Roy, Chandreyee; Monsivais, Daniel; Bhattacharya, Kunal; Dunbar, Robin I.M.; Kaski, Kimmo

Morningness–eveningness assessment from mobile phone communication analysis

scientific reports

OPEN

Morningness–eveningness assessment from mobile phone communication analysis

Chandreyee Roy[1✉], Daniel Monsivais[1], Kunal Bhattacharya[1,2], Robin I. M. Dunbar[3] & Kimmo Kaski[1,4]

Human behaviour follows a 24-h rhythm and is known to be governed by the individual chronotypes. Due to the widespread use of technology in our daily lives, it is possible to record the activities of individuals through their different digital traces. In the present study we utilise a large mobile phone communication dataset containing time stamps of calls and text messages to study the circadian rhythms of anonymous users in a European country. After removing the effect of the synchronization of East-West sun progression with the calling activity, we used two closely related approaches to heuristically compute the chronotypes of the individuals in the dataset, to identify them as morning persons or "larks" and evening persons or "owls". Using the computed chronotypes we showed how the chronotype is largely dependent on age with younger cohorts being more likely to be owls than older cohorts. Moreover, our analysis showed how on average females have distinctly different chronotypes from males. Younger females are more larkish than males while older females are more owlish. Finally, we also studied the period of low calling activity for each of the users which is considered as a marker of their sleep period during the night. We found that while "extreme larks" tend to sleep more than "extreme owls" on the weekends, we do not observe much variation between them on weekdays. In addition, we have observed that women tend to sleep even less than males on weekdays while there is not much difference between them on the weekends.

Human beings are known to be diurnal in nature that is characterized by a period of activity during the day and a period of inactivity during the night. These rhythmic activities are entrained to the light-dark cycles of the solar clock that occur due to the earth's rotation around the sun[1,2]. This rhythmicity is also affected by the social constraints of living in a society, for example going to work on time, while being generally aligned according to the solar clock. The discovery of artificial light has had a considerable impact on the daily activities of human beings, thus eventually affecting their sleeping patterns[3,4]. However, the physiological activities and behaviour of humans are well known to follow a circadian rhythm that is closely related to their individual chronotypes.

The chronotype varies among different individuals and is dependent on various factors like gender, age, and genetics, among others[5–8]. Individuals having early chronotypes rise early in the morning and sleep early as well[9]. They are well known in the literature as "larks". On the other hand, late chronotypes wake up late as well as sleep late and are befittingly known as "owls". The rest of the population falls within this spectrum from larks to owls, identified by their individual chronotypes[10]. Identification of chronotypes is an important issue, because an individual's productivity could depend on the synchrony between her inherent chronotype and her daily work-life timings. We might expect a lark to be more productive during the morning and an owl to be more productive during the evening. Workplaces mostly have schedules that are biased towards early chronotypes and not for the late ones. This can cause sleep deprivation and poor eating habits in the latter which can further lead to health complications[11–13].

Traditionally, studies concerning identification of chronotypes have been done using the Munich Chrono-Type Questionnaire (MCTQ)[14–16]. This questionnaire, the first of its kind, consists of a set of unique questions relating to an individual's sleep-wake cycle along with iconic supporting drawings that help clarify differences between the time an individual decides to go to sleep and the actual time of falling asleep. Other questionnaires like the Morningness–Eveningness Questionnaire (MEQ)[9], that looks into the sleep/wake time preferences of the

[1]Department of Computer Science, Aalto University School of Science, Espoo, Finland. [2]Department of Industrial Engineering and Management, Aalto University School of Science, Espoo, Finland. [3]Department of Experimental Psychology, University of Oxford, Oxford, UK. [4]The Alan Turing Institute, London, UK. ✉email: chandreyee.roy@aalto.fi

individuals instead of their actual sleep/wake timings have also been used in this context. These survey studies have been used worldwide and in different populations to study the changes in sleep-wake behaviour across age groups (adolescents, adults, and others)[5,17], as well as in different social and psychological settings[18]. Individuals are found to be of early chronotypes in adolescent stages of their lives and gradually changing to late chronotypes through their teenage years reaching a maximum around 20 years of age. After 20, they have been observed to gradually change back to behaving as early chronotypes[5]. Additionally, gender differences of chronotypes[19] have also been studied using MEQ in which they concluded the existence of different synchronization patterns for men and women. While these surveys are excellent tools to understand human behaviour, they are generally restricted in terms of sample sizes, memory of the participants, and what is socially and societally expected.

With the advent of the digital age and its rapid development over the years, humans have been increasingly becoming dependent on technology for their daily needs. This has led to them leaving traces of their activity online in the digital world, which in turn can give us considerable insight into their daily activities. Data from mobile phone communication records containing call time stamps and GPS locations along with duration of the voice calls and text messages sent by anonymized users portray periods of activity and inactivity by individuals and consequently are useful for studying their chronotypes. Additionally, data analysis studies harnessing data from the mobile phone call detail records of a very large number of users provide close to accurate description of the dynamics involved in human social behaviour[20–25]. Access to these kinds of large population datasets enables us to study the social networks formed by humans and relationships formed by them in the networks[26–29]. It can also be used to study migration patterns[30] of the individuals and more recently, it has been used to study the behaviour of people during the COVID -19 pandemic as well[31].

Since mobile phone communication datasets clearly display the circadian rhythms of human activity by looking at the frequency of calls made by an individual during a 24 h cycle[32], one can broadly determine when an individual is active or inactive. Studies show that the calling activity of individual users on average follow a bimodal distribution where users are active twice during the day with the two peaks in the frequency of calls occurring in the morning and in the evening, respectively. Thus, one can identify the chronotype of an individual by inspecting these rhythmic cycles of calling activity. For example, studies by Aledavood and co-authors[33,34] used data from smartphones of volunteers to identify larks and owls as well as their social networks. They also observed that the personal networks maintained by owls are larger than those maintained by larks.

In our current study, we have used the call detail records of a large population-level dataset to observe the morning and evening calling activities of the users living in a European country during the years 2007, 2008 and 2009. Our aim is to construct a chronotype directly from these activities collected during the entire 24-h cycle of all seven days of the week instead of only looking at mid-sleep times only on weekends.

Materials and methods

Individual mobile phone Call Details Records.   The dataset used in this study comprises the Call Details Records (CDRs) from individuals living in a southern European country, which had mobile phone subscription with a specific service provider that had 20% of the market share in that country. It results from the merging of 3 separate subsets (January–December 2007[32,35], January–December 2008, and January–December 2009), altogether covering a three year period. The data-sets were anonymized before being handed over by the service provider, such that the true identity of the individual is unknown and each individual is described by a unique identifier (id-number). The CDRs lists all the outgoing calls made by each individual during a three year period, and each entry includes, the id-numbers associated with the caller and the callee, the time and date when the communication event happened, as well as the type of communication event (call or text message)[36]. The data-set includes also user-contract data-sets with some demographic information (age, gender, and registered postal code) of the individuals who were subscribers of the service provider in at least one of the three periods. Over the three year period, different individuals start a new subscription and/or terminate the contract with the service provider, but of the order of six hundred thousand individuals remained loyal to the service provider (i.e. their contracts started before January 1st 2007 and were still active on December 31st, 2009). From this set of loyal subscribers whose demographic information was available (some users have missing entries or they contain typos), we chose 11,178 individuals, who made at least 100 hundred calls/text messages each year and with a total number of calls/sms not exceeding 5000 calls (to exclude possible calls centers or subscription sellers).

Geographical grouping based on individual's location.   Using the information of the postal code available in the user-contract files, we split individuals into 5 groups, each one falling inside a longitudinal band enclosing their geographical location. However, as a signed non-disclosure agreement (NDA) prohibit us to disclose the country where the service provider offered, the actual values delimiting each selected geographical longitudinal band are masked. From here on, the longitudinal values will be reported from a reference point located near the easternmost part of the country, which will work as the zero reference. Thus, five latitudinal bands L_I-L_V of widths 2.5°, 3.05°, 2.75°, 2.2° and 2.8° are defined, separated by exclusion bands of width 0.2 degrees, and with the first band L_I (the reference point) the easternmost, and the subsequent bands located to the west progressively until the last longitudinal band L_V lies in the westernmost part of the region (13.3°-wide). Table 1 lists the number of individuals in each longitudinal band, as well, gender and age distribution information of the population on each band. The widths of the longitudinal bands are adjusted such that the number of people in each band are roughly of the same order.

The reason for this longitudinal splitting is to take into account of the dependence of the human chronotype on the East-West progression of the Sun, as has been shown in earlier studies[1,35]. In Fig. 1, for each of the five longitudinal bands L_I-L_V, the calling activity of the analyzed individuals during weekday nights (Mondays to Thursdays) aggregated over the 3 year period is shown. There, a clear shift between the calling activity distribution

Feature	Longitudinal band				
	L_I	L_{II}	L_{III}	L_{IV}	L_V
Limits	$[0°, -2.3°]$	$[-2.5°, -5.35°]$	$[-5.55°, -8.1°]$	$[-8.3°, -10.3°]$	$[-10.5°, -13.1°]$
Width	2.5°	3.05°	2.75°	2.2°	2.8°
Individuals	2031	2589	2386	2816	1356
Females	1108	1479	1338	1564	795
Males	923	1110	1048	1252	561
Young (18-35)	637	891	854	928	443
Mid-age (35-60)	1017	1272	1177	1425	700
Old (60-78)	377	426	355	463	213

Table 1. Demographic information of the population.
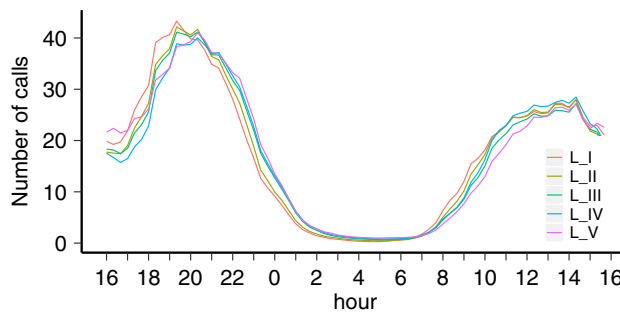
Figure 1. Aggregated individual calling activity for weekdays in 5 geographical regions. For each the 5 longitudinal bands L_I–L_V, the calling activity of the analyzed individuals for all 24-h over-the-night periods during weekdays (Mondays to Thursdays) aggregated over the 3 year period. The activity in the easternmost band L_I (red line) is noticeably shifted towards early hours compared to the activity in the westernmost band L_V (magenta line). The bands limits are as follows: $L_I : [0°, -2.3°]$, $L_{II} : [-2.5°, -5.35°]$, $L_{III} : [-5.55°, -8.1°]$, $L_{IV} : [-8.3°, -10.3°]$, and , $L_V : [-10.5°, -13.1°]$.

of each region can be seen, with the easternmost band L_I starting and ending its calling activity around 45 minutes earlier than the westernmost band L_V.

Results

Mid-sleep time and sleep duration. In order to determine individual's daily periods of inactivity, we analyze separately the number of events (calls/text message) that each individual made on different days of the week for over the 3-year period. As we are interested in determining the mid-sleep time, we determine the calling activity taking place on each night of the week, such that we split it in seven 24-h periods each one starting at 4:00 pm (e.g. 4:00 pm Monday, and ending at 3:59 pm on the next day which is Tuesday). From here on, we refer to these periods as "nights", with, for example, Saturday night implying the time period between Saturday 4:00 pm and Sunday 3:59 pm. In addition, the first four nights of the week (Monday to Thursdays) are also aggregated into a 24-h long period named "weekday night", which is a standard way to refer to in chronotype studies to workdays.

Using these definitions, we aggregated the weekly calling activity of each individual over the 3-year period on the corresponding period of the week (Weekday, Friday, Saturday and Sunday nights). In Fig. 2, we show the aggregated calling activity of one individual for the four night periods studied.

The bimodal distribution shown in Fig. 2 for one individual is present in almost all the individuals' profiles, and the consistent bimodality shown in the average calling activity of the population at 5 different longitudinal bands (see Fig. 1) is a reflection of this generality. In Fig. 3 we plot the calling activity patterns of a sample of users in one of the 5 latitudinal regions (arranged in an actigram-like representation) to show that the bimodality of the daily overnight calling activity is consistent. This bimodal pattern will be used to approximate each individual's calling activity by a Gaussian Mixture Model (GMM)[37], which has recently been used to describe human activity from CDRs[32,38].

A GMM with two modes (Gaussians) used as an approximation of the calling activity is given by:

$$F(t) = \frac{a_0}{\sigma_E \sqrt{2\pi}} e^{\frac{1}{2\sigma_E^2}(t-\bar{t}_E)^2} + \frac{a_0}{\sigma_M \sqrt{2\pi}} e^{\frac{1}{2\sigma_M^2}(t-\bar{t}_M)^2}, \tag{1}$$

where $\bar{t}_E$ and σ_E are the mean and the standard deviation of the Gaussian located in the left (evening) and $\bar{t}_M$ and σ_M the corresponding values for the one located in the right (morning of the following day).

3
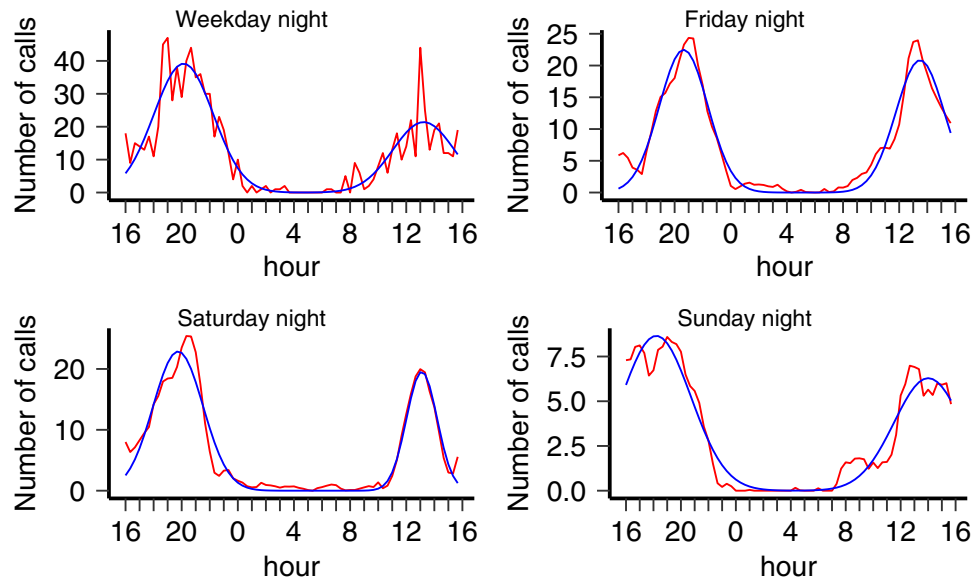
Figure 2. Aggregated calling activity of one individual for 4 different overnight periods. The calling activity as a function of the time of the day (red lines), for four different overnight periods: Weekdays (top-left), Fridays (top-right), Saturdays (bottom-left) and Sundays (bottom-right). Blue lines are smoothed Gaussian curves. For this individual, the minimum of activity typically occurs around 5:00 am.
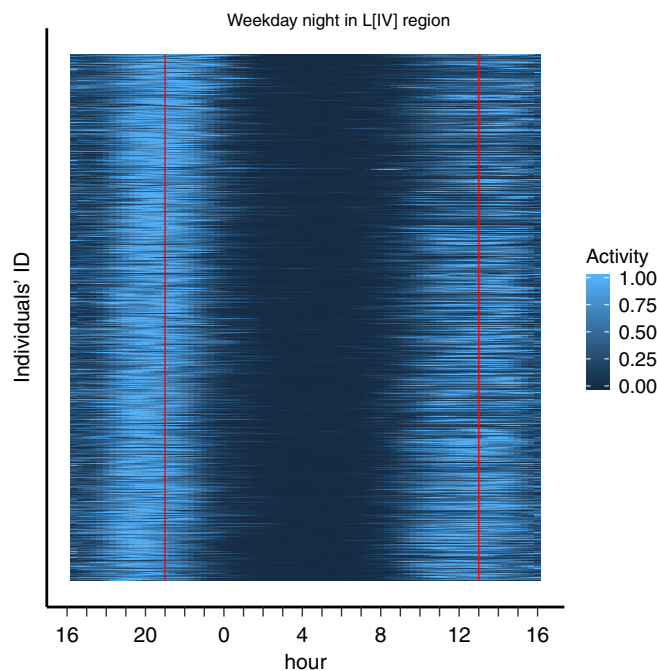
Figure 3. Actigram of individual calling activity during weekdays. An example of an actigram showing the calling activity for weekdays' overnight periods of 1560 individuals chosen from the longitudinal band L_{IV} (similar patterns exist for the other bands). Each individual calling activity correspond to a horizontal line in the actigram, and each individual activity is scaled into the interval [0, 1] such that the most active periods are represented by light regions and periods without activity by dark regions. To show that individuals' period of low activity (chronotypes) are not homogeneous, we ordered the presented individual activity profiles according to time shift between each profile and the mean over the population's activity profiles (i.e. using the time shift that maximizes the cross-correlation between the individual profile and the mean). In this representation, "larks" or morning people appeared in the bottom section of the actigram, with their activity profile clearly shifted towards early hours, whereas "owls" or evening people (top section of the plot) are clearly shift towards late hours.

The means $\bar{t}_E$, $\bar{t}_M$ and the standard deviations σ_E, σ_M given by the approximations can be used to describe the relevant quantities of each individual activity pattern, namely the sleeping duration T_{LCA} and the mid-sleep time T_{mid}. Assuming that the period of sleeping is bounded by the period when the calling activity falls to a minimum, we can approximate the sleeping duration or the period of low calling activity T_{LCA} of the day of the week d by the width of the area between the activity modes, that is,

$$T_{LCA}^d = (\bar{t}_M - \sigma_M) + (24 - (\bar{t}_E + \sigma_E)). \tag{2}$$

Similarly, the mid-sleep time T_{mid}^d of the day of the week d is taken as the midpoint between the calling activity modes, thus

$$T_{mid}^d = (\bar{t}_M - \sigma_M + \bar{t}_E + \sigma_E)/2 - 12. \tag{3}$$

.

Morningness–eveningness classification.   The means $\bar{t}_E$ and $\bar{t}_M$, and the standard deviations σ_E and σ_M calculated using Eq. (1) are not always well defined because of the randomness of individuals' calling activities. Therefore, after filtering out the outliers in the dataset we consider a total of 11,178 individuals for our analysis with numbers 2031, 2589, 2386, 2816 and 1356 in longitudinal bands L_I, L_{II}, L_{III}, L_{IV} and L_V, respectively. The definitions of the weekly overnight periods (Weekday-, Friday-, Saturday-, and Sunday-night) have associated same number of mid-sleep times T_{mid}^d, which can be determined for each individual from the calling activity. These four mid-sleep times can be used to assess the tendency of an individual to have early (morningness) or late (eveningness) schedules.

In general, the mid-sleep times of any individual depends on the day of the week, such that the mid-sleep times on weekdays occur usually earlier than on weekends. However, when comparing between individuals, one can expect that the set of mid-sleep times from a morning person, occurs in general earlier than those of an evening person, thus we can use this expected difference for chronotype classification. The correlations between $T_{mid}^{Weekday}$ and T_{mid}^{Friday}, $T_{mid}^{Saturday}$ and T_{mid}^{Sunday} are 0.65, 0.54 and 0.67, respectively, which corroborates that in general individuals having early schedules on weekdays have also early schedules on weekends.

In spite of the differences found between the different chronotypes that each individual has for different days of the week, the individual has a consistent type (morningness or eveningness) relative to other individuals in the population. Comparing the chronotypes between individuals, those having earlier schedules on weekdays have also earlier schedules on weekends, and similarly for those having later schedules. We use this consistent order between daily chronotypes between individuals to assess their morningness–eveningness. The four possible mid-sleep times (on Weekdays, Fridays, Saturdays and Sundays) are assigned to a 4-dimensional vector

$$\mathbf{T}_{mid} = \left\{ T_{mid}^{Weekday}, T_{mid}^{Friday}, T_{mid}^{Saturday}, T_{mid}^{Sunday} \right\},$$

and this vector will be used to assess the chronotype.

Next we apply Principle Component Analysis (PCA) in the space of vectors $\mathbf{T}_{mid}$ of the population to get a better representation of the chronotype vectors. The loadings of the T_{mid}^d on the first principal component (**PC1**) has been provided in the SI. We have plotted a summary of the negative of the **PC1** (see Supplementary Table S1) in a box-plot on the left in Fig. 4a for the populations in the five longitudinal bands to exhibit the East-West progression of the mean values. A positive **PC1** can be interpreted as an individual having a later mid-sleep time and a negative value can be interpreted for her to have an earlier mid-sleep time. These values can thus be used to understand the morningness–eveningness of an individual in the population. We observe in Fig. 4a, that the mean of the **PC1** decreases from L_V to L_I, which would imply that there are more larks in the Eastern part of the country than in the Western part, which, however, could be misleading[35]. In order to remove this possible artefact, we have computed a multiple linear regression model of T_{mid}^d values with latitude and longitude of the users as independent variables. The coefficients of the latitude and longitude computed from the model along with their p-values have been summarized in Supplementary Table S2. We observe that the longitude is most significant for all T_{mid}^d while there is a very small dependence on latitude. Therefore, we considered the residuals computed from the regression model and have again applied a PCA on them. On the right of Fig. 4a, we have shown a summary of the new first principal component (**PC1**$_{reg}$) for the vector $\mathbf{T}_{mid}$ in a box-plot which shows that the effect of East-West progression has been removed.

We observe that the distribution of **PC1**$_{reg}$ or the p chronotype has a small skewness of 0.12 and is very slightly leptokurtic with a kurtosis value of 3.09, in comparison with the Gaussian distribution having skewness 0 and kurtosis 3. Since the distribution is positively skewed, we expect there to be slightly more owls than larks present in the population. The individuals can be divided into five clusters, inline with the standard classification in the literature of the morningness–eveningness into five groups (see for example Adan and Natale[19], with slightly different nomenclature namely, definitely morning-type, moderately morning, neither-type, moderately evening-type, and definitely evening-type). In Fig. 4b we have divided the users into these clusters using the means (m) and standard deviation (σ) of the distribution of **PC1**$_{reg}$ or the p chronotype. Hence, the individuals grouped by the partitions: $\{-\infty, m - 2\sigma, m - \sigma, m + \sigma, m + 2\sigma, +\infty\}$, are accordingly named as extreme larks (violet), larks (blue), third birds (green), owls (yellow) and extreme owls (red), and comprise 2.22%, 13.05%, 69.04%, 12.84%, and 2.85% of the population, respectively.

Furthermore, we have computed a PCA on the T_{mid} values for the weekdays and weekends separately to observe changes in the behavioural traits of larks and owls. The first principal component obtained from PCA on mid-sleep times on weekdays ($PC1_{reg}^{Weekday}$) accounts for 83.7% of the variance in the data and the one obtained
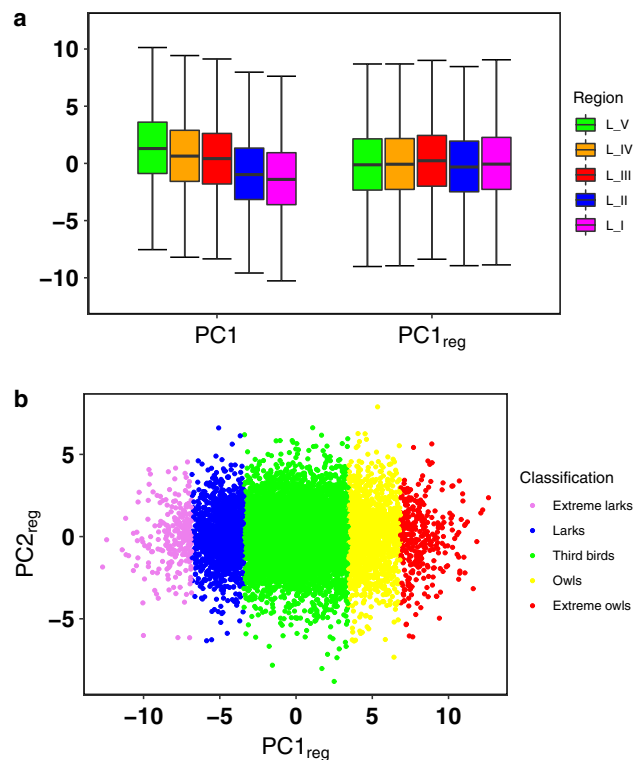
Figure 4. Principal component analysis (PCA) on the four mid-sleep times T_{mid}^d. (**a**) On the left a summary of the first principal component (**PC1**) obtained from PCA computed on the four sets of T_{mid}^d for all users in five different longitudinal bands of the country. On the right, we have shown a summary of the first principal component (**PC1**$_{reg}$ or p chronotype) obtained after applying PCA to the residuals obtained from a regression analysis of T_{mid}^d using the latitude and longitude as independent variables. The green, orange, red, blue and magenta bars represent the westernmost, western, middle, eastern and the easternmost parts of the country, respectively. The black horizontal lines in the middle represent the median of the distribution. The box plot includes all the values within the range of the 25th and 75th percentile and the end of the whiskers represent the maximum and the minimum scores excluding outliers. (**b**) The first two principal components (**PC1**$_{reg}$ and **PC2**$_{reg}$) of the PCA on regressed values of the chronotypes in vector $\mathbf{T}_{mid}$. The distribution of the **PC1**$_{reg}$ or p chronotype is slightly skewed (skewness = 0.12), and leptokurtic (kurtosis = 3.09). Therefore, one can divide the individuals into five groups using the mean and standard deviation of the distribution called extreme larks, larks, third birds, owls and extreme owls according to their chronotype and are represented in the figure by the colours violet, blue, green, yellow and red, from left to right, respectively.

from PCA on mid-sleep times on weekends ($PC1_{reg}^{Weekend}$) account for 81.8% of the variance in the data. The distribution of the PC1s obtained are more leptokurtic and more skewed than the distribution of the p chronotype. $PC1_{reg}^{Weekday}$ and $PC1_{reg}^{Weekend}$ are observed to have a kurtosis of {3.26, 3.25} and skewness of {0.18, 0.16}, respectively. One can then classify the five different groups of people (extreme larks, larks, third birds, owls, extreme owls) using the same method described in previous paragraph. In Fig. 5 we show a Venn diagram that depicts the joint distribution of the PC1 distributions considering only the larks and the owls (including the extreme larks and extreme owls but excluding the third birds). Here we observe that approximately 7.1% of larks and 7.4% of owls of the total population show the same behavioural traits on both weekdays and weekends. Around and 8% and 7.4% of larks on weekdays and weekends respectively change to third birds. Similarly 7.6% and 8.0% of owls convert to third birds on weekdays and weekends, respectively. A very small percentage of the population, 0.1% of the larks and 0.3% of the owls, on weekdays and weekends change to the opposite behaviour, i.e. larks become owls and vice versa.

Model for morningness–eveningness assessment using factor analysis. In general, the individual mid-sleep times are different on different days of the week, with the earliest mid-sleep time occurring in weekdays and the latest on Saturdays (around 1 h difference on average). When comparing mid-sleep times between individuals we observe the following. The relative order between individuals is, in general, the same regardless of the period of the week analysed, such that individuals belonging to the group with earlier chronotypes on weekdays (relative to the whole population), also belong to the groups with earlier chronotypes on weekends.

Based on the above observation we have attempted to compute a chronotype score using a factor analysis for all the users in the population that can reflect the morningness or eveningness of an individual. The first maximum in the activity shown in Fig. 1, considering the night centered approach, represents an average peak in the evening activity (EA) of a user on a particular day of the week and the second peak represents the morning
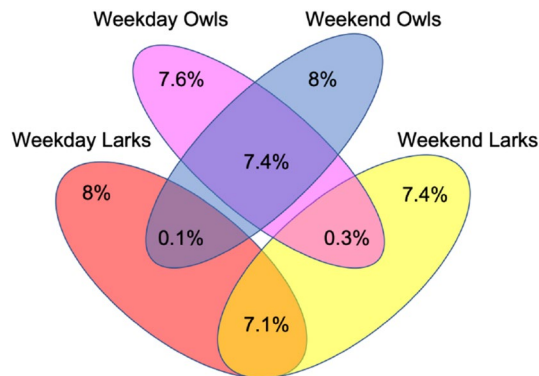
Figure 5. Venn diagram to show changes in larkish and owlish behaviour. The joint distribution of the larks and owls (excluding the third birds) obtained from classification of the chronotypes obtained by computing a PCA on T_{mid} values for weekends and for weekdays separately and presented in the form of a Venn diagram. The colours red and pink represent all the larks and all the owls on the weekdays (including extreme larks and extreme owls). The blue and yellow represent all owls and all larks on weekends (including extreme larks and extreme owls). We observe that 7.1% of the population who are larks remain the same be it weekdays or weekends and similarly for 7.4% of the population that are owls. A very few percentage of the population (0.1% for larks and 0.3% for owls) changes their behaviour drastically to the opposite kind depending on whether it is a weekend or weekday. We also observe that 8% for larks and 7.6% for owls on weekdays; and 7.4% for larks and 8% for owls on weekends convert to third birds in the population.
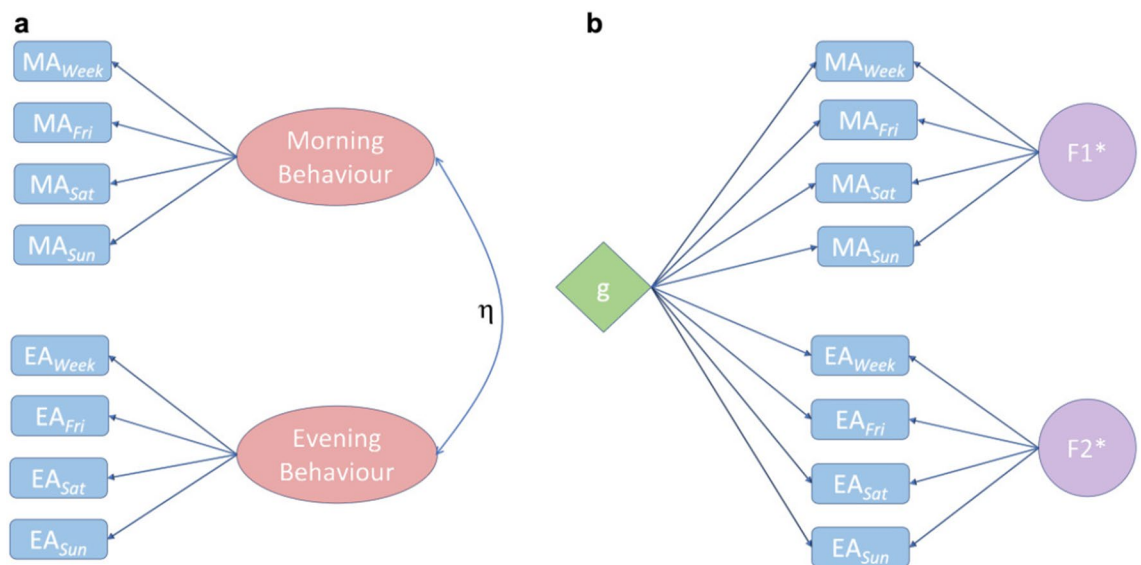
Figure 6. Factor analysis of the chronotype of an individual using CDRs: (**a**) An exploratory factor analysis (EFA) of the peak locations of the morning activity (MA) and evening activity (EA) on the seven days of the week shows the emergence of two underlying latent factors (Morning behaviour and Evening behaviour of an individual) that governs the outcome of MA and EA, separately. The correlation between the two latent factor is computed to be $\eta = 0.31$. (**b**) A higher order model (Bifactor model) is used to find an underlying general g factor that can be used to understand the chronotype of an individual. $F1^*$ and $F2^*$ are the "group" factors that affect the MAs and EAs, separately. The loadings of the factors on the observables for both the figures have been specified in the two separate tables (Supplementary Table S3 and Supplementary Table S4) in the SI. Any cross-loadings with values below 0.3 have been ignored.

activity (MA). These pairs of observables for each day have been computed from the data for each individual and is denoted by MA_d and EA_d where d stands for Weekdays, Fridays, Saturdays and Sundays. Generally, the MAs of individuals are mostly constrained due to social obligations, like reaching their workplaces on time. In the evening, they are more relaxed and can follow their own individual chronotypes. Therefore, we hypothesize that the morning and evening behaviour of an individual are different from each other and these can be used to assess their chronotype. We have considered an exploratory factor analysis (EFA)[39,40] on the sets of observables for $\{MA_d\}$ and $\{EA_d\}$ as shown in Fig. 6a to explore underlying latent variables that affect an individual's morning

and evening activities. The EFA is a technique used to identify conceivable underlying constructs within the observables and is distinct from PCA that is usually employed to reduce the dimensions in the data.

We first assess the factorability of the data by carrying out the following tests. The Kaiser–Meyer–Olkin test for factor adequacy in the model gives a score of 0.74[41]. In addition we use the Bartlett's test for sphericity[42] to check any redundancy between the observables that are summarized with fewer number of the latent variables and it gives a $\chi^2 = 32344.32$ with $p < 0.0001$. Both tests indicate a favourable use of an EFA. The EFA was conducted using the "psych" package[43] with an oblimin rotation and a maximum likelihood method. A two factor structure was supported by a scree plot, the Kaiser criterion, relevant factor loadings, as well as the interpretability of the dimensions. The mean communality value was slightly above 0.50 ranging from 0.23 to 0.79. This agrees with our aforementioned hypothesis and accordingly we have observed that all the MAs are loaded on one factor (Morning behaviour) and all the EAs are loaded on another factor (Evening Behaviour) as depicted in Fig. 6a. The correlation between the two factors is $\eta = 0.31$. A mediocre score of 0.59 for the unidimensionality computed on this model containing all the MAs and the EAs further supports our claim that there exists more than one latent factor. A score close to 1.0 would have indicated a single factor explaining the behaviour during an entire day. Therefore, this would imply that the individuals behave rather differently in the morning from that in the evening. The individual factor loadings on the observables are summarised in Supplementary Table S3 in the SI. The cross-loadings of the factors have not been shown in this figure since their values are less than 0.3. The model fit indices for the EFA, namely, comparative fit index (CFI), Tucker–Lewis Index (TLI), and the root mean square error of approximation (RMSEA), have been computed to be 0.91, 0.80, and 0.14, respectively. While the values of CFI and TLI indicate a good fit of the data in our model, we get a high value for the RMSEA[44].

Next, we have carried out an exploratory bifactor analysis (EBA)[45] on the model to determine scores for a single construct like the chronotype of an individual that would reflect the morningness or eveningness of the person even when the data is multidimensional. The bifactor models are useful in representing hierarchical latent structures in the data as the first-order factors[46]. It computes the factor scores for a general factor g, which loads directly onto all the observables in the model and also produces group factors that distinguish between the groups formed among the observables. We have used "omega" function from the package "psych", which does a factor analysis followed by an oblique rotation and extracts the general factor using Schmid–Leiman transformation[47]. The tests of reliability in our model ω_t and ω_h[48] are computed to be 0.84 and 0.39, respectively. Here ω_t accounts for the total variance in the data due to the general factor g and the group factors together, whereas ω_h accounts for the proportion of variance in the data due to the general factor only. In Fig. 6b we show that the all the observables or items are loaded on the general factor (g), which represents the chronotype of an individual. The factors represented by $F1^*$ and $F2^*$ are group factors or *nuisance dimensions-factors* that measure responses of the observables that are not considered by the g factor. The loadings of all the factors in this analysis have been summarized in Supplementary Table S4 in SI.

The mean factor scores obtained for the morningness behaviour, eveningness behaviour, and the g chronotype from the models discussed above are found to behave in a way similar to the principal component in the left of Fig. 4a when plotted as a function of the longitudinal bands from West to East. As discussed previously in the case of the PCA we have computed a multiple linear regression model of the observables with the latitude and longitude of the users as independent variables (details in Supplementary Table S5 of SI) to remove the geographical dependence in the data. We have considered the residuals computed from the regression model for all the items in our analysis. We have applied an EFA and EBA on the residuals and the corresponding plots are shown in Supplementary Fig. S1 in SI.

Age and gender dependence of the chronotype. The factor scores obtained from our models can be interpreted as an indicator of a user's chronotype. Individuals having negative scores are considered to be larkish or morning-type and those having positive scores to be owlish or evening-type. The higher the value of the scores, the more extreme larkish or extreme owlish behaviour of an individual is expected to be. Fig. 7 shows the average factor scores of (a) Morning Behaviour—MB, (b) Evening Behaviour—EB and (c) the g chronotype as a function of the users' age and gender. The factor scores for g chronotype in Fig. 7c for both the genders are found to be decreasing with age indicating that younger individuals (from 18 to 35 years old) are more owlish in nature. Furthermore, it is also observed that males in the younger age cohorts tend to have higher factor scores than women indicating that they are more owlish than females. However, after 35 there is a crossover and the females are observed to be more owlish in nature than males. For the mid age cohorts (between 35 to 60 years old), we observe a peak in the factor scores. Finally, older age cohorts (above 60 years) are found to behave like larks with women still having later chronotypes than men. The age and gender dependence of the p chronotype computed using PCA discussed previously is also shown in Fig. 7d. The variation of p chronotype is observed to be qualitatively similar to the one observed with g chronotype. We have also observed a high correlation between the two chronotypes and it is found to be 0.76.

Dependence of the period of low calling activity on the chronotype. The period of low calling activity T_{LCA}, which can be interpreted as a representation of an individual's sleep duration during the night time, has been calculated using Eq. (2). Furthermore, we have calculated the average sleep duration on weekend ($T_{LCA} - Weekend$) and weekdays ($T_{LCA} - Weekday$) separately and Fig. 8a shows their variation as a function of the g chronotype. We find that the users sleep more on the weekends than on weekdays and this result is consistent with the previous findings[49]. Moreover, larks are observed to sleep more than owls on weekends. This is because both larks and owls tend to align themselves according to their own chronotype as there are no social constraints governing their schedules. Since the larks tend to follow the solar clock they tend to sleep more than owls. On weekdays, extreme owls have the same sleep duration as the larks which implies that they are not able
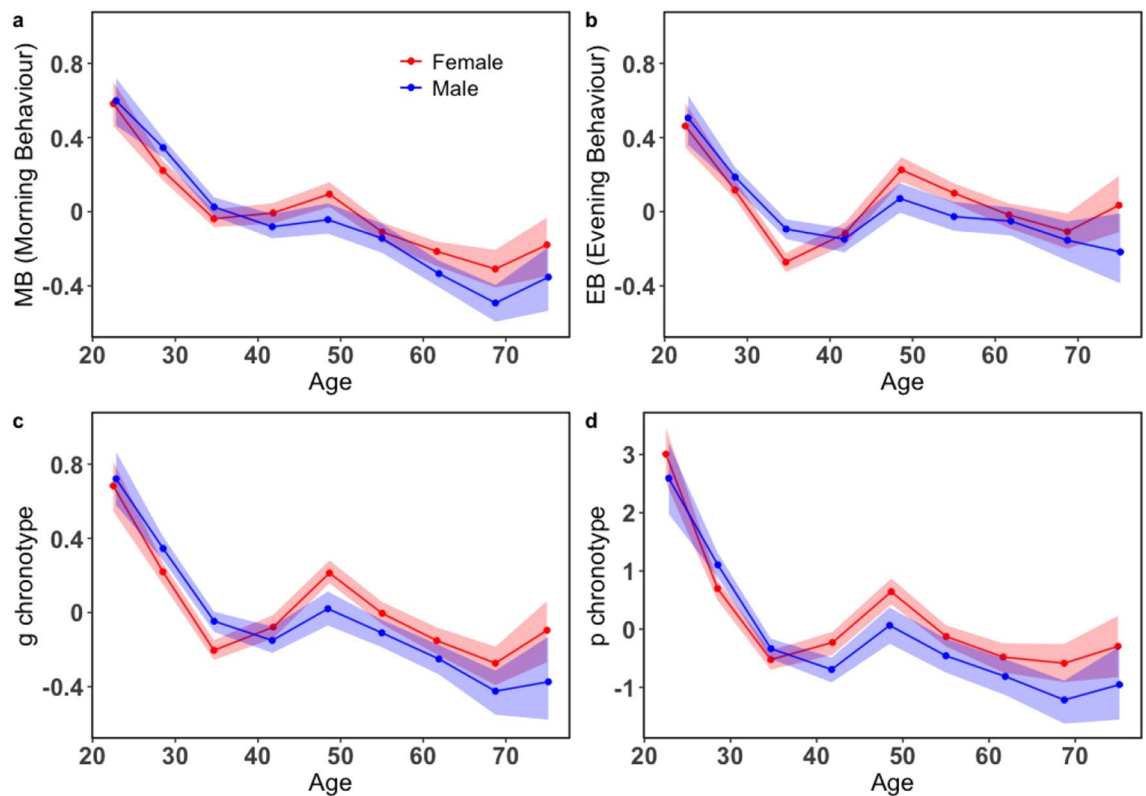
Figure 7. The mean of the factor scores and first principal component computed from PCA for different age and gender cohorts: The variation of the average factor scores of the (**a**) morning behaviour, (**b**) evening behaviour and (**c**) the general g chronotype along with (**d**) the first principal component (**PC1**$_{reg}$) from the PCA renamed as p chronotype is plotted as a function of age and gender. Each point is calculated separately with the shaded regions in all the three plots representing the bootstrap 95% confidence intervals. The curve in red is for females and blue for males.
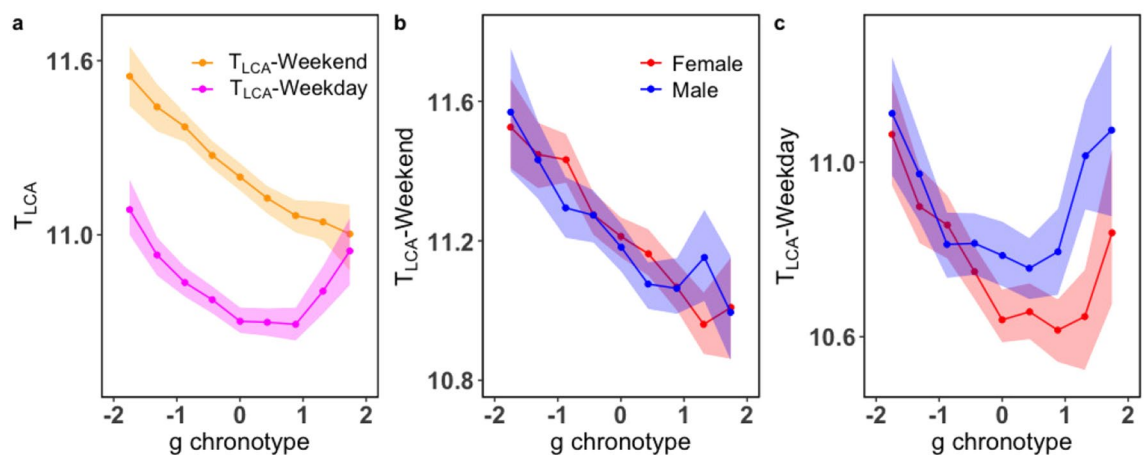
Figure 8. Period of low calling activity of users of different chronotypes: (**a**) The period of low calling activity T_{LCA}^{d} is the width of the region between the activity modes, calculated according to Eq. (2) for different days (*d*) of the week. The average of T_{LCA}^{d} on weekends $T_{LCA} - Weekend$ and weekdays $T_{LCA} - Weekday$ is plotted as a function of the general (g) chronotype in orange and magenta, respectively. (**b**) $T_{LCA} - Weekend$ and (**c**) $T_{LCA} - Weekday$ are plotted as a function of the chronotype for both genders. For each chronotype, $T_{LCA} - Weekend$ and $T_{LCA} - Weekday$ has been calculated separately with the shaded regions in all three plots representing the bootstrap 95% confidence intervals. The red plot is for females and blue for males in both (**b**) and (**c**).

to keep up with the social constraints like work schedules and end up oversleeping. Figure 8b shows the sleep duration on weekends as function of the chronotype and the gender. We do not observe any significant differences between the two genders on the weekends. However, in Fig. 8c we find that owlish males sleep more on weekdays than owlish females. The average age of the females in this regime falls in the age cohort of 40 to 60 year old. This could be a reason for the less sleep duration as they are more active than males around this age due to reasons already discussed in the previous section.

Discussion

In this study, we have utilized mobile phone communication data of a population in a European country to study chronotypes of the service users. We have shown that the chronotype of an individual can be broadly identified through two related yet distinct statistical approaches. In our first approach, we have used PCA to heuristically calculate a composite score for the different chronotypes using the mid-sleep times on weekdays along with the weekends as well. The first principal component from this analysis is composed of all the mid-sleep times with all positive loadings. We observed a slightly leptokurtic distribution with a small skewness for the computed values. Using the mean and the standard deviation of the distribution, we divided the users into the following five different clusters—third birds, the users in the centre (69.0%), larks (13.1%), owls (12.8%), extreme larks (2.2%), and extreme owls (2.9%). Moreover, if one were to do the PCA for weekdays and weekends separately, the joint distribution obtained from the two first principal components indicate changes in the behavioural traits of the individuals. While some of the larks and owls behave the same on weekends and weekdays, some have also been observed to change to the third birds category. We also observed very small percentages of larks to convert to owls and vice versa.

Using a second approach, i.e. an EFA, which assumed chronotype to be a latent trait, we also found that the morning activities and evening activities of the users are governed by two separate factors, namely the morning behaviour and the evening behaviour. The morning behaviour of the users is usually more constrained due to the society following a strict schedule for offices, schools, etc. and most chronotypes try to align themselves accordingly. However, this is not the case in their evening behaviour since the individuals are more flexible with their evening schedules. Therefore, we see a more pronounced change in the morning behaviour than in the evenings as depicted in Fig. 7a,b. Furthermore, it is seen that older cohorts usually do not follow a strict schedule since they are mostly retired from work and consequently tend to follow their inherent chronotypes. In contrast, the younger cohorts need to follow a stricter social timetable and thus exhibit a vastly different behaviour than the older cohorts.

Traditionally, the chronotypes have been calculated using the mid-sleep time of the individuals on weekends only since they are assumed to follow their inherent chronotypes freely during these days of the week[50]. However, through our study we have shown that the g chronotype computed using an EBA is also an appropriate method to study the morningness and eveningness of an individual. It is a higher order version of the EFA that is able to compute a general factor g that is directly related to all the observables in the data. Hence, this general factor renamed as g chronotype from EBA is able to capture all the effects of the activities of an individual on all days of the week. Although we do not have a separate method through which we could validate our measured chronotype but the fact that similar observables converge to the same factor (MAs loading in a single factor) and dissimilar observables (MAs and EAs) loading on different factors supports our measurement of the chronotype to a reasonable extent.

We found that the younger cohorts tend to have later chronotypes, which gradually change to earlier chronotype with increasing age, similar to the results of the earlier survey study[50]. However, the variation in the chronotype reduces considerably for age groups above 40. In addition, we have observed a small peak for mid-age cohorts that could be a direct influence of the lifestyle led by most of the individuals in this age group. Most of them have to connect with both their children (young age cohorts) and parents (old age cohorts) who usually live in separate accommodations. Thus, the peak can be assumed to be a manifestation of their calling activity needed to maintain their social interactions with both the age groups. The individuals in the older age cohorts (60 years and above) are most likely not adhering to regular work schedules and so, they tend to follow their inherent chronotype, which is more aligned with their biological and the solar clock. This could be a reason for them to follow a more larkish behaviour. These changes can also be attributed to other factors like hormonal changes during an individual's life span that affects their sleeping patterns[5]. Additionally, women above the age of 40 are found to show more owlish chronotype when compared to men. This trait may be a direct cause of societal responsibilities, like child care, that are usually assumed to be taken predominantly by women[51].

Furthermore, we have observed similar behaviour for both the g chronotype and the p chronotype computed from EBA and PCA, respectively, as depicted in Fig. 7d. Also, a separate chronotype(m) computed from an EFA on mid-sleep times for all days is found to show the same variation as the g chronotype (see Supplementary Fig. S2 in SI). Using the g chronotype we are able to demonstrate that the chronotypes identified by directly taking into account and combining several observables of human activity, instead of a derived quantity like the mid-sleep time, can also be used to distinguish between the morningness and the eveningness of individuals. Moreover, our results agree with the previous findings using traditional methods like the MCTQ and MEQ questionnaires[1,5,52–55]. Using a period of the users' low calling activity as markers of their sleep duration[35] we find that on average all chronotypes sleep more on weekends than on weekdays[49] and in both cases larks are generally found to sleep more than owls. On weekends, larks go to sleep earlier than owls and so they have a longer sleeping period. The shorter sleeping periods observed for owls may be a cause for sleep deprivation occurring among them, which can further lead to health issues[11–13]. However, on the weekdays we observe that extreme owls have sleep duration similar to extreme larks suggesting that on weekdays the former may have difficulties in observing work schedules.

Finally, we conclude that our results obtained by combining data from mobile phone communication of individuals during a 24 h day-night cycle, one can form a detailed understanding of their chronotypes. These kinds of studies using mobile phone service subscribers' CDRs, demographic, and location information provide a novel and time-wise longitudinal perspective to the circadian rhythms of individuals. Our data-driven approach adds and complements the questionnaire based studies and findings in them, as it avoids possible shortcomings in terms of sample size and dependence on the memory of the participants. Such data could also be used to characterize different other aspects of the sleep-wake behaviour of individuals such as the degree of awareness of their morning or evening nature[56–58]. This could require the inclusion of more items pertaining to the shape of the distributions of calling activity. In the recent past, the rapid adoption of newer modes of digital communication and different smart devices like fitness trackers that can keep track of an individual's sleep routines, body temperatures and various other activities throughout the day, have largely supplemented the usage of mobile phones. Therefore, we believe that the approach of combining digital data from multiple channels of communication to assess the chronotype of an individual as a reflective or a latent trait using unsupervised or supervised models would be extremely worthwhile and timely in fields such as mobile health and medicine[59,60].

Data availibility

References

1. Roenneberg, T., Kumar, C. & Merrow, M. The human circadian clock entrains to sun time. *Curr. Biol.* **17**, R44–R45. https://doi.org/10.1016/j.cub.2006.12.011 (2007).
2. Roenneberg, T., Daan, S. & Merrow, M. The art of entrainment. *J. Biol. Rhythms* **18**, 183–194. https://doi.org/10.1177/0748730403018003001 (2003).
3. Sack, R. L. *et al.* Circadian rhythm sleep disorders: Part I, basic principles, shift work and jet lag disorders. *Sleep* **30**, 1460–1483. https://doi.org/10.1093/sleep/30.11.1460 (2007).
4. Duffy, J. F. & Czeisler, C. A. Effect of light on human circadian physiology. *Sleep Med. Clin.* **4**, 165–177. https://doi.org/10.1016/j.jsmc.2009.01.004 (2009).
5. Roenneberg, T. *et al.* A marker for the end of adolescence. *Curr. Biol.* **17:24**, R1038. https://doi.org/10.1016/j.cub.2004.11.039 (2004).
6. Archer, S. N. *et al.* A length polymorphism in the circadian clock gene per3 is linked to delayed sleep phase syndrome and extreme diurnal preference. *Sleep* **26**, 413–415. https://doi.org/10.1093/sleep/26.4.413 (2003).
7. Toh, K. L. *et al.* An hper2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* **291**, 1040–1043. https://doi.org/10.1126/science.1057499 (2001).
8. Koskenvuo, M., Hublin, C., Partinen, M., Heikkilä, K. & Kaprio, J. Heritability of diurnal type: A nationwide study of 8753 adult twin pairs. *J. Sleep Res.* **16**, 156–162. https://doi.org/10.1111/j.1365-2869.2007.00580.x (2007).
9. Horne, J. A. & Östberg, O. A self-assessment questionnaire to determine morningness–eveningness in human circadian rhythms. *Int. J. Chronobiol.* **4**(2), 97–110 (1976).
10. Gale, C. & Martyn, C. Larks and owls and health, wealth, and wisdom. *BMJ* **317**, 1675–1677. https://doi.org/10.1136/bmj.317.7174.1675 (1998).
11. Jones, S. E. *et al.* Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat. Commun.* **10**, 343. https://doi.org/10.1038/s41467-018-08259-7 (2019).
12. Merikanto, I. *et al.* Evening types are prone to depression. *Chronobiol. Int.* **30**(5), 719–725. https://doi.org/10.3109/07420528.2013.784770 (2013).
13. Facer-Childs, E. R., Boiling, S. & Balanos, G. M. The effects of time of day and chronotype on cognitive and physical performance in healthy volunteers. *Sports Med. Open* **4**, 47. https://doi.org/10.1186/s40798-018-0162-z (2018).
14. Roenneberg, T., Wirz-Justice, A. & Merrow, M. Life between clocks: Daily temporal patterns of human chronotypes. *J. Biol. Rhythms* **18**(1), 80–90. https://doi.org/10.1177/0748730402239679 (2003).
15. Roennenberg, T. Having trouble typing? What on earth is chronotype?. *J. Biol. Rhythms* **30**(6), 487–491. https://doi.org/10.1177/0748730415603835 (2015).
16. Shahid, A., Wilkinson, K., Marcu, S. & Shapiro, C. M. Munich chronotype questionnaire (MCTQ). In *STOP, THAT and One Hundred Other Sleep Scales* (Springer, 2011). https://doi.org/10.1007/978-1-4419-9893-4_58.
17. Jung, H. L., In, S. K., Seong, J. K., Wei, W. & Jeanne, F. D. Change in individual chronotype over a lifetime: A retrospective study. *Sleep Med. Res.* **2**, 48–53. https://doi.org/10.17241/smr.2011.2.2.48 (2011).
18. Rodrigues, P. F. *et al.* Morningness–eveningness preferences in Portuguese adolescents: Adaptation and psychometric validity of the h&o questionnaire. *Person. Individ. Differ.* **88**, 62–65. https://doi.org/10.1016/j.paid.2015.08.048 (2016).
19. Adan, A. & Natale, V. Gender differences in morningness–eveningness preference. *Chronobiol. Int.* **19**, 709–720. https://doi.org/10.1081/CBI-120005390 (2002).
20. Blondel, V. *et al. Mobile Phone Data for Development: Analysis of Mobile Phone Datasets for the Development of Ivory Coast* (MIT Media Lab, 2013).
21. Eagle, N. & Pentland, A. Reality mining: Sensing complex social systems. *Person. Ubiquit. Comput.* **10**, 255–268. https://doi.org/10.1007/s00779-005-0046-3 (2006).
22. Pentland, A. *Social Physics: How Social Networks Can Make Us Smarter* (Penguin, 2015).
23. Onnela, J. P. *et al.* Structure and tie strengths in mobile communication networks. *PNAS* **104**(18), 7332–7336. https://doi.org/10.1073/pnas.0610245104 (2007).
24. Bhattacharya, K. & Kaski, K. Social physics: Uncovering human behaviour from communication. *Adv. Phys. X* **4:1**, 1527723. https://doi.org/10.1080/23746149.2018.1527723 (2018).
25. Miritello, G. *et al.* Time as a limited resource: Communication strategy in mobile phone networks. *Soc. Netw.* **35**(1), 89–95. https://doi.org/10.1016/j.socnet.2013.01.003 (2013).
26. Fudolig, M. I. D., Bhattacharya, K., Monsivais, D., Jo, H. H. & Kaski, K. Link-centric analysis of variation by demographics in mobile phone communication patterns. *PLoS One* **15**(1), e0227037. https://doi.org/10.1371/journal.pone.0227037 (2020).
27. Fudolig, M. I. D., Monsivais, D., Bhattacharya, K., Jo, H. H. & Kaski, K. Different patterns of social closeness observed in mobile phone communication. *J. Comput. Soc. Sci.* **3**, 1–17. https://doi.org/10.1007/s42001-019-00054-8 (2020).

28. Bhattacharya, K. *et al.* Network of families in a contemporary population: Regional and cultural assortativity. *EPJ Data Sci.* **7**, 9. https://doi.org/10.1140/epjds/s13688-018-0137-9 (2018).
29. Fudolig, M. I. D., Monsivais, D., Bhattacharya, K., Jo, H.-H. & Kaski, K. Internal migration and mobile communication patterns among pairs with strong ties. *EPJ Data Sci.* **10**, 1–21 (2021).
30. Ghosh, A. *et al.* Migration patterns of parents, children and siblings: Evidence for patrilocality in contemporary Finland. *Popul. Space Place* **25**, e2208 (2019).
31. Grantz, K. H. *et al.* The use of mobile phone data to inform analysis of covid-19 pandemic epidemiology. *Nat. Commun.* **11**, 4961. https://doi.org/10.1038/s41467-020-18190-5 (2020).
32. Monsivais, D., Bhattacharya, K., Ghosh, A., Dunbar, R. I. M. & Kaski, K. Seasonal and geographical impact on human resting periods. *Sci. Rep.* **7**, 10717. https://doi.org/10.1038/s41598-017-11125-z (2017).
33. Aledavood, T., Lehmann, S. & Saramäki, J. Social network differences of chronotypes identified from mobile phone data. *EPJ Data Sci.* **7**, 46. https://doi.org/10.1140/epjds/s13688-018-0174-4 (2018).
34. Aledavood, T., Kivimäki, I., Lehmann, S. & Saramäki, J. A non-negative matrix factorization based method for quantifying rhythms of activity and sleep and chronotypes using mobile phone data. *arXiv preprint* arXiv:2009.09914 (2020).
35. Monsivais, D., Ghosh, A., Bhattacharya, K., Dunbar, R. & Kaski, K. Tracking urban human activity from mobile phone calling patterns. *PLoS Comput. Biol.* **13**(11), e1005824. https://doi.org/10.1371/journal.pcbi.1005824 (2017).
36. There were nine billion calls made in total and one billion text messages sent by all the users in the dataset. So only one-tenth of the total number of the calling activity considered comprise only of text messages. Since the fraction is quite small we do not expect there to be any significant bias in the data for people who prefer texting more than calling.
37. Reynolds, D. Gaussian mixture models. In *Encyclopedia of Biometrics* Vol. 741 (eds Li, S. Z. & Jain, A. K.) (Springer, 2015). https://doi.org/10.1007/978-1-4899-7488-4_196.
38. Aubourg, T., Demongeot, J., Renard, F., Provost, H. & Vuillerme, N. How to measure circadian rhythms of activity and their disruptions in humans using passive and unobtrusive capture of phone call activity. *Stud. Health Technol. Inform.* **264**, 1631–1632. https://doi.org/10.3233/shti190569 (2019).
39. Fabrigar, L. R. & Wegener, D. T. *Exploratory Factor Analysis* (Oxford University Press, 2012).
40. Costello, A. B. & Osborne, J. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Pract. Assess. Res. Eval.* **10**, 7. https://doi.org/10.7275/jyj1-4868 (2005).
41. Pett, M. A., Lackey, N. R., & Sullivan, J. J. *Making sense of factor analysis: The use of factor analysis for instrument development in health care research.* (sage, 2003).
42. Bartlett, M. S. Tests of significance in factor analysis. *Br. J. Psychol.* **3**, 77–85. https://doi.org/10.1111/j.2044-8317.1950.tb00285.x (1950).
43. Revelle, W., & Revelle, M. W. Package 'psych'. *The comprehensive R archive network*, **337**, 338 (2015).
44. Thompson, B. *Exploratory and Confirmatory Factor Analysis* (American Psychological Association, 2004).
45. Jennrich, R. I. & Bentler, P. M. Exploratory bi-factor analysis. *Psychometrika* **76**(4), 537–549. https://doi.org/10.1007/s11336-011-9218-4 (2011).
46. Reise, S., Moore, T. & Haviland, M. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *J. Person. Assess.* **92**(6), 544–559. https://doi.org/10.1007/s11336-011-9218-4 (2010).
47. Schmid, J. & Leiman, J. M. The development of hierarchical factor solutions. *Psychometrika* **22**, 53–61. https://doi.org/10.1007/BF02289209 (1957).
48. Revelle, W. (2009). An introduction to psychometric theory with applications in R. https://personality-project.org/r/book/Chapter7.pdf.
49. Foster, R. G. & Roenneberg, T. Human responses to the geophysical daily, annual and lunar cycles. *Curr. Biol.* **18**(17), R784–R794. https://doi.org/10.1016/j.cub.2008.07.003 (2008).
50. Fischer, D., Lombardi, D. A., Marucci-Wellman, H. & Roenneberg, T. Chronotypes in the US—Influence of age and sex. *PLoS One* **12**(6), e0178782. https://doi.org/10.1371/journal.pone.0178782 (2017).
51. Ghosh, A., Monsivais, D., Bhattacharya, K., Dunbar, R. I. M. & Kaski, K. Quantifying gender preferences in human social interactions using a large cellphone dataset. *EPJ Data Sci.* **8**(9), 89–95. https://doi.org/10.1140/epjds/s13688-019-0185-9 (2019).
52. Duffy, J. F., Rimmer, D. W. & Czeisler, C. A. Association of intrinsic circadian period with morningness–eveningness, usual wake time, and circadian phase. *Behav. Neurosci.* **115**(4), 895–899. https://doi.org/10.1037/0735-7044.115.4.895 (2001).
53. Henson, J. *et al.* Physical behaviors and chronotype in people with type 2 diabetes. *BMJ Open Diabetes Res. Care.* **8**(1), e001375. https://doi.org/10.1136/bmjdrc-2020-001375 (2020).
54. Juda, M., Vetter, C. & Roenneberg, T. Chronotype modulates sleep duration, sleep quality, and social jet lag in shift-workers. *J. Biol. Rhythms* **28**(2), 141–151. https://doi.org/10.1177/0748730412475042 (2013).
55. Panjeh, S. *et al.* What are we measuring with the morningness–eveningness questionnaire? exploratory factor analysis across four samples from two countries. *Chronobiol. Int.* **38**(2), 234–247. https://doi.org/10.1080/07420528.2020.1815758 (2021).
56. Ogińska, H. Can you feel the rhythm? A short questionnaire to describe two dimensions of chronotype. *Person. Individ. Differ.* **50**, 1039–1043. https://doi.org/10.1016/j.paid.2011.01.020 (2011).
57. Randler, C., Díaz-Morales, J. F., Rahafar, A. & Vollmer, C. Morningness–eveningness and amplitude-development and validation of an improved composite scale to measure circadian preference and stability (messi). *Chronobiol. Int.* **33**, 832–848. https://doi.org/10.3109/07420528.2016.1171233 (2016).
58. Rodrigues, P. F. *et al.* Initial psychometric characterization for the Portuguese version of the morningness–eveningness-stability-scale improved (messi). *Chronobiol. Int.* **35**, 1608–1618. https://doi.org/10.1080/07420528.2018.1495646 (2018).
59. Onnela, J. P. & Rauch, S. L. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* **41**(7), 1691–1696. https://doi.org/10.1038/npp.2016.7 (2016).
60. Huguet, A. *et al.* A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. *PLoS One* **11**(5), e154248. https://doi.org/10.1371/journal.pone.0154248 (2016).

Acknowledgements

Author contributions

All authors contributed in developing the scope of the paper. C.R. and D.M. developed the numerical code and made the figures and tables. C.R., D.M. and K.B. wrote the first draft of the text. All authors contributed in interpretation, discussion, analysis of results and text improvement. All authors reviewed the manuscript.

Competing interest

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-93799-0.

Correspondence and requests for materials should be addressed to C.R.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.