To Challenge the Morning Lark and the Night Owl: Using Smartphone Sensing Data to Investigate Day–Night Behaviour Patterns

RAMONA SCHOEDEL[1]*, FLORIAN PARGENT[1], QUAY AU[2], SARAH THERES VÖLKEL[3], TOBIAS SCHUWERK[4], MARKUS BÜHNER[1] and CLEMENS STACHL[5]

[1]Psychological Methods and Assessment, Department of Psychology, LMU Munich, Munich, Germany
[2]Computational Statistics, Department of Statistics, LMU Munich, Munich, Germany
[3]Media Informatics Group, Institute of Informatics, LMU Munich, Munich, Germany
[4]Developmental Psychology, Department of Psychology, LMU Munich, Munich, Germany
[5]Department of Communication, Stanford University, USA

Abstract: For decades, day–night patterns in behaviour have been investigated by asking people about their sleep–wake timing, their diurnal activity patterns, and their sleep duration. We demonstrate that the increasing digitalization of lifestyle offers new possibilities for research to investigate day–night patterns and related traits with the help of behavioural data. Using smartphone sensing, we collected in vivo data from 597 participants across several weeks and extracted behavioural day–night pattern indicators. Using this data, we explored three popular research topics. First, we focused on individual differences in day–night patterns by investigating whether 'morning larks' and 'night owls' manifest in smartphone-sensed behavioural indicators. Second, we examined whether personality traits are related to day–night patterns. Finally, exploring social jetlag, we investigated whether traits and work weekly day–night behaviours influence day–night patterns on weekends. Our findings highlight that behavioural data play an essential role in understanding daily routines and their relations to personality traits. We discuss how psychological research can integrate new behavioural approaches to study personality. © 2020 The Authors. European Journal of Personality published by John Wiley & Sons Ltd on behalf of European Association of Personality Psychology

INTRODUCTION

Are there times of day when you do not use your smartphone at all? Most likely at night. As our everyday companions, smartphones can provide much information about people's day–night patterns (Harari et al., 2016). So far, behavioural manifestations of the underlying circadian system like sleep–wake timing, diurnal activity, or sleep duration have mainly been assessed via self-reports (Adan et al., 2012). However, self-reports about behaviour are known to differ from actual records of behaviour (Baumeister et al., 2007; Gosling et al., 1998). Emphasizing this dilemma, Lauderdale et al. (2008) correlated behaviourally assessed sleep duration with self-reports and concluded that people systematically misjudge it. An alternative approach is to collect actigraphy-based data to study sleep behaviour: movements and environmental factors like ambient brightness are recorded with wristbands and are jointly converted to indicators for sleep–wake timing by special algorithms (e.g.

Križan & Hisler, 2019; Tonetti et al., 2016; Vitale et al., 2015). Regarding the trade-off between measurement accuracy and ecological validity, another interesting complement for studying sleep behaviour could be the use of smartphone sensing data. These data cannot provide a direct measurement of sleep–wake phases, but only periods of nightly inactivity of smartphone use in which physiological sleep occurs. In contrast to actigraphy, these measurements do not take body signals such as movements or pulse into account. However, first studies have indicated that smartphone data provide useful information about sleep–wake timing as smartphones are meanwhile considered to be part of new sleeping habits (Chen et al., 2013; Min et al., 2014; Lin et al., 2019; Borger et al., 2019). Borger et al. (2019) have shown that indicators for sleep onset and offset derived via actigraphy and smartphone touch interactions are highly correlated. In addition, independence from sensors worn on the body also offers advantages in terms of ecological validity. With the help of commercially available smartphones, behavioural indicators for sleep–wake timing can be collected efficiently and unobtrusively in everyday life over a more extended period, even for large samples. To illustrate this, we use smartphone-sensed indicators for sleep–wake timing to investigate traits related to day–night patterns. For this purpose, we chose to study three frequently researched questions, which we will introduce in the following sections.

*Correspondence Ramona Schoedel, Leopoldstraße 13, 80802 Munich, Germany.

Email: ramona.schoedel@psy.lmu.de

Individual differences in behavioural day–night patterns

The human circadian system has been studied for decades by interdisciplinary research teams. The most prominent finding across all research disciplines is that individuals show stable differences in day–night patterns, a stable trait that is often referred to as the *chronotype* (e.g. Adan et al., 2012; Cavallera & Giudici, 2008; Roenneberg et al., 2003). Literature frequently describes two extremes: the *morning type* ('morning lark') wakes up and goes to bed early, feels fit after getting up, and performs best early in the day. The *evening type* ('night owl') wakes up and goes to bed later, feels tired after waking up, and performs best towards the end of the day (for extensive reviews, see Adan et al., 2012; Cavallera & Giudici, 2008; Takano et al., 2014). The chronotype has been argued to be a genetically predisposed trait with various biological manifestations like body temperature or hormone levels (Bailey & Heitkemper, 2001; Horne & Östberg, 1976; Roenneberg et al., 2003; Katzenberg et al., 1998). In addition, chronotype should be distinguished from sleep duration, which has been argued to be an independent trait (Roenneberg et al., 2007).

Based on the distinction between variable-centred and person-centred personality assessment (Asendorpf, 2003), one might assume that chrono 'types' refer to distinct groups of individuals with similar manifestations in chronotype-related behaviours. However, Putilov (2017) points out in his review that researchers have not yet reached an agreement on the number and content of underlying dimensions, the resultant number of types, and whether the conceptualization as types makes sense at all (Roenneberg et al., 2003). Two different operationalizations of chronotype are most prominent in the literature (see Table 1).

Dating back to Horne and Östberg (1976), chronotype is described as *circadian* or *morningness–eveningness preferences*. The term 'circadian typology' is often used synonymously and shows the emphasis on the categorization of chronotypes in this research tradition (e.g. Adan et al., 2012; Lipnevich et al., 2017). In comparison, Roenneberg et al. (2015) accentuate the chronotype as a continuous variable and describe it as a trait reflecting the *phase of entrainment*, which represents individual differences in the synchronization of the internal circadian rhythm to environmental factors (e.g. light/dark cycle, diurnal temperature curve, social interaction). Despite their different understanding of the underlying construct of chronotype (Roenneberg, 2015), both operationalizations have been found to be strongly correlated (Zavada et al., 2005). In the present study, we take the structural ambiguity of chronotype as our starting point to investigate how smartphone sensing data reflecting day–night activity patterns could help to inform chronotype research, as operationalized both in the Horne–Östberg and in the Roenneberg tradition.

In the Horne–Östberg tradition, the Morningness–Eveningness Questionnaire (MEQ; Horne & Östberg, 1976) still represents the gold standard for chronotype assessment (Putilov, 2017). The MEQ asks for circadian preferences and categorizes people according to *ad hoc* specified cut-off values (Horne & Östberg, 1976). In the development of the MEQ, neither the grouping nor the factorial structure was investigated. Cut-off values were determined using a small but not representative sample (Caci et al., 2009). Meanwhile, various derivates and short scales of the MEQ have been published (Adan et al., 2012; Putilov, 2017). Assumptions on the underlying structure of circadian preferences range from a continuum with two extremes (Natale & Cicogna, 2002; Tonetti et al., 2016), over two dimensions (morningness and eveningness as separate dimensions Lipnevich et al., 2017) to a multidimensional construct with up to four factors (Adan et al., 2012; Randler et al., 2016; Caci et al., 2009). Recently, Preckel et al. (2019) have published pioneering work on a typology of circadian preferences providing empirical evidence on the possible number of types. In an adolescent sample, they found evidence for four types resulting from the combination of the two independent dimensions of morningness and eveningness preference. Joining this search for structure, we translate the questionnaire items typically used to determine the Horne–Östberg chronotype into behavioural smartphone sensing equivalents. Smartphone usage variables can

Table 1. Description of the two most popular approaches to chronotype

Feature	Horne–Östberg chronotype	Roenneberg chronotype
Assessment	Morningness–Eveningness Questionnaire (MEQ) by Horne & Östberg (1976)	Munich Chronotype Questionnaire (MCTQ) by Roenneberg et al. (2003)
Chronotype as	Time of day preferences	Phase of entrainment
Items	Ask for imagined free days: preferred sleeping times, preferred times for mental/physical activity, subjective feeling in the morning/evening, self-reported chronotype	Ask for both free and work days habitual sleeping times
Determination of chronotype	Cut-off values classify participants according to their 19-item sum score	Midpoint of sleep for free days without alarm clock usage
Emphasized structure	Four dimensions (peak time, morning affect, retiring, rising) according to Caci et al. (2009)	Continuous variable

The structure for the Horne–Östberg chronotype refers to the original chronotype assessment with the MEQ. However, several derivates of the MEQ have been developed and there is no consensus in research about the factorial structure of the chronotype approximated by the assessment of circadian preferences. Solutions range from one to four dimensions.

approximate many of them. Following Putilov's (2017) recommendation to consider behavioural markers for circadian preferences, we investigate whether we can find types of individuals with similar smartphone usage patterns indicating circadian preferences. Finally, we explore the factorial structure of the behavioural indicators.

In the Roenneberg tradition, freely chosen sleep–wake timing is considered the best approximation of the internal circadian rhythm. Therefore, sleep–wake habits for both work and free days are assessed while controlling for alarm clock usage (Roenneberg et al., 2003; Roenneberg et al., 2015). In this taxonomy, the midpoint between sleep onset and offset determines the chronotype. This reference point for sleep has proven to coincide with nocturnal melatonin production, which in turn controls sleep–wake timing (Terman et al., 2001; Roenneberg et al., 2003; Roenneberg et al., 2007; Roenneberg et al., 2015). In this context, the Munich Chronotype Questionnaire (MCTQ), which has been repeatedly validated by behavioural (actigraphy) and biological (melatonin, cortisol) circadian system markers, is primarily used (Roenneberg et al., 2003; Roenneberg et al., 2007). Only recently, Lin et al. (2019) took up the idea to determine the Roenneberg chronotype by using smartphone sensing data and provided first indications that there is a considerable overlap between sleeping times assessed via smartphones and self-reports. However, their algorithm for characterizing a *digital chronotype* does not explicitly correspond to Roenneberg's chronotype criteria, as they did not differentiate between work and free days and were restricted to the use of a very limited range of data (screen and notification events Lin et al., 2019). We propose a more fine-grained algorithm for determining a smartphone sensing-based proxy by using only free days without alarm clock usage. To explore our smartphone chronotype, we look at descriptives and correlational analyses that were presented by Roenneberg's group to describe the MCTQ-based chronotype. For example, Roenneberg et al. (2007) found that sleep duration depends on chronotype if analysed separately for work and free days and that chronotype is related to age and gender.

Behavioural day–night patterns and personality traits

Important research questions are associations between day–night patterns, personality, and demographics. Different aspects of day–night behaviour have been addressed in this context. For example, the *morningness preference* has been linked to personality. Higher values in this dimension indicate a preference for getting up and going to bed early, feeling fit in the morning, and achieving peak performance earlier in the day (Lipnevich et al., 2017). The most established findings in meta-analyses are that conscientiousness and agreeableness are positively related to morningness (Tsaousis, 2010; Lipnevich et al., 2017). No or only small relationships in a specified direction can be found for neuroticism and openness (Adan et al., 2012; Tsaousis, 2010; Lipnevich et al., 2017). Negative relationships between morningness and extraversion were found, but only if the trait extraversion was described with Eysenck's three-factor model (Adan et al., 2012; Tsaousis, 2010). Using the five-factor model, this

association is almost zero (Tsaousis, 2010). For the sake of completeness, please note that morningness has also been found to be related to personality styles or, more precisely, with thinking and behaving styles (Díaz-Morales, 2007). Furthermore, age has been robustly related to morningness. Shifts towards eveningness in adolescence and towards morningness with increasing age (at around 50) have been reported (e.g. Adan et al., 2012; Cavallera & Giudici, 2008). Regarding gender, a meta-analysis has found that the preference for morningness is slightly higher for women compared with men (Randler, 2007). However, complex interactions between age and gender have been reported in previous literature. For example, girls at the age of 13 and 14 have a lower tendency towards morningness than their male counterparts (Mateo et al., 2012), and their peak towards eveningness is earlier (e.g. Adan et al., 2012). In addition, Randler and Engelke (2019) have shown a complex interaction between age and gender with regard to morningness preferences: young women were more and older women less morning oriented than young or older men.

In addition, associations between *sleep duration* and personality traits have been investigated, but findings have been ambiguous so far. For example, there is some evidence that individuals with higher values in neuroticism report to sleep longer (Duggan et al., 2014). According to Križan and Hisler (2019), neuroticism is not related to the mean sleep duration but positively related to the intraindividual variation in sleep duration. Some studies reported correlations between sleep duration and conscientiousness, agreeableness, or openness but not extraversion (Randler, 2008; Križan & Hisler, 2019). In contrast, other researchers did not find any evidence that sleep duration and big five personality traits are associated (Gray & Watson, 2002; Randler et al., 2017; Sutin et al., 2019). Sleep duration decreases with age (Randler, 2008) but was not found to be related to gender (Randler et al., 2017).

In summary, past research provides some evidence for associations between personality traits and day–night behaviour, but past findings are inconsistent. One possible reason for this could be that the majority of studies (except Križan & Hisler, 2019; Sutin et al., 2019) asked participants about their habits but did not include any behavioural measures of sleep. Not only might people differ in their ability to estimate their sleep duration, personality traits themselves might play a role in the evaluation of their day–night behaviours. To circumvent this issue here, we use data from smartphone sensing to derive indicators for sleep–wake behaviour and to consequently investigate their relationship with big five personality traits on factor and facet level. Additionally, we explore *sleep continuity*, which has been defined as a measure of how well people fall asleep and sleep through (Ohayon et al., 2017). Recent actigraphy-based research has found, for example, that conscientiousness and extraversion were negatively related to behavioural indicators of sleep continuity, such as wake after sleep onset. In contrast, higher scores in neuroticism were associated with more wakening (Sutin et al., 2019). As a rough smartphone-based approximation measure, we look at two aspects of sleep continuity: how often and for

how long people check their smartphones during the night. Additionally, we analyse smartphone activity logs to explore how *alarm clock usage*—particularly 'snoozing'—is related to personality.

Intraindividual and interindividual differences in day–night patterns: The social jetlag

Finally, we explore the so-called *social jetlag* hypothesis (e.g. Adan et al., 2012; Wittmann et al., 2006). Roenneberg et al. (2007) surveyed the sleep habits of more than 55 000 people using the MCTQ and found that sleep behaviour differs for work-free days versus workdays. Specifically, their findings suggest that people, on average, go to bed and awake earlier on work than on free days. Furthermore, the proportion of sleep onset and offset is smaller for workdays than for free days. It has been suggested that this effect is induced by social obligations (Wittmann et al., 2006). Thus, the pairing of late bedtimes with consistent wake-up times leads to a sleep deficit for a week. As a consequence, sleep is compensated on weekends (Roenneberg et al., 2015). This misalignment of the internal biological and the external social clock is associated with health risk behaviours (e.g. increased body mass index and smoking Roenneberg et al., 2012; Wittmann et al., 2006). According to Wittmann et al. (2006) and Roepke and Duffy (2010), late chronotypes are particularly affected by the social jetlag as they stay up until late at night but have to get up early to go to work or to pursue other social obligations on the following day. The assessment of individuals' daily routines through the analysis of smartphone activity logs for several weeks allows us to investigate compensatory nightly rest by considering intraindividual and interindividual factors. Using these indicators, we want to explore whether the smartphone-sensed proxies for sleep duration on weekends and respective weeks are related and whether interindividual factors like the Roenneberg chronotype, demographics, and personality traits have an impact.

Rationale

Our study aims to reinvestigate selected topics regarding day–night pattern-related traits by using smartphone sensing data. Because we use a new type of data in this field of research, this is exploratory work. A handful of studies have started to use smartphone data in this context (e.g. Chen et al., 2013; Min et al., 2014; Lin et al., 2019). However, these studies have mostly been limited in terms of sample size and types of sensing data.

Here, we show how behavioural records from smartphones can be used to investigate individual differences in day–night patterns, how they relate to personality traits, and how they are influenced by intraindividual and interindividual factors. Besides the examination of whether 'morning larks' and 'night owls' manifest in indicators of sleep–wake timing and diurnal activity patterns, we explore the smartphone-based operationalization of the Roenneberg chronotype. We investigate the associations of day–night behaviour patterns and personality traits. Finally, we illustrate how continuously logged behavioural data can be used to investigate the contribution of both intraindividual and interindividual factors to predict indicators for sleep behaviour on weekends, using the social jetlag hypothesis as an example.

METHOD

Our analyses are based on data collected within the long-time project *PhoneStudy* (Stachl et al., 2018). This ongoing interdisciplinary research project at LMU Munich uses the continuously developed smartphone sensing application *PhoneStudy* for Android smartphones for collecting natural smartphone usage behaviours in the field. Data about app usage, calling activity, general phone usage (e.g. calendar, music, power supply), and connectivity (e.g. Bluetooth, WiFi) are logged whenever the respective events occur. GPS data are usually recorded once every 15 minutes. Data are synchronized hourly to the back end server via Secure Sockets Layer (SSL) encryption, whenever a WiFi connection is available. The responsible institutional review board and data protection office approved the project and all associated studies. All materials and aggregated data can be found in our open science framework project (OSF; Schoedel et al., 2020).[1] To protect the data privacy rights of our participants, the raw sensing data cannot be made available due to their granularity.

Description of data set

We combined data resulting from three studies conducted between 2014 and 2018. In Table 2, we show some basic information about the included studies. Despite some marginal differences, data collection procedures of all studies followed the same principle: after giving informed consent, participants were asked to install the *PhoneStudy* app for at least 30 days on their private smartphones and to complete several questionnaires before, during, or after the smartphone logging period. Participants were mostly recruited in the university context via flyers, mailings lists, social media, and personal contact in Munich, Germany. For more detailed information about study procedures, see also Stachl et al. (2017); Harari et al. (2019); Schuwerk et al. (2019); Schoedel et al. (2018); and Stachl et al. (2019).

We applied several exclusion criteria to our initial data set of 743 participants. We excluded participants with fewer than 21 days of sensing data, more than 50% missing values across all variables, and if questionnaire data were not available. We included data from a maximum of 32 days of continuous logging. This resulted in a final sample size of 597 (61% women). As recruitment took place in the university context, participants were, on average well educated (71% with a high school and 20% with a university degree). With a mean age of 23.56 years ($SD = 6.55$; $Min = 18$, $Max = 72$), the sample was skewed towards younger participants (18–21: 39%; 22–25: 34%; 26–30: 12%; 31–40: 5%; 41 and older: 3%). For a more detailed description of the sample, according to studies, see Table 3.

[1]https://osf.io/a4h3b/

Table 2. Description of data sets used in the study

Data set	References	N	Study period	Compensation
1	Stachl et al. (2017), Harari et al. (2019)	132 (137)	09/2014– 08/2015	Individualized personality profile and 30 € or course credits
2	Schuwerk et al. (2019)	240 (245)	08/2016– 08/2017	Up to 35 € and lottery (smartphone or tablet worth 400 €)
3	Schoedel et al. (2018)	225 (361)	10/2017– 01/2018	Individualized personality profile and user activity feedback, course credits, and lottery (10 × 50 €)

N indicates the size of the sample of the respective study after application of our inclusion criteria. The total number of subjects per study is given in parentheses.

Table 3. Description of the sample according to studies

Data set	N	Age	Education	Students	Employment status
1	132	23.61 (4.73)	No qualification: 0.00% Secondary school: 3.79% High school: 65.15% University: 31.06%	No data available	No data available
2	240	22.94 (4.57)	No qualification: 0.00% Secondary school: 9.58% High school: 72.50% University: 17.92%	73.50%	No data available
3	225	24.20 (8.86)	No qualification: 0.44% Secondary school: 8.88% High school: 72.44% University: 16.44%	77.33%	Unemployed: 4.89% In training: 24.89% Minor employm.: 41.33% Part-time: 10.67% Full-time: 15.56% Other: 0.88%

N indicates the size of the samples according to studies. The column Age presents the mean value, and standard deviations are given in parentheses. As procedures slightly varied across studies, not all demographic variables are available for all data sets. The category *other* in the column Employment Status comprises retraining and pension.

Measures

Self-report measures

We administered various self-report questionnaires. However, we limit our report to the ones used in our statistical analyses. Besides demographics, personality traits were assessed with the Big Five Structure Inventory (BFSI Arendasy, 2009). Each of the big five factors—openness, conscientiousness, extraversion, agreeableness, and emotional stability—was measured on respectively six subscales (Table 8). Participants were asked to rate 300 personality describing adjectives and short phrases on a 4-point Likert scale with the labels *untypical for me*, *rather untypical for me*, *rather typical for me*, and *typical for me*. Compared with the widely used structure inventory NEO-PI-R (Costa & McCrae, 2008), the BFSI is supposed to have better psychometric properties: Cronbach *α* values (ranging between 0.72 and 0.92) are partly higher, and subscales are unidimensional in the original paper (Arendasy, 2009). In addition, the BFSI should be less dependent on the participant's reading comprehension ability as it uses short and simple items (Arendasy, 2009). The construction of the BFSI does not follow the classical test theory, but the item response theory framework. Accordingly, the BFSI has been developed in conformity with the partial credit model (Masters, 1982),

which is a probabilistic model describing an individual's observable score on a single item as the result of the functional relationship between the individual's latent trait value (person parameter) and latent item thresholds, which indirectly determine item difficulty (item parameter Arendasy, 2009). Correspondingly, we used the person parameter estimates as personality scores in all our analyses.

Day–night behavioural measures

Raw smartphone sensing data are sequences of timestamped event data. Whenever a usage event happens, a data entry specified by several event characteristics (e.g. date, study day, details about the event like app package name or type of call) is created. To get an idea of the raw data structure, see also the supplemental codebook (Schoedel et al., 2020). To investigate the research questions specified above, we created variables by reviewing the literature and translating behavioural sleep indicators into smartphone sensing behaviours. Based on our smartphone sensing data, we computed proxy variables to estimate sleep-related behaviours. Please note that our variables are likely to overestimate actual sleep as the last smartphone usage event in the evening has to be before the physiological onset of sleep, and the first smartphone usage event in the morning occurs with delay after waking up. As smartphone sensing data are prone to

logging errors, we extracted robust behavioural estimators when appropriate for the respective variable (Kafadar, 2003; Rousseeuw & Croux, 1993). To stay within the scope of this article, we only summarize our procedure and the engineered variables in the following sections. However, note that variable extraction is usually the most complex and time-consuming task in analyses of smartphone sensing data, and the process includes many researchers' degrees of freedom. For transparency, we provide all code in our OSF project, and the variable extraction procedure is described in detail in the supplemental codebook (Schoedel et al., 2020).

General indicators for sleep-related behaviours. We computed the following variables daily while distinguishing between days during the week versus the weekend (Roenneberg et al., 2007). Based on the algorithm specified in Table 7, we determined the *first and last events* according to individual study days and calculated mean and intraindividual variation variables. We defined the smartphone proxy for sleep duration, *nightly inactivity*, as the period between the last event of the day and the first event of the following day. To explore social jetlag, we calculated the average daily inactivity during the night for weekdays and weekends for all study weeks individually.

In addition, we translated two aspects of sleep continuity, sleep fragmentation and waking up after bed, into smartphone usage behaviour by calculating the average number and duration of *checking events* at night. At this point, we would like to point out that our measures do not fully meet the definition of sleep fragmentation and wake after sleep onset by Ohayon et al. (2017). Hence, our measurements only give a rough estimate, taking into account the occurrence of very short smartphone checking events during the nightly inactivity period of smartphone use, which was not part of a more extended usage period in the evening and the next morning. Accordingly, we defined nightly checking events as short periods of less than 2 minutes of smartphone usage during otherwise nightly inactivity. Due to the lack of empirical data in the literature, we have set this threshold value considering that smartphone usage of fewer than 2 minutes might be caused by less significant actions such as checking the clock during the night.

Finally, we calculated some variables related to using the smartphone as an alarm clock: the *mean point of time of alarm app ringing*, the mean daily number, and duration of *snoozing events* (snoozing was defined as the repetition of alarm app events in the morning).

Horne–Östberg chronotype variables. To operationalize circadian preferences in terms of smartphone usage behaviour, we computed variables following the items of the MEQ (Table 1). We translated preferred sleeping times as *mean points of time of the first and the last smartphone usage event on weekends*, as weekends are likely to be organized freely. Following this assumption, we also specified preferred times for activity as diurnal smartphone activity patterns. In this context, we distinguished between different behavioural categories: social communication (social media/communication app usage, calls, and texting), entertainment (browser, gaming, music/video, and news app usage), and general smartphone usage (all active

smartphone usage events). To take into account the distribution of usage events throughout the day, we computed the first quartile, the median, and the third quartile of usage events according to the behavioural categories for each day. In other words, we extracted timestamps that indicate when *25%, 50%, and 75% of the daily events of the respective usage category* took place. Then we computed the mean across all study days for each of the three quantiles. Finally, to depict the subjective feeling of sleepiness in the morning, we considered the *mean number and duration of snoozing events during the week* to indicate how readily people get up in the morning.

Roenneberg chronotype variables. Similar to the assessment of the chronotype using the MCTQ, we calculated the *midpoint of sleep* (*MSF*), which is the mean halfway point in time between the last event of a day and the first event of the next day for free (weekend) days without alarm app usage. In addition, we determined the *corrected midpoint of sleep* (MSF_{corr}), which has been proposed by Roenneberg et al. (2007) to correct for the sleep debt collected during the week. According to them the MSF_{corr} is better suited for estimating the true underlying chronotype.

Data analysis

Clustering

In the following, we give a short overview of the applied methods. More detailed information can be found in Appendix A. To investigate whether participants can be assigned to groups of similar smartphone usage behaviours indicating circadian preferences, we used clustering as an unsupervised machine learning method. We applied the commonly used *k-means* clustering algorithm with the Euclidean distance as proximity measure. Clustering aims to reduce complexity by finding meaningful structures within the data. According to their similarity in a predefined set of variables, participants are clustered in within-homogeneous groups that are well separated from participants of other clusters (Tan et al., 2006). However, one disadvantage of clustering algorithms is that they sometimes identify random and, therefore, nonreplicable structures (Tan et al., 2006). In line with the literature, we address this problem by using a data-driven approach to determine the number of clusters (Tibshirani & Walther, 2005) and by evaluating the stability and validity of the identified clusters based on bootstrapped metrics (Hennig, 2007; 2008; Tan et al., 2006). We followed the recommendations of Hennig (2018) and used 100 bootstrap iterations. For evaluating cluster stability, we considered the Jaccard coefficient (*JC*, indicates stability if values exceed 0.85) and the criteria of *recovery* and *dissolution*, which count how often each cluster has been successfully recovered and dissolved across all bootstrap iterations (Hennig, 2007; 2008). For evaluating the internal validity of clusters we looked at metrics indicating how similar participants within each cluster are (within-compact) and how different participants from different clusters are (between-separated): the ratio of average within- and between-cluster distances (*wb.ratio* Tan et al., 2006), the *silhouette* coefficient

(Rousseeuw, 1987), and the *Dunn* index (Dunn, 1974; Halkidi et al., 2001). Clusters are within-compact and between-separated if the ratio of distances is small, the silhouette index is close to 1, and the Dunn index is high (Tan et al., 2006; Hennig, 2018). As the *k-means* algorithm cannot handle missing values, we used the multivariate imputation by chained equations technique and specified a random forest imputation model (MICE, van Buuren & Groothuis-Oudshoorn, 2011).

Exploratory factor analysis
To explore the factorial structure of our smartphone-based proxy for the Horne–Östberg chronotype, we conducted an exploratory factor analysis based on the averaged correlation matrix of the imputed data sets. We determined the number of factors using the *empirical* Kaiser criterion, which has been shown to perform well for short scales (Braeken & Van Assen, 2017).

Multilevel modelling
Measures for nightly inactivity of smartphone usage were repeatedly measured across several study weeks. Considering the intraindividual data dependency, we used multilevel regression modelling with behavioural measures on a weekly basis reflecting level 1 variables that were nested within individuals (level 2). Therefore, we specified a *random-intercept-random-slope model* predicting the mean nightly inactivity duration on weekends based on the mean nightly inactivity duration of the respective preceding workweek (level 1). The averaged nightly inactivity duration, the Roenneberg chronotype, the big five traits, age, and gender, were included as predictors on level 2.

Regarding data preprocessing, we were faced with the challenge of selecting one path from a series of plausible steps. To do justice to these many researcher degrees of freedom and to increase research transparency, we follow the suggestion of Steegen et al. (2016) and present a *multiverse analysis*: for each possible combination of plausible preprocessing steps, a 'new' data set is constructed, and the same multilevel model is estimated for each of those data sets. The multiverse analysis illustrates how much the results depend on the choice of specific preprocessing steps or vice versa, which results are robust across all preprocessing options (Steegen et al., 2016; Simonsohn et al., 2015). Our preprocessing choices include the *coding of the weekend* (Friday to Sunday versus Friday to Monday), the selection of the *number of repeated measurements* (3 versus 4 weeks), the handling of *outliers* (median versus winsorization), and the handling of *missing values* (listwise deletion versus multiple imputation). A detailed description of the alternatives for each decision can be found in supplemental method section in Appendix A. Combining all described decisions resulted in $2 \times 2 \times 2 \times 2 = 16$ choice combinations (see left side in Figure 4).

We used the uncorrected version of the Roenneberg chronotype as a predictor, as we explicitly control for a nightly inactivity deficit in the multilevel model. Gender was dummy coded (0 = male, 1 = female), and all continuous predictor variables were z-standardized based on the grand mean. The level 1 predictor duration of nightly inactivity during the week was centred around the individual mean, which in turn was entered as level 2 predictor (Curran & Bauer, 2011). For a more detailed description of the equation of the multilevel model, we refer the interested reader to the supplemental method section in Appendix A.

Statistical software
All data preprocessing and analyses were conducted using R 3.5.0 (R Core Team, 2018). We used *packrat* (Ushey et al., 2018) for package management. For extracting behavioural variables, we mainly used the R packages *dplyr* (Wickham et al., 2019) and *fxtract* (Au, 2019). Multiple imputation was done by using the package *mice* (van Buuren & Groothuis-Oudshoorn, 2011). In addition, we used the following packages to conduct our main analyses: *fpc* for clustering (Hennig, 2018), *psych* for exploratory factor analysis (Revelle, 2018), and *lme4* and *lmerTest* for multilevel modelling (Bates et al., 2015; Kuznetsova et al., 2017). For data visualization, we applied *ggplot2* (Wickham, 2016) and *corrplot* (Wei & Simko, 2017) and created *raincloud plots* (Allen et al., 2019). The complete list of used R packages can be found in our OSF project (Schoedel et al., 2020).

RESULTS

Descriptives

We recorded a mean of 22 547 events ($SD = 24\,368$) for each participant across the whole study period. Participants had on average smartphone records for 21 ($SD = 1.57$) weekdays and 8 ($SD = 0.92$) weekend days. The mean number of logs per study day was 765 ($SD = 804.70$). As can be seen in Table 4, the average time of first and last smartphone usage was later for weekends than weekdays, and the duration of nightly inactivity was about 20 minutes longer on weekends than on weekdays. However, the mean number and duration of checking events during the night were similar for weekends and weekdays. A total of 91% of our participants used alarm clock apps in the morning, at 7.19 AM on average during the week and about 30 minutes later on weekends. Note that 38% of participants did not use alarm clock apps on any weekend during the entire study period. The number and duration of snoozing events were similar for weekdays and weekends. Descriptive statistics for big five personality traits can be found in Table 8 in the Appendix.

Individual differences in behavioural day–night patterns

Person-centred and variable-centred structure of the Horne–Östberg chronotype
In the first step, we determined the number of clusters. Following the suggestions of Tibshirani and Walther (2005), we looked for solutions resulting in a prediction strength above 0.80. Doing so, in 49 out of 50 imputed data sets, the data-driven proposed number for clustering based on smartphone proxies for circadian preferences was 1. However, decreasing the prediction strength criterion to a value

Table 4. Descriptive statistics for day–night behaviour patterns

Variable	Week		Weekend		Cohens' d [CI₉₅%]
	Mean	SD	Mean	SD	
Mean first event week	7.89	1.31	8.96	1.30	0.82 [0.70, 0.93]
Mean last event week	23.15	1.23	23.79	1.42	0.49 [0.37, 0.60]
Mean duration nightly inactivity week (h)	8.68	1.20	9.02	1.45	0.26 [0.14, 0.37]
Mean number checking events week	5.59	3.97	5.61	5.37	0.00 [-0.11, 0.12]
Mean duration checking events week (s)	26.07	26.12	25.29	40.72	-0.02 [-0.14, 0.09]
Mean first alarm event week	7.19	1.29	7.47	1.68	0.19 [0.06, 0.33]
Mean number snoozing week	1.33	1.76	1.33	2.04	0.00 [-0.13, 0.13]
Mean duration snoozing week (min)	23.26	23.61	23.89	34.26	0.02 [-0.11, 0.16]

The coefficients for first and last events represent times of the day. The decimal places indicate the percentage of a full hour. For example, 7.89 means 7:53 AM or 23.15 means 11:09 PM

of 0.75 yielded a 2-cluster solution for all imputed data sets. Although the recommended predictive power was slightly missed, we further investigated $k-means$ clustering with $k = 2$. The averaged bootstrapped performance measures for the cluster-wise stability assessment show that each component of the 2-cluster solution turned out to be highly stable (cluster 1 $_{n = 296}$: $JC = 0.94$, $dissolved = 0$, $recovered = 100$; cluster 2 $_{n = 301}$: $JC = 0.93$, $dissolved = 0$, $recovered = 100$). However, the internal cluster validation coefficients indicated that the two clusters were poorly separable from each other and were not compact in themselves ($wb.ratio = 0.73$, $silhouette = 0.25$, $Dunn = 0.06$). To get a better understanding of the identified structure in the daily smartphone usage timing, descriptive statistics of the variables that were considered for clustering are displayed in Table 5. On average, participants assigned to cluster 2 had later first and last smartphone usage events on weekends and the daily 25%, 50%, and 75% timestamps for general, social interaction, and entertainment usage events on weekends were on average about 2 hours later. The mean number of snoozing events

was similar in both groups, but participants of cluster 2 on average snoozed approximately 3.5 minutes longer. As an external criterion, we considered the smartphone-based Roenneberg chronotype. The mean midpoint of sleep was $M = 3.90$ ($SD = 1.15$) for cluster 1 and $M = 5.19$ ($SD = 1.38$) for cluster 2.

To return to the question of whether we found different groups of individuals with similar smartphone usage patterns indicating circadian preferences, we refer to Table 5. Effect sizes for variables indicating sleep–wake timing are large, suggesting that participants assigned to cluster 2 have noticeable back-shifted diurnal smartphone usage patterns in comparison with participants assigned to cluster 1. Figure 1 shows, however, that the distributions of the two cluster groups overlap. A considerable proportion of participants could not be clearly assigned to one of the two clusters. Accordingly, the distribution based on the entire sample was not bimodal but only unimodal.

In the second step, we also explored the factorial structure of the smartphone-based proxies for the Horne–Östberg

Table 5. Descriptive statistics for smartphone usage indicating circadian preferences by clusters

Variable	Cluster 1		Cluster 2		Cohens' d [CI₉₅%]
	Mean	SD	Mean	SD	
First/last events on weekends					
Mean time of the first event	8.35	1.16	9.58	1.13	1.07 [0.90, 1.25]
Mean time of the last event	23.09	1.18	24.51	1.29	1.15 [0.97, 1.32]
Mean on weekends daily timestamp of					
25% general usage	12.28	1.29	14.38	1.37	1.57 [1.39, 1.76]
50% general usage	15.34	1.30	17.62	1.20	1.82 [1.63, 2.01]
75% general usage	18.37	1.34	20.62	1.17	1.79 [1.60, 1.98]
25% social interaction usage	12.51	1.28	14.47	1.22	1.57 [1.38, 1.76]
50% social interaction usage	15.38	1.34	17.53	1.16	1.72 [1.53, 1.91]
75% social interaction usage	18.18	1.53	20.29	1.16	1.56 [1.37, 1.74]
25% entertainment usage	12.91	1.76	15.22	2.02	1.21 [1.03, 1.39]
50% entertainment usage	15.07	1.86	17.70	1.75	1.46 [1.27, 1.64]
75% entertainment usage	17.25	2.05	20.03	1.74	1.47 [1.28, 1.65]
Snoozing events on weekdays					
Mean number of snoozing events	1.31	1.88	1.35	1.65	0.02 [-0.15, 0.20]
Mean duration of snoozing events	21.53	22.01	24.91	24.97	0.14 [-0.03, 0.32]

Except the snoozing variables, the coefficients represent times of the day and the corresponding standard deviations are given in hours. The decimal places indicate the percentage of a full hour. The mean daily timestamp of 25% general usage indicates that 25% of all activities on a given day had happened at this point in time. The mean number of snoozing events means the daily mean absolute frequency and the snoozing duration is in minutes.
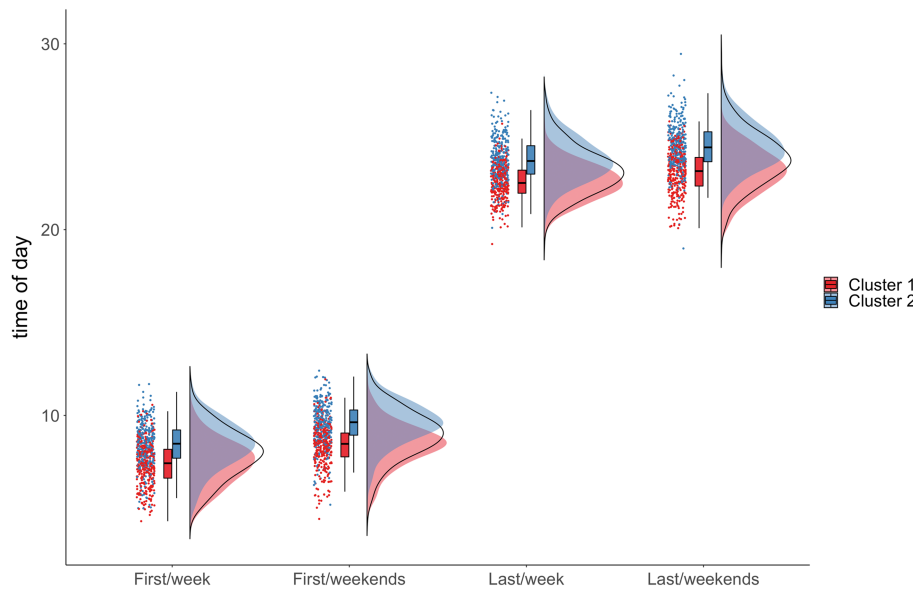
Figure 1. Plots displaying the distribution of mean daily first and last events on weekdays versus weekends by cluster. The black line shows the distribution based on the total sample. The ordinate axis goes beyond midnight, because last events after midnight were added to 24. An event at 26 therefore means it happened at 2.00 AM.

Table 6. Exploratory factor analysis of the smartphone-sensed circadian preferences

Variable	F1	F2	F3	U
Mean time of the first event on weekends	0.09	0.06	**0.49**	0.67
Mean time of the last event on weekends	**0.54**	0.03	0.04	0.66
Mean daily timestamp of 25% general usage on weekends	0.05	0.10	**0.84**	0.16
Mean daily timestamp of 50% general usage on weekends	**0.44**	0.17	**0.47**	0.20
Mean daily timestamp of 75% general usage on weekends	**0.74**	0.14	0.11	0.23
Mean daily timestamp of 25% social interaction usage on weekends	0.26	-0.01	**0.68**	0.30
Mean daily timestamp of 50% social interaction usage on weekends	**0.63**	-0.01	**0.38**	0.22
Mean daily timestamp of 75% social interaction usage on weekends	**0.88**	0.03	0.03	0.19
Mean daily timestamp of 25% entertainment usage on weekends	-0.20	**0.77**	**0.33**	0.24
Mean daily timestamp of 50% entertainment usage on weekends	0.02	**0.99**	0.00	0.01
Mean daily timestamp of 75% entertainment usage on weekends	**0.34**	**0.77**	-0.17	0.21
Mean daily number of snoozing events on weekdays	0.26	0.01	-0.24	0.94
Mean daily duration of snoozing events on weekdays	0.27	0.00	-0.19	0.94
F2	0.46	1.00		
F3	0.52	0.47	1.00	

Maximum likelihood factor analysis, obliquely rotated (oblimin) with three factors. Loadings greater than the amount of 0.30 are in bold. The correlations between the factors are displayed at the bottom of the table. F1 = Factor 1; F2 = Factor 2, F3 = Factor 3; U = Uniqueness.

chronotype. The empirical Kaiser criterion suggested a 3-factorial solution accounting for 62% of the variance. The obliquely (oblimin) rotated factor matrix is displayed in Table 6. Factor 1 explained 23% of the variance and comprised behavioural indicators describing markers for later diurnal smartphone usage. In contrast, the behavioural variables loading high on factor 3 (19% variance explanation) described markers characteristic for early diurnal smartphone usage. The 50% timestamps for daily (general and social interaction) smartphone usage considerably loaded on both, factors 1 and 3. Finally, factor 2 explained 20% of the variance and reflected behavioural indicators of

smartphone usage for entertainment purposes independent of the time of the day. The two snoozing items did not load considerably on any factor. All factors were correlated (see Table 6).

The Roenneberg chronotype and its correlates
The smartphone-based midpoint of sleep (MSF) and the sleep debt corrected version MSF_{corr}, which both indicate the Roenneberg chronotype, were approximately unimodally symmetrically distributed (see Figure 2). As no weekends without alarm clock usage were available for some participants, their *MSF* could not be computed. Therefore, the
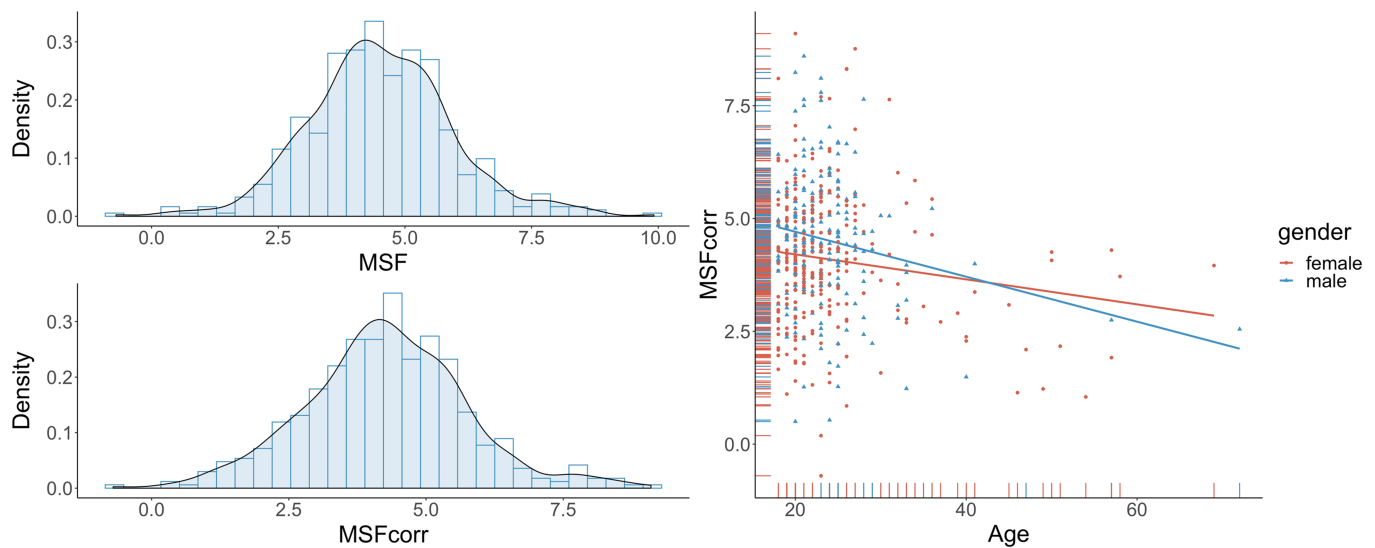
Figure 2.   Plots displaying the distributions of the local time of the midpoint of sleep (*MSF*) and its sleep debt corrected version (*MSF*$_{corr}$) and its relationship with age divided by gender.

following results are based on a subsample of $n = 497$ participants. On average, the mean *MSF* was at 4.52 AM ($SD = 1.42$) and the *MSF*$_{corr}$ slightly earlier at 4.26 AM ($SD = 1.47$). The *MSF* and *MSF*$_{corr}$ ranged between 0.75 PM and 9.91 AM. The *MSF* was weakly negatively related to nightly inactivity duration during the weeks ($r = -0.13$, $CI_{95\%}$ [-0.21, -0.04]) as well as the weekends ($r = -0.11$, $CI_{95\%}$ [-0.20, -0.03]). As suggested by Roenneberg et al. (2007), we used the *MSF*$_{corr}$ for investigating the relationship of chronotype and demographics. Age ($r = -0.16$, $CI_{95\%}$ [-0.24, -0.07]) and gender ($r = -0.15$, $CI_{95\%}$ [-0.23, -0.06]) were both negatively related to the corrected midpoint of sleep, indicating that older and female participants had on average earlier chronotype values. However, the age correlation should be interpreted with caution, as the plot on the right side of Figure 2 indicates that it was probably caused by data points of older participants of whom we only had few in the sample ($Q_3 = 25$). The correlation disappears ($r_s = -0.03$, $CI_{95\%}$ [-0.12, 0.06]) when computing the Spearman correlation, which is only based on ranks.

Day–night behaviours and personality traits

Because our analysis of relationships between behavioural day–night patterns and personality is exploratory, we do not perform any hypothesis tests, nor do we speculate about correlations on a variable-by-variable basis. Instead, based on the correlation plot displayed in Figure 3, we want to show the general result pattern and address some conspicuities. Overall, Spearman correlations ranged between $r_s = -0.24$ (mean time of last events during the week and sense of duty) and $r_s = 0.15$ (mean time of the first event on weekends and carefreeness). As can be seen in Figure 3, the most striking aspect is that conscientiousness and its facets (except competence) were related to various day–night behaviours. First, more conscientious people on average had earlier mean and less varying daily points of time of first and last

smartphone usage events both during weeks and on weekends. Furthermore, their duration of nightly inactivity varied less on weekdays and they had lower values on the Roenneberg chronotype. Finally, individuals with higher values on the facet sense of duty snoozed on average less often and shorter on weekdays.

Further but less coherent patterns in Figure 3 can be seen for openness, extraversion, and emotional stability. For example, openness to imagination showed some positive relations to day–night behavioural indicators. Openness to value and norm system was associated positively with the mean number and duration of snoozing events, especially on weekdays. Higher extraversion was related to longer smartphone checking events during nights on weekdays. Furthermore, carefreeness as a facet of emotional stability was associated positively with later day–night activity patterns. Regarding demographics, female participants' first use on weekends and general last use was on average earlier. Accordingly, they also had lower Roenneberg chronotype values. However, no correlations of considerable size were found for age.

Using multilevel modelling to explore social jetlag

To investigate social jetlag, we explored compensatory sleep on weekends approximated as nightly inactivity duration by multilevel modelling. The duration of nightly inactivity on weekends was predicted by the duration of nightly inactivity during the week and the interindividual variables Roenneberg chronotype, big five personality traits, age, gender, and the averaged individual mean duration of nightly inactivity. The results are presented in the 12 panels in Figure 4, which show the estimates and their 95% confidence intervals across all multiverse data sets for each predictor in the model. Some aspects were evident across all data sets. There were no relationships between the nightly inactivity duration on weekends and the variables Roenneberg
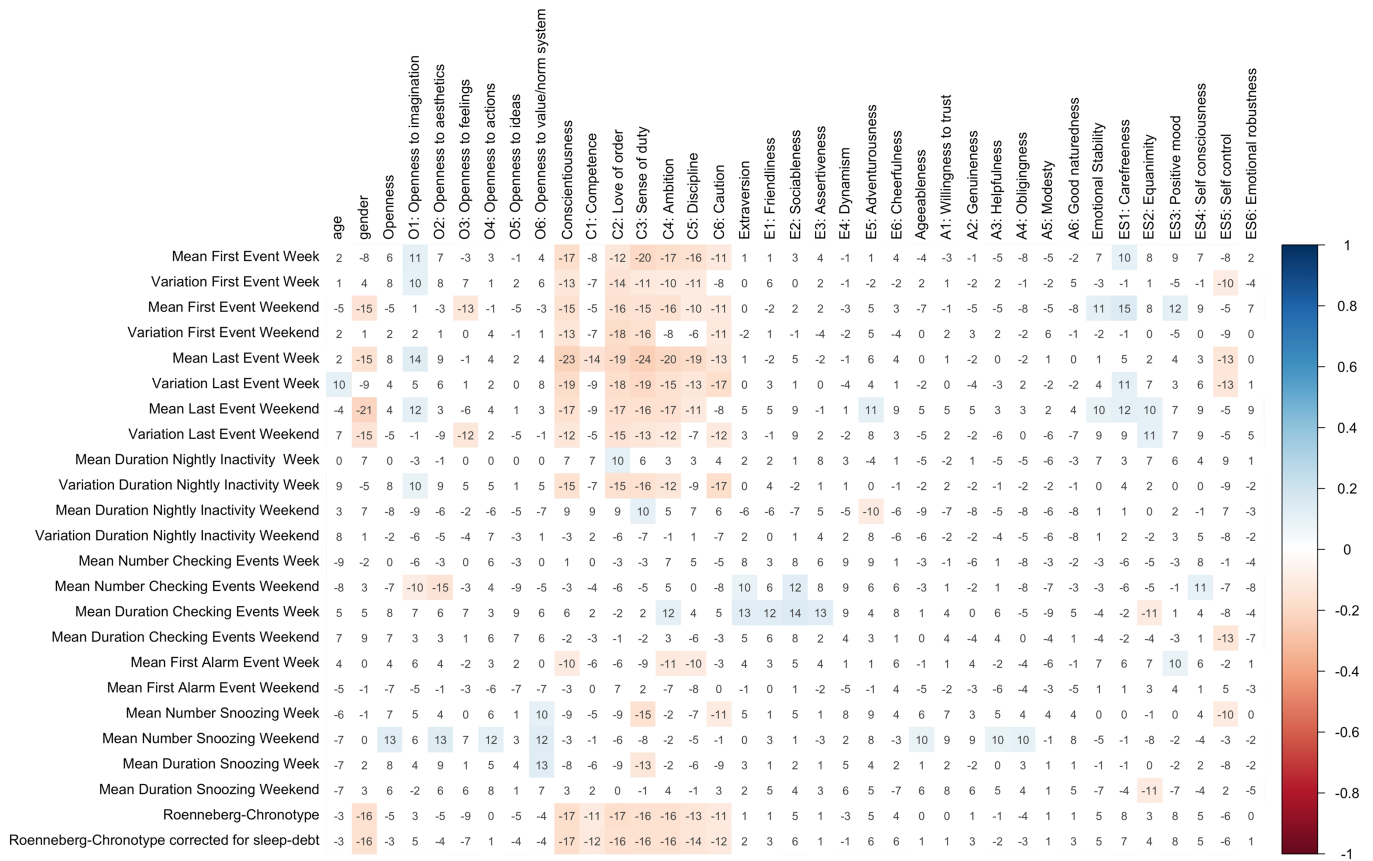
Figure 3. Pairwise complete Spearman correlations between smartphone-sensed day–night activities for weekdays versus weekends and personality traits. Male participants were coded as 0. As not all participants used alarm clock apps, the sample size for respective correlations was reduced ($n_{week} = 506$, $n_{weekend} = 371$). The colour of the squares indicates the direction and the strength of the respective correlations. For better readability, correlations are presented as percentage (e. g. a value of 3 means $r_s = 0.03$). Additionally, only correlations with greater absolute values than 0.10 are highlighted in colour.

chronotype, openness, extraversion, agreeableness, emotional stability, and the interaction between the Roenneberg chronotype and the nightly weekday inactivity. Second, the averaged nightly inactivity duration across the study weeks (level 2) was positively associated with the nightly inactivity period on weekends. Nevertheless, estimates for the individual nightly inactivity duration on weekdays (level 1) and conscientiousness, age, and gender (all level 2) varied across the multiverse data sets. Depending on the preprocessing steps, individuals with longer nightly inactivity duration on weekdays in the corresponding week, higher conscientiousness, higher age, and male gender had, on average, longer nightly inactivity periods on weekends.

As can be seen in Figure 4, some patterns can be identified in the multiverse results across different variables: the coding of the weekend seemed to have an influence. In conditions in which the weekend was coded as nights between Friday and Monday, the mean duration of nightly inactivity on weekends was, on average, lower compared with the conditions in which weekends were coded as nights between Friday and Sunday. Also, for gender, a pattern can be determined depending on the coding of the weekend. For conscientiousness, estimates in conditions including 3 weeks were, on average, higher than conditions comprising 4 weeks. Regarding the average duration of nightly inactivity during the week (level 2), estimates were higher when winsorized and imputed.

To get a better understanding of the results concerning social jetlag, we calculated an additional multiverse analysis. For this purpose, we considered a variant of the multilevel model without personality traits and demographics as covariates. As results did not considerably differ and not to go beyond the scope of this paper, they can be found as a supplementary analysis in our OSF project.

DISCUSSION

We investigated three prominent research questions related to common behavioural day–night patterns by using smartphone sensing data. First, we focused on individual differences in day–night activity patterns. Based on behavioural indicators of circadian preferences, we explored the structure underlying our smartphone proxy for the Horne–Östberg chronotype. Regarding the search for a smartphone chronotype, we found nondiscrete groups of individuals with similar diurnal smartphone usage patterns. In addition, our smartphone-based proxy for the Horne–Östberg chronotype turned out to be a multidimensional construct. In addition, we presented an algorithm for computing the chronotype as defined by Roenneberg et al. (2003). We used smartphone-based indicators for the midpoint of sleep and
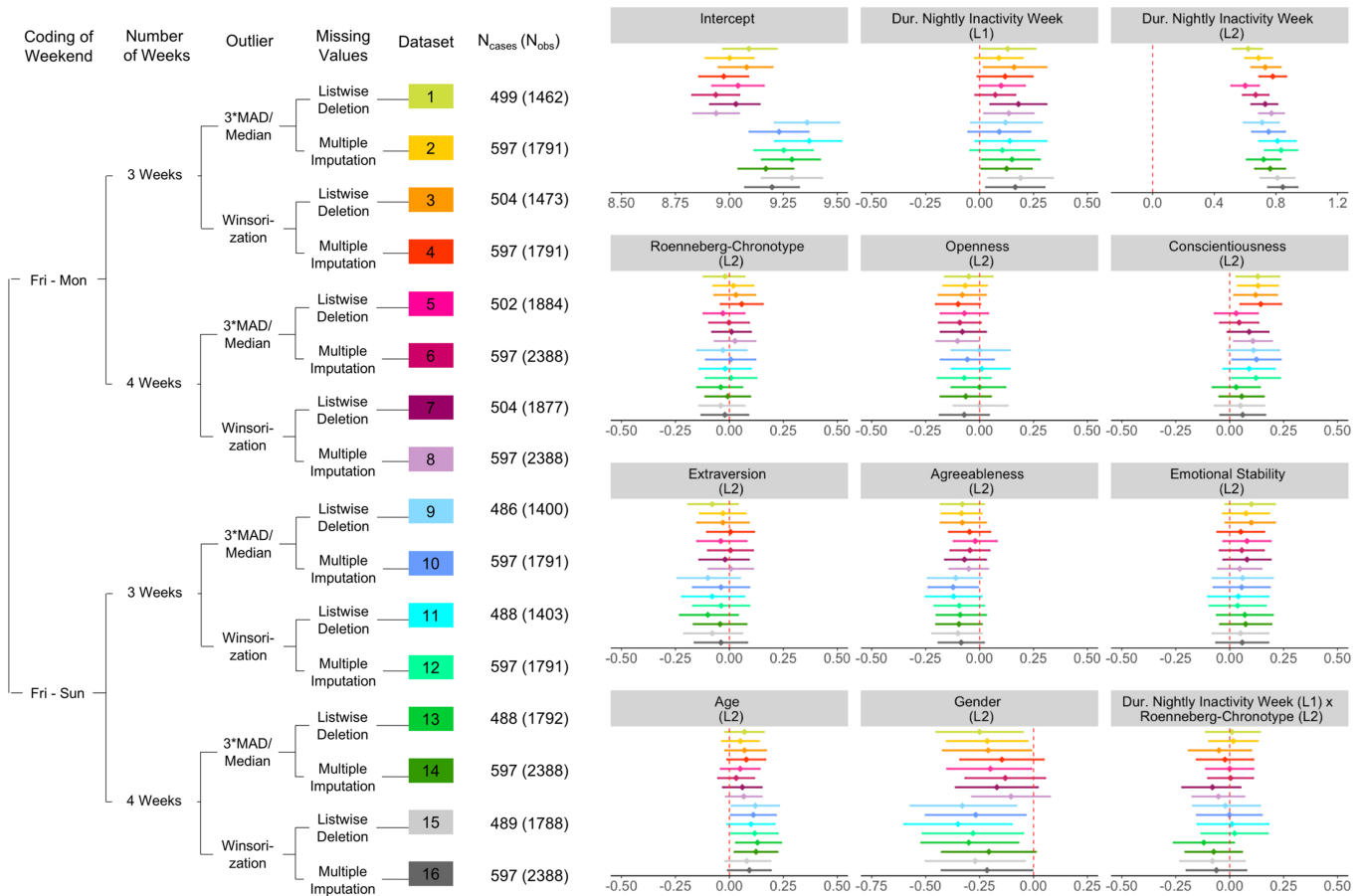
Figure 4. The decision tree on the left side shows how the multiverse of 16 data sets was created. The 12 panels on the right display the estimates and their 95% confidence intervals for the intercept and each predictor, resulting from multilevel modelling across the multiverse of 16 data sets. L1 = level 1 predictors (*z*-standardized and person-mean-centred); L2 = level 2 predictors (*z*-standardized, except gender). Male participants were coded as 0. The individual mean of the level 1 predictor was additionally entered as level 2 predictor. Each data set and the corresponding model are coded the same colour.

found associations with age, gender, and duration of nightly inactivity. Regarding personality traits, we found associations of conscientiousness with smartphone-sensed indicators for day–night behaviour. Finally, we explored social jetlag by examining whether people were inactive longer during weekend nights if they accumulated a deficit of nightly inactivity during the preceding workweek while controlling for individual differences. Our findings suggest that nightly inactivity duration on weekends was mainly related to individuals' general level of nightly inactivity across all study weeks. We will critically discuss our results in the following sections. Because our research was explorative, explanations drawn *post hoc* should not be easily generalized but be confirmed by preregistered hypotheses testing in future studies.

Smartphone sensing in the context of behavioural day–night patterns

Individual differences in day–night activity patterns
In contrast to previous research based on self-reports, we used smartphone-sensed behavioural data to investigate the

structure of chronotype and to inform both the variable-centred and the person-centred approach to chronotype. Emphasizing chronotype as a continuous dimension reflecting circadian habits, Roenneberg et al. (2003) have suggested computing the midpoint of sleep. Instead of assessing these habits by questionnaires (e.g. Roenneberg et al., 2003; Roenneberg et al., 2007), we followed Lin et al. (2019) and used smartphone sensing data to determine a smartphone equivalent for the Roenneberg chronotype. We compared our resulting measure with the findings reported by Roenneberg et al. (2007) and found similar descriptive parameters (distribution, mean) and associations with external criteria like gender and sleep duration during the week. In accordance with our assumption that smartphone-based sleep–wake timing indicators overestimate sleep times, the range of values was slightly larger for our measure. Regarding age and chronotype, we found a negative correlation, which was caused by a few older participants with lower chronotype values. However, because the age composition of our sample was highly skewed towards younger participants, we do not want to over-interpret this finding. A nonmatching result was that whereas Roenneberg et al. (2007) found a positive correlation between chronotype and

sleep duration on weekends, we found a negative association. Roenneberg et al. (2007) argued that later chronotypes sleep longer on weekends because they collect a sleep debt during the workweek. In contrast to the large representative sample of their epidemiological study, our sample consisted mainly of students who are more likely to have fewer social obligations during the week than people who have a 9 AM to 5 PM job. Accordingly, compared with nights during the week, our participants' nightly inactivity (indicating sleep) did not differ considerably on weekends. Therefore, one interpretation of our results could be that students have the opportunity to be more flexible in their daily routines during the week following their chronotype. Therefore, late chronotypes do not need disproportionately more sleep on weekends. Accordingly, previous studies have shown that many students report napping after lunch during the week (Vela-Bueno et al., 2008). These naps could serve to use both the weekend and the week for sleep compensation (Gradisar et al., 2008). In line with our interpretation, students with late chronotypes have been found to nap more extensively than students with early chronotypes (Zimmermann, 2011). Please note that this is only our *post hoc* interpretation and further confirmatory research using behavioural data to study the interplay of sleep duration, chronotype, and work schedules is needed..

Keeping the focus on variable-centred trait assessment (Asendorpf, 2003), but following the Horne–Östberg tradition, we operationalized circadian preferences as diurnal smartphone usage behaviours and explored the underlying factorial structure. We found three correlated dimensions reflecting early use of the smartphone during the day, late use of the smartphone during the day, and entertainment usage. In comparison, findings of previous studies investigating the structure of self-reported chronotype have resulted in one to four factors (e.g. Caci et al., 2009; Lipnevich et al., 2017; Natale & Cicogna, 2002). In their recent meta-analysis, Lipnevich et al. (2017) concluded that the preferences for morningness versus eveningness are not the extreme poles of one dimension but two interdependent dimensions. Accordingly, our two correlated dimensions reflecting early and late diurnal smartphone usage activity align with their findings. Regarding our factor entertainment usage, we think that this could be regarded as a methodological artefact, as the content entertainment might have overlaid the diurnal character of the respective behavioural circadian indicators.

Dimensional approaches to personality, such as the two described above, offer the advantage to focus on individual differences. However, in contrast to person-centred approaches, they are not able to describe the structure of traits within persons (e.g. Asendorpf & van Aken, 1999; Asendorpf, 2003). In addition, types might have an advantage for applied purposes as the classification as 'morning larks' or 'night owls' is widely anchored in the popular science literature and scientific research. Therefore, besides examining dimensionality, we also explored the existence of types of individuals with similar diurnal smartphone usage patterns by using unsupervised machine learning. We found two groups that showed earlier versus later smartphone usage over the day. As the effect sizes show, these two groups considerably differed in indicators of diurnal smartphone usage

patterns. However, our results also indicate that despite the high average group differences, a large number of participants could not easily be assigned to one of these two groups, which overlapped considerably in the behavioural indicators used. Therefore, we asked ourselves whether we should call the structure we found types. In previous chronotype literature, types had often been considered as empirically validated, if the resulting groups subsequently proved to be different concerning external criteria (e.g. body temperature, electroencephalography (EEG) recordings Horne & Östberg, 1976; Putilov et al., 2015). In contrast, we did not determine any cut-off values but searched for nonrandom structures in the data. Only recently, Preckel et al. (2019) followed a similar approach identifying four chronotypes in an adolescent sample. However, because circadian preferences change with age (Roenneberg et al., 2007), and our sample was older, and we focused on smartphone-sensed rather than self-reported circadian habits, we argue that the results are not fully comparable.

From a statistical point of view, the existence of types is only justified if underlying variables are multimodally distributed (Hicks, 1984; Fleiss et al., 1971), which was not the case for our behavioural day–night indicators. However, previous research in the social sciences has revealed that nonoverlapping types hardly exist for human behaviours (Meehl, 2004; Costa Jr et al., 2002). Accordingly, Asendorpf and van Aken (1999) distinguish between discrete and nondiscrete types in the context of personality research. Thus, the criteria for defining types are not uniformly defined and applied in the literature. Our results are in line with this argument. Even if there were discrete underlying chronotype groups, it is unlikely that they would appear so clearly in everyday behavioural indicators due to social obligations and societal demands. Nevertheless, the identified nondiscrete groups in our study can be a good starting point towards a smartphone-based behavioural proxy of chronotype operationalized as circadian preferences. Future research should replicate the structure in diurnal smartphone usage indicators across different samples and use external validity criteria.

Conscientiousness and differences in behavioural day–night patterns

In contrast to the majority of previous studies, we used behavioural markers for day–night activity patterns to investigate associations with personality traits and demographics. In line with past studies showing women's preference for morningness (Randler, 2007), women in our study were earlier in the day, and their day–night activity timing varied less. Besides, our results were consistent with previous research showing a majorly coherent pattern of day–night activity and conscientiousness (Adan et al., 2012; Lipnevich et al., 2017) but less clear relations for other big five personality traits (e.g. Gray & Watson, 2002; Randler et al., 2017). Precisely, highly conscientious participants on average showed lower and less varying sleep–wake timing indicators and lower Roenneberg chronotype values. Following questionnaire-based research (Adan et al., 2012; Lipnevich et al., 2017; Tsaousis, 2010; Križan & Hisler, 2019), our

results indicate that more conscientious people on averagen are active earlier during the day and have longer nightly rest periods on weekends. Compared with findings from a meta-analysis ($r = .33$ according to Tsaousis, 2010), our correlations were smaller. However, our findings show that more conscientious people, who describe themselves as dutiful, ambitious, and disciplined (Arendasy, 2009), also act accordingly in everyday life (e.g. getting up early in the morning, longer nightly rest on weekends). Accordingly, Spears et al. (2019) found in a recent longitudinal study that conscientiousness was associated with mortality risk after 10 years and that this association was mediated by sleep duration as an everyday expression of behaviour.

In contrast to previous findings, conscientiousness and emotional stability were not related to indicators for sleep continuity, but extraversion was (Križan & Hisler, 2019; Sutin et al., 2019; Sella et al., 2020). These recent studies measured sleep continuity using actigraphy and therefore used completely different operationalizations of the related indicators sleep fragmentation and wake up after bed (Križan & Hisler, 2019; Sutin et al., 2019; Sella et al., 2020). For example, Sella et al. (2020) defined sleep fragmentation as the number of awakenings exceeding a certain duration. In contrast to actigraphy, smartphone sensing does not provide continuous measurement of wakefulness but approximates this measure via active smartphone usage. This requires the determination of a specific threshold value to classify smartphone usage either as part of a continuous usage phase belonging to the last or first event of the day or as a short usage event during the period of otherwise nightly inactivity. Determining a threshold value according to this principle, our approach has two significant drawbacks. First, using 2 minutes as a threshold was a subjective decision due to the lack of empirical data from previous literature. Second, the derived variable checking duration is restricted in its variance by a maximum value of 2 minutes. Consequently, individual differences in the actual wake after sleep onset might be masked by our smartphone-based operationalization, which in turn could explain the differences in findings compared with actigraphy.

n addition, we did not find some of the relationships that have previously been reported. For example, in our data, we did not find associations between a preference for morningness and agreeableness (Adan et al., 2012; Tsaousis, 2010) or age (Adan et al., 2012). As already discussed in the previous section, our results regarding age should be interpreted with caution due to the restricted variability of age in our sample. Overall, the differing findings could result from the usage of actual behavioural variables in contrast to self-reported preferences in most previous studies. Additionally, differences with past studies might not be surprising considering that previous questionnaire-based research is not clear either (e.g. Duggan et al., 2014; Gray & Watson, 2002). Besides, to the best of our knowledge, we have been the first to explore differences in alarm clock app usage. Our results provide first indications about the relation of snoozing behaviour and personality facets (sense of duty

and openness to value and norm system). They should be further investigated in future research.

Individual differences in compensatory nightly inactivity on weekends

To explore social jetlag, we investigated which intraindividual and interindividual factors predict the duration of nightly inactivity of smartphone usage (assumed to indicate sleep duration) on weekends. To explore this research question and to get an impression of the robustness of our estimates, we created a multiverse of 16 data sets resulting from combining different choices of plausible preprocessing steps. In the following, we focus only on those aspects that have been demonstrated across all data sets. Individuals who had higher overall levels of smartphone inactivity during nights on weekdays were also inactive longer on weekend nights. Even though our inactivity measure is not identical to sleep, our results indicate that individuals differ in their nightly rest duration. These findings support the notion that sleep duration is an independent trait (Ferrara & De Gennaro, 2001; Roenneberg et al., 2007). In contrast to the assumptions of social jetlag (Roenneberg et al., 2015; Wittmann et al., 2006), we neither found compensatory nightly inactivity on weekends nor any impact of the Roenneberg chronotype. As already discussed in the section above, our sample was highly skewed towards students. Thus, maybe their social obligations during the week are less pronounced, and therefore, we could not find their need for compensatory sleep on weekends. In addition, previous studies often used self-reports to investigate social jetlag (e.g. Wittmann et al., 2006; Roenneberg et al., 2012). Even though participants are instructed to indicate their habits for the last 4 weeks (Roenneberg et al., 2003), their answers might be biased towards a more general judgment of sleep–wake timing or influenced by short-term experiences like the sleep behaviour of the previous night. In contrast, we looked at behavioural snippets of 3 or 4 concrete weeks.

Finally, our multiverse analysis showed that the results depend on the selected preprocessing steps. Especially for the predictors age, gender, and conscientiousness, the size of the estimates differed depending on the constructed data sets. Our study therefore points to two problems. First, for behavioural indicators extracted from smartphone sensing data, the definition of the weekend and the number of weeks included made a difference to the results. Future research in the field of smartphone sensing should, therefore, carefully explore and report whether decisions made in the preprocessing have an impact on the results. Second, our study highlights the issue of selective reporting in research articles (Simonsohn et al., 2015; Steegen et al., 2016). We could just as well have reported only one of the paths and the results of the corresponding model, and the choice of each path would have been equally plausible. However, depending on the preprocessing decisions, we might or might not have emphasized the effect of conscientiousness or gender or age at this point. In line with Simonsohn et al. (2015) and Steegen et al. (2016), we argue that

decisions that might affect the results should be made transparent.

Limitations and outlook

Our study exemplifies the usage of smartphone sensing data in the research field of behavioural day–night patterns. Strictly speaking, the assessment of day–night structures in everyday life and, therefore, sleep-wake phases would require the collection of EEG data (Shambroom et al., 2012). For reasons of efficiency, self-report questionnaires have so far been used to approximate sleep-related behaviours. We propose smartphone sensing as an alternative to collect proxies for these behaviours. However, our approach has some limitations.

First, similar to questionnaires (Lauderdale et al., 2008), our behavioural markers are only proxies for actual sleep–wake timing. In our data set, only app, phone, screen, and notification events were available to determine the nightly inactivity period. Thus, actual sleep times were estimated based on active smartphone usage behaviours. However, for improving the accuracy of smartphone-based sleep–wake indicators, it would be helpful to include sensor data that do not require active usage, for example, brightness and ambient noise (Min et al., 2014). An even better estimate of sleep could be obtained by integrating the idea of actigraphy into the smartphone sensing approach. Meanwhile, many commercial wearables, which can also be used conveniently during bedtime, offer an open interface to integrate motion and physiological data like heart rate variability or galvanic skin response into research apps used for smartphone sensing.

Second, we defined new behavioural variables, which we extracted from smartphone sensing data. Although we derived our variables from previous literature, we had many degrees of freedom. Which period is defined as a weekend? What does active smartphone usage mean? How can daily values be aggregated? These questions are only a few examples for the vast amount of decisions we had to make during data preprocessing. To make this process as transparent as possible, we provide an extensive codebook and analyse a multiverse of data sets where appropriate. However, the researcher community should develop a common standard for sensing data so that the results obtained do not depend on the respective data preprocessing decisions in individual studies.

One further limitation of our study was the skewed sample. In comparison with previous epidemiological studies, it was skewed in terms of age and occupation. As age and work schedules are related to sleep–wake timings (Adan et al., 2012), future studies using smartphone sensing data should use more representative samples.

Finally, in our study, we only focused on smartphone sensing data. Although resulting indicators cannot be equated one-to-one with physiological sleep, smartphone sensing can nevertheless unobtrusively collect data in the field over a long period. This is very beneficial as far as day–night habits are investigated. However, in research focusing on constructs like sleep quality (Križan & Hisler,

2019), it is essential to measure a possible mismatch between behavioural sleep indicators in contrast to individual perceptions and feelings about sleep–wake timings. Consequently, the integration of the experience sampling method (e.g. Takano et al., 2014) could help to gain further interesting insights in individual differences into behavioural day–night patterns. Future studies could additionally benefit from combining actigraphy and smartphone sensing. Both methods assess actual behaviour but highlight different aspects of day–night activity patterns (Borger et al., 2019). In summary, we do not want to discuss whether self-reports, smartphone sensing, or actigraphy are better suitable for depicting actual behavioural day–night patterns. We think that all data collection approaches have their place and could be very fruitfully combined to gain better insights into human day–night behaviour patterns.

CONCLUSION

We used smartphone sensing data to extract behavioural variables usually assessed by self-reports in the context of day–night behaviours. Our study contributes to gain new insights into traits related to day–night behaviour patterns. First, we investigated two prominent operationalizations of chronotype: based on indicators for sleep–wake timing and diurnal activity, we found two overlapping groups of smartphone-based 'morning larks' and 'night owls' and two correlated dimensions that were similar to previously reported questionnaire-based factors. By computing a smartphone-based proxy, we presented a smartphone-sensed measure for the Roenneberg chronotype. Second, conscientiousness was related to earlier day schedules. In addition, we found individuals to differ in their overall level of nightly rest. We argue that it is important to understand individual differences in behavioural day–night patterns, as they previously have been found to be related to individuals' well-being and health. This work demonstrates that smartphone sensing provides an efficient and ecologically valid tool that can help to foster this understanding.

ACKNOWLEDGEMENTS

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

REFERENCES

Adan, A., Archer, S. N., Hidalgo, M. P., Di Milia, L., Natale, V., & Randler, C. (2012). Circadian typology: A comprehensive review. *Chronobiology International*, *29*(9), 1153–1175. https://doi.org/10.3109/07420528.2012.719971

Allen, M., Poggiali, D., Whitaker, K., Marshall, T., & Kievit, R. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Res*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.1

Arendasy, M. (2009). *BFSI: Big-Five Struktur-Inventar (Test & Manual)*. Mödling: Austria: SCHUHFRIED GmbH.

Asendorpf, J. B. (2003). Head-to-head comparison of the predictive validity of personality types and dimensions. *European Journal of Personality*, *17*(5), 327–346. https://doi.org/10.1002/per.492

Asendorpf, J. B., & van Aken, MarcelA. G. (1999). Resilient, overcontrolled, and undercontroleed personality prototypes in childhood: Replicability, predictive power, and the trait-type issue. *Journal of Personality and Social Psychology*, *77*(4), 815–832. https://doi.org/10.1037/0022-3514.77.4.815

Au, Q. (2019). *fxtract: Feature extraction from grouped data*. https://github.com/QuayAu/fxtract

Bailey, S. L., & Heitkemper, M. M. (2001). Circadian rhythmicity of cortisol and body temperature: Morningness–eveningness effects. *Chronobiology International*, *18*(2), 249–261. https://doi.org/10.1081/CBI-100103189

Basagaña, X., Barrera-Gómez, J., Benet, M., Antó, J. M., & Garcia-Aymerich, J. (2013). A framework for multiple imputation in cluster analysis. *American Journal of Epidemiology*, *177*(7), 718–725. https://doi.org/10.1093/aje/kws289

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Borger, J. N., Huber, R., & Ghosh, A. (2019). Capturing sleep–wake cycles by using day-to-day smartphone touchscreen interactions. *NPJ Digital Medicine*, *2*(1), 1–8.

Braeken, J., & Van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods*, *22*(3), 450–466. https://doi.org/10.1037/met0000074

Caci, H., Deschaux, O., Adan, A., & Natale, V. (2009). Comparing three morningness scales: Age and gender effects, structure and cut-off criteria. *Sleep Medicine*, *10*(2), 240–245. https://doi.org/10.1016/j.sleep.2008.01.007

Cavallera, G., & Giudici, S. (2008). Morningness and eveningness personality: A survey in literature from 1995 up till 2006. *Personality and Individual Differences*, *44*(1), 3–21. https://doi.org/10.1016/j.paid.2007.07.009

Chen, Z., Lin, M., Chen, F., Lane, N. D., Cardone, G., Wang, R., …, & Campbell, A. T.(2013). *Unobtrusive sleep monitoring using smartphones*. In *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare*, pp. 145–152. https://doi.org/10.4108/icst.pervasivehealth.2013.252148

Costa Jr, P. T., Herbst, J. H., McCrae, R. R., Samuels, J., & Ozer, D. J. (2002). The replicability and utility of three personality types. *European Journal of Personality*, *16*(S1), S73–S87. https://doi.org/10.1002/per.448

Costa, P. T., & McCrae, R. R. (2008). The Revised Neo Personality Inventory (NEO-PI-R), *The SAGE handbook of personality theory and assessment*, pp. 179–198. https://doi.org/10.4135/9781849200479.n9

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619. https://doi.org/10.1146/annurev.psych.093008.100356

Díaz-Morales, J. F. (2007). Morning and evening-types: Exploring their personality styles. *Personality and Individual Differences*, *43*(4), 769–778. https://doi.org/10.1016/j.paid.2007.02.002

Duggan, K. A., Friedman, H. S., McDevitt, E. A., & Mednick, S. C. (2014). Personality and healthy sleep: The importance of conscientiousness and neuroticism. *PloS ONE*, *9*(3), e90628. https://doi.org/10.1371/journal.pone.0090628

Dunn, J. C. (1974). Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics*, *4*(1), 95–104. https://doi.org/10.1080/01969727408546059

Ferrara, M., & De Gennaro, L. (2001). How much sleep do we need? *Sleep Medicine Reviews*, *5*(2), 155–179. https://doi.org/10.1053/smrv.2000.0138

Fleiss, J. L., Lawlor, W., Platman, S. R., & Fieve, R. R. (1971). On the use of inverted factor analysis for generating typologies. *Journal of Abnormal Psychology*, *77*(2), 127.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, *74*(5), 1337–1349. https://doi.org/10.1037/0022-3514.74.5.1337

Gradisar, M., Wright, H., Robinson, J., Pain, S., & Gamble, A. (2008). Adolescent napping behavior: Comparisons of school week versus weekend sleep patterns. *Sleep and Biological Rhythms*, *6*(3), 183–186. https://doi.org/10.1111/j.1479-8425.2008.00351.x

Gray, E. K., & Watson, D. (2002). General and specific traits of personality and their relation to sleep and academic performance. *Journal of Personality*, *70*(2), 177–206. https://doi.org/10.1111/1467-6494.05002

Grund, S., Lüdtke, O., & Robitzsch, A. (2018). Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*, *21*(1), 111–149. https://doi.org/10.1177/1094428117703686

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On clustering validation techniques. *Journal of Intelligent Information Systems*, *17*(2-3), 107–145. https://doi.org/10.1023/A:1012801612483

Harari, G. M., Müller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., Rentfrow, P. J., Campbell, A. T., & Gosling, S. D. (2019). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of Personality and Social Psychology*. Advance online publication, https://doi.org/10.1037/pspp0000245

Hennig, C. (2007). Cluster-wise assessment of cluster stability. *Computational Statistics & Data Analysis*, *52*(1), 258–271. https://doi.org/10.1016/j.csda.2006.11.025

Hennig, C. (2008). Dissolution point and isolation robustness: Robustness criteria for general cluster analysis methods. *Journal of Multivariate Analysis*, *99*(6), 1154–1176. https://doi.org/10.1016/j.jmva.2007.07.002

Hennig, C. (2018). *fpc: Flexible procedures for clustering*. R package version 2.1-11.1, https://CRAN.R-project.org/package=fpc

Hicks, L. E. (1984). Conceptual and empirical analysis of some assumptions of an explicitly typological theory. *Journal of Personality and Social Psychology*, *46*(5), 1118–1131. https://doi.org/10.1037/0022-3514.46.5.1118

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms. *International Journal of Chronobiology*, *4*(2), 97–110.

Katzenberg, D., Young, T., Finn, L., Lin, L., King, D. P., Takahashi, J. S., & Mignot, E. (1998). A clock polymorphism associated with human diurnal preference. *Sleep*, *21*(6), 569–576. https://doi.org/10.1093/sleep/21.6.569

Križan, Z., & Hisler, G. (2019). Personality and sleep: Neuroticism and conscientiousness predict behaviourally recorded sleep years

later. *European Journal of Personality*, *33*, 133–153. https://doi.org/10.1002/per.2191

Kuznetsova, A., Brockhoff, P. B., & Christensen, RuneH. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lipnevich, A. A., Credé, M., Hahn, E., Spinath, F. M., Roberts, R. D., & Preckel, F. (2017). How distinctive are morningness and eveningness from the big five factors of personality? A meta-analytic investigation. *Journal of Personality and Social Psychology*, *112*(3), 491. https://doi.org/10.1037/pspp0000099

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. https://doi.org/10.1007/bf02296272

Mateo, M. J. C., Diaz-Morales, J. F., Barreno, C. E., Prieto, P. D., & Randler, C. (2012). Morningness–eveningness and sleep habits among adolescents: Age and gender differences. *Psicothema*, *24*(3), 410–415.

Meehl, P. E. (2004). What's in a taxon? *Journal of Abnormal Psychology*, *113*(1), 39. https://doi.org/10.1037/0021-843X.113.1.39

Min, J.-K., Doryab, A., Wiese, J., Amini, S., Zimmerman, J., & Hong, J. I. (2014). Toss'n'turn: Smartphone as sleep and sleep quality detector. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 477–486. https://doi.org/10.1145/2556288.2557220

Natale, V., & Cicogna, P. (2002). Morningness–eveningness dimension: Is it really a continuum? *Personality and Individual Differences*, *32*(5), 809–816. https://doi.org/10.1016/S0191-8869(01)00085-X

Ohayon, M., Wickwire, E. M., Hirshkowitz, M., Albert, S. M., Avidan, A., Daly, F. J., …, & Hazen, N. (2017). National sleep foundation's sleep quality recommendations: First report. *Sleep Health*, *3*(1), 6–19. https://doi.org/10.1016/j.sleh.2016.11.006

Preckel, F., Fischbach, A., Scherrer, V., Brunner, M., Ugen, S., Lipnevich, A. A., & Roberts, R. D. (2019). Circadian preference as a typology: Latent-class analysis of adolescents' morningness/eveningness, relation with sleep behavior, and with academic outcomes. *Learning and Individual Differences*, *78*, 101725. https://doi.org/10.1016/j.lindif.2019.03.007

Putilov, A. A. (2017). Owls, larks, swifts, woodcocks and they are not alone: A historical review of methodology for multidimensional self-assessment of individual differences in sleep-wake pattern. *Chronobiology international*, *34*(3), 426–437. https://doi.org/10.1080/07420528.2017.1278704

Putilov, A. A., Donskaya, O. G., & Verevkin, E. G. (2015). How many diurnal types are there? a search for two further "bird species". *Personality and Individual Differences*, *72*, 12–17. https://doi.org/10.1016/j.paid.2014.08.003

R Core Team (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science. *Perspectives on Psychological Science*, *11*(6), 838–854. https://doi.org/10.1177/1745691616650285

Kafadar, K. (2003). John tukey and robustness. *Statistical Science*, *18*(3), 319–331. https://doi.org/10.1214/ss/1076102419

Lauderdale, D. S., Knutson, K. L., Yan, L. L., Liu, K., & Rathouz, P. J. (2008). Self-reported and measured sleep duration. *Epidemiology*, *19*(6), 838–845. https://doi.org/10.1097/ede.0b013e318187a7b0

Lin, Y.-H., Wong, B.-Y., Lin, S.-H., Chiu, Y.-C., Pan, Y.-C., & Lee, Y.-H. (2019). Development of a mobile application (app) to delineate "digital chronotype" and the effects of delayed chronotype by bedtime smartphone use. *Journal of Psychiatric Research*, *110*, 9–15. https://doi.org/10.1016/j.jpsychires.2018.12.012

Newman, D. A. (2014). Missing data. *Organizational Research Methods*, *17*(4), 372–411. https://doi.org/10.1177/1094428114548590

Randler, C. (2007). Gender differences in morningness–eveningness assessed by self-report questionnaires: A meta-analysis. *Personality and Individual Differences*, *43*(7), 1667–1675. https://doi.org/10.1016/j.paid.2007.05.004

Randler, C. (2008). Morningness–eveningness, sleep–wake variables and big five personality factors. *Personality and Individual Differences*, *45*(2), 191–196. https://doi.org/10.1016/j.paid.2008.03.007

Randler, C., Díaz-Morales, J. F., Rahafar, A., & Vollmer, C. (2016). Morningness–eveningness and amplitude–development and validation of an improved composite scale to measure circadian preference and stability (messi). *Chronobiology international*, *33*(7), 832–848. https://doi.org/10.3109/07420528.2016.1171233

Randler, C., & Engelke, J. (2019). Gender differences in chronotype diminish with age: A meta-analysis based on morningness/chronotype questionnaires. *Chronobiology International*, *36*(7), 888–905. https://doi.org/10.1080/07420528.2019.1585867

Randler, C., Schredl, M., & Göritz, A. S. (2017). Chronotype, sleep behavior, and the big five personality factors. *Sage Open*, *7*(3), 1–9. https://doi.org/10.1177/2158244017728321

Revelle, W. (2018). *psych: Procedures for psychological, psychometric, and personality research*. R package version 1.8.12, https://CRAN.R-project.org/package=psych

Roenneberg, T. (2015). Having trouble typing? What on earth is chronotype? *Journal of Biological Rhythms*, *30*(6), 487–491. https://doi.org/10.1177/0748730415603835

Roenneberg, T., Allebrandt, K. V., Merrow, M., & Vetter, C. (2012). Social jetlag and obesity. *Current Biology*, *22*(10), 939–943. https://doi.org/10.1016/j.cub.2012.03.038

Roenneberg, T., Keller, L. K., Fischer, D., Matera, J. L., Vetter, C., & Winnebeck, E. C. (2015). Human activity and rest in situ. *Methods in Enzymology*, *552*, 257–283. https://doi.org/10.1016/bs.mie.2014.11.028

Roenneberg, T., Kuehnle, T., Juda, M., Kantermann, T., Allebrandt, K., Gordijn, M., & Merrow, M. (2007). Epidemiology of the human circadian clock. *Sleep Medicine Reviews*, *11*(6), 429–438. https://doi.org/10.1016/j.smrv.2007.07.005

Roenneberg, T., Wirz-Justice, A., & Merrow, M. (2003). Life between clocks: Daily temporal patterns of human chronotypes. *Journal of Biological Rhythms*, *18*(1), 80–90. https://doi.org/10.1177/0748730402239679

Roepke, S. E., & Duffy, J. F. (2010). Differential impact of chronotype on weekday and weekend sleep timing and duration. *Nature and Science of Sleep*, *2*, 213–220. https://doi.org/10.2147/NSS.S12572

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, *88*(424), 1273–1283. https://doi.org/10.2307/2291267

Schoedel, R., Au, Q., Völkel, S. T., Lehmann, F., Becker, D., Bühner, M., Bischl, B., Hussmann, H., & Stachl, C. (2018). Digital footprints of sensation seeking. *Zeitschrift für Psychologie*, *226*(4), 232–245. https://doi.org/10.1027/2151-2604/a000342

Schoedel, R., Pargent, F., Au, Q., Völkel, S., Schuwerk, T., Bühner, M., & Stachl, C. (2020). To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day-night behavior patterns. *European Journal of Personality*, *34*, 733–752. https://doi.org/10.1002/per.2258

Schuwerk, T., Kaltefleiter, L. J., Au, J.-Q., Hoesl, A., & Stachl, C. (2019). Enter the wild: Autistic traits and their relationship to mentalizing and social interaction in everyday life. *Journal of Autism and Developmental Disorders*, *49*(10), 4193–4208. https://doi.org/10.1007/s10803-019-04134-6

Sella, E., Carbone, E., Toffalini, E., & Borella, E. (2020). Personality traits and sleep quality: The role of sleep-related beliefs. *Personality and Individual Differences*, *156*, 109770. https://doi.org/10.1016/j.paid.2019.109770

Shah, A. D., Bartlett, J. W., Carpenter, J., Nicholas, O., & Hemingway, H. (2014). Comparison of random forest and parametric imputation models for imputing missing data using mice: A caliber study. *American Journal of Epidemiology*, *179*(6), 764–774. https://doi.org/10.1093/aje/kwt312

Shambroom, J. R., Fábregas, S. E., & Johnstone, J. (2012). Validation of an automated wireless system to monitor sleep in healthy adults. *Journal of Sleep Research*, *21*(2), 221–230. https://doi.org/10.1111/j.1365-2869.2011.00944.x

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). *Specification curve: Descriptive and inferential statistics o all reasonable specifications*. SSRN, https://doi.org/10.2139/ssrn.2694998

Spears, S. K., Montgomery-Downs, H. E., Steinman, S. A., Duggan, K. A., & Turiano, N. A.(2019). Sleep: A pathway linking personality to mortality risk. *Journal of Research in Personality*, *81*, 11–24. https://doi.org/10.1016/j.jrp.2019.04.007

Stachl, C., Au, Q., Schoedel, R., Buschek, D., Völkel, S., Schuwerk, T., …, & Bühner, M. (2019). *Behavioral patterns in smartphone usage predict big five personality traits*. https://doi.org/10.31234/osf.io/ks4vd

Stachl, C., Hilbert, S., Au, J.-Q., Buschek, D., De Luca, A., Bischl, B., Hussmann, H., & Bühner, M. (2017). Personality traits predict smartphone usage. *European Journal of Personality*, *31*(6), 701–722. https://doi.org/10.1002/per.2113

Stachl, C., Schoedel, R., Au, Q., Völkel, S., Buschek, D., Hussmann, H., …, & Bühner, M. (2018). *The phonestudy project*. https://doi.org/10.17605/osf.io/ut42y

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, *11*(5), 702–712. https://doi.org/10.1177/1745691616658637

Sutin, A. R., Gamaldo, A. A., Stephan, Y., Strickhouser, J. E., & Terracciano, A. (2019). Personality traits and the subjective and objective experience of sleep. *International Journal of Behavioral Medicine*. https://doi.org/10.1007/s12529-019-09828-w

Takano, K., Sakamoto, S., & Tanno, Y. (2014). Repetitive thought impairs sleep quality: An experience sampling study. *Behavior Therapy*, *45*(1), 67–82. https://doi.org/10.1016/j.beth.2013.09.004

Tan, P.-N., Steinbach, M., & Kumar, V. (2006). *Cluster analysis: Basic concepts and algorithms*.

Terman, J. S., Terman, M., Lo, E.-S., & Cooper, T. B. (2001). Circadian time of morning light administration and therapeutic response in winter depression. *Archives of General Psychiatry*, *58*(1), 69–75. https://doi.org/10.1001/archpsyc.58.1.69

Tibshirani, R., & Walther, G. (2005). Cluster validation by prediction strength. *Journal of Computational and Graphical Statistics*, *14*(3), 511–528. https://doi.org/10.1198/106186005X59-243

Tonetti, L., Pascalis, V. D., Fabbri, M., Martoni, M., Russo, P. M., & Natale, V. (2016). Circadian typology and the alternative five-factor model of personality. *International Journal of Psychology*, *51*(5), 332–339. https://doi.org/10.1002/ijop.12170

Tsaousis, I. (2010). Circadian preferences and personality traits: A meta-analysis. *European Journal of Personality*, *24*(4), 356–373. https://doi.org/10.1002/per.754

Ushey, K., McPherson, J., Cheng, J., Atkins, A., & Allaire, J. (2018). *packrat: A dependency management system for projects and their r package dependencies*. R package version 0.4.9-3, https://CRAN.R-project.org/package=packrat

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3), 1–67. https://www.jstatsoft.org/v45/i03/

Vela-Bueno, A., Fernandez-Mendoza, J., Olavarrieta-Bernardino, S., Vgontzas, A. N., Bixler, E. O., de la Cruz-Troca, J. J., Rodriguez-Muñoz, A., & Oliván-Palacios, J. (2008). Sleep and behavioral correlates of napping among young adults: A survey of first-year university students in madrid, spain. *Journal of American College Health*, *57*(2), 150–158. https://doi.org/10.3200/jach.57.2.150-158

Vitale, J. A., Roveda, E., Montaruli, A., Galasso, L., Weydahl, A., Caumo, A., & Carandente, F.(2015). Chronotype influences activity circadian rhythm and sleep: Differences in sleep quality between weekdays and weekend. *Chronobiology International*, *32*(3), 405–415. https://doi.org/10.3109/07420528.2014.986273

Wei, T., & Simko, V. (2017). *R package corrplot: Visualization of a correlation matrix*. R package version 0.84, https://github.com/taiyun/corrplot

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. http://ggplot2.org

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *Dplyr: A grammar of data manipulation*. R package version 0.8.0.1, https://CRAN.R-project.org/package=dplyr

Wittmann, M., Dinich, J., Merrow, M., & Roenneberg, T. (2006). Social jetlag: misalignment of biological and social time. *Chronobiology International*, *23*(1-2), 497–509. https://doi.org/10.1080/07420520500545979

Zavada, A., Gordijn, M. C., Beersma, D. G., Daan, S., & Roenneberg, T. (2005). Comparison of the Munich Chronotype Questionnaire with the Horne--Östberg's Morningness–Eveningness Score. *Chronobiology International*, *22*(2), 267–278. https://doi.org/10.1081/CBI-200053536

Zimmermann, L. K. (2011). Chronotype and the transition to college life. *Chronobiology International*, *28*(10), 904–910. https://doi.org/10.3109/07420528.2011.618959

APPENDIX A: SUPPLEMENTAL METHOD

Measures

Table A1. Description of the algorithm for detecting nightly inactivity

Step	Description
1	Exclude passive smartphone events (GPS logs, notifications, and related screen events)
2	Exclude active usage events lasting shorter than two minutes and label them as checking behavior
3	Search for the maximum distance between consecutive events
4	Label the starting point of the maximum distance as last event of the day and the end point as first event of the next day

To avoid longer periods of inactivity being detected during the day, the time frame for maximum distance detection was limited to 6.00 PM to 2.00 PM of the following day. We defined and filtered checking behaviour, because we wanted to exclude less significant actions like checking the clock or notification texts.

Clustering

K-Means Algorithm

For clustering, we used the *k-means* algorithm, which is one of the most frequently used algorithms for clustering (Tan et al., 2006). In the following section, we only describe the basic principles behind *k-means* clustering and refer the interested reader to Tan et al. (2006) for a detailed explanation. After the user has defined the expected number of clusters *k*, *k* points in the sample data are randomly determined and represent initial centroids. In the second step, all remaining data

points are assigned to the centroid for which the Euclidean distance is lowest. Afterward, the centroids in each of the k clusters are updated by calculating the arithmetic mean of all points in the respective clusters. Step by step, the procedures are repeated as long as the centroids do not change anymore, which indicates that the grouping structure in the data has been identified. As the centroid represents the data points within the clusters, *k-means* clustering is also often referred to as prototype-based or partitional clustering (Tan et al., 2006).

Evaluation Metrics

To ensure cluster validity, we took several steps to find non-random structures in our data. The first step is to determine the appropriate number of clusters. Tibshirani and Walther (2005) proposed to reframe clustering as a supervised prediction problem by splitting the data into a training and a test set and estimating the number of pairwise cases that are assigned to the same cluster in the test set based on centroids of the training set. The associated prediction strength measure defined by Tibshirani and Walther (2005) can be used to determine an optimal number of clusters. Another important aspect is cluster stability (Hennig, 2007). If clusters disappear when data are slightly modified, they are not regarded as stable and consequently might reflect only random structure. Hennig (2007) therefore suggests bootstrapping the data and considering the Jaccard coefficient (*JC*) for each cluster separately. The *JC* gives the proportion of data points (participants) that are assigned to the same cluster across the bootstrapped iterations, thus expressing the similarity of cluster solutions across bootstrapped data sets on a cluster-wise basis (Hennig, 2007). Further descriptive measures of cluster stability are the criteria of *recovery* and *dissolution,* which count how often each cluster has been successfully recovered and dissolved across all bootstrap iterations (Hennig, 2007; 2008). As recommended by Hennig (2018), we used 100 bootstrap replications and interpreted clusters as stable if the *JC* exceeded values above 0.85.

Imputation of Missing Values

Based on a variable-by-variable procedure, missings are replaced by values of a conditional distribution, which results from estimating imputation models using the remaining variables of the data set (van Buuren & Groothuis-Oudshoorn, 2011). We chose the random forest as an imputation algorithm as it has been proven useful for complex, incomplete data problems (Shah et al., 2014). To reduce the imputation bias caused by stochastic variation, we specified 50 imputation models. For each of the resulting 50 data sets, we performed a separate cluster analysis and report the mean/modus of the performance coefficients and cluster membership across data sets (Basagaña et al., 2013).

Multilevel Modelling

Decisions in the Multiverse

For constructing the data multiverse (Steegen et al., 2016), we considered the following decisions concerning preprocessing steps:

Descision 1: Coding of Weekend. In an earlier draft of the manuscript, we defined the weekend not as a period from *Friday to Sunday*, but from *Friday to Monday*. We found it challenging to decide whether Sunday evening and the following night still belong to the weekend or whether it is more of a weekday in terms of sleep–wake behaviour. In sleep research, the nights from Friday to Saturday and from Saturday to Sunday are considered as weekends traditionally. Because on Monday, one usually has to attend to social obligations again, sleep behaviour during the night from Sunday to Monday is assumed not to be chosen as freely and used to balance the weekly sleep deficit as the other two weekend nights (Roenneberg et al., 2007). Despite the standard in sleep research, we want to include both variants in our multilevel modelling and thus make our research process transparent.

Decision 2: Number of Weeks. We considered the number of repeated measurements to be plausible as both *3* and *4 weeks* because we noticed during the aggregation of the raw timestamped event data that some participants had only partially participated in the last weekend (e.g. only on Saturday, no longer on Sunday).

Decision 3: Outliers. For the handling of outliers, we found two points of view plausible. First, smartphone sensing-derived variables are usually susceptible to distortion due to data errors, which do not matter if enough data are aggregated using robust measures over a longer period. However, as for week-based variables, only a few single data points can be summarized; outliers due to data errors are more problematic. Therefore, we identified outliers as cases *deviating more than three times the mean absolute deviation from the median* and replaced them by the *person-specific median* of the corresponding variable. Second, the identification of outliers arising from the underlying smartphone usage behaviour can be emphasized. In this case, it would be plausible to use a method for outlier handling that limits the variability of the smartphone indicators less than using the median. To cover this aspect, we used *winsorization* as the second alternative.

Decision 4: Missing Values. Dealing with missing values in multilevel models is a challenging task. Traditionally, *listwise deletion* has been used, which uses only complete observations for estimating the model (e.g. Newman, 2014). Besides the disadvantage of the reduced sample and power, results are likely to be biased if the incomplete observations differ systematically from complete observations (Newman, 2014; Grund et al., 2018). An alternative approach to deal with missing data is to apply *multiple imputation*. However, in the context of multilevel models, this is not a trivial task as the imputation model itself should consider the multilevel structure. Current methods and software implementations are reaching their limits if more complicated use cases like random slopes or cross-level interactions are included in the model (Grund et al., 2018). For our analyses, we used the multivariate imputation by chained equations technique and implemented a random slope imputation model with group-level variables as proposed by Grund et al. (2018). Please note the imputation bias because we were unable to

integrate cross-level interactions with existing software implementations. In addition, Grund et al. (2018) point out that this area of research is still ongoing and that there are no clear recommendations for dealing with missing data in use cases such as ours.

Model Description

To comprehensibly illustrate the multilevel model used for the multiverse analysis, we present the pseudo-model equation using the lmer syntax of the *lme4 package* in R (Bates et al., 2015). We specified a *random-intercept-random-slope model* predicting the mean duration of nightly inactivity on weekends based on the mean nightly inactivity duration during the previous week (level 1). Chronotype, the big five traits, age, and gender were included as level 2 predictors. The level 1 predictor duration of nightly inactivity during the week was person centred and the individual mean was entered as level 2 predictor (Curran & Bauer, 2011). Besides, the cross-level interaction of the mean nightly inactivity duration during the previous week and chronotype was added:

$$
\begin{aligned}
\text{Nightly Inactivity}_{weekend} \sim\; & 1 \\
& + \text{Nightly Inactivity}_{week}(\text{L1, z, pc}) \\
& + \text{Chronotype (L2, z, gc)} \\
& + \text{Nightly Inactivity}_{week}(\text{L2, z, gc}) \\
& + \text{Openness (L2, z, gc)} \\
& + \text{Conscientiousness (L2, z, gc)} \\
& + \text{Extraversion (L2, z, gc)} + \text{Agreeableness (L2, z, gc)} \\
& + \text{Emotional Stability (L2, z, gc)} + \text{Age (L2, z, gc)} \\
& + \text{Gender (L2, dc)} + \text{Nightly Inactivity}_{week}(\text{L1, z, pc}) \\
& * \text{Chronotype (L2, z, gc)} + (1 + \text{Nightly Inactivity}_{week} \\
& (\text{L1, z, pc}) | \text{userid}),
\end{aligned} \tag{A1}
$$

where L1 denotes predictors on level 1, L2 denotes predictors on level 2, z denotes that predictors were *z*-standardized, pc denotes that predictors were person-mean-centred, gc denotes that predictors were grand-mean-centred, and dc denotes that gender was dummy coded (0 = male, 1 = female).

APPENDIX B: SUPPLEMENTAL RESULTS

Big Five Personality Traits

Table B1. Descriptive statistics of personality factors and facets

Variable	M	SD	alpha CI95%
Openness	-0.05	0.71	[0.93, 0.94]
O1: Openness to imagination	1.28	1.41	[0.84, 0.87]
O2: Openness to aesthetics	0.37	1.29	[0.85, 0.88]
O3: Openness to feelings	2.05	2.09	[0.91, 0.93]
O4: Openness to actions	1.35	1.4	[0.84, 0.87]
O5: Openness to ideas	1.65	1.42	[0.82, 0.86]
O6: Openness to value/norm system	0.9	1.02	[0.73, 0.79]
Conscientiousness	-0.09	0.74	[0.95, 0.96]
C1: Competence	0.84	1.22	[0.76, 0.82]
C2: Love of order	1.1	1.58	[0.87, 0.90]
C3: Sense of duty	1.93	1.41	[0.80, 0.85]
C4: Ambition	1.83	1.68	[0.86, 0.89]
C5: Discipline	1.45	1.46	[0.81, 0.86]
C6: Caution	1.51	1.34	[0.80, 0.84]
Extraversion	-0.01	0.74	[0.95, 0.96]
E1: Friendliness	1.45	1.29	[0.80, 0.84]
E2: Sociableness	1.3	1.74	[0.89, 0.92]
E3: Assertiveness	0.45	1.38	[0.84, 0.87]
E4: Dynamism	1.2	1.59	[0.85, 0.88]
E5: Adventurousness	0.45	1.49	[0.88, 0.91]
E6: Cheerfulness	1.97	1.64	[0.86, 0.89]
Ageeableness	-0.06	0.75	[0.92, 0.94]
A1: Willingness to trust	0.4	1.43	[0.86, 0.89]
A2: Genuineness	1.01	0.94	[0.61, 0.70]
A3: Helpfulness	1.65	1.38	[0.77, 0.82]
A4: Obligingness	1.17	1.31	[0.81, 0.85]
A5: Modesty	0.77	1.13	[0.79, 0.84]
A6: Good naturedness	2.1	1.77	[0.84, 0.88]
Emotional Stability	-0.03	0.71	[0.93, 0.94]
ES1: Carefreeness	0.12	1.3	[0.82, 0.86]
ES2: Equanimity	0.57	1.07	[0.78, 0.83]
ES3: Positive mood	0.95	1.43	[0.84, 0.88]
ES4: Self consciousness	0.66	1.18	[0.83, 0.86]
ES5: Self control	0.64	1	[0.74, 0.81]
ES6: Emotional robustness	0.65	1.19	[0.80, 0.85]

N = 597; Alpha CI95% = 95% bootstrapped confidence intervals for Cronbach alpha coefficients.