Typical Phone Use Habits:
Intense Use Does Not Predict Negative Well-Being

Kleomenis Katevas[*]
Imperial College London
London, UK
k.katevas@imperial.ac.uk

Ioannis Arapakis
Telefónica Research
Barcelona, Spain
ioannis.arapakis@telefonica.com

Martin Pielot
Telefónica Research
Barcelona, Spain
martin.pielot@telefonica.com

ABSTRACT

Not all smartphone owners use their device in the same way. In this work, we uncover broad, latent patterns of mobile phone use behavior. We conducted a study where, via a dedicated logging app, we collected daily mobile phone activity data from a sample of 340 participants for a period of four weeks. Through an unsupervised learning approach and a methodologically rigorous analysis, we reveal five generic phone use profiles which describe at least 10% of the participants each: *limited use*, *business use*, *power use*, and *personality- & externally induced problematic use*. We provide evidence that intense mobile phone use alone does not predict negative well-being. Instead, our approach automatically revealed two groups with tendencies for lower well-being, which are characterized by nightly phone use sessions.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

Clustering; User groups; Mobile phone use; Well-being; Personality traits; User study; Unsupervised-learning; Machine-learning; Multi-level modelling

INTRODUCTION

The average person in the US spends approximately three hours per day on the mobile phone, while young people (18-24 years) spend significantly more[1]. However, mobile phone overuse may lead to undesirable outcomes. Oulasvirta *et al.* [28] suggested that repetitive inspection of dynamic content such as notifications from social media apps may lead to phone overuse. Other studies have reported the association of phone overuse with sleeping disorders, increased stress, depression, social isolation and decline in academic performance [18, 23, 24, 25].

Detecting problematic phone use is a topic that has been previously examined in the research community using a variety of methods such as self-reported questionnaires [3, 27], that are known to be expensive, subjective, and prone to errors [38]. A different supervised learning based approach, presented by Shin and Dey [39], predicts these behaviours using data logs collected from the user's mobile device [39]. This solution was based on relatively small training sample that requires an existing ground truth and is hard to generalize. In addition, it tends to treat all users as homogeneous entities, neglecting that different mobile user profiles exist.

It is important to highlight that observing an extensive mobile phone use does not directly imply that a problematic use also exists. For example, extensive phone use by a business person during work could be considered as normal behaviour compared to late night phone use of a student, a behaviour that could easily affect the student's academic performance. According to Billieux [4], "disentangling a person's mobile phone user profile is a preliminary step necessary to avoid regrouping dissimilar behaviors into a deceptive label of mobile phone addiction".

The aim of this work is twofold: (1) uncover typical patterns of mobile phone use from a wide range of features, and (2) investigate whether clear indications of mobile phone overuse emerges spontaneously in some of these clusters.

To this end, we investigate an unsupervised approach for understanding the different types of mobile use behaviour in the daily routines of 340 participants. For an average duration of 4 weeks we collected a wide range of mobile sensor data using a custom made Android app that participants installed on their personal devices.

The main contributions of this work are:

- An unsupervised learning framework for deriving natural groups of users that is rigorously validated through a wide set of metrics and statistical measures.

- A taxonomy of five types of mobile phone users (*limited use*, *business use*, *power use*, and *personality- & externally induced problematic use*) that provides rich and explanatory insights with respect to mobile phone use patterns, user demographics (per class), and personality traits such as the emotional user's state (Bad/Good, Tense/Calm, Tired/Awake, Bored), Big Five Personality Test (Big5), Personal Health Questionnaire Depression Scale (PHQ-8) and Boredom Susceptibility Scale (SSS-BS).

- Evidence that there are specific types of phone use that associated with negative well-being, such as nightly phone use sessions. In contrast, overall usage intensity did not predict negative well-being.

In this work, we consider a fairly large group of users (contrary to other studies) that is not bounded by a specific age range or profession, *i.e.*, our sample has some diverse characteristics and thus we believe it achieves a higher degree of external validity, compared to other studies that employed smaller samples of participants. To our knowledge, we are the first to derive generic mobile phone user profiles and investigate their interaction with behavioural and mental health measures.

BACKGROUND AND RELATED WORK

Definition of Problematic Phone Use
In this work, we use the term *problematic phone use* to refer to the family of behaviors with undesirable effects on emotional well-being. Previous work has used a wide range of other terms, including problematic cellular phone use [44], dysfunctional use [5], maladaptive use [2], and compulsive use & technostress [24]. Some researchers even coined the term *smartphone addiction* [36, 34, 17, 5]. For example, Park and Lee [29] argued that mobile phone overuse has similar characteristics as impulse control disorders (ICD). Problematic phone use could be considered as one form of technological addictions. Since 2013, the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)[2] recognizes behavioural, non-substance related addictions and recommends further research into existing technological addictions for later inclusion. Kwon *et al.* [23] give examples of smartphone addiction, such as not paying attention in traffic or not being able to concentrate in class. However, no clear uncontested definition exists for this term. With *problematic phone use*, we refer to any type of mobile phone use that negatively affects its user's emotional well-being.

Undesirable Side-Effects of Mobile Phone Use
Previous work reported a wide range of undesirable side-effects of problematic mobile phone use on emotional well-being, *e.g.*, causing psychological distress [2], depression [44, 40], anxiety [18, 25], sleep disorders [18, 40], and reduced satisfaction with life [25]. Beranuy *et al.* [2] reported an association between heavy mobile phone use and psychological distress in 404 undergraduate college students. Jerano *et al.* [18] examined the effects of internet and cell phone use in 337 college students studying in Spain. They reported an association of heavy Internet use with high anxiety, and high cell-phone use with high anxiety and insomnia. Lee *et al.* [24] found a correlation of compulsive smartphone use and psychological traits such as locus of control, social interaction anxiety, materialism and the need for touch. Lepp *et al.* [25] found a negative correlation of mobile phone use and texting with academic performance, and a positive correlation to anxiety. Yen *et al.* [44] examined the association between problematic mobile phone use and depression in a large number of 10,191 adolescent students in Taiwan. They reported that adolescents who had significant depression were more

[2]http://www.dsm5.org/

likely to have four or more symptoms of problematic mobile phone use. In a questionnaire-based study with 300 university students, Samaha and Hawi [37] found correlation between smartphone addiction and stress, a negative correlation with academic performance and a mediated negative correlation with satisfaction with life.

Predictors of Problematic Phone Use
Socio-demographic factors have been found to predict problematic mobile phone use: female mobile phone users are more likely to develop dependence-related symptoms with regards to mobile phones [6, 2, 24, 18]. Furthermore, younger mobile phone users tend to be more likely to exhibit symptoms of dependency to mobile phone use [3, 44].

Personality traits can also be predictors of problematic phone use. In particular, problematic phone use has been correlated to neuroticism [12], anxiety [18, 17], distress [2], and social extraversion [17]. Billieaux *et al.* [6] reported that impulsivity is related to the number and duration of the calls made daily, while urgency appeared to be the strongest predictor of problematic use. Low self-esteem has also been linked with dysfunctional use [3, 4].

Further, two studies found that patterns of phone use themselves can predict whether participants would score above a threshold on questionnaires related to undesirable emotional well-being. Shin and Dey [39] found that the number of apps used per day, the ratio of SMSs to calls, the number of event-initiated sessions, the number of apps used per event initiated session, and the length of non-event-initiated sessions are the best predictors for scoring high on a questionnaire about behavioral addiction. Lee *et al.* [24] found that *e.g.*, using the phone a lot, longer period of phone use during mornings and evenings, time spent browsing the web, and "highly skewed usage pattern with respect to a few frequently used apps" predict whether phone users will fall into the high-risk group of the *Smartphone Addiction Proneness Scale for Adults* [20].

Both mentioned works group people into a "normal" and a "problematic" group depending on the questionnaire based scores. This approach ensures the presence of an existing group of problematic users. Thus, the used machine-learning algorithm will do its best to identify those groups. However, in reality, problematic phone users might not stand out too much from normal users. Therefore, it remains unclear whether problematic mobile phone use is a fringe phenomenon, or whether it has manifested itself in common patterns of phone use.

Types of Mobile Phone Use
In order to put problematic phone use into context, it is important to know the typical mobile phone use behavior. Typical patterns of phone use have widely been reported in recent, related work. For example, Andrews *et al.* [1] reported the mobile use patterns of 19 staff members of the University of Lincoln for a period of 2 weeks, by examining the screen on/off events collected using an Android app. They found that their participants used the phone 84.68 times each day (SD = 55.23) and spent 5.05 hours each day on it. On the basis of logs from 29,279 devices. Hintze *et al.* [15] observed

that the average mobile phone user has 60 sessions per day and uses the device for about 1.5 hours per day. Mathur *et al.* [26] analyzed four broad facets of smartphone behavior in India. They found that compared to Western users, Indian users exhibit more nightly use and charge their phones more frequently.

However, not all people use their mobile phone the same way. Such general statistics as reported by the work above may not do justice to all types of mobile phone users. Thus, previous work has attempted to identify clusters of mobile phone use behavior through unsupervised-learning approaches. The rationale was that these approaches would reveal the naturally-existing, most dominant types of phone use.

Do *et al.* [10] explored typical patterns of phone use. They model patterns of phone use by representing app launches by a bag-of-application model, this is, modelling how frequently it would be observed to launch different types of apps (voice, SMS, internet, camera, gallery) during four time frames: night, morning, afternoon and evening. While they identify common topics of phone use, they do not disclose common types of users.

Zhao *et al.* [45] examined the app use behaviours in a large scale study with 106,762 Android users. Using an unsupervised learning approach, they distinguished 382 types of users based on their app usage traces (*i.e.* weights of how much each app category has been used per user). They profiled the 3 biggest clusters ranked by size, and named *Night communicators*, *Screen checkers*, and *Evening learners* – showing that distinct patterns of usage exist. This work was limited by the absence of more broad features, such as app usage duration, phone calling interaction, or generic phone usage, information that could lead into a more precise clustering of user behaviour.

Visuri *et al.* [42] analyzed patterns of phone use to improve the scheduling of quantified-self self-report questionnaires. On the basis of pre-study questionnaire and the device use of 48 participants, they identified five clusters: (a) the casual user who was most passive and typically female; (b) the social chatterer who was rather young and frequently used communication apps; (c) a balanced cluster with no clear characteristics; (d) the night owl who tended to use the phone during the night and frequently used communication apps; and (e) the work on-the-go user who was an active phone user that quickly responded to interruptions. Clusters 6 and 7 were discarded, because they had no distinct features and only few users. While not their primary outcome, the results of Visuri *et al.* indicate that mobile phone users can be separated into a handful of meaningful clusters.

None of these previous works investigated to what extent these types of behaviors are associated with problematic phone use and undesirable effects on emotional well-being.

EXPERIMENTAL METHODOLOGY

The first goal of this work is to identify clusters of *typical* patterns of phone use. The objective is to keep our natural groupings of users as high-level and general as possible so that they can be easy for interpretation and further analysis.

Treating every person as individual exceeds our capacity to effectively design better-adapted solutions. Hence, we seek to identify a few, broad clusters that cover the most dominant patterns of mobile phone use. To this end, we introduced the constraint that clusters of typical phone use should at least contain 10% of the study population.

The second goal is to explore whether problematic phone use spontaneously emerges as a cluster. If problematic phone use is a real and wide-spread problem which is clearly related to how the mobile phone is used, we would expect that it clearly stands out as a cluster as well. This would provide evidence for the existence of problematic phone use as a relevant problem to the society. To this end, we apply an unsupervised-learning pipeline onto a data set of mobile phone use logs.

Participants

Our data set contains mobile phone use logs from 340 Android phone users, recruited through a specialized agency for an average duration of four weeks during summer 2016. We requested a sample that matches the gender and age distribution of Spain, the country of study. The only restriction was that people were required to own an Android phone as Android phones account for the large majority ($\sim$90%) of the smartphone share in the country of study. The participants' ages ranged from 18 to 66 years ($M = 37.850$, $SD = 11.025$), with a balanced gender split (53.24% female and 46.76% male).

Data

The data was collected through an app which was running in the background while passively collecting rich sensor data about the user's context and phone use. Participants registered the app to listen for accessibility events, which allowed it to log which apps they were using. In addition, the app collected information about the user's phone use, including the screen status (on, off, unlocked), phone calls (incoming, outgoing), data activity (kb/sec), ringer mode (default, vibration, silent), battery level and number of photos taken. Basic demographics such as age and gender were self-reported at the beginning of the study and included in the data set.

Experience sampling

In addition to sensor-data logging, the application collected self-report via experience sampling. The application created about 10 to 15 notifications per day. The notification text prompted users to report their emotional state by asking "How do you feel?".

Clicking on such a notification opened a questionnaire with 4 Likert-scale items about the emotional state of the user. More specifically, the user reported whether they are feeling: Bad/Good, Tense/Calm, and Tired/Awake, using a five-point scale (reduced into a three-point scale, *i.e.* Bad/Neutral/Good), as well as feeling Bored using a boolean scale (No/Yes).

The timing of the delivery was semi random to capture a wide range of scenarios. Between 22:00 and 8:00 in participants' local time, the notifications were silent and without vibration. After 6 responses of a given day, notifications were suspended and resumed only on the next day. If the participants did not respond within 10 minutes, the notification was removed from

the notifications tray. To be eligible for the compensation, participants were required to respond to at least two of these questionnaires for a minimum of 21 days.

In addition to asking participants to report their mood, on the bottom of the questionnaires, the users were offered additional things to do. Amongst them was an option to "learn more about yourself" that led to psychometric questionnaires sequentially selected from the following: Big Five Personality Test (Big5), Personal Health Questionnaire Depression Scale (PHQ-8) and Boredom Susceptibility Scale (SSS-BS). The Big5 [19] scale consists of 20 questions that evaluates the user's personality in five domains: *Openness* (to experience), *Conscientiousness*, *Extraversion*, *Agreeableness* and *Neurocism*, using a scale from 0 (low) to 24 (high). PHQ-8 [22] is an 8-item questionnaire that is used to detect the presence and severity of depression in patients, according to DSM-5. The scale ranges between 0 (low likelihood) and 24 (high likelihood). SSS-BS [46] measures aversion to routine, repetition and dull people, using a scale that ranges between 0 (low boredom propensity) and 66 (high boredom propensity). Responding to these questionnaires was voluntary. Hence, we only collected responses for a subset of the participants.

The case study has been approved by the legal department of our institution. Taking user security and privacy seriously, all communication with our back-end service was secured using SSL encryption (TLS v1.2) and all data collected was stored in our database using pseudonymization, being impossible to identify and reach out participants.

Features
Phone use directly reflects in the number of phone unlocks, app launches, calls received etc. Since our focus was to investigate typical phone use, we computed the daily median of 23 of such features (*e.g.* daily number of phone sessions, median duration of calls, battery drain etc.), as well as three features that characterize the device status during the data collection period (*i.e.* status of the ringer mode). We chose daily features over hourly or weekly as we believed that hourly features would be too fine-grained and weekly too coarse-grained information. Moreover, we chose the median as a measure of central tendency since most of the scores of the features are not normally distributed.

One defining property of overuse and addictive behavior is that it leads us to override and ignore our basic physiological needs, such as sleep. This, in turn, leads to negative outcomes, such as reduced well-being [13]. Thus, we separately investigate phone use that took place during night time (*i.e.* 0:00 – 5:59) as a proxy for interactions that were likely to have interrupted sleep. All interactions happening between 6:00 and 23:59 will be referred to as "day" features, interactions between 0:00 and 5:59 as "night" features for the remainder of this work. We ended up with 26 features in total (see Table 3) that belong to the following categories of mobile sensor events:

Screen sessions – 4 features capturing the number of times a user unlocked the screen, as well as the duration of phone activity, during the day and night separately. These features indicate the general amount and frequency of phone use.

Phone calls – 8 features representing the number and duration of outgoing and incoming calls as an indication of perceived dependence on the phone [6].

App launches – 8 features modeling the number of app launches per day & night of 4 categories: *Social* – as frequently checking social media apps has been previously associated with negative outcomes [28], *Messaging* – as an indication of personal and direct communication, *Email* – as an indicator of professional communication, and *Games* – as an indication of time killing in recreation activities. We excluded other type of apps as general phone use can still be captured by the 'screen sessions' feature listed above.

Network data activity – captures the average daily download volume (in kb/sec), serving as a proxy to the degree of phone use intensity. Previous works have shown that data activity was an important predictor of negative effects in well-being such as boredom [30].

Photos taken – models the daily number of pictures taken with the phone. Extensive use of the phone's camera can indicate that users are occupied with interesting, picture-worthy activities and/or that they are communicating activities to friends via photo messages or social media posts. Moreover, the popular selfie (self-portrait photograph) trend has been recently associated with narcissism, self-perception of leadership, vanity and exhibitionistic propensities [43]. Note that only the quantity of pictures taken per day was accessed, without looking at the content of the pictures itself.

Battery drain – represents the percentage of the device's battery drain levels per hour. High battery drain indicates heavier use of the phone (*e.g.* engagement with games, communication using video calling apps etc.).

Ringer mode – a one-hot-encoded feature representing the fraction of time that a device was set to normal, silent and vibrate ringer mode during the study. These features are an indication of how tolerant a user is to disruptions.

Unsupervised Learning
The main objective of this analysis is to divide the users in our data set into natural groups that reflect salient patterns of mobile phone use. To this end, we employed unsupervised learning to derive generic profiles of mobile phone users, *i.e.*, profiles that abstract from characteristics of an individual to behaviours and habits shared by the members of each group.

Given that there are no predefined classes, the outcome of a clustering process is not (always) deterministic and may result in a different partitioning of the data, depending on the specific criteria used each time (*e.g.*, number of desirable clusters or type of clustering algorithm). Since we do not know *a-priori* which clustering algorithm and configuration settings will perform better for the examined domain, we opted for several representative options. The reason was to avoid the common pitfall of resorting to a single clustering algorithm that may not be able to produce proper partitions, due to the nature of the given data set.

To understand which clustering algorithm performs best, we evaluated the result produced by each clustering algorithm

using several internal validity metrics. Since each validity metric can result in a different ranking of the clustering configurations, we reached a consensus for the best clustering setup through rank aggregation. This gave us a super list which was as "close" as possible to all individual ordered lists simultaneously, with the top positions granted to the clustering configurations that performed better amongst many validity metrics.

Preprocessing

For the purposes of the unsupervised learning task, we aggregated our data per user and per day, and computed the median values for all extracted features. Given that some of the non-agglomerative clustering algorithms that we considered (Table 1) work by optimizing a distance criterion (*e.g.*, Euclidean distance), we normalized our features, which homogenized their relative importance in the clustering experiments.

Clustering Algorithms

Table 1 summarizes the clustering algorithms and the ranges of values we used to parametrize them. More specifically, we considered centroid-, connectivity-, density-, and distribution-based clustering algorithms. *Centroid-based* clustering algorithms derive the notion of similarity by the closeness of a data point to the centroid of the clusters. These algorithms run iteratively to find the local optima and the the number of clusters required in advance. *Hierarchical-based* (aka *connectivity-based*) clustering algorithms work by forming an initial pair of clusters and then recursively consider whether it is worth splitting each one further; this type of clustering algorithms produce a hierarchy that can be represented as a binary tree (*i.e.*, dendrogram). *Density-based* clustering algorithms search the data space for areas of varied density of data points in the data space. They isolate various different density regions and assign the data points within these regions in the same cluster. Last, the *distribution-based* clustering algorithms do not place data instances categorically in one cluster or the other; instead, they assign them a certain probability of belonging to each cluster.

The set of clustering algorithms × their parameter settings (see Table 1) results in 135 possible clustering configurations. From this set, we excluded those clustering configurations that produced clusters with less than 10% of the users, namely `Mean-shift` and `DBSCAN`. These clustering algorithms almost always resulted in as many clusters as there are data points and, thus, were unfit for our analysis. In what follows, we compute several cluster-validity criteria to assess the quality of the clustering and establish whether the identified groups reflect typical and distinguishable types of phone use.

Measuring Cluster Validity

Due to the unsupervised nature of clustering methods, results may vary greatly in terms of the data partitioning. Therefore, an important challenge in cluster analysis is the evaluation of the quality of the results. The "correctness" of clustering algorithm results is verified using a process known as cluster validation [14]. Three main approaches exist for this purpose: (1) external criteria, (2) internal criteria, and (3) relative criteria. In this work, we evaluated the results of our clustering

experiments using the second type of validity criteria (internal), which consider only quantities and features inherent to the data. This choice was based on the larger availability of clustering validity criteria that do not rely on an external ground truth.

If multiple internal validity criteria are being used, there is a risk that some of these optimize a similar cost function. This can be a problem if the type of cost function is not suitable for the data set: if criteria with sub-optimal cost functions dominate the set of evaluation criteria, they will bias the overall evaluation and lead to a poor choice of the final clustering. To mitigate this, we removed the subset of criteria that optimized very similar cost functions or produced highly-correlated rankings. This is similar to the ensemble learning approach, where instead of relying on one machine-learning algorithm to make decisions, we would combine the output of several different models by learning an ensemble of models and using them in combination. This process resulted in to a selection of 13 widely-used criteria that are shown in Table 2.

Selecting Validity Measures

In previous section, we addressed how we measure the quality of the produced clustering configurations. Prior to applying the rank aggregation algorithm, we downsized our internal validity metrics to a subset that shows low consistency among the produced rankings. This is because (1) using all available internal validity metrics can slow down rank aggregation and (2) including highly correlated internal validity metrics can bias the results towards a specific type of cost function and eventually affect their quality. To this end, we computed all $\frac{n(n-1)}{2}$ pairwise Spearman correlations of the orderings. This gave a complete weighted graph $G = (V, E)$: vertices are the validity measures and each edge $e = (v_i, v_j)$ is weighted with the correlation value between v_i and v_j. Given G, we selected the k vertices, such that the resulting k-clique has the minimum edge-weight. The minimum edge-weight clique problem is NP-Hard [33]. Therefore, we used the greedy approximation shown in Algorithm 1. The output contained the k validity measures with the lowest overall agreement.

Rank Aggregation

Given the k validity measures with high consistency, we wanted to identify the n clustering configurations that were better (*i.e.*, were ranked higher) than the rest. More specifically, we needed those clustering configurations that had a better standing in as many lists as possible. This is a typical instance of *multiple-winner voting problem*, where we are given a set of ordered preferences as input (ordered clustering configurations per measure in our case), and we want to select n winners. For our analysis, we chose rank aggregation [11], which has ties to other known techniques (e.g., *Borda Count*, *Condorcet*, etc.), but is better at filtering out noise. Furthermore, we used the R implementation *RankAggreg* [41] with the settings recommended by [31], which is an efficient implementation of the algorithm despite of the problem's NP-Hardness.

We can formalize our goal within the framework of the following optimization problem: We want to find a super list which would be as "close" as possible to all individual ordered lists simultaneously.

Table 1. Clustering Algorithm Parameters.

Category	Algorithm	Parameters	Value
Centroid	K-means	# Clusters	[2, 5]
Connectivity	Agglomerative	# Clusters	[2, 5]
		Linkage type	*ward, complete, average*
		Affinity	*euclidean, l1, l2, manhattan, cosine*
	Spectral	# Clusters	[2, 5]
		Kernel type	*nearest neighbors, rbf*
Density	DBSCAN	Neighborhood size	*1e-6, 1e-5, 1e-4, 5e-3, 1e-3, 4e-2, 3e-2, 2e-3, 1e-4*
		Min samples	34 (10% of the user set)
	Mean-shift	Bandwidth	Auto using *0.1, 0.25, 0.5, 0.75, 0.9* quantiles
Distribution	Gaussian Mixtures	# Components	[2, 5]
		Covariance type	*full, tied, diagonal, spherical*
		Convergence threshold	*1e-4, 1e-3, 1e-2, 1e-1*

Table 2. Internal validity criteria used for evaluating the quality of the produced clusters

Algorithm	Rule	Algorithm	Rule
Banfeld Raftery	*min*	Point Biserial	*max*
C index	*min*	Davies-Bouldin	*min*
Dunn	*max*	Ray-Turi	*min*
Gamma	*max*	SD-Scat	*min*
log(BGSS/WGSS)	*min diff*	Silhouette	*max*
McClain-Rao	*min*	Xie-Beni	*min*
PBM	*max*		

The objective function that satisfies it is the following:

$$\Phi(\delta) = \sum_{i=1}^{m} w_i d(\delta, L_i) \qquad (1)$$

where δ is a proposed ordered list of length $k = |L_i|$, w_i is the importance weight associated with list L_i, d is a distance function, and L_i is the ith ordered list. The idea is to find δ^* which would minimize the total distance between δ^* and L_i's:

$$\delta^* = \arg\min \sum_{i=1}^{m} w_i d(\delta, L_i) \qquad (2)$$

In our case, the items of the lists are the clustering configurations and each ranked list L_i is a internal validity metric. We ranked the lists in ascending order according to the score that each clustering configuration achieved, and subsequently produced their aggregated ranking using the Cross-Entropy Monte Carlo algorithm [9, 35]. The Cross-Entropy Monte Carlo algorithm is an iterative procedure for solving difficult combinatorial problems in which it is computationally not feasible to find the solution directly. We applied a non-weighted aggregation that optimized a distance criterion, *e.g.* Kendall's tau and Spearman's Footrule distance, and allowed for a far more objective and automated assessment of the clustering results. The result was a list with the top positions granted to the clustering configurations that performed better than others, for the internal validity metrics that we considered. The top-ranked clustering configuration was the `Spectral` with rbf kernel, for $k = 5$, which we used in the subsequent analysis.

Multilevel Modelling

Research studies that span across longer periods of time require a different way of thinking about design and analysis (*e.g.*, equating time as an independent variable, accounting for correlated errors) than do cross-sectional designs, even to the

Algorithm 1 Greedy approximation algorithm for selecting the internal validity measures with lowest agreement.

Input: Weighted Graph $G = (V, E)$, int k
Output: Max-Edge Weight Subgraph G_k, $|G_k| = k$
1: $G_k \leftarrow \emptyset$
2: $e_1 \leftarrow$ Select edge with maximum weight.
3: Add vertices of e_1 to G_k
4: **while** ($|G_k| < k$) **do**
5:    **for** $v_i \in (G \; G_k)$ **do**
6:       $s_i \leftarrow \sum_{v_j \in G_k} weight(v_i, v_j)$
7:    **end for**
8:    Select v_j with max s_i
9:    $G_k \leftarrow (G_k \cup v_j)$
10: **end while**
11: **return** G_k

extent that the data sets must be structured differently. To this end, we analyze the mood and psychometric questionnaire data we collected throughout our study by means of mixed multilevel models, a method that allows dealing with some specificities of repeated measures data (violation of sphericity, nested data, autocorrelation of residuals) more effectively compared to traditional statistical methods such as ANOVA [21]. Mixed models involve looking at how individuals (or units, groups, organizations, etc.) change over time and whether there are differences in patterns of change.

For our analysis, we used the R package *nlme* [32]. Our approach to the model construction was the one described in [8, 7, 21] by fitting a series of two-level models, in which measurements for each condition (level 1) were nested within participants (level 2). Initially, for our level 1 analysis, we began with a random intercept model for all of our predictors (ESM, PHQ-8, etc.) and estimated the intra-class correlation coefficient (ICC) to determine the strength of the non-independence of the observations. The unconditional means model provides between-group and within-group variance estimates in the form of τ_{00} and σ^2, respectively. To determine whether τ_{00} is significant, we compared -2 log likelihood values between (1) a model with a random intercept, and (2) a model without one.

For our repeated measures data (*e.g.*, ESM), the next step for progressing from a regression model to a growth model (via model comparisons) was to establish a simple model without any random effects to serve as a baseline. This step involved modeling the fixed relationship between time and the corresponding dependent variable. We began by fitting a linear relationship and progressively added more complicated rela-

Table 3. Descriptive statistics (Mean, Std.) for the produced clusters of mobile phone users

| Feature	C1	C2	C3	C4	C5	All
f_1: App Launched Day: # Social	3.64 (6.65)	4.54 (7.29)	16.34 (31.17)	9.40 (12.86)	9.33 (14.06)	8.43 (17.28)
f_2: App Launched Day: # Messaging	17.85 (15.88)	45.41 (33.27)	44.83 (55.30)	40.35 (30.75)	39.20 (29.78)	35.12 (35.82)
f_3: App Launched Day: # Email	1.83 (3.25)	4.72 (6.78)	8.52 (10.38)	2.99 (3.86)	4.47 (9.35)	4.26 (7.44)
f_4: App Launched Day: # Games	0.49 (2.07)	0.59 (1.56)	4.40 (7.97)	1.10 (2.82)	2.43 (6.35)	1.77 (5.04)
f_5: App Launched Night: # Social	0.11 (0.56)	0.47 (1.31)	2.19 (6.96)	1.27 (3.34)	0.80 (2.14)	0.91 (3.63)
f_6: App Launched Night: # Messaging	0.22 (0.63)	0.59 (1.61)	2.90 (9.60)	3.37 (7.30)	2.11 (4.43)	1.74 (5.78)
f_7: App Launched Night: # Email	0.00 (0.00)	0.00 (0.00)	0.33 (0.64)	0.02 (0.13)	0.00 (0.00)	0.07 (0.31)
f_8: App Launched Night: # Games	0.01 (0.10)	0.10 (0.36)	0.22 (0.64)	0.15 (0.81)	0.13 (0.39)	0.11 (0.50)
f_9: Battery Drain (% per hour)	6.14 (5.03)	9.63 (5.60)	18.16 (11.28)	12.01 (6.61)	11.54 (6.21)	11.11 (8.31)
f_{10}: Calls Day: # Incoming	0.66 (0.85)	4.34 (2.41)	0.91 (0.92)	0.67 (0.87)	0.94 (1.54)	1.25 (1.79)
f_{11}: Calls Days: Incoming Duration (min)	0.60 (0.95)	6.85 (5.57)	0.88 (1.59)	0.43 (0.79)	1.05 (3.06)	1.54 (3.33)
f_{12}: Calls Day: # Outgoing	0.33 (0.61)	4.46 (3.91)	0.48 (0.68)	0.51 (1.01)	0.82 (1.43)	1.03 (2.14)
f_{13}: Calls Day: Outgoing Duration (min)	0.37 (1.02)	8.52 (7.11)	0.51 (1.07)	0.75 (1.94)	1.06 (2.03)	1.68 (3.97)
f_{14}: Calls Night: # Incoming	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.01 (0.06)	0.00 (0.03)
f_{15}: Calls Night: Incoming Duration (min)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.03)	0.00 (0.01)
f_{16}: Calls Night: # Outgoing	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
f_{17}: Calls Night: Outgoing Duration (min)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
f_{18}: Network Activity: Received (kB/s)	0.13 (0.27)	0.10 (0.16)	0.29 (0.42)	0.11 (0.16)	0.11 (0.26)	0.15 (0.28)
f_{19}: Photos: # Taken	0.65 (1.19)	0.74 (0.99)	1.47 (2.25)	0.85 (1.47)	0.81 (1.25)	0.89 (1.52)
f_{20}: Ringer Mode: % Normal	0.93 (0.10)	0.85 (0.20)	0.77 (0.22)	0.29 (0.22)	0.22 (0.19)	0.63 (0.35)
f_{21}: Ringer Mode: % Silent	0.02 (0.05)	0.06 (0.11)	0.07 (0.11)	0.55 (0.23)	0.07 (0.11)	0.14 (0.23)
f_{22}: Ringer Mode: % Vibrate	0.05 (0.08)	0.09 (0.16)	0.17 (0.18)	0.14 (0.15)	0.71 (0.18)	0.23 (0.29)
f_{23}: Sessions Day: # Sessions	22.36 (14.42)	43.67 (25.09)	47.90 (29.67)	37.11 (22.43)	39.26 (24.08)	36.23 (24.68)
f_{24}: Sessions Day: Duration (min)	73.45 (57.06)	177.74 (152.01)	293.06 (224.24)	162.85 (146.01)	164.43 (118.75)	164.76 (161.32)
f_{25}: Sessions Night: # Sessions	0.97 (1.27)	1.24 (1.57)	3.10 (3.92)	2.69 (4.68)	2.33 (3.97)	2.01 (3.39)
f_{26}: Sessions Night: Duration (min)	2.47 (4.70)	12.75 (39.51)	63.68 (129.82)	26.56 (81.57)	22.34 (68.50)	24.20 (77.89)
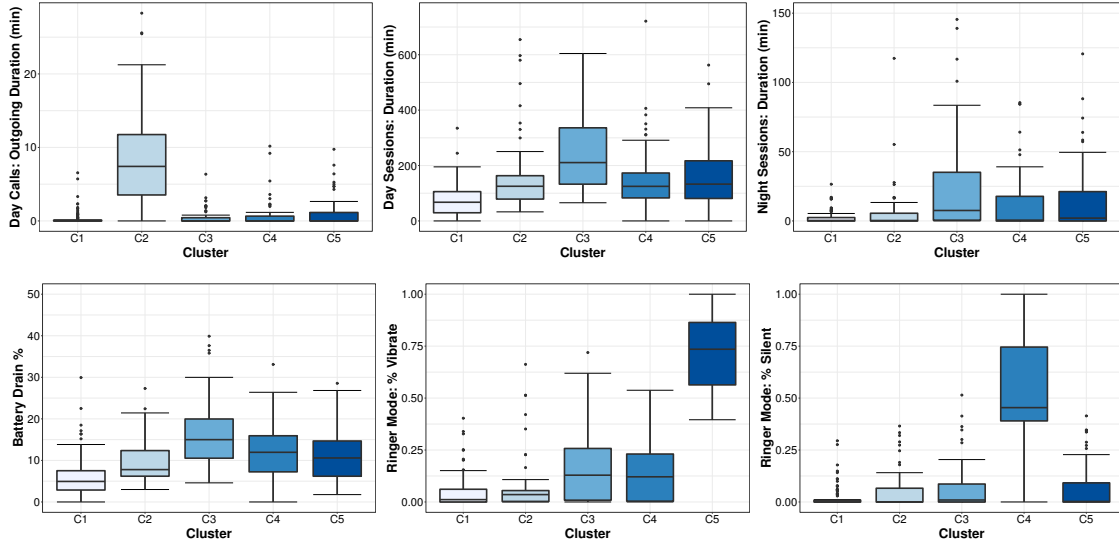
Figure 1. Examples of score distributions of highly discriminatory features across predicted clusters

tionships such as quadratic, cubic, etc. The results indicated whether time is significantly related to the dependent variable.

A potential limitation of the previous model is that it assumes that the relationship between time and the dependent variable is constant for all individuals. An alternative model is one that allows slopes to randomly vary. Therefore, a random slope model was tested by adding the linear effect of time as a random effect. The results of this analysis indicated whether a model that allows the slope between time and the dependent variable to randomly vary fits the data better than a model that fixes the slope to a constant value for all individuals.

The fourth step in developing the level-1 model involved assessing the error structure of the model since significance tests may be dramatically affected if error structures are not properly specified. Therefore, we determined whether a model fit improves by incorporating (a) an autoregressive structure with serial correlations and (b) heterogeneity in the error structures.

Table 4. Summary of the models for the ESM:Tense-Calm scale

	Model 2	Model 3	Model 4	Model 5
Fixed factors				
Intercept	-.59***	-.59***	-.57***	-.57***
Cluster:2	.07	.07	.09	.09
Cluster:3	.00	.00	-.00	-.00
Cluster:4	.11*	.12*	.12*	.13*
Cluster:5	.11*	.12*	.14*	.14*
isNight:Yes	–	-.01	–	-.01
Cluster:2 × isNight:Yes	–	.01	–	.02
Cluster:3 × isNight:Yes	–	-.02	–	-.02
Cluster:4 × isNight:Yes	–	-.06	–	-.06
Cluster:5 × isNight:Yes	–	-.01	–	-.01
WorkingDay:No	–	–	-.05***	-.05***
Cluster:2 × WorkingDay:No	–	–	-.07**	-.07**
Cluster:3 × WorkingDay:No	–	–	.03	.03
Cluster:4 × WorkingDay:No	–	–	-.04	-.04
Cluster:5 × WorkingDay:No	–	–	-.07**	-.07**
Random factors				
Intercept	.35	.35	.35	.35
Residuals	.54	.54	.54	.54

No. Observations: 31396; No. Groups: 340

$^*p < .05;\ ^{**}p < .01;\ ^{***}p < .001$

Table 5. Summary of the models for the ESM:Tired-Awake scale

	Model 2	Model 3	Model 4	Model 5
Fixed factors				
Intercept	.28***	.31***	.26***	.28***
Cluster:2	-.11	-.12	-.11	-.12
Cluster:3	-.05	-.03	-.03	-.01
Cluster:4	-.12	-.10	-.11	-.10
Cluster:5	-.13*	-.12	-.13*	-.12
isNight:Yes	–	-.58***	–	-.58***
Cluster:2 × isNight:Yes	–	.10	–	.09
Cluster:3 × isNight:Yes	–	.03	–	.03
Cluster:4 × isNight:Yes	–	.08	–	.08
Cluster:5 × isNight:Yes	–	.12**	–	.12*
WorkingDay:No	–	–	.09***	.09***
Cluster:2 × WorkingDay:No	–	–	-.00	.00*
Cluster:3 × WorkingDay:No	–	–	-.06**	-.05
Cluster:4 × WorkingDay:No	–	–	-.01	-.01
Cluster:5 × WorkingDay:No	–	–	-.01	.01
Random factors				
Intercept	.39	.39	.39	.39
Residuals	.68	.67	.68	.67

No. Observations: 31373; No. Groups: 340

$^*p < .05;\ ^{**}p < .01;\ ^{***}p < .001$

Table 6. Summary of the models for the ESM:Valence scale

	Model 2	Model 3	Model 4	Model 5
Fixed factors				
Intercept	.74***	.74***	.74***	.74***
Cluster:2	-.02	-.02	-.02	-.03
Cluster:3	-.05	-.05	-.04	-.04
Cluster:4	-.11*	-.10*	-.11*	-.11*
Cluster:5	-.11*	-.11*	-.13*	-.12*
isNight:Yes	–	-.04	–	-.04*
Cluster:2 × isNight:Yes	–	.00	–	.00
Cluster:3 × isNight:Yes	–	.01	–	.01
Cluster:4 × isNight:Yes	–	.01	–	.01
Cluster:5 × isNight:Yes	–	.00	–	.00
WorkingDay:No	–	–	.00	.00
Cluster:2 × WorkingDay:No	–	–	.02	.02
Cluster:3 × WorkingDay:No	–	–	-.01	-.01
Cluster:4 × WorkingDay:No	–	–	.02	.02
Cluster:5 × WorkingDay:No	–	–	.04*	.04*
Random factors				
Intercept	.32	.32	.32	.32
Residuals	.44	.44	.44	.44

No. Observations: 31789; No. Groups: 340

$^*p < .05;\ ^{**}p < .01;\ ^{***}p < .001$

Table 7. Summary of the models for the ESM:Boredom scale

	Model 2	Model 3	Model 4	Model 5
Fixed factors				
Intercept	.09***	.09***	.09***	.08***
Cluster:2	-.00	-.00	-.00	-.00
Cluster:3	.02	.03	.03	.03
Cluster:4	.05*	.05*	.04	.04
Cluster:5	.06**	.06**	.06**	.06**
isNight:Yes	–	.03**	–	.03**
Cluster:2 × isNight:Yes	–	-.02	–	-.03
Cluster:3 × isNight:Yes	–	-.04**	–	-.04**
Cluster:4 × isNight:Yes	–	-.04**	–	-.04**
Cluster:5 × isNight:Yes	–	.00	–	-.00
WorkingDay:No	–	–	.01	.01
Cluster:2 × WorkingDay:No	–	–	.03	.03**
Cluster:3 × WorkingDay:No	–	–	-.00	-.00
Cluster:4 × WorkingDay:No	–	–	.03	.03**
Cluster:5 × WorkingDay:No	–	–	.00	.00
Random factors				
Intercept	.14	.14	.14	.14
Residuals	.26	.26	.26	.26

No. Observations: 32307; No. Groups: 340

$^*p < .05;\ ^{**}p < .01;\ ^{***}p < .001$

For building the level-2 model we considered for each of the growth parameters that contain meaningful variability (*e.g.*, intercept, linear) the appropriate individual difference predictors, as well as their interactions. More specifically, we fitted a series of models for each mood and psychometric scale, while other factors were added as predictors, to assess if the magnitude of each scale is related to the cluster assignment when controlling for the effects of those predictors. Besides the cluster assignment independent variable (Model 2), the models accounted for the time window (day vs. night) of the observations (Model 3), whether its a working day (Model 4), as well as all the aforementioned predictors and their interactions altogether (Model 5).

A PROPOSED TAXONOMY OF MOBILE PHONE USERS

Table 3 shows the means and standard deviations for the features we used to derive the clusters of mobile phone users. As a reference, we also provide the same descriptive statistics for all users in our study.

In this section, we examine the characteristics of each cluster and draw insights on how different user profiles interact with mobile phone technology based on the clusters' structural description (selected examples of cluster variations are shown in Figure 1). Following the findings of our multilevel modelling (presented in Tables 4-9), we also highlight potential implications on the emotional well-being of mobile phone users. Note that we report only statistically significant findings at $p < .05$, $p < .01$ and $p < .001$, for which we have corrected the level of significance using the Holm's [16] method. For data that meet the parametric assumptions we applied the one-way independent ANOVA test, whereas for all other cases we used the Kruskal-Wallis test. When a difference was found to be significant, we followed that finding by applying the independent t-test or the Wilcoxon signed-rank test, respectively.

Table 8. Summary of the models for the PHQ-8 scale				
	C1	**C2**	**C3**	**C4**	**C5**

	C1	**C2**	**C3**	**C4**	**C5**
Fixed factors					
Intercept	5.40***	8.50***	6.85***	10.50***	9.00***
Cluster:1	–	-3.09	-1.44	-5.09**	-3.59*
Cluster:2	3.09	–	1.64	-2.00	-.50
Cluster:3	1.44	-1.64	–	-3.64*	-2.14
Cluster:4	5.09**	2.00	3.64**	–	1.50
Cluster:5	3.59*	.50	2.14	-1.50	–
Random factors					
Intercept	4.11	3.96	4.11	4.11	4.11

No. Observations: 65; No. Groups: 63

$^*p < .05; ^{**}p < .01; ^{***}p < .001$

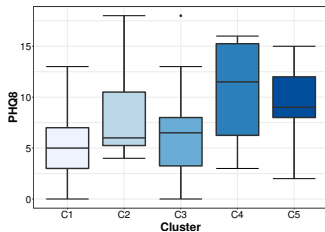**Figure 2. PHQ-8 scores per cluster**

Considering the significant differences across the clusters, we now conclude to a set of five user profiles.

Cluster 1 – Limited Use (Baseline)
99 participants fall into this cluster, with 54.55% of its members being female and 45.45% male. Their age ranges from 18 to 66 ($M = 42.4$, $SD = 10.2$).

In comparison to the other clusters, its members score low in almost all usage categories. They tend to keep their ringer mode in the normal setting, a behaviour that might be related to not being bothered too much by incoming calls and notifications.

Since this cluster shows the lowest amount of phone use, we use it as a baseline for the other clusters to compare the impact that different forms of intensified use have on emotional well-being.

Cluster 2 – Business Use
45 participants fall into the Business Use cluster, where 44.44% of its members are female and 55.56% are male. Their age ranges from 26 to 62 ($M = 43.2$, $SD = 8.7$).

Members of this cluster stand out by their significantly more frequent use of phone calls – incoming and outgoing. While phone calls are comparably frequent, members of this cluster have comparably fewer nightly use sessions and app launches. The ringer mode is typically set to normal, indicating that hearing the phone is important to them. Even though such phone use patterns could be seen in several different user categories, we believe that these patterns are indicative of business use, hence we used such label.

In terms of well-being, members of this cluster did not stand out much. Only during the weekend, they were found to report significantly lower tense arousal. When accounting for night-time and day of the week, we found that during the weekend, members of the cluster reported higher levels of boredom than the baseline cluster.

Table 9. Summary of the models for the Big5: Neuroticism scale				
	C1	**C2**	**C3**	**C4**	**C5**

	C1	**C2**	**C3**	**C4**	**C5**
Fixed factors					
Intercept	10.95***	11.22***	12.14***	9.88***	13.58***
Cluster:1	–	-.27	-1.19	1.06	-2.63
Cluster:2	.27	–	-.92	1.33	-2.35
Cluster:3	1.19	.92	–	2.25*	-1.43
Cluster:4	-1.06	-1.33	-2.25*	–	-3.70**
Cluster:5	2.63**	2.35	1.43	3.70**	–
Random factors					
Intercept	4.22	4.22	4.22	4.22	4.22

No. Observations: 166; No. Groups: 164

$^*p < .05; ^{**}p < .01; ^{***}p < .001$

Cluster 3 – Power Use
67 participants fall into the Power Use cluster, with 56.72% of the members being female and 43.28% male. Their age ranges from 19 to 56 ($M = 37.4$, $SD = 9.2$).

Members of this cluster stand out by their increased session duration & number of nightly sessions, battery use, and mobile data use. During the day, they launched a significantly higher number of email-, game-, and to a lesser extent social media apps. Messaging apps, in contrast, are used less often during the day compared to other clusters. During the night, members of this cluster show the highest use of email apps and a somewhat increase use of messaging apps. While the ringer mode is typically set to normal mode, members of this cluster had the highest variance in ringer mode setting.

Despite the high level of mobile phone use, members of the Power Use cluster do not stand out negatively in any of the well-being related factors. Compared to the baseline, they are more awake during the weekend and report lower levels of boredom during the night. Compared to Cluster 4 mentioned below, they scored lower in terms of depression (PHQ-8) and neuroticism (Big5).

Cluster 4 – Personality-Induced Problematic Phone Use
62 participants fall into this cluster, with 58.06% of its members being female and 41.94% male. Their age ranges from 18 to 62 ($M = 33.1$, $SD = 11.9$).

Members of this cluster stand out by an increased number and length of sessions during night time, and an increased use of email- and messaging apps during the night. The ringer is typically set to silent mode. Members of this cluster scored worst in terms of well-being. Compared to the baseline cluster, they reported significantly higher levels of tense-arousal, boredom, and lower valence. In contrast, when accounting for night-time and non-working days, the significant differences regarding tense-arousal and valence disappeared. Bored, however, was significantly lower during night-time. These findings indicate a tendency towards experiencing more stress and boredom during working hours. Members of this cluster further tended to be more neurotic / less emotionally stable than members of other clusters. Finally, members of this cluster scored significantly higher on the PHQ-8 questionnaire than Limited and Power Users, indicating a tendency towards experiencing depression-related symptoms.

Cluster 5 – Externally-Induced Problematic Phone Use
67 participants fall into this cluster, with 49.25% of its members being female and 50.75% male. Their age ranges from 18 to 58 ($M = 32.4$, $SD = 10.0$).

Like the previous cluster, members of this cluster tend to have more and longer phone use sessions during night-time. The main difference to Cluster 4 is that during night-time, only the use of messaging apps is comparably higher. In contrast, the use of email apps is lower. Finally, the ringer mode of these users is typically set to vibrate.

There were significant effects of membership of this cluster on the emotional self-reports. Compared to the baseline cluster, its members reported significantly higher tense arousal, lower energetic arousal, lower valence, and higher levels of boredom. During night-time, however, energetic arousal was significantly lower – an indication of being more tired – and the other effects subsided. During the weekend, tense arousal was significantly lower, valence was significantly higher, and significant effects on energetic arousal and boredom disappeared.

Also, members of this cluster scored significantly higher in terms of emotional stability compared to the baseline cluster as well as Cluster 4. Further, PHQ-8 depression scores (Figure 2) were significantly higher than those of the baseline cluster, however, the effect was not as pronounced as in Cluster 4.

While members of this cluster tend to be stressed during working time, they seem to better compensate during non-working hours: they are being tired during the night and happy during the weekend. This finding is corroborated by higher emotional stability. We interpret the main difference that members of this cluster have a stressful daytime, but are more affected by external factors rather than internal factors, and therefore better cope with the stressful weekdays.

DISCUSSION
We identified five clusters of phone use: Limited Use, Business Use, Power Use, Personality-Induced Problematic Phone Use, and Externally-Induced Problematic Phone Use. We used the Limited Use cluster as baseline to compare the other clusters against. Business Users tend to use the mobile phone as phone primarily. Power Users use extensively the smartphone capabilities. Personality-Induced Problematic Phone Use and Externally-Induced Problematic Phone Use both stand out by their night sessions and not having the ringer in normal mode.

Heavy phone use does not predict negative well-being
Surprisingly, we found that extensive use of the phone alone did not predict negative emotional well-being. Neither Business Users nor Power Users, who scored highest in phone calling and session duration, respectively, scored negatively in terms of affect or personality. Only members of the Clusters 4 and 5 – who were using the phone overall less than Power Users – scored significantly higher on the PHQ-8 depression scale. Thus, rather than using the phone a lot, it's comparably frequent and long night-time use that predicts negative effects on emotional states. Members of the Clusters 4 and 5 – those

with more and longer night-time sessions – felt tenser, worse, and more bored, in particular during working days.

This finding contrasts with previous work. For example, Thomee *et al.* [40] found that high mobile phone use was associated with sleep disturbances and symptoms of depression in Sweden. Yen *et al.* [44] also reported an association of high mobile phone use, outgoing phone calls and text messages with depression. However, our results indicate that nightly sessions are better predictors rather than overall use intensity.

Difference between clusters with negative well-being
The main difference between the Clusters 4 and 5 of people who struggle with negative emotional well-being is that negative effects are more focused onto working hours, but tend to subside or even reverse during night-time and weekend. The main difference to Cluster 4 in the phone use habits is that they tend to keep the phone in vibration mode, and that while they use it during night-time, they avoid doing email, which we consider work-related. We hypothesize that members of Cluster 5 seem to find it easier to disconnect from stressful working days. This is corroborated by the comparably high scores in emotional stability. The use of the vibration mode indicates that the stress might rather be induced by external factors (*e.g.* stressful job) rather than internal factors (*e.g.* emotional instability).

Limitations
One limitation of our study is its focus on Android phone users. Even though this results in the large majority (~90%) of users in the country of study, there is a small percentage of users that was not included. Moreover, this study focuses on specific emotional scales to reflect negative well-being such as PHQ-8. More scales that measures anxiety, addiction or boredom should be examined in future extension of this work.

CONCLUSIONS
On the basis of phone use logs from 340 Android users, we identified five types of typical phone use: limited use, business use, power use, and personality- & externally induced problematic use. While previous work largely predicted that high intensity phone use predicts negative outcomes, membership of the Power Use cluster did not predict negative well-being. Instead, members of the two clusters associated with negative well-being had only average phone use. Moreover, nightly use of messaging and having the ringer mode in a setting other than normal mode is what set the two problematic-use clusters apart from the other three clusters. This finding indicates that instead of vilifying phone use in general, we need awareness that the *type* of phone use might be a much stronger indicator than the overall phone-use intensity. This highlights the need for a more nuanced approach to studying and understanding the underlying mental problems, and not let ourselves fall into the old moral "new technology is bad" panic[3]. Future work should attempt to replicate / falsify our findings that nightly use rather than overall use is a predictor of negative well-being. Finally, further future work should explore why the ringer mode is such a strong indicator of well-being.

---

[3]https://www.engadget.com/2018/02/09/new-tech-addictions-are-mostly-just-old-moral-panic/

REFERENCES

1. Sally Andrews, David A. Ellis, Heather Shaw, and Lukasz Piwek. 2015. Beyond Self-Report: Tools to Compare Estimated and Real-World Smartphone Use. *PLOS ONE* 10, 10 (10 2015), 1–9. DOI: `http://dx.doi.org/10.1371/journal.pone.0139004`

2. Marta Beranuy, Ursula Oberst, Xavier Carbonell, and Ander Chamarro. 2009. Problematic Internet and Mobile Phone Use and Clinical Symptoms in College Students: The Role of Emotional Intelligence. *Comput. Hum. Behav.* 25, 5 (Sept. 2009), 1182–1187. DOI: `http://dx.doi.org/10.1016/j.chb.2009.03.001`

3. Adriana Bianchi and James G Phillips. 2005. Psychological predictors of problem mobile phone use. *CyberPsychology & Behavior* 8, 1 (2005), 39–51.

4. Joel Billieux. 2012. *Current Psychiatry Reviews* 8, 4 (2012), 299–307.

5. Joël Billieux, Pierre Philippot, Cécile Schmid, Pierre Maurage, Jan De Mol, and Martial Van der Linden. 2015. Is dysfunctional use of the mobile phone a behavioural addiction? confronting symptom-based versus process-based approaches. *Clinical psychology & psychotherapy* 22, 5 (2015), 460–468.

6. Joël Billieux, Martial Van der Linden, and Lucien Rochat. 2008. The role of impulsivity in actual and problematic use of the mobile phone. *Applied Cognitive Psychology* 22, 9 (2008), 1195–1210.

7. Paul Bliese. 2017. *Multilevel Modeling in R (v2.6)*. `https://cran.r-project.org/doc/contrib/Bliese_Multilevel.pdf`

8. P. D. Bliese and R. E. Ployhart. 2002. Growth modeling using random coefficient models: Model building, testing, and illustrations. *Organizational Research Methods* 5, 4 (2002), 362–387.

9. P. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein. 2005. A tutorial on the cross-entropy method. *Annals of Operations Research* 134, 1 (2005), 19–67.

10. Trinh-Minh-Tri Do and Daniel Gatica-Perez. 2010. By Their Apps You Shall Understand Them: Mining Large-scale Patterns of Mobile Phone Usage. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia (MUM '10)*. ACM, New York, NY, USA, Article 27, 10 pages. DOI: `http://dx.doi.org/10.1145/1899475.1899502`

11. C. Dwork, R. Kumar, M. Naor, and D. Sivakumar. 2001. Rank aggregation methods for the Web. In *Proc. IW3C2 Intl. Conf. on World Wide Web*. 613–622.

12. Alexandra Ehrenberg, Suzanna Juckes, Katherine M White, and Shari P Walsh. 2008. Personality and self-esteem as predictors of young people's technology use. *Cyberpsychology & behavior* 11, 6 (2008), 739–741.

13. Ingrid Nesdal Fossum, Linn Tinnesand Nordnes, Sunniva Straume Storemark, Bjørn Bjorvatn, and Ståle Pallesen. 2014. The Association Between Use of Electronic Media in Bed Before Going to Sleep and Insomnia Symptoms, Daytime Sleepiness, Morningness, and Chronotype. *Behavioral Sleep Medicine* 12, 5 (2014), 343–357.

14. Maria Halkidi, Yannis Batistakis, and Michalis Vazirgiannis. 2001. On Clustering Validation Techniques. *Journal of Intelligent Information Systems* 17, 2-3 (Dec. 2001), 107–145.

15. Daniel Hintze, Philipp Hintze, Rainhard D. Findling, and René Mayrhofer. 2017. A Large-Scale, Long-Term Analysis of Mobile Device Usage Characteristics. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 2, Article 13 (June 2017), 21 pages. DOI: `http://dx.doi.org/10.1145/3090078`

16. Sture Holm. 1979. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* 6, 2 (1979), 65–70. `http://www.jstor.org/stable/4615733`

17. Fu-Yuan Hong, Shao-I. Chiu, and Der-Hsiang Huang. 2012. A model of the relationship between psychological characteristics, mobile phone addiction and use of mobile phones by Taiwanese university female students. *Computers in Human Behavior* 28, 6 (2012), 2152 – 2159. DOI:`http://dx.doi.org/10.1016/j.chb.2012.06.020`

18. Cristina Jenaro, Noelia Flores, Marìa Gómez-Vela, Francisca González-Gil, and Cristina Caballo. 2007. Problematic internet and cell-phone use: Psychological, behavioral, and health correlates. *Addiction Research & Theory* 15, 3 (2007), 309–320.

19. Oliver P John and Sanjay Srivastava. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* 2, 1999 (1999), 102–138.

20. D. Kim, Y. Jung, J. Lee, M. Kim, Y. Lee, E. Kang, C. Keum, and J. Nam. 2012. Development of Smartphone Addiction Proneness Scale for Adults: Self-Report. *The Korean Journal of Counseling* (2012).

21. Sean D. Kristjansson, John C. Kircher, and Andrea K Webb. 2007. Multilevel models for repeated measures research designs in psychophysiology: An introduction to growth curve modeling. *Psychophysiology* 44, 5 (2007), 728–736. DOI: `http://dx.doi.org/10.1111/j.1469-8986.2007.00544.x`

22. Kurt Kroenke, Tara W Strine, Robert L Spitzer, Janet BW Williams, Joyce T Berry, and Ali H Mokdad. 2009. The PHQ-8 as a measure of current depression in the general population. *Journal of affective disorders* 114, 1 (2009), 163–173.

23. Min Kwon, Joon-Yeop Lee, Wang-Youn Won, Jae-Woo Park, Jung-Ah Min, Changtae Hahn, Xinyu Gu, Ji-Hye Choi, and Dai-Jin Kim. 2013. Development and Validation of a Smartphone Addiction Scale (SAS). *PLOS ONE* 8, 2 (02 2013), 1–7. DOI: `http://dx.doi.org/10.1371/journal.pone.0056936`

24. Uichin Lee, Joonwon Lee, Minsam Ko, Changhun Lee, Yuhwan Kim, Subin Yang, Koji Yatani, Gahgene Gweon, Kyong-Mee Chung, and Junehwa Song. 2014. Hooked on Smartphones: An Exploratory Study on Smartphone Overuse Among College Students. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2327–2336. DOI: `http://dx.doi.org/10.1145/2556288.2557366`

25. Andrew Lepp, Jacob E. Barkley, and Aryn C. Karpinski. 2014. The relationship between cell phone use, academic performance, anxiety, and Satisfaction with Life in college students. *Computers in Human Behavior* 31, Supplement C (2014), 343 – 350. DOI: `http://dx.doi.org/10.1016/j.chb.2013.10.049`

26. Akhil Mathur, Lakshmi Manasa Kalanadhabhatta, Rahul Majethia, and Fahim Kawsar. 2017. Moving Beyond Market Research: Demystifying Smartphone User Behavior in India. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 82 (Sept. 2017), 27 pages. DOI:`http://dx.doi.org/10.1145/3130947`

27. Lisa J Merlo, Amanda M Stone, and Alex Bibbey. 2013. Measuring problematic mobile phone use: development and preliminary psychometric properties of the PUMP scale. *Journal of addiction* 2013 (2013).

28. Antti Oulasvirta, Tye Rattenbury, Lingyi Ma, and Eeva Raita. 2012. Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing* 16, 1 (01 Jan 2012), 105–114. DOI: `http://dx.doi.org/10.1007/s00779-011-0412-2`

29. Bong-Won Park and Kun Chang Lee. 2011. *The Effect of Users' Characteristics and Experiential Factors on the Compulsive Usage of the Smartphone*. Springer Berlin Heidelberg, Berlin, Heidelberg, 438–446. DOI: `http://dx.doi.org/10.1007/978-3-642-20998-7_52`

30. Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When Attention is Not Scarce - Detecting Boredom from Mobile Phone Usage. In *Proc. UbiComp '15 (UbiComp '15)*. ACM, 825–836. DOI: `http://dx.doi.org/10.1145/2750858.2804252`

31. V. Pihur, S. Datta, and S. Datta. 2007. Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach. *Bioinformatics* 23, 13 (2007), 1607–1615.

32. Jose Pinheiro, Douglas Bates, Saikat DebRoy, Deepayan Sarkar, and R Core Team. 2017. *nlme: Linear and Nonlinear Mixed Effects Models (v3.1-131)*. `https://CRAN.R-project.org/package=nlme`

33. S. S. Ravi, D. J. Rosenkrantz, and G. K. Tavyi. 1994. Heuristic and Special Case Algorithms for Dispersion Problems. *Operations Research* 42, 2 (1994), 299–310.

34. James A. Roberts, Chris Pullig, and Chris Manolis. 2015. I need my smartphone: A hierarchical model of personality and cell-phone addiction. *Personality and Individual Differences* 79, Supplement C (2015), 13 – 19. DOI:`http://dx.doi.org/10.1016/j.paid.2015.01.049`

35. R. Y. Rubinstein and D. P. Kroese. 2005. *The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning*. Springer.

36. Mohammad Salehan and Arash Negahban. 2013. Social networking on smartphones: When mobile phones become addictive. *Computers in Human Behavior* 29, 6 (2013), 2632 – 2639. DOI: `http://dx.doi.org/10.1016/j.chb.2013.07.003`

37. Maya Samaha and Nazir S. Hawi. 2016. Relationships Among Smartphone Addiction, Stress, Academic Performance, and Satisfaction with Life. *Comput. Hum. Behav.* 57, C (April 2016), 321–325. DOI: `http://dx.doi.org/10.1016/j.chb.2015.12.045`

38. Florence Samkange-Zeeb, Gabriele Berg, and Maria Blettner. 2004. Validation of self-reported cellular phone use. *Journal of Exposure Science and Environmental Epidemiology* 14, 3 (2004), 245.

39. Choonsung Shin and Anind K. Dey. 2013. Automatically Detecting Problematic Use of Smartphones. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13)*. ACM, New York, NY, USA, 335–344. DOI:`http://dx.doi.org/10.1145/2493432.2493443`

40. Sara Thomée, Annika Härenstam, and Mats Hagberg. 2011. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults - a prospective cohort study. *BMC Public Health* 11, 1 (31 Jan 2011), 66. DOI: `http://dx.doi.org/10.1186/1471-2458-11-66`

41. Susmita Datta Vasyl Pihur, Somnath Datta. 2017. *RankAggreg: Weighted Rank Aggregation (v0.6.2)*. `https://cran.r-project.org/web/packages/RankAggreg/RankAggreg.pdf`

42. Aku Visuri, Niels van Berkel, Chu Luo, Jorge Goncalves, Denzil Ferreira, and Vassilis Kostakos. 2017. Predicting Interruptibility for Manual Data Collection: A Cluster-based User Model. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '17)*. ACM, New York, NY, USA, Article 12, 14 pages. DOI: `http://dx.doi.org/10.1145/3098279.3098532`

43. Eric B Weiser. 2015. # Me: Narcissism and its facets as predictors of selfie-posting frequency. *Personality and Individual Differences* 86 (2015), 477–481.

44. Cheng-Fang Yen, Tze-Chun Tang, Ju-Yu Yen, Huang-Chi Lin, Chi-Fen Huang, Shu-Chun Liu, and Chih-Hung Ko. 2009. Symptoms of problematic cellular phone use, functional impairment and its association with depression among adolescents in Southern Taiwan. *Journal of adolescence* 32, 4 (2009), 863–873.

45. Sha Zhao, Julian Ramos, Jianrong Tao, Ziwen Jiang, Shijian Li, Zhaohui Wu, Gang Pan, and Anind K. Dey. 2016. Discovering Different Kinds of Smartphone Users Through Their Application Usage Behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 498–509. DOI:http://dx.doi.org/10.1145/2971648.2971696

46. Marvin Zuckerman. 1979. Sensation seeking: Beyond the optimal level of arousal. 1979. (1979).