

# SCIENTIFIC REPORTS



OPEN

## Seasonal and geographical impact on human resting periods

Daniel Monsivais<sup>1</sup>, Kunal Bhattacharya<sup>1</sup>, Asim Ghosh<sup>1</sup>, Robin I. M. Dunbar<sup>1,2</sup> & Kimmo Kaski<sup>1,2</sup>

We study the influence of seasonally and geographically related daily dynamics of daylight and ambient temperature on human resting or sleeping patterns using mobile phone data of a large number of individuals. We observe two daily inactivity periods in the people's aggregated mobile phone calling patterns and infer these to represent the resting times of the population. We find that the nocturnal resting period is strongly influenced by the length of daylight, and that its seasonal variation depends on the latitude, such that for people living in two different cities separated by eight latitudinal degrees, the difference in the resting periods of people between the summer and winter in southern cities is almost twice that in the northern cities. We also observe that the duration of the afternoon resting period is influenced by the temperature, and that there is a threshold from which this influence sets in. Finally, we observe that the yearly dynamics of the afternoon and nocturnal resting periods appear to be counterbalancing each other. This also lends support to the notion that the total daily resting time of people is more or less conserved across the year.

In modern societies, daily activities are marked by the entrainment of the human circadian clock, which is strongly influenced by two different cues, each one following a 24-hour period but with a different offset. The first cue is related to environmental and sun-based events, and therefore being subjected to seasonal variations due to the yearly movement of earth around the sun. The second cue follows a local civil time, where social and economical factors impose restrictions on the timing and routine of human activities<sup>1</sup>. Regardless of the nature of the external cue participating in the entrainment of the human circadian clock, any change in it has a direct influence on various aspects of human life<sup>2,3</sup>.

For human circadian rhythms, particularly for the sleep wake cycle (SWC), there are various biological<sup>4</sup>, sociological<sup>5</sup> and environmental<sup>6</sup> cues affecting the entrainment of the circadian clock, sometimes with undesirable effects on the mental and physical health of individuals<sup>7,12</sup>. Due to its importance, the SWC has been studied in recent years from different perspectives, trying to understand and identify what are the processes and pace-makers governing its dynamics<sup>13,14</sup>. Broadly speaking the current understanding of the generation and maintenance of SWC is that sleep is governed by two main mechanisms consisting of the 24-hour circadian system and awake-dependent homeostatic build-up of sleep pressure. Then sleep itself feeds back to regulate the circadian system and homeostatic sleep pressure build-up. Light plays a key role in synchronising and pace-making the circadian cycle or rhythm, gated by the SWC in suppressing the melatonin production in the pineal gland, regulating alertness. Apart from these, the daily social activities also play a role in driving SWC.

In general, the current research on the human SWC has focused on experiments on small groups under controlled situations<sup>8,11</sup>, and on studies based on questionnaires<sup>6,15,16</sup>, mainly the Munich Chronotype Questionnaire (MCTQ). The use of questionnaires for studying SWC are proved to be efficient when assessing sleep wake cycle, but have some limit to the domain of its applicability<sup>17</sup>. Also, it has been suggested that humans living in modern societies are subjected to new environments, like electrical lighting<sup>18,19</sup>, and exposure to these has disrupted and changed people's natural sleeping habits, although there is no agreement in the field about these possible changes<sup>19–23</sup>. Studies of the sleeping patterns of people living in pre-industrial societies<sup>20</sup> have shown that their sleeping times are similar to those of modern societies, and that temperature could play an important role in the dynamics of sleep. In the recent past the presence of new communication technologies as well as the accessibility to large-scale techno-social datasets ('Big data') have allowed the study of human behavior from diverse perspectives applying reality mining techniques. In particular, mobile phone call detail records (CDRs) have

<sup>1</sup>Department of Computer Science, Aalto University School of Science, P.O. Box 15400, FI-00076 AALTO, Espoo, Finland. <sup>2</sup>Department of Experimental Psychology, University of Oxford, South Parks Rd, Oxford, OX1 3UD, United Kingdom. Correspondence and requests for materials should be addressed to D.M. (email: [daniel.monsivais-velazquez@aalto.fi](mailto:daniel.monsivais-velazquez@aalto.fi))

been analysed to study mental health<sup>24</sup>, social networks<sup>25–28</sup>, sociobiology<sup>29–33</sup>, behavior of cities<sup>34, 35</sup>, as well as mobility<sup>36–41</sup>.

In the recent past, data collected from sensors and logs from mobile phone have been used to study circadian rhythms<sup>29, 32, 38, 42–44</sup>. Aledavood *et al.*<sup>29</sup> studying mobile phone calling patterns together with data collected from questionnaires, described the characteristic time at which the peak of alertness of a small cohort of 24 students occurs. In another study, comparing various medium size (around 100, 1200, 700 and 2400 participants) datasets of mobile phone records and emails, Aledavood *et al.*<sup>45</sup> showed each individual level has a distinctive, persistent daily communication pattern that strongly differs from the average pattern.

Users of the mobile phone network have specific time intervals when their calling activity ceases and it is expected that these periods are characterised by users taking rest. Each day shows two periods of low calling activity, the first after lunch time around 4:00 p.m., and the second coinciding with the night period, centered around 4:00 a.m. In turn, these two periods delimit two regions of high activity, one peaking around noon and the second one around 8:00 p.m. (see Fig. 1B). The daily calling activity described by these four periods follows a complex dynamics across the year and along different geographical zones, allowing it to be used to provide insight into the dynamics of human resting or sleeping pattern.

## Results

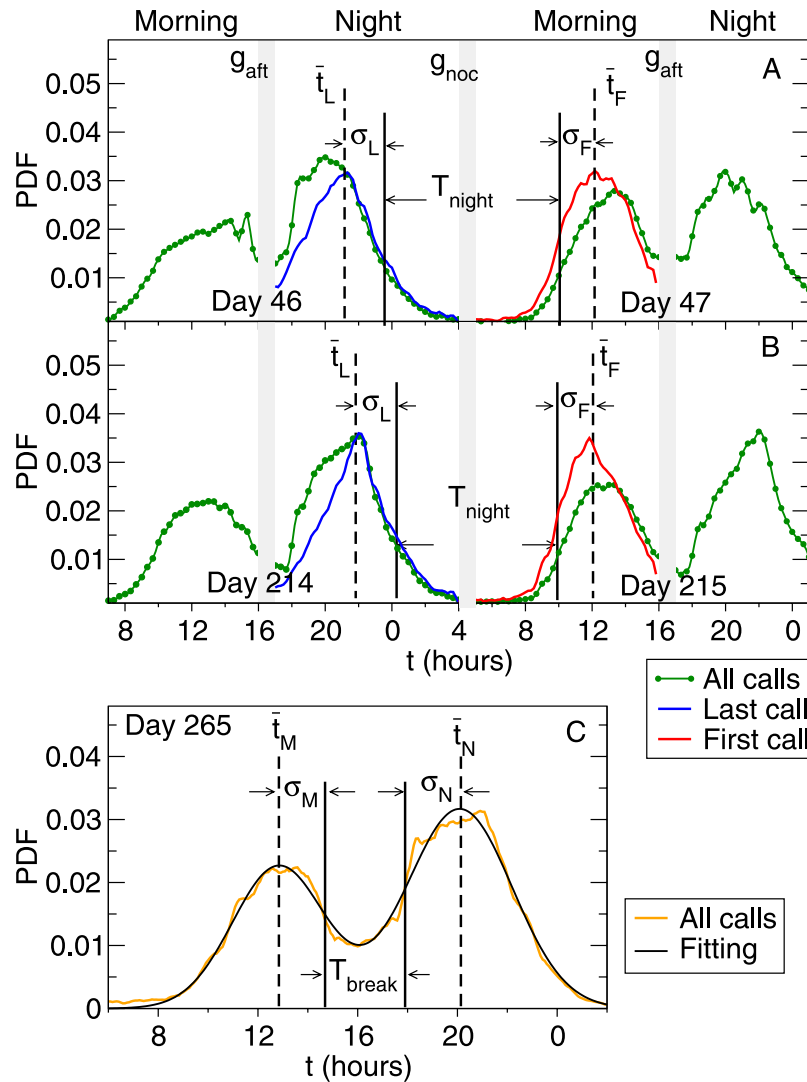
We analyse a dataset containing anonymised CDRs corresponding to a 12 month period in 2007 from a mobile phone service provider having subscribers in a number of cities in a country located in the Southern Europe subregion of the United Nations geoscheme<sup>46</sup>. The dataset contains more than 3 billion calls between 50 million unique identifiers, from which 10 million were associated with individuals having a contract with the company in question. The remaining identifiers belong to the subscribers of other companies or land-lines. Each call in the dataset involves at least one subscriber. For the majority of the subscribers, the age, gender, postal code, and location of the most accessed cell tower (MACT) are available, and we include into the analysis only those subscribers (termed as ‘users’ from here on) whose demographic information is complete. A user is considered to ‘live in a city’ if the following three geographical locations – the associated city centre, the location of the MACT and the centre of the postal code are sufficiently close (details in SI). We choose cities having more than a hundred thousand inhabitants in the year 2007, such that our final analysis takes into account a set of 36 cities with around 1 million users in total.

The calling activity of an entire city depends on a number of factors, but it can be described in terms of two variables: the time of the day, and the date of the year. From the dataset, we calculate the probability distribution  $p_{all}(t, d)$  of finding an outgoing call at time  $t$  of a day  $d = 1, \dots, 365$  by a caller living in a particular city. We define a ‘day’ starting from 4:00 a.m. of a calendar day and running to 3:59 a.m. of the next calendar day. The reason of this choice is to clearly distinguish between different daily calling activity periods, separating them into two parts: those events occurring before the observed termination of the calling activity during the night and those occurring after that. The time when the calling activity falls to its minimal observed value in general occurs around 4:00 a.m., and we choose this natural bound as the beginning/end point of the calling activity ‘day’, delimiting the calling distributions  $p_{all}$  of different days.

In Fig. 1A we show  $p_{all}(t, d)$  (green curve) during the two different pairs of consecutive days,  $d = 46–47$  (marking mid-February) and  $d = 214–215$  (the beginning of August) of 2007, for a city with around six hundred thousand inhabitants. The distribution  $p_{all}(t, d)$  turns out to be bimodal with the first mode corresponding to the calls made during the morning, peaking around noon, and the second mode being related to calls during the evening, reaching its maximum around 8:00 p.m. Such bimodal patterns are found for all the days around the year and for all the cities included in this study. These two peaks are naturally delimited by two regions when the activity falls to a minimum, one between 4:00 p.m. and 5:00 p.m., associated mainly with the time after lunch, and a second at the end/beginning of the day, between 4:00 a.m. and 5:00 a.m., which lies inside the normal sleeping period. We will refer to the former as the afternoon calling activity minimum  $g_{aft}$  and to the latter as the nocturnal calling activity minimum  $g_{noc}$ .

In order to study the periods of low activity, we split the day into two non-overlapping periods – ‘morning’ and ‘night’, each 11 hours long, delimited by  $g_{noc}$  and  $g_{aft}$ . Here we define the ‘morning’ as the time period between 5:00 a.m. and 3:59 p.m., and ‘night’ the time period between 5:00 p.m. and 3:59 a.m. on the following calendar day. Each ‘morning’, a user can make a number of calls but to focus on the time when the calling activity starts we consider only the first call made during the ‘morning’ for each user and construct the associated probability distribution of the time of the first call  $P_F(t, d)$ . Similarly, we find the last call made by every user during the ‘night’ and construct the corresponding probability distribution for the time of the last call  $P_L(t, d)$ . It should be emphasized that in this study we include only the calls made by the users (outgoing calls), excluding all the incoming calls, because these may not depend on the activity pattern of the users. In Fig. 1A we compare the probability distribution for all the calls  $P_{all}(t, d)$  with the corresponding distributions  $P_L(t, d)$  and  $P_F(t, d)$  for the times of the last call (blue) and first call (red), respectively, for two different pairs of consecutive days (during winter and summer) for the particular city with a population over six hundred thousand. The shapes of the distributions  $P_L(t, d)$  and  $P_F(t, d)$  depicted in Fig. 1A appear to be preserved for all the days and cities we have studied. For some particular days, mainly related to holidays and festivities, the corresponding data were filtered out from the analysis due to their atypical behavior (see Fig. A1 in SI).

From here on our assumption is that the two periods of low calling activity, characterise two commonly observed daily periods of low human activity (or resting periods). We identify the period of low calling activity occurring in the night, with the nightly sleep. There are results reported in the literature showing that analyses based on smart-phone usage describe the nocturnal sleeping period, thus giving support to our approach<sup>42, 44</sup>. In these studies as well as in ours, the underlying methods rely on estimating the sleep periods from the inactivity

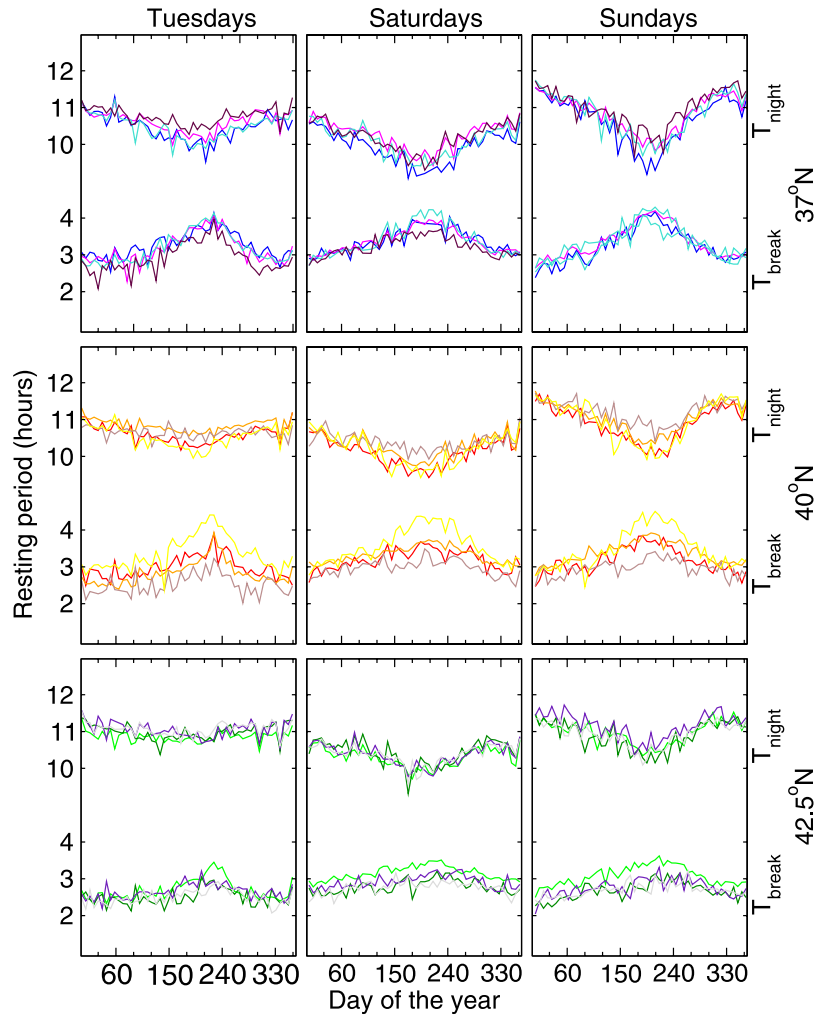


**Figure 1.** (Top panel) Probability distribution functions (PDF) of outgoing calls at time  $t$  of the day in a city for two sets of consecutive days of 2007:  $P_{all}$  when all the outgoing calls are included (Green),  $P_L$  when only the last outgoing calls in the *night* are included (Blue), and  $P_F$  when only the first outgoing calls in the *morning* are included (Red). From  $P_L$  and  $P_F$ , their means,  $\bar{t}_L$  and  $\bar{t}_F$ , and their standard deviations  $\sigma_L$  and  $\sigma_F$ , respectively, are calculated and used to define the period of low calling activity (PLCA) as the region bounded by  $\bar{t}_L + \sigma_L$  and  $\bar{t}_F - \sigma_F$  and determine its width  $T_{night}$  as the time interval between  $\bar{t}_L + \sigma_L$  and  $\bar{t}_F - \sigma_F$ . For the day 46 (middle of February, set A),  $T_{night} \approx 10.5$  hours, whilst for the day 214 (early August, set B),  $T_{night} \approx 9.5$  hours. (Note that  $P_{all}$  is delimited by the nocturnal calling gap  $g_{noc}$  (between 4:00 a.m. and 5:00 a.m.),  $P_L$  lies between the diurnal calling gap  $g_{aft}$  (from 4:00 p.m. to 5:00 p.m.) and  $g_{noc}$ , whilst  $P_F$  is bounded by  $g_{noc}$  and  $g_{aft}$ ). (Bottom panel) Probability distribution of outgoing calls during the day 265 (Sunday, orange line) in a city (of label 2), fitted by a superposition of two normal (Gaussian) distributions (black line), the first centered round noon and the second one in the evening (8 p.m.). The fitting is done with the mean values  $t_M$  and  $t_N$ , and the standard deviations  $\sigma_M$  and  $\sigma_N$ , corresponding to the noon and evening-centered activity modes, respectively (see details in the text). The afternoon break period  $T_{break}$  is defined as  $(t_N - \sigma_N) - (t_M + \sigma_M)$ .

of the type of device from which the data is collected. Nonetheless, both methods have limitations on accurately estimating the sleep periods which vary depending on the nature of collected data.

For the case of the afternoon period of low calling activity, it has been shown in different studies<sup>47–51</sup> that there is decay in the performance and intensity level of certain behavioral/physiological human activities, coinciding with the rising of the sleep propensity in the sleep/wake regulation. This period occurring during the afternoon has been associated with napping and/or resting<sup>49–52</sup>, and in general it is reported to occur between 3:00 p.m. and 6:00 p.m.<sup>48, 49, 51</sup>, which coincides with the afternoon period of low calling activity. As these periods overlap, we assume that the period of low calling activity is a good indicator of the resting time.

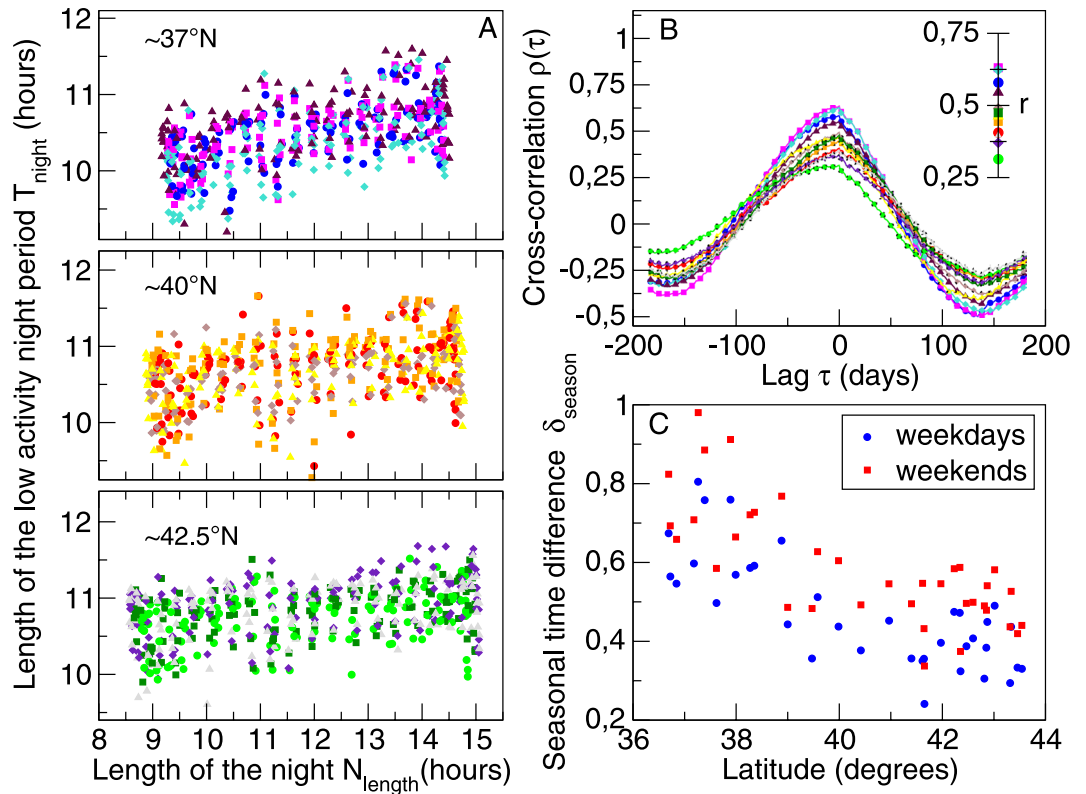
**Influence of latitude in seasonal variability of low-activity period.** As regards to the calling activity presented in Fig. 1A, we expect the sleeping hours to be mainly concentrated in the night when the calling activity



**Figure 2.** Periods of low calling activity or of resting:  $T_{break}$  and  $T_{night}$  for 12 different cities across 2007 for 3 different days of the week (Tuesdays, Saturdays, and Sundays in the left, central, and right column, respectively). Four cities are located in each one of the three different latitudinal bands (top)  $37^{\circ}\text{N}$ , (middle)  $40^{\circ}\text{N}$ , and (bottom)  $42.5^{\circ}\text{N}$ . For cities lying around  $37^{\circ}\text{N}$ , the color line associated with their times series are blue, magenta, turquoise and maroon; for cities at  $40^{\circ}\text{N}$ , the colors are red, orange, brown and yellow; and for cities around  $42.5^{\circ}\text{N}$  the colors are green, dark green, indigo and gray. Inside each one of the nine plots, the annual behavior of  $T_{break}$  (bottom lines) and  $T_{night}$  (upper lines) are shown, for the four different cities located at each band. On every plot,  $T_{break}$  and  $T_{night}$  show an opposite seasonal variations, with dynamics that appear to counterbalance each other, particularly on Sundays.

falls to a minimum. This nocturnal period of resting (NPR) is quantified in the following fashion. From distributions  $P_L$  and  $P_F$  for each city, we calculate the two means  $\bar{t}_L(d)$  and  $\bar{t}_F(d)$  to determine the times when the last call of the day and first call of the following morning were made, respectively. Thus the NPR is defined to be the region bounded by  $\bar{t}_L(d)$  and  $\bar{t}_F(d + 1)$ . In order to estimate the duration of the period when the calling activity has practically ceased, we determine the width of NPR,  $T_{night}$ , by taking into account the widths of the distributions  $P_F$  and  $P_L$ , determined by the standard deviations  $\sigma_F$  and  $\sigma_L$ , respectively. Hence, we can write  $T_{night}(d) = 24 - (\bar{t}_L(d) + \sigma_L) + (\bar{t}_F(d + 1) - \sigma_F)$  (see Fig. 1A). This definition for the width of NPR does not require any introduction of arbitrary cut-off parameters.

We have calculated  $P_L$  and  $P_F$  for 12 cities lying within one of three latitudinal bands, centered at  $37^{\circ}\text{N}$ ,  $40^{\circ}\text{N}$ , and  $42.5^{\circ}\text{N}$ , in such a way that each band contains 4 cities. In Fig. 2 we show the yearly variation of  $T_{night}$  for these 12 cities. The plots show that the width  $T_{night}$  changes across the year, being longest near the winter solstice and shortest six months later, near the summer solstice. Comparing the different plots corresponding to  $T_{night}$ , the following two characteristics should be noted. First, cities lying in the same latitude band show similar  $T_{night}$  curves, differing only by a vertical offset. This implies that although each city has a characteristic period of low activity, its yearly variation is very similar to that in other cities at the same latitude. Secondly, the difference between the highest (near winter solstice) and lowest values (near summer solstice) of  $T_{night}$  changes from latitude to latitude, being larger in southern cities and showing that there is an external factor that influences  $T_{night}$  with different intensity at different latitudes.



**Figure 3.** Relation between the period of low activity  $T_{\text{night}}$  and the length of the night  $N_{\text{length}}$ : (Panel A) Scatter plots of  $T_{\text{night}}$  against  $N_{\text{length}}$  for 12 cities located at one of the three different latitudinal bands  $37^\circ\text{N}$  (blue, magenta, turquoise and maroon),  $40^\circ\text{N}$  (red, orange, brown and yellow), and  $42.5^\circ\text{N}$  (green, dark green, indigo and gray), such that in each set the points represent the pairs  $(N_{\text{length}}, T_{\text{length}})$  across the 365 days of 2007 (only the Tuesdays, Saturdays and Sundays are shown). (Panel B) The normalized cross-correlation  $r$  between  $T_{\text{night}}$  and  $N_{\text{length}}$  as a function of the time lag  $\tau$ . Across the year  $T_{\text{night}}$  and  $N_{\text{length}}$  time-series are synchronized, as their cross-correlation  $\rho(\tau)$  reaches its maximum for the time lag  $\tau = 0$ , thus  $T_{\text{night}}$  and  $N_{\text{length}}$  are in phase and as depicted in the inset, the correlation between them is for all cases very high, i.e.  $r > 0.3$ ,  $p < 0.001$ . The time-series were normalized (see Methods) and all the days of the week were taken into account. (Panel C) Latitudinal dependence of the net change  $\delta_{\text{season}}$  between  $\bar{T}_{\text{night}}$  at winter solstice (maximum) and at summer solstice (minimum). For cities in the south, this difference is larger than for the cities in the north. Each point corresponds to one of the 36 studied cities calculated as averages of  $\bar{T}_{\text{night}}$  for weekdays from Mondays to Thursdays (blue) and for weekends from Fridays to Sundays (red).

As a matter of fact, the shape of  $T_{\text{night}}(d)$  throughout the year resembles the yearly behavior of the length of the night  $N_{\text{length}}(d)$ , defined here as the time between the sunset and the sunrise for each city. The value of  $N_{\text{length}}$  peaks around the winter solstice ( $d = 356$ ) and decreases monotonically until it reaches a minimum value around the summer solstice ( $d = 172$ ), coinciding with the behavior of  $T_{\text{night}}$ . In the scatter plots of  $T_{\text{night}}$  vs.  $N_{\text{length}}$  (Fig. 3A) we can see a noticeably linear dependence between these quantities, with the southern cities ( $\approx 37^\circ\text{N}$ ) showing the strongest effect. We use a linear regression,  $\bar{T}_{\text{night}} = \beta \bar{N}_{\text{length}} + \alpha$ , to quantify the yearly variation of  $\bar{T}_{\text{night}}$  as a function of  $\bar{N}_{\text{length}}$ . Here  $\bar{T}_{\text{night}}$  is defined as the average value of  $T_{\text{night}}$  from Monday to Thursday each week, to characterize typical weekdays and to reduce the fluctuations of  $T_{\text{night}}$ . We carry out the regression for 36 different cities (including the aforementioned 12 cities) located between the  $36^\circ\text{N}$  and  $44^\circ\text{N}$ , for weekdays. Similarly, to characterize typical weekend days, we have calculated  $T_{\text{night}}$  averaged across Fridays, Saturdays and Sundays. In order to study the effect of latitude on the seasonal variation of the NPR, we have calculated the net change  $\delta_{\text{season}}$  between  $\bar{T}_{\text{night}}$  for the winter solstice (maximum) and for the summer solstice (minimum), given by  $\delta_{\text{season}} = \beta \cdot [N_{\text{length}}(d = 356) - N_{\text{length}}(d = 172)]$ . This simple definition provides a way to calculate how the latitude affects the range where the period of low activity varies. For the 12 cities studied, we find that  $N_{\text{length}}$  and  $T_{\text{night}}$  are highly correlated, as can be seen in the inset in Fig. 3B, with the correlation  $r$  in the intervals  $[0.54, 0.62]$ ,  $[0.40, 0.48]$ , and  $[0.31, 0.47]$ , for the cities lying on the  $37^\circ\text{N}$ ,  $40^\circ\text{N}$ , and  $42.5^\circ\text{N}$  latitudinal bands, respectively, with a confidence level greater than 99% in all cases. Besides this high correlation between  $N_{\text{length}}$  and  $T_{\text{night}}$ , we explore if there is a time lag between their corresponding time-series. For this, we calculate the normalized cross-correlation  $\rho$  between the time-series for the same 12 cities as above (see Methods). In Fig. 3B it can be seen that  $\rho$  reaches its maximum when the time lag  $\tau$  is zero, meaning that these time-series vary in the same phase. In Fig. 3C we show  $\delta_{\text{season}}$  for each of the 36 cities at their own latitudes. Remarkably, the time difference  $\delta_{\text{season}}$  between winter and summer is always at least 15 minutes larger for the southern cities compared to the northern cities. For weekends we



observed a difference of 30 minutes, with almost a ratio 2:1 (30 minutes in the northernmost cities against one hour in the southernmost ones). The slopes of the linear regression obtained for each one of the 36 cities can be seen in the SI, Fig. A4. Cities around the same latitude have similar slopes ( $\beta$ s), and these slopes are larger for cities at the southern latitudes. In studies based on MCTQ, Allebrandt *et al.*<sup>53</sup> and Kantermann *et al.*<sup>54</sup> have reported that sleep duration variation across the year is around 20 minutes, for a sample of  $\approx 55,000$  people from Central Europe and a sample  $\approx 58,000$  people from Estonia, respectively. In contrast, in studies using wrist band activity sensors applied on subjects still living in hunter-gatherer societies (San and Tsimane), Yetish *et al.*<sup>20</sup> reported that sleep duration of these tribes living in urban-free environments changes around 50 minutes across the year.

**Ambient temperature effects on the duration of diurnal resting period.** For all the 36 cities included in this study and for almost every day,  $P_{all}$  shows apart from the nocturnal period of resting or predominantly night-time sleeping period, another low activity period centered around the diurnal gap  $g_{aft}$ , ranging from 4:00 p.m. to 5:00 p.m. (Fig. 1C). The dynamics of this diurnal afternoon period of resting (DPR), associated with a decrease in human activity, can also be traced from the  $P_{all}$  dynamics though here the decrease is less dramatic than in the nocturnal case. We find that  $P_{all}$  can be described by the sum of two normal distributions, with one of the activity peaks centered around the noon, and the other in the evening around 8:00 p.m. as seen in Fig. 1C. Unlike in the NPR case where almost all the activity of people ceases due to them sleeping, the DPR shows a more moderate decrease in activity between these two peaks. This is associated with only a fraction of people having naps or taking rest. The length of this low activity period DPR can be measured with the fitting function given by  $F(x) = \frac{a_0}{\sigma_M \sqrt{2\pi}} e^{-0.5((t-\bar{t}_M)/\sigma_M)^2} + \frac{a_N}{\sigma_N \sqrt{2\pi}} e^{-0.5((t-\bar{t}_N)/\sigma_N)^2}$ . Here  $\bar{t}_M$  and  $\sigma_M$  are the mean and the standard deviation of the noon-centered distribution and  $\bar{t}_N$  and  $\sigma_N$  the corresponding values for the evening-centered one. Hence we define the duration of the DPR or the afternoon break period to be  $T_{break}$  as  $(\bar{t}_N - \sigma_N) - (\bar{t}_M + \sigma_M)$  (Fig. 1C). The yearly change in  $T_{break}$  for 12 cities located at one of three different latitudinal bands 37°N, 40°N, or 42.5°N is depicted in Fig. 2.

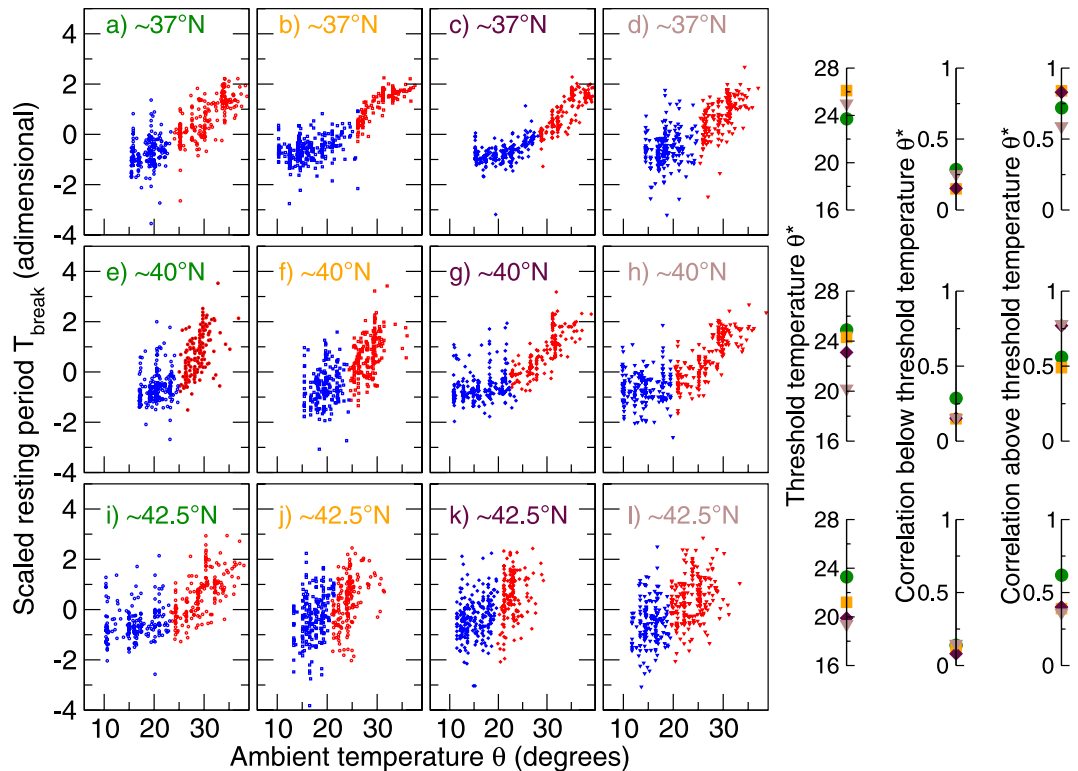
In order to study the influence of the daily temperature on the dynamics of  $T_{break}$ , we present it as a scatter plot with the maximum daily temperature  $\theta$  for the 12 above mentioned cities, as depicted in Fig. 4. It can be seen that  $T_{break}$  follows two different dynamics, distinguishable by a threshold temperature  $\theta^*$ , below which  $T_{break}$  seems not to be strongly influenced by the temperature and above which there is a clear linear trend (see Fig. A5 in the SI for a comparison between the time-series). To make this distinction more visible and to estimate the threshold, we apply a clustering algorithm to split the data-points into two sets, each one containing those points representing the dynamics below or above the threshold temperature. As the scatter of the data points is not compact and they are not spread along two separate regions with a well defined border, the density-based clustering algorithms<sup>57,58</sup> are not well-suited for determining a turning point in the dynamics. Instead we chose to use a spectral clustering technique based on the Shi-Malik algorithm<sup>55,56</sup>, where a graph is built from the data points and mapped into a low-dimensional eigen-space, where the graph is partitioned based on the similarities of the eigenvectors. We use an implementation of this algorithm found in a standard machine learning library<sup>59</sup>. This spectral clustering technique requires to set the number of clusters that data points must be split into. In order to search for the above-mentioned turning-point in the dynamics we set the number of clusters to two, which corresponds to the visual evidence (see Fig. 4) of the presence of two different behaviors of the afternoon resting period. An evaluation of different choice of parameters is shown in the SI (Figs. A8–A9). The results for each city can be seen in Fig. 4, in the form of two sets (red and blue points), with a threshold temperature ranging from 18° to 25° (Fig. 4). Once each set is split into two subsets, we calculate the correlation between  $T_{break}$  and  $\theta$  for the set of cities (Fig. 4, right panels). We observe a weak correlation when the temperature is below the threshold, i.e. for almost all of the 12 cities  $r \leq 0.25$  and  $p \approx 0.05$ . On the other hand, when the temperature is above the threshold, we observe the correlations to be strong, i.e.  $r \geq 0.5$  and  $p \approx 0.0001$ , which indicates a strong influence of ambient temperature on the afternoon break period  $T_{break}$ .

Here we have shown that both, the daily NPR and DPR (i.e. the nocturnal and diurnal resting periods), as well as the characteristic times  $T_{night}$  and  $T_{break}$  follow specific dynamics showing seasonal and geographical variations. As depicted in Fig. 2 these resting periods seem to follow opposite dynamics across the year such that when added up, the resulting total daily resting period shows only a small variation over the year (see Fig. A6 in the SI). This lends support to the notion that the total daily resting time of people is more or less conserved.

## Discussion

At the individual level the SWC is known to depend on various physiological and social factors that make it difficult to identify the precise causes influencing its dynamics. However, studying the yearly variation of the NPR averaged over a big population gives us quite unique insight into its complex behavior and helps us to identify some of the key influencing factors. The SWC is bounded inside the NPR and we expect that both entities follow similar dynamics but with somewhat different onset and termination times. We have shown that the duration of the NPR, characterized by the width  $T_{night}$ , closely follows the seasonal variation in the duration of the night. Moreover, we have observed that the duration of NPR is influenced by the latitude of the city in question. The difference between  $T_{night}$  in winter and in summer for the people in cities located at 36°N was found to be almost twice as much as that for the people in cities located 8° further north. It seems that the activity of people in southern cities are most affected by the length of the night (or conversely length of the daylight), despite the fact that in northern locations the length of the night changes faster and spans a bigger time interval than more southern latitudes.

Human physiology and its hormonal regulation is expected to govern the dynamics of the SWC at the individual level. In particular, the melatonin hormone is known to follow a circadian rhythm<sup>60</sup> and has been linked



**Figure 4.** Relation between the afternoon break period  $T_{break}$  and the maximum daily temperature  $\theta$ , for 12 different cities located at one of the three different latitudinal bands: (top)  $37^{\circ}\text{N}$ , (middle)  $40^{\circ}\text{N}$ , and (bottom)  $42.5^{\circ}\text{N}$ . (Left panels) For each city (a to l), a point represents a pair  $(T_{break}, \theta)$  for one of the 365 days of 2007. The scatter plots show that  $T_{break}$  behaves differently for different temperature ranges (blue points correspond to “cold” or “cool” days and red points to “warm” or “hot” days), separated by a threshold temperature  $\theta^*$  below which  $T_{break}$  seems to stay more or less constant and above it increasing with temperature. (Right panel) For each of the three latitudinal bands variabilities of: the threshold temperature  $\theta^*$  (first column) and the correlation  $r$  between  $T_{break}$  and  $\theta$  when the temperature is below (second column) and above (third column) the threshold temperature  $\theta^*$ . (The color labeling (green, orange, indigo or brown) of cities in each latitudinal band is the same in both panels). The correlation are calculated using points above the temperature thresholds, turning out to be strong in all cases with values  $>99\%$ . For temperatures below the threshold, the correlation is weak with  $p$ -values in general greater than 0.05. The points were clustered as below (blue) and above (red) the threshold by using a spectral clustering technique based on the Shi-Malik algorithm<sup>55,56</sup>.

to the SWC. For individuals, melatonin secretion commences during the evening between 8 p.m. and 10 p.m., peaking between 2:00 a.m. and 4:00 a.m. However, the onset and the duration of melatonin secretion during the circadian cycle is in turn related with the exposure to light and darkness, and considerable research has been done<sup>61–64</sup> showing how changes in the onset and length of the light exposure disrupts the secretion of melatonin, as well as the SWC<sup>65</sup>. For people living at latitudes far away from the equator, where the seasonal changes in the onset and length of daylight are large, it has been shown that their melatonin cycle is seasonally altered and this perturbation gives rise to disruptions in their SWC<sup>66</sup>. Thus it could be possible that the melatonin cycle is participating in the entrainment of the SWC (and the NPR) with the duration of the daylight across the year. On the other hand, the diurnal period of resting DPR is found to depend rather strongly on the temperature, making its duration during hot days longer, but showing no noticeable variation during cold or cool days. Surprisingly, however, this increase in the daily diurnal resting time is found to be counterbalanced by the decrease of the daily nocturnal resting time. Hence it seems that on warm or hot days a bigger fraction of the total resting period is taken by the afternoon break, thereby reducing the homeostatic pressure for sleeping and consequently the need for nocturnal resting period.

The evidence of the trade-off between the afternoon and nighttime rest or sleep at latitudes closer to the tropics has implications for our understanding of the time budgets of tropical hunter-gatherers who have also been reported to take afternoon ‘naps’ for about a quarter of a day<sup>20</sup>. Our results suggest that, in high temperatures tropical habitats, humans find it difficult to remain active during the afternoon when ambient temperatures are at their highest, much as in the case of for most monkeys and apes<sup>67,68</sup>. Figure 4 suggests that there is a critical threshold at which rest becomes necessary and that this occurs at ambient temperatures somewhere between  $20$ – $25^{\circ}\text{C}$  (though ‘naps’ of significant length are probably not common until ambient temperatures exceed  $\approx 30^{\circ}$ ). This has important implications in that it removes a significant amount of time from the active working day,

thereby reducing the time that can be devoted to ecologically or economically more important activities. Such constraints have imposed significant limitations on our species' ability to evolve into complex societies<sup>69</sup>.

## Methods

**Time binning and smoothing of distributions.** To generate each probability distributions, we divide the temporal axis in 5 minutes bins, in such a way that each distribution  $P_{all}$  contains 268 points, whilst  $P_L$  and  $P_F$  contain 132 points each one. In some cases, mainly for cities with small population, their associated probability distributions show fluctuations due to small sample size, then we apply a smoothing process to each distribution in order to reduce the noise, using the Savitsky-Golay<sup>70</sup> algorithm (7 points and degree 4).

**Data filtering.** From our dataset, two locations could be associated to each user. First one is the location of the most accessed cell tower (MACT) by the mobile device. From the zip code of the domicile of the user we get the second one, namely the location of the centre of the postal zone (postal location). These two locations were used when determining which users should be included in the analysis. We discarded those users with at least one of these two locations missing.

To reduce the noise due to poor sampling, we chose from the whole set of towns included in the mobile phone network only those cities with more than hundred thousand inhabitants in the studied year, and lying between latitudes 36°N and 44°N, which accounts for 36 cities. For each selected city, we encircled it with a 30 km diameter circle, and use the center of the circle as the associated geographical location of the city. All the studied cities fit inside a 30 km diameter circle.

Finally, we choose only those users who live in one of these cities. To ensure this into some extent, we impose the next two restrictions:

- The distance between the city location and at least one of the associated user locations (MACT and postal) should be less than 15 km, and
- The distance between the MACT and postal locations of the user must be less than 30 km.

There is no way to verify that the user actually lives in the associated city, but one could expect that almost every user lives in the same city as the one specified by the domicile, and that the MACT is indicative of their usual location. Thus the imposed restrictions seem to be a good way to ensure that a big fraction of the included users are correctly assigned to their actual city.

From the demographic information available, one can find the age of the users ranges from eighteen to more than hundred years. In this work we choose only users with age between 30 and 75 years old, because the calling pattern in younger people (mainly in the 18 to 25 years range) is more erratic than for older people. Older people (75+) were excluded the analysis due to the sparse calling activity that many of them had in the CDRs. After the previous filtering, the included users into the analysis were 925, 135 individuals.

**Cross-Correlation.** To measure the possible time delay between the time-series  $T_{night}(d)$  and  $N_{length}(d)$  we calculate the cross-correlation  $\rho \equiv \rho(\tau)$  between them, finding the time delay at which the cross-correlation reaches its maximum, meaning that the time-series are in the best possible alignment (in phase).

The (normalized) cross-correlation  $\rho \equiv \rho_{XY}(\tau)$  for a time delay  $\tau$  between two time-series  $X(t), Y(t)$ , is given by

$$\rho_{XY}(\tau) = \frac{1}{\sigma_X \sigma_Y} E[(X(t) - \bar{X})(Y(t + \tau) - \bar{Y})], \quad (1)$$

where  $E[\ ]$  denotes the expectation value, and  $\bar{X}, \bar{Y}$  and  $\sigma_X, \sigma_Y$ , are the average values and the standard deviations of  $X$  and  $Y$ , respectively.

Using the discrete form of  $X \equiv \{X_0..X_{n-1}\}, Y \equiv \{Y_0..Y_{n-1}\}$ , with  $n$  the number of samples, and time delay  $\tau$  as an integer between  $-n/2$  and  $n/2$ , then the cross-correlation reads as follows

$$\rho_{XY}(\tau) = \frac{1}{n - \tau} \sum_{i=\max(0,\tau)}^{\min(n-1+\tau)} \left( \frac{X_i - \bar{X}}{\sigma_X} \right) \left( \frac{Y_{i+\tau} - \bar{Y}}{\sigma_Y} \right), \quad (2)$$

where  $\bar{X} = \sum X_i/n, \bar{Y} = \sum Y_i/n, \sigma_X = (\bar{X^2} - \bar{X}^2)^{0.5}$ , and  $\sigma_Y = (\bar{Y^2} - \bar{Y}^2)^{0.5}$ , the average and standard deviation of  $X$  and  $Y$ , calculated accordingly the usual definitions. For a given  $\tau$ , the lower and upper bound of the summation as well as the denominator ( $n-\tau$ ) are chosen in such a way that only the overlapping part of the time-series is contributing to calculation of the cross-correlation.

**Data availability.** The datasets analysed during the current study are not publicly available due to a signed non-disclosure agreement. The dataset contains sensitive information of the subscribers, particularly age, gender, postal address and the location of the most accessed cell tower. Calling time distributions from which all the results shown in this work were generated are available from the last author on reasonable request.

## References

1. Aschoff, J. & Wever, R. Human circadian rhythms: a multioscillatory system. *Federation proceedings* **35**, 2326–2332 (1976).
2. Lange, T., Dimitrov, S. & Born, J. Effects of sleep and circadian rhythm on the human immune system. *Annals of the New York Academy of Sciences* **1193**, 48–59 (2010).
3. Sou  tre, E. *et al.* Circadian rhythms in depression and recovery: evidence for blunted amplitude as the main chronobiological abnormality. *Psychiatry research* **28**, 263–278 (1989).



4. Czeisler, C. A. *et al.* Bright light resets the human circadian pacemaker independent of the timing of the sleep-wake cycle. *Science* **233**, 667–671 (1986).
5. Grandin, L. D., Alloy, L. B. & Abramson, L. Y. The social zeitgeber theory, circadian rhythms, and mood disorders: review and evaluation. *Clinical psychology review* **26**, 679–694 (2006).
6. Roenneberg, T., Kumar, C. J. & Mewes, M. The human circadian clock entrains to sun time. *Current Biology* **17**, R44–R45 (2007).
7. Lewy, A. J., Lefler, B. J., Emens, J. S. & Bauer, V. K. The circadian basis of winter depression. *Proceedings of the National Academy of Sciences* **103**, 7414–7419 (2006).
8. Orzeł-Gryglewska, J. Consequences of sleep deprivation. *International journal of occupational medicine and environmental health* **23**, 95–114 (2010).
9. Evans, J. A. & Davidson, A. J. Health consequences of circadian disruption in humans and animal models. *Prog Mol Biol Transl Sci* **119**, 283–323 (2013).
10. Möller-Levet, C. S. *et al.* Effects of insufficient sleep on circadian rhythmicity and expression amplitude of the human blood transcriptome. *Proceedings of the National Academy of Sciences* **110**, E1132–E1141 (2013).
11. Davies, S. K. *et al.* Effect of sleep deprivation on the human metabolome. *Proceedings of the National Academy of Sciences* **111**, 10761–10766 (2014).
12. Santhi, N. *et al.* Sex differences in the circadian regulation of sleep and waking cognition in humans. *Proceedings of the National Academy of Sciences* **113**, E2730–E2739 (2016).
13. Wright, K. P., Hughes, R. J., Kronauer, R. E., Dijk, D.-J. & Czeisler, C. A. Intrinsic near-24-h pacemaker period determines limits of circadian entrainment to a weak synchronizer in humans. *Proceedings of the National Academy of Sciences* **98**, 14027–14032 (2001).
14. Hofstra, W. A. & de Weerd, A. W. How to assess circadian rhythm in humans: a review of literature. *Epilepsy & Behavior* **13**, 438–444 (2008).
15. Horne, J. A. & Ostberg, O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology* **4**, 97–110 (1975).
16. Roenneberg, T., Wirz-Justice, A. & Mewes, M. Life between clocks: daily temporal patterns of human chronotypes. *Journal of biological rhythms* **18**, 80–90 (2003).
17. Levandovski, R., Sasso, E. & Hidalgo, M. P. Chronotype: a review of the advances, limits and applicability of the main instruments used in the literature to assess human phenotype. *Trends in psychiatry and psychotherapy* **35**, 3–11 (2013).
18. Wright, K. P. *et al.* Entrainment of the human circadian clock to the natural light-dark cycle. *Current Biology* **23**, 1554–1558 (2013).
19. de la Iglesia, H. O. *et al.* Ancestral sleep. *Current biology: CB* **26**, R271–2 (2016).
20. Yetish, G. *et al.* Natural sleep and its seasonal variations in three pre-industrial societies. *Current Biology* **25**, 2862–2868 (2015).
21. Pever, J. & Horner, R. L. Human behavior: Sleep in hunter-gatherer societies. *Current Biology* **25**, R1133–R1135 (2015).
22. Yetish, G. *et al.* Response to de la iglesia *et al.* *Current Biology* **26**, R273–R274 (2016).
23. Ekirch, A. R. Segmented sleep in preindustrial societies. *Sleep* **39**, 715–716 (2016).
24. Torous, J., Staples, P. & Onnela, J.-P. Realizing the potential of mobile mental health: new methods for new data in psychiatry. *Current psychiatry reports* **17**, 1–7 (2015).
25. Kovanen, L., Kaski, K., Kertész, J. & Saramäki, J. Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proceedings of the National Academy of Sciences* **110**, 18070–18075 (2013).
26. Eagle, N., Pentland, A. S. & Lazer, D. Inferring friendship network structure by using mobile phone data. *Proceedings of the national academy of sciences* **106**, 15274–15278 (2009).
27. Jiang, Z.-Q. *et al.* Calling patterns in human communication dynamics. *Proceedings of the National Academy of Sciences* **110**, 1600–1605 (2013).
28. Blondel, V. D., Decuyper, A. & Krings, G. A survey of results on mobile phone datasets analysis. *EPJ Data Science* **4**, 1–55 (2015).
29. Aledavood, T. *et al.* Daily rhythms in mobile telephone communication. *PLoS one* **10**, e0138098 (2015).
30. David-Barrett, T. *et al.* Communication with family and friends across the life course. *arXiv preprint arXiv:1512.09114* (2015).
31. Bhattacharya, K., Ghosh, A., Monsivais, D., Dunbar, R. I. M. & Kaski, K. Sex differences in social focus across the life cycle in humans. *Royal Society Open Science* **3**, doi:10.1098/rsos.160097 (2016).
32. Aledavood, T. *et al.* Channel-specific daily patterns in mobile phone communication. In *Proceedings of ECCS 2014*, 209–218 (Springer, 2016).
33. Wang, Q., Gao, J., Zhou, T., Hu, Z. & Tian, H. Critical size of ego communication networks. *EPL (Europhysics Letters)* **114**, 58004 (2016).
34. Louail, T. *et al.* From mobile phone data to the spatial structure of cities. *Scientific reports* **4** (2014).
35. Sun, L., Axhausen, K. W., Lee, D.-H. & Huang, X. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences* **110**, 13774–13779 (2013).
36. Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* **327**, 1018–1021 (2010).
37. Sevtsuk, A. & Ratti, C. Does urban mobility have a daily routine? learning from the aggregate data of mobile networks. *Journal of Urban Technology* **17**, 41–60 (2010).
38. Stopczynski, A. *et al.* Measuring large-scale social networks with high resolution. *PLoS one* **9**, e95978 (2014).
39. Pappalardo, L. *et al.* Returners and explorers dichotomy in human mobility. *Nature communications* **6** (2015).
40. Jiang, S. *et al.* The timegeo modeling framework for urban motility without travel surveys. *Proceedings of the National Academy of Sciences* 201524261 (2016).
41. Pappalardo, L. & Simini, F. Modelling individual routines and spatio-temporal trajectories in human mobility. *arXiv preprint arXiv:1607.05952* (2016).
42. Abdullah, S., Matthews, M., Murnane, E. L., Gay, G. & Choudhury, T. Towards circadian computing: early to bed and early to rise makes some of us unhealthy and sleep deprived. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, 673–684 (ACM, 2014).
43. Christensen, M. A. *et al.* Direct measurements of smartphone screen-time: Relationships with demographics and sleep. *PLoS one* **11**, e0165331 (2016).
44. Cuttone, A. *et al.* SensibleSleep: A Bayesian model for learning sleep patterns from smartphone events. *PLoS one* **12**, e0169901 (2017).
45. Aledavood, T., Lehmann, S. & Saramäki, J. Digital daily cycles of individuals. *Frontiers in Physics* **3**, 73 (2015).
46. Standard country or area codes for statistical use. <https://unstats.un.org/unsd/methodology/m49/>. Accessed: 2017-08-01.
47. Mittle, M. M. *et al.* Catastrophes, sleep, and public policy: consensus report. *Sleep* **11**, 100 (1988).
48. Broughton, R. & Mullington, J. Circasemidian sleep propensity and the phase-amplitude maintenance model of human sleep/wake regulation. *Journal of sleep research* **1**, 93–98 (1992).
49. Broughton, R. J. Scn controlled circadian arousal and the afternoon 'nap zone'. *Sleep Res Online* **1**, 166–178 (1998).
50. Bes, F., Jobert, M. & Schulz, H. Modeling napping, post-lunch dip, and other variations in human sleep propensity. *Sleep* **32**, 392–398 (2009).
51. Strogatz, S. H., Kronauer, R. E. & Czeisler, C. A. Circadian pacemaker interferes with sleep onset at specific times each day: role in insomnia. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **253**, R172–R178 (1987).
52. Mittle, M. Two-peak 24-hour patterns in sleep, mortality, and error. In *Sleep and health risk*, 65–77 (Springer, 1991).
53. Allebrandt, K. V. *et al.* Chronotype and sleep duration: the influence of season of assessment. *Chronobiology international* **31**, 731–740 (2014).

54. Kantermann, T., Juda, M., Merrow, M. & Roenneberg, T. The human circadian clock's seasonal adjustment is disrupted by daylight saving time. *Current Biology* **17**, 1996–2000 (2007).
55. Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* **22**, 888–905 (2000).
56. Von Luxburg, U. A tutorial on spectral clustering. *Statistics and computing* **17**, 395–416 (2007).
57. MacKay, D. J. *Information theory, inference and learning algorithms* (Cambridge university press, 2003).
58. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise. *Kdd* **96**, 226–231 (1996).
59. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
60. Arendt, J. Melatonin, circadian rhythms, and sleep. *New England Journal of Medicine* **343**, 1114–1116 (2000).
61. Brown, G. M. Light, melatonin and the sleep-wake cycle. *Journal of Psychiatry and Neuroscience* **19**, 345 (1994).
62. Dawson, D. & Encel, N. Melatonin and sleep in humans. *Journal of pineal research* **15**, 1–12 (1993).
63. Broadway, J., Arendt, J. & Folkard, S. Bright light phase shifts the human melatonin rhythm during the antarctic winter. *Neuroscience letters* **79**, 185–189 (1987).
64. Vondrašová, D., Hájek, I. & Illnerová, H. Exposure to long summer days affects the human melatonin and cortisol rhythms. *Brain research* **759**, 166–170 (1997).
65. Wehr, T. A. Effect of seasonal changes in daylength on human neuroendocrine function. *Hormone Research in Paediatrics* **49**, 118–124 (1998).
66. Wehr, T. A. The durations of human melatonin secretion and sleep respond to changes in daylength (photoperiod). *The Journal of Clinical Endocrinology & Metabolism* **73**, 1276–1280 (1991).
67. Korstjens, A. H., Lehmann, J. & Dunbar, R. Resting time as an ecological constraint on primate biogeography. *Animal Behaviour* **79**, 361–374 (2010).
68. Dávid-Barrett, T. & Dunbar, R. I. Bipedality and hair loss in human evolution revisited: The impact of altitude and activity scheduling. *Journal of human evolution* **94**, 72–82 (2016).
69. Dunbar, R. *Human evolution: A Pelican introduction* (Penguin UK, 2014).
70. Savitzky, A. & Golay, M. J. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* **36**, 1627–1639 (1964).

## Acknowledgements

AG and KK acknowledge project COSDYN, Academy of Finland (Project No. 276439) for financial support. DM, KB, and KK acknowledge EU HORIZON 2020 FET Open RIA project (IBSEN) No. 662725. DM also acknowledges CONACYT (Mexico), grant No. 383907. RD acknowledges European Research Council for the Advanced Investigator Grant No. 295663. Authors thank AL Barabási for the dataset used in this research.

## Author Contributions

K.B., A.G. and D.M. carried out the analysis of the data. All the authors were involved in designing the project and the preparation of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-11125-z](https://doi.org/10.1038/s41598-017-11125-z)

**Competing Interests:** The authors declare that they have no competing interests.

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017