleep behaviors. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 121–132.

[8] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 173–182.

[9] Eun Kyoung 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 SIGCHI Conference on Human Factors in*

Computing Systems. 1143–1152.

[10] Yong K Choi, George Demiris, Shih-Yin Lin, Sarah J Iribarren, Carol A Landis, Hilaire J Thompson, Susan M McCurry, Margaret M Heitkemper, and Teresa M Ward. 2018. Smartphone applications to support sleep self-management: review and evaluation. *Journal of Clinical Sleep Medicine* 14, 10 (2018), 1783–1790.

[11] James Clawson, Jessica A Pater, Andrew D Miller, Elizabeth D Mynatt, and Lena Mamykina. 2015. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 647–658.

[12] Massimiliano de Zambotti, Aimee Goldstone, Stephanie Claudatos, Ian M Colrain, and Fiona C Baker. 2018. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. *Chronobiology international* 35, 4 (2018), 465–476.

[13] Massimiliano de Zambotti, Leonardo Rosas, Ian M Colrain, and Fiona C Baker. 2019. The sleep of the ring: comparison of the O¯URA sleep tracker against polysomnography. *Behavioral sleep medicine* 17, 2 (2019), 124–136.

[14] Neil J Douglas, Stephen Thomas, and Mohammed A Jan. 1992. Clinical value of polysomnography. *The Lancet* 339, 8789 (1992), 347–350.

[15] Daniel A Epstein, Monica Caraway, Chuck Johnston, An Ping, James Fogarty, and Sean A Munson. 2016. Beyond abandonment to next steps: understanding and designing for life after personal informatics tool use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1109–1113.

[16] Jordan Etkin. 2016. The hidden cost of personal quantification. *Journal of Consumer Research* 42, 6 (2016), 967–984.

[17] Barney G Glaser, Anselm L Strauss, and Elizabeth Strutzel. 1968. The discovery of grounded theory; strategies for qualitative research. *Nursing research* 17, 4 (1968), 364.

[18] Jonathan Grudin and John Pruitt. 2002. Personas, participatory design and product development: An infrastructure for engagement. In *Proc. PDC*, Vol. 2.

[19] Shahab Haghayegh, Sepideh Khoshnevis, Michael H Smolensky, Kenneth R Diller, and Richard J Castriotta. 2019. Accuracy of wristband Fitbit models in assessing sleep: systematic review and meta-analysis. *Journal of Medical Internet Research* 21, 11 (2019), e16273.

[20] Rikke Hagensby Jensen, Jesper Kjeldskov, and Mikael B. Skov. 2018. Assisted Shifting of Electricity Use: A Long-Term Study of Managing Residential Heating. *ACM Trans. Comput.-Hum. Interact.* 25, 5, Article 25 (2018), 33 pages. https://doi.org/10.1145/3210310

[21] L Jeon and Joseph Finkelstein. 2015. Consumer sleep tracking devices: a critical review. *Digital Healthcare Empowering Europeans: Proceedings of MIE2015* 210 (2015), 458.

[22] Evangelos Karapanos, Rúben Gouveia, Marc Hassenzahl, and Jodi Forlizzi. 2016. Wellbeing in the making: peoples' experiences with wearable activity trackers. *Psychology of well-being* 6, 1 (2016), 4.

[23] Ping-Ru T Ko, Julie A Kientz, Eun Kyoung Choe, Matthew Kay, Carol A Landis, and Nathaniel F Watson. 2015. Consumer sleep technologies: a review of the landscape. *Journal of Clinical Sleep Medicine* 11, 12 (2015), 1455–1461.

[24] Bhanu Prakash Kolla, Subir Mansukhani, and Meghna P Mansukhani. 2016. Consumer sleep tracking devices: a review of mechanisms, validity and utility. *Expert review of medical devices* 13, 5 (2016), 497–506.

[25] Ian Li, Anind K Dey, and Jodi Forlizzi. 2011. Understanding my data, myself: supporting self-reflection with ubicomp technologies. In *Proceedings of the 13th international conference on Ubiquitous computing*. 405–414.

[26] Zilu Liang and Bernd Ploderer. 2016. Sleep tracking in the real world: a qualitative study into barriers for improving sleep. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. 537–541.

[27] Zilu Liang and Bernd Ploderer. 2020. How does Fitbit measure brainwaves: A qualitative study into the credibility of sleep-tracking technologies. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–29.

[28] Zilu Liang, Bernd Ploderer, Wanyu Liu, Yukiko Nagata, James Bailey, Lars Kulik, and Yuxuan Li. 2016. SleepExplorer: a visualization tool to make sense of correlations between personal sleep data and contextual factors. *Personal and Ubiquitous Computing* 20, 6 (2016), 985–1000.

[29] Wanyu Liu, Bernd Ploderer, and Thuong Hoang. 2015. In bed with technology: challenges and opportunities for sleep tracking. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*. 142–151.

[30] Richard MacManus. 2015. *Trackers: How Technology is Helping Us Monitor & Improve Our Health*. David Bateman.

[31] David F Mastin, Jeff Bryson, and Robert Corwyn. 2006. Assessment of sleep hygiene using the Sleep Hygiene Index. *Journal of behavioral medicine* 29, 3 (2006), 223–227.

[32] Milad Asgari Mehrabadi, Iman Azimi, Fatemeh Sarhaddi, Anna Axelin, Hannakaisa Niela-Vilén, Saana Myllyntausta, Sari Stenholm, Nikil Dutt, Pasi Liljeberg, and Amir M Rahmani. 2020. Sleep Tracking of a Commercially Available Smart Ring and Smartwatch Against Medical-Grade Actigraphy in Everyday Settings: Instrument Validation Study. *JMIR mHealth and uHealth* 8, 11 (2020), e20465.

[33] Jochen Meyer, Merlin Wasmann, Wilko Heuten, Abdallah El Ali, and Susanne CJ Boll. 2017. Identification and classification of usage patterns in long-term activity tracking. In *Proceedings of the 2017 CHI conference on human*

factors in computing systems. 667–678.

[34] Anne B. Newman, F. Javier Nieto, Ursula Guidry, Bonnie K. Lind, Susan Redline, Eyal Shahar, Thomas G. Pickering, and Stuart F. Quan for the Sleep Heart Health Study Research Group. 2001. Relation of Sleep-disordered Breathing to Cardiovascular Disease Risk Factors : The Sleep Heart Health Study. *American Journal of Epidemiology* 154, 1 (07 2001), 50–59. https://doi.org/10.1093/aje/154.1.50

[35] National Advisory Board on Research Ethics. 2019. *Ethical principles of research in the humanities and social and behavioural sciences and proposals for ethical review*. Retrieved March 22, 2022 from https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf

[36] June J Pilcher, Douglas R Ginter, and Brigitte Sadowsky. 1997. Sleep quality versus sleep quantity: relationships between sleep and measures of health, well-being and sleepiness in college students. *Journal of psychosomatic research* 42, 6 (1997), 583–596.

[37] Amon Rapp and Federica Cena. 2016. Personal informatics for everyday life: How users without prior self-tracking experience engage with personal data. *International Journal of Human-Computer Studies* 94 (2016), 1–17.

[38] Ruth Ravichandran, Sang-Wha Sien, Shwetak N Patel, Julie A Kientz, and Laura R Pina. 2017. Making sense of sleep sensors: How sleep sensing technologies support and undermine sleep health. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 6864–6875.

[39] John Rooksby, Mattias Rost, Alistair Morrison, and Matthew Chalmers. 2014. Personal tracking as lived informatics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1163–1172.

[40] Ibrahim Sadek, Antoine Demarasse, and Mounir Mokhtari. 2019. Internet of things for sleep tracking: wearables vs. nonwearables. *Health and Technology* (2019), 1–8.

[41] Anita Valanju Shelgikar, Patricia F Anderson, and Marc R Stephens. 2016. Sleep tracking, wearable technology, and opportunities for research and clinical care. *Chest* 150, 3 (2016), 732–743.

[42] Patrick C Shih, Kyungsik Han, Erika Shehan Poole, Mary Beth Rosson, and John M Carroll. 2015. Use and adoption challenges of wearable activity trackers. *IConference 2015 Proceedings* (2015).

[43] Grace Shin, Yuanyuan Feng, Mohammad Hossein Jarrahi, and Nicci Gafinowitz. 2019. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. *JAMIA open* 2, 1 (2019), 62–72.

[44] John M Shneerson. 2009. *Sleep medicine: a guide to sleep and its disorders*. John Wiley & Sons.

[45] Katta Spiel, Fares Kayali, Louise Horvath, Michael Penkler, Sabine Harrer, Miguel Sicart, and Jessica Hammer. 2018. Fitter, Happier, More Productive? The Normative Ontology of Fitness Trackers. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI EA '18)*. Association for Computing Machinery, New York, NY, USA, Article alt08, 10 pages. https://doi.org/10.1145/3170427.3188401

[46] Statista. [n.d.]. *Percentage of the global population that used a mobile app or fitness tracking device to track their health as of 2016, by age*. https://www.statista.com/statistics/742448/global-fitness-tracking-and-technology-by-age/

[47] Katarzyna Stawarz, Anna L. Cox, and Ann Blandford. 2015. Beyond Self-Tracking and Reminders: Designing Smartphone Apps That Support Habit Formation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2653–2662. https://doi.org/10.1145/2702123.2702230

[48] Lie Ming Tang, Jochen Meyer, Daniel A Epstein, Kevin Bragg, Lina Engelen, Adrian Bauman, and Judy Kay. 2018. Defining adherence: making sense of physical activity tracker data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–22.

[49] Daniel Trommler, Christiane Attig, and Thomas Franke. 2018. Trust in activity tracker measurement and its link to user acceptance. *Mensch und Computer 2018-Tagungsband* (2018).

[50] Norifumi Tsuno, Alain Besset, and Karen Ritchie. 2005. Sleep and depression. *The Journal of Clinical Psychiatry* (2005).

[51] Niels van Berkel, Chu Luo, Denzil Ferreira, Jorge Goncalves, and Vassilis Kostakos. 2015. The curse of quantified-self: an endless quest for answers. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*. 973–978.

[52] Matthew Walker. 2017. *Why we sleep: Unlocking the power of sleep and dreams*. Simon and Schuster.

[53] Mark Whooley, Bernd Ploderer, and Kathleen Gray. 2014. On the integration of self-tracking data amongst quantified self members. (2014).

[54] Rayoung Yang, Eunice Shin, Mark W Newman, and Mark S Ackerman. 2015. When fitness trackers don't'fit' end-user difficulties in the assessment of personal tracking device accuracy. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 623–634.