MyTraces: Investigating Correlation and Causation between Users' Emotional States and Mobile Phone Interaction

ABHINAV MEHROTRA, University College London, UK
FANI TSAPELI, University of Birmingham, UK
ROBERT HENDLEY, University of Birmingham, UK
MIRCO MUSOLESI, University College London and The Alan Turing Institute, UK

Most of the existing work concerning the analysis of emotional states and mobile phone interaction has been based on correlation analysis. In this paper, for the first time, we carry out a causality study to investigate the causal links between users' emotional states and their interaction with mobile phones, which could provide valuable information to practitioners and researchers. The analysis is based on a dataset collected *in-the-wild*. We recorded 5,118 mood reports from 28 users over a period of 20 days.

Our results show that users' emotions have a causal impact on different aspects of mobile phone interaction. On the other hand, we can observe a causal impact of the use of specific applications, reflecting the external users' context, such as socializing and traveling, on happiness and stress level. This study has profound implications for the design of interactive mobile systems since it identifies the dimensions that have causal effects on users' interaction with mobile phones and vice versa. These findings might lead to the design of more effective computing systems and services that rely on the analysis of the emotional state of users, for example for marketing and digital health applications.

CCS Concepts: •**Human-centered computing → Empirical studies in ubiquitous and mobile computing;** *HCI design and evaluation methods;* •**Computing methodologies** → *Causal reasoning and diagnostics;*

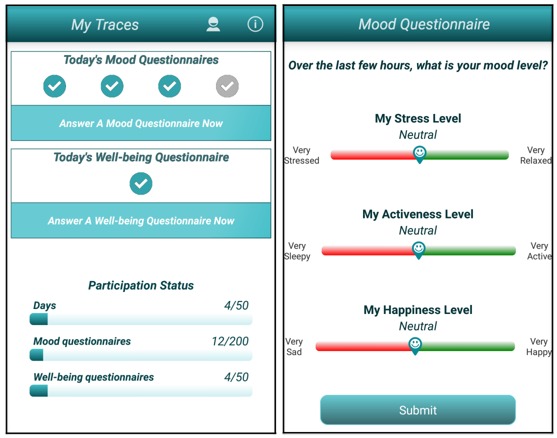
Additional Key Words and Phrases: Mobile Sensing, Causality Analysis.

ACM Reference format:
Abhinav Mehrotra, Fani Tsapeli, Robert Hendley, and Mirco Musolesi. 2017. MyTraces: Investigating Correlation and Causation between Users' Emotional States and Mobile Phone Interaction. *PACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 3, Article 83 (September 2017), 21 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Smartphones are part of our everyday lives and it is not surprising that, in the past years, researchers and practitioners have investigated several aspects of our interaction with these devices. Examples include the characterization of application usage behavior [18], users' attentiveness and receptivity to notifications [28] and mobile communication (via calls and SMSs) patterns [42]. However, until now, only a handful of studies have investigated the relationship between users' mood and their mobile phone interaction behavior [7, 22, 29, 34]. Indeed, going beyond the study of physical interactions with smartphones and exploring the *emotional* interactions with them is a fascinating emerging area in ubiquitous computing. Previous studies, such as [22], have demonstrated that communication and application usage patterns can be exploited to statistically infer the daily mood of a user. Another study [7] investigated the correlation between the application usage patterns and users' mood, sleep and irritability states. Some studies [29, 34] have also shown that users' cognitive states

2017. 2474-9567/2017/9-ART83 $15.00
DOI: 10.1145/nnnnnnn.nnnnnnn



(a) (b)

Fig. 1. MyTraces application: (a) main screen, (b) mood questionnaire.

(such as feeling bored and engagement with other tasks) correlate with their receptivity to information delivered through mobile phones.

However, our goal is to go beyond "simple" correlation studies and try to extract and quantify *causation relationships*. To the best of our knowledge, only [54] has considered the problem of causality using mobile sensor data, in particular to study the impact of physical activities (such as sitting, walking and so on) on stress levels. However, the authors of [54] investigate only these high-level activities and not day-to-day (micro-)interactions with both phone and notifications, such as click patterns, reactions to notifications and so on.

More generally, detecting causal links between users' emotional states and their interaction with mobile phones could provide valuable information to developers for designing more effective "emotion-aware" applications and to social scientists for studying certain aspects of human behavior. For example, tracking users' communication through phone calls or text messages as well as social media applications enables practitioners and researchers to understand whether remote communication has an impact on users' emotional state or whether people change their communication patterns according to their emotional state. In addition, understanding whether users' emotional state influences their receptivity to mobile phone notifications could allow the estimation of an optimal time for delivering certain type of information. This is of key importance for marketing [2] and digital health [5, 6] applications. Moreover, the results of this work might be used for developing more effective positive behavior interventions based on mobile phones [19]. In general, the interaction with mobile phones becomes, in a sense, a source of *secondary signals* for quantifying the user's emotional states.

In this paper, to the best of our knowledge, we present the first causality study[1] concerning the user's emotional states and mobile phone interaction behavior. In order to carry out our investigation, we designed and developed an application called MyTraces (Figure 1) that uses an experience sampling method (ESM) approach to collect users' emotional state levels reported by them during different times of the day and continuously logs their interaction with mobile phones. More specifically, the application collects information about three emotional

[1]The causal analysis presented in this paper shows the potential causal effect in the absence of other confounders and, in any case, it provides evidence of strong dependency between variables. Our purpose is to examine the dependencies among the examined observable factors that are expression of the underlying causes that cannot be quantified directly (e.g., socialization could be the underlying cause of happiness rather than the use of a social app).

states including activeness, happiness and stress levels on a 5-point Likert scale as well as different aspects of phone interaction including not only application usage and communication patterns as in the existing literature, but also, for the first time, micro-interactions with phone and notifications, which are used to derive a variety of metrics.

The key contributions of this paper can be summarized as follows:

- For the first time we examine how phone usage and users' interaction with notifications associates with emotional states using a dataset collected in-the-wild.
- We present the first in-depth causality analysis that investigates a series of causal relationships between users' emotional states and mobile phone interaction.
- We discuss the lessons learnt from this causality analysis including its inherent limitations and a series of questions in this field for the research community that, in our opinion, have been opened by this study.

We believe that this work has profound implications for the design of interactive mobile systems. In fact, for the first time, this paper goes beyond the investigation of simple associations between users' interaction and their emotional states, and therefore their experience. In this work we provide a first initial characterization of the causal links between users' emotional states and their interaction with mobile devices. Indeed, these findings can be applied to the design of more effective interactive applications considering also the emotional states of the users.

2 OUR APPROACH

2.1 Definition of User Behavior Metrics

In this section we introduce a series of metrics that are derived by quantifying users' emotional states and their interaction with mobile phones. These metrics represent the basis of our correlation and causality analysis that we will present in the following section. Some metrics are indeed classic indicators widely used for this class of studies in the ubiquitous computing community [7, 22], while others, such as the metrics related to phone usage in terms of notification and screen interaction, are introduced for the first time in this work.

2.1.1 Emotional States. Most of the previous studies [22, 37] have considered the Circumplex mood model with two dimensions namely, valence and arousal [38]. An alternative model was presented by Schimmack and Rainer in [43]. According to their proposal the arousal state can be split into two dimensions: tense arousal and energetic arousal. The authors justify this split with the fact that the energetic arousal is influenced by a circadian rhythm (i.e., it corresponds to activity in brain cells that regulate organisms' sleep-wake cycle), whereas tense arousal does not show a similar circadian rhythm. Therefore, in our study we split "arousal" into tense arousal (stressed-relaxed) and energetic arousal (sleepy-active).

Consequently, we consider three aspects of emotional states that are captured during the day:

- **activeness level**: a state of being aroused and physiological readiness to respond [35, 51];
- **happiness level**: a state of positiveness and joy that is derived from external and momentary pleasures [44];
- **stress level**: a negative state of being under high mental pressure [45].

The levels of these emotional states are computed on a 5 point-based Likert scale, where 1 indicates the lowest level and 5 the highest level.

2.1.2 Phone Interaction Metrics. We now describe how we compute four different types of metrics capturing users' interaction with their mobile phones in terms of notification, phone usage, application usage and communication patterns. In order to compute these metrics we rely on three classes of data (described in Table 2):

Group	Metric	Description
Notification	Count	Total number of notifications clicked.
	Acceptance %	Percentage of notifications clicked out of total arrived.
	% Handled (Other Device)	Percentage of notifications that are not handled on phone out of total notifications arrived.
	Average Seen Time (ST)	Average of the seen time of all notifications.
	Average Decision Time (DT)	Average of the decision time of all notifications.
	Average Response Time (RT)	Average of the response time of all notifications.
Phone Usage	All App Launch Count	Number of times applications are launched.
	All App Unique Count	Number of applications launched.
	All App Usage Time	Time duration for which applications were used.
	Sig App Launch Count	Number of times significant applications are launched.
	Sig App Unique Count	Number of significant applications launched.
	Sig App Usage Time	Time duration for which applications were used.
	Non-Sig Launch Count	Number of times non-significant applications are launched.
	Non-Sig Unique App Count	Number of non-significant applications launched.
	Non-Sig App Usage Time	Time duration for which applications were used.
	Phone Usage Time	Time duration for which phone was used.
	Single Click Count	Number of single clicks on the phone screen.
	Long Click Count	Number of long clicks on the phone screen.
	Unlock Count	Number of times the phone was unlocked.
Application Usage	Launch Count	Number of times applications are launched.
	Usage Time	Time duration for which applications were used.
Communication	Call Count	Number of calls.
	Call Unique Count	Number of calls to unique contacts.
	Call Average Duration	Average of the durations of all calls.
	SMS Count	Number of SMSs sent.
	SMS Unique Count	Number of SMSs sent to unique contacts.
	SMS Average Length	Average body length of all SMSs sent.
	SMS Sent to Received Ratio	Ratio of sent to received SMSs counts.

Table 1. Description of phone interaction metrics.

notification, phone usage and communication. All metrics (see Table 1) are computed for each user on an hourly basis for all days.

Phone Usage Metrics. We derive 3 metrics by using the phone usage data for representing the information about the user's application usage behavior: *App Launch Count*, *App Unique Count* and *App Usage Time*. We compute these metrics for all applications as well as significant and non-significant applications. Here, significant applications refer to regularly used applications. In order to identify such applications, we compute the average launch count for all applications and the applications which are launch more frequently than this average are considered as significant applications. All applications that do not fall into the category of significant applications (i.e., applications that are not used regularly) are labelled as non-significant applications. Note that this separation might provide us with some interesting insights about the use of rarely used applications (i.e., non-significant applications) in specific situations.

Moreover, we compute 4 additional metrics that represent the basic information about the user's interaction with the phone: *Phone Usage Time*, *Single Click Count*, *Long Click Count* and *Unlock Count*. It is worth noting that the "phone usage time" metric indicates the overall time period during which the user was engaged with the phone including all applications and the home screen.

Notification Metrics. We use the notification and phone usage data to compute six metrics that represent aggregate information about the user's receptivity and attentiveness to notifications. For three of these metrics (*Average Seen Time*, *Average Decision Time* and *Average Response Time*) we use three time-based terms: *seen, decision and response time*.

Here, *seen time* is the time period from the notification arrival until the time the notification was seen by the user. *Decision time* is the time period from the moment a user saw a notification until the time they acted upon it (by clicking, launching its corresponding app or swiping to dismiss). *Response time* is the sum of the seen and decision time periods (i.e., the overall time from the notification arrival until the moment the notification was acted upon).

Moreover, one of the metric "*% Handled (Other Device)*" represents the information about the user's engagement with other devices. In order to infer whether a notification is handled or not (i.e., handled on some other device), we assume that a notification is automatically removed from the notification bar of the phone if it was delivered on some other device and the user has already interacted with it on that device.

Application Usage Metrics. In order to investigate users' application usage behavior for different apps, we divide them using the categories defined at the Google Play store [3]. The nine categories include business, communication, game, lifestyle, music, productivity, social, tool and travel applications. Finally, we compute two metrics (*Launch Count* and *Usage Time*) for each of the nine application categories.

Communication Metrics. We use the call logs from the communication data to extract the information about the user's calling behavior. We compute 3 call-based metrics (*Call Count*, *Call Unique Count* and *Call Average Duration*) for the incoming, outgoing, missed and rejected calls. Note that we could not compute the average time duration for missed and rejected calls because such calls always have time duration equal to zero.

Moreover, we compute 4 SMS-based metrics (*SMS Count*, *SMS Unique Count*, *SMS Average Length* and *SMS Sent to Received Ratio*) only for the messages that are sent by the user. We do not compute these metrics for incoming SMSs because mobile phones do not provide the user with any control for handling such incoming messages. Thus, a user's behavior could not be captured by any feature of the incoming SMSs.

2.1.3 Context-based Metrics. We use the context data to compute two metrics: (i) duration of the interval during which the user performs different activities; (ii) duration of the interval the user spends at different places. We compute both metrics on an hourly basis for each day. Moreover, for each hour of a day we also capture weather metrics including temperature and humidity. It is worth noting that these metrics are used as confounding variables for the causality analysis as discussed in the next section.

2.2 Correlation Analysis

In this section we describe the methodology we followed in order to study the relationships between emotional states (i.e., activeness, happiness and stress) and phone interactions. In order to quantify this association, we compute the individual-based Kendall's rank correlation coefficients. We consider the absolute values of these coefficients because we are interested in the strength of the relationships between the variables. We then compute the average of these coefficient values. We rely on Fisher's method [16] for combining the p-values of individual-based correlation analyses.

The correlation analysis is performed between the emotional state at the current hour (i.e., hour of the day in which the information about the user's emotion is acquired) and the values of the examined metrics for three different time intervals: *preceding, current* and *next hour*. Consequently, the final number of data samples in our analysis is equal to the total number of emotional state reports provided by the participants.

It is worth noting that a large body of psychological studies have shown that emotions persist for a few hours [8, 9]. Therefore, in our study we ask users to report their emotional states in the previous hours (see Figure 1.b). More specifically, these questionnaires were administered four times a day at intervals of three hours. However, for our analysis we consider emotional states in an hour long period since we use the behavioral metrics of previous and next hours and using a longer periods of emotional states would introduce overlaps of the information given the fact that the emotions are reported every three hours. We discuss the data collection process in detail in the next section.

The correlation results are presented as a correlation matrix plot. In this matrix the y-axis indicates the phone interaction metrics and the x-axis indicates the type of emotions that are correlated with the metric computed for the specific hour. Here, the *hour* is represented by the numeric value on the x-axis labels. For instance, in Figure 3 the box in the first row (*Acceptance Percentage*) and the first column (*-1 Activeness*) presents the coefficient for the correlation of the activeness level with the acceptance percentage of notifications, computed by using data related to the $current - 1$ hour. Here, the current hour refers to the hour in which a user reported their emotional state. We set the significance level α for the correlation results to 0.05 and non-significant correlation coefficients are indicated by the white boxes in the correlation plots.

2.3 Causality Analysis

2.3.1 Overview. Correlation analysis reveals the relationship between emotional state and the phone interaction metrics. However, a pure correlation between two variables does not necessarily imply the existence of a causal influence as the values of both the examined variables may be associated with other factors, i.e., they can be "explained" by other factors. For instance, users commuting a long distance may report reduced activeness level and they may also spend most of their time playing games with the smartphone. In this case, a correlation between activeness level and game applications may be observed. However, the values of both variables are strongly influenced by user activity (i.e., commuting in vehicle). Therefore, we perform the causality analysis between the variables that are significantly correlated.

In order to be able to claim that an observed relationship between variables X and Y (i.e., X influences Y) is causal, the variable X should always temporally precede Y and there should be no other explanation for the association observed between them. Here, X is called the *treatment* variable and Y is the *outcome* variable. In case there is a third variable Z (called *confounding* variable), which influences both X and Y, the observed association between X and Y might be spurious and attributed solely to Z. Thus, in order to conduct a valid causal inference analysis it is necessary to control any confounding variables.

According to Rubin's framework, causal inference analysis can be conducted by comparing potential outcomes [40]. To understand this, let us consider U as a set of units (e.g., the samples of the variable) that are denoted by u. Where u has been exposed to the treatments $X \in [0, 1]$ to give the output values as $Y0(u)$ and $Y1(u)$ respectively. Now, the average effect of the treatment for all units $u \in U$ can be estimated as $E\{Y_1(u) - Y_0(u)\}$.

However, the fundamental problem of causal inference is that we cannot observe both Y_0 and Y_1 for the same unit. Therefore, we use the widely used matching design approach [39, 47] according to which the impact of a treatment variable X on an outcome variable Y can be assessed by comparing samples or units with similar values of Z (i.e., the observed confounding variables). For example, if we want to assess the impact of a treatment X on the stress level (i.e., Y) and we consider the time that they spend at home as a confounding variable (i.e., Z), we can compare their stress levels only for the observations when the users have spent similar amount of time at home (i.e., the distance between Z values is close to zero). Here, the similarity between Z values can be computed by using any distance measure such as the Euclidean distance.

The purpose of using the matching design is to find the "most similar" pairs of units and then compute the treatment effect. To understand this, let us denote with U and V the sets of units that have received a treatment 0 and 1, respectively. Matching design attempts to find an optimal set of paired units $(u, v) \in P$, with $u \in U$ and $v \in V$ such that the distance between paired units on their confounding variables is minimum. Then, the average treatment effect (ATE) is estimated as follows:

$$E\{Y_0(u) - Y_1(v)\}_{(u,v)\in P} \tag{1}$$

Note that several approaches for finding optimal pairs of units have been proposed [50]. A detailed review of such approaches is out of the scope of this study.

Although the treatment variable is generally supposed to be binary, the traditional matching framework has been extended to support also continuous values for treatment variables [23]. In this case, units cannot be split into treated and control groups; instead, every unit of the study can be matched to any other. Then, the objective of the applied matching method is to match units with minimum difference of their confounding variables and maximum difference of their treatment values. After creating a set P of matched units, the average treatment effect is estimated as:

$$E\left\{\frac{Y(u) - Y(v)}{X(u) - X(v)}\right\}_{(u,v)\in P} \tag{2}$$

In this study, we will apply the matching design framework for continuous treatments [23] in order to analyze:

(1) the impact of phone usage on participants' emotional state. In this case, the treatment variable is one of the phone interaction metrics presented in Table 1 and the outcome variable is one of the three emotional state-based metrics. *Since a causal link cannot exist between two uncorrelated variables and the treatment needs to precede temporally the outcome variable, we conduct this analysis only on variables for which a statistically significant correlation between emotional state and the previous hour's phone interaction metric has been observed.*

(2) the impact of participants' emotional state on their interaction with their smartphone. In this case, the treatment variable is one of three emotional state-based metrics and the outcome variable is one of the phone interaction metrics. *Similarly, we conduct this analysis only on variables for which a statistically significant correlation between emotional state and next hour's phone interaction metric has been observed.*

A separate causality study is conducted for each pair of variables. It is worth noting that distinguishing cause from effect is an open issue in causality analysis. With respect to this specific type of study, understanding whether the reported mood precedes temporally the examined activity is very hard. For this reason, in our work, we use the last reported emotional state as confounding variable. Thus, this allows us to test if the emotional state of the individual prior to the observed activity influences the outcome variable (for example phone usage) or not. However, the emotional state may change after being reported and before the observed activity. Unfortunately, in such cases the temporal precedence cannot be captured. At the same time, as previous studies suggest, the emotions do not fluctuate frequently [9], therefore we believe that this issue will not significantly impact our study. Finally, given the limited amount of data per individual, it is only possible to perform a causality analysis across users.

It is also worth noting that according to the results presented in [55], the method used for causality analysis performs better than other existing linear methods, such as linear regression. Moreover, we cannot apply ANOVA methods since their key assumptions such as linearity and normality do not hold for our dataset. For example, variables such as notification response time and app usage are very skewed.

2.3.2 Confounding Variables. A crucial step of our analysis is the selection of confounding variables. There are several factors that could influence both users' emotional state and some of the phone interaction metrics that have been previously discussed. For instance, in order to examine the impact of using game applications on the user's emotional state we may need to control for other types of applications that were used in the same period. The amount of time that users spend playing mobile phone games may correlate with the amount of time

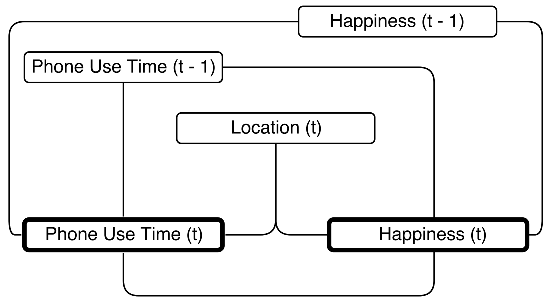


Fig. 2. Dependencies of *Phone Usage Time* and *Happiness* with the metrics of our study. The lines indicate the association and therefore no direction is reported. The treatment and outcome variables are shown in the boxes with bold lines.

they spend using other applications due to the fact that they are feeling bored at that moment and wants to kill time by using different applications. Thus, an observed change in the user's emotional state may be influenced by the use of another application rather than a game application. Furthermore, users' context modalities such as location and activity may also influence their emotion [54].

We apply the methodology described in [54, 55] in order to select an appropriate set of confounding variables. More specifically, we examine the correlations of all the previously presented metrics with the treatment and outcome variables of each study. Metrics that correlate with both the treatment and the outcome variables will be included in the set of confounding variables. Moreover, the examined variables may be autocorrelated. For example, a participant who is experiencing a stressful situation will probably report high stress levels at multiple consecutive times. Hence, the previous user behavior (i.e. interaction with the phone, activity and location) and emotional state should be examined as possible confounding variables. We conduct partial autocorrelation tests for the treatment and outcome variables in order to discover dependencies with their previous values. We also examine the correlation of the treatment and outcome variables with lagged values of the confounding variables conditional to the shorter lagged correlated values. For example, in Figure 2 we depict the dependencies of the treatment variable *Phone Usage Time* and the outcome variable *Happiness* with our metrics. We found that only *Location* correlates with the examined variables (i.e., both *Phone Usage Time* and *Happiness*). We also found that our variables (*Phone Usage Time* and *Happiness* level) at time t correlate with the preceding values (i.e., variables at time $t-1$). However, they do not correlate with preceding values of *Location*, but conditional to the current *Location* (i.e., *Location* at time t). Hence, the set of confounding variables will include only the *Location* and the one-lagged values of *Phone Usage Time* and *Happiness* variables.

Since mobile phone usage patterns may vary significantly for different users we require the matched samples to belong to the same user. Moreover, the phone usage may depend on the sampling time. For example, most users will probably use their phone less in the early morning (e.g., 4:00 am) compared to afternoon hours. Consequently, we cannot derive valid conclusions if we compare samples of different time intervals. Thus, we split a day into four time intervals: early morning (00:00-06:00), morning (06:00-12:00), afternoon (12:00-18:00) and evening (18:00-24:00) and we allow samples to be matched only if they are reported during the same time interval. Then, we apply optimal matching [31] in order to find optimal pairs of samples. We use as distance metric between samples the Mahalanobis distance [26] weighted by the difference on the treatment variable values as described in [23]. After finding the optimal pairs, we use Equation 2 in order to estimate the average treatment effect. If there is no effect, ATE should be close to zero. We use the t-test to examine if the observed value is significantly different from zero.

Data Type	Features
Notification	Arrival time, seen time and removal time, alert type (sound, vibrate and flashing LED), user's response (click or dismiss), sender application name and notification title.
Context	Physical activity and location.
Communication	Time, type and sender/recipient of calls and SMSs.
Phone Usage	Lock/unlock event, single click, long click, scrolls and usage time of all foreground applications (including home screen).

Table 2. Classes of data used for computing phone interaction metrics and context-based features.

Finally, we should stress that our study controls only for the observed confounding variables (i.e., the metrics described in Section 2.1.3) and could be biased in case of missing confounders. This is a known limitation of all causality studies based on observational data. Although several unobserved factors could influence user emotional state, bias could be induced only by those factors that influence also users interaction with their phones. By including a large number of metrics in our study and by controlling also for the previous values of emotional state and phone usage, we minimize this bias. However, the possibility that unobserved factors could influence the validity of our results cannot be eliminated. Thus, the causal analysis presented in this paper shows the potential causal effect in the absence of other confounders and in any case it provides evidence of strong dependency between variables.

2.3.3 Evaluation of the Causality Framework. As discussed earlier, it is worth underlining that the focus of our study is not on the design of a new causality framework. In fact, we adopt the causality framework presented by Tsapeli et al. in [55], in which the authors propose a causal inference method for time-series data based on matching-design techniques that does not require any assumptions about the functional form of the relationships among the variables. The method is extensively evaluated on synthetic data in scenarios with both linear and non-linear dependencies and with varying number of confounding variables. According to results presented in [55] this method is more effective in avoiding false positive conclusions than existing approaches. Furthermore, this method has already been applied to mobile sensing data in order to study the impact of daily activities on stress [54].

3 DATA COLLECTION

3.1 Overview

In order to study the influence of emotional states on the user's mobile interaction behavior, we conducted an *in-the-wild* study. More specifically, we developed an Android app called MyTraces (shown in Figure 1) that runs in the background to unobtrusively and continuously collect users' mobile phone interaction logs and the context information (as listed in Table 2).

The MyTraces application relies on the Android's Notification Listener Service [1] to log interaction with notifications. It uses Google's Activity Recognition API [4] to obtain the information about the user's physical activity classified as walking, bicycling, commuting on vehicle or still. Moreover, the application samples GPS data in an adaptive sensing fashion as described in [10]. In order to cluster the GPS data we apply the clustering algorithm presented in [54]. For each clustered location we assign one of the following labels: *home*, *work* or *other*. We assign the *home* label to the place where a user spends the majority of the night hours (defined as the time interval between 20:00 to 08:00). We consider *work* place as the second most significant place (i.e., the place where users spend most of their time apart from home). All other places are labeled as *other*.

To acquire the data about the user's emotional states (activeness, happiness and stress level) throughout the day, we rely on the experience sampling method (ESM) [13]. As shown in Figure 1.b, users can register their mood through a sliding bar. This bar uses a 5 point-based Likert scale where 1 indicates the lowest level and 5 the highest level. Every day a mood questionnaire is triggered at four random times in every three hour time window between 8.00 am and 11.00 pm by using the phone's time (i.e., local time zones). We chose this time window so that the participants do not feel annoyed by being asked to respond to the surveys early in the morning and late at night. In case a questionnaire is dismissed or not responded to within 30 minutes from its arrival time, the application triggers another alert after 30 minutes.

Since the higher values of activeness and happiness levels indicate a positive emotion, we measured the stress level according to a negative scale that means lower value would indicate a high level of stress. This way we make the scale of all emotional states consistent (i.e., the lower values refer to negative emotion and higher values to positive emotion). Therefore, we reverse the scale by subtracting each response value from 6. So, if for example the response is 5 (i.e., very low stress), we subtract it from 6 to rescale it to 1. Thus, with the reversed scale the lower value will refer to lower stress level and the higher value would indicate higher stress level.

It is worth noting that users were asked to report their emotional states during the past hour (see Figure 1.b) rather at the instance of responding to an ESM. As discussed earlier, we employ this approach of using such a longer period for querying emotional states because previous psychological studies have shown that emotions persist for a few hours [8, 9].

Additionally, we also collected data about the weather at users' location during the day on an hourly basis. This data consists of features such as temperature and humidity. In order to obtain this data we rely on the Weather Underground's History API [2]. It is worth noting that we use the History API because this data was collected after the data collection from mobile phones.

3.1.1 Recruitment of the Participants. The MyTraces application was published on Google Play Store and has been available to the general public for free since 4th January 2016. It was advertised through different channels: academic mailing lists, Twitter, Facebook and Reddit. In order to attract more participants for our study, we committed to give incentives to the participants for replying to the questionnaires for a minimum of 30 days. We committed to select (through a lottery) one winner of a Moto 360 Smartwatch and 20 winners of an Amazon voucher.

In order to ensure privacy compliance, the MyTraces application goes through a two-level user agreement to access the user's critical data. Firstly, the user has to give explicit permission as required by the Android operating system for capturing application usage, notifications and user's interaction with mobile phone (such as clicks, long-clicks and scrolls). Secondly, the application shows a list of information that is collected and asks for the user's consent. Furthermore, the study was performed in accordance with our institution's ethical research procedure and the consent form itself for the data collection was reviewed by our institution's Ethics Review Board.

3.2 Dataset

We consider the data collected from 4th January 2016 to 1st July 2016. In this period the application was installed by 104 users. However, many users did not actively respond to the mood questionnaires and some uninstalled the application after a few days. Therefore, we selected a subset of the data for the analysis by considering only the users who ran the application for at least 20 days and responded to at least 50% of the mood questionnaires in order to have a a sample sufficiently large to be statistically significant. Consequently, there are 28 users who satisfied these constraints. Note that we do not have information about the demographics of these participants because it was not asked during the study for privacy reasons.

[2] www.wunderground.com/weather/api/d/docs?d=data/history

Our final dataset (i.e., the subset of active users) comprises 9 million phone usage events, 5,118 responses to mood questionnaires, more than 9 million mobile interaction logs and 2 million context samples.

4 RESULTS

As discussed earlier, the entire correlation analysis is performed considering the emotional state (namely, activeness, happiness and stress) recorded during a given hour period h and the values of the examined metrics for three different time intervals: *preceding* hour period ($h-1$), *current* hour period (h) and *next* hour period (h+1).

4.1 When is the Causality Analysis Performed?

If a statistically-significant and moderate correlation[3] is observed between an emotional state and a metric for *next hour*, we perform a causality analysis to quantify the impact of the emotional state on that metric. Moreover, we also perform a causality analysis in the other direction, i.e., we quantify the impact of behavioral metrics on emotional states. This is performed when there is a significant and moderate correlation between an emotional state and a metric for the *preceding hour*.

It is worth noting that though we observe statistically significant but weak correlation of emotional states with some of the phone interaction metrics we do not perform a causality analysis for them. This is because the weaker correlation implies that the causality will also be weaker and, for this reason, such results would not be robust.

4.2 Emotional States and Notifications

In this section we present the results of the correlation and causality analysis for the reported emotional states and notification metrics.

The key findings of this section are:

- People's activeness level has a significant association with the seen and decision time of notifications that arrive in the *next hour*.
- In stressful situations people become more attentive, this results in the reduction of notification response time.

In Figure 3 we show the correlation coefficients that are computed to assess the relationship between emotional states (activeness, happiness and stress) and notification metrics. The results show that the activeness level moderately correlates with the average ST (i.e., seen time) and DT (i.e., decision time) of notifications that arrive in the *next hour*. This indicates that the users' awareness and pace for reacting to notifications is linked with their activeness level. This is in a sense expected, since a user who might be less energetic would delay their response to notifications. Moreover, we also observe a moderate association between stress level and the average RT of notifications that arrive in the *next hour*. We believe that this correlation exists because users become more alert while performing a complex and stressful task.

In order to investigate if there is any causal link here, we perform causality analysis to quantify the impact of activeness on the average ST and DT of notifications, and the impact of stress on the average RT. Table 3 presents the mean difference (i.e., average treatment effect (ATE)) as described in Equation 2. The results indicate that the activeness level has no impact on the average ST and DT. This means that there is another variable (such as time

[3] A correlation is usually considered statistically-significant if the p-value is less than the significance level α (i.e., p-value <0.05) and moderate if the correlation coefficient is greater than 0.2.

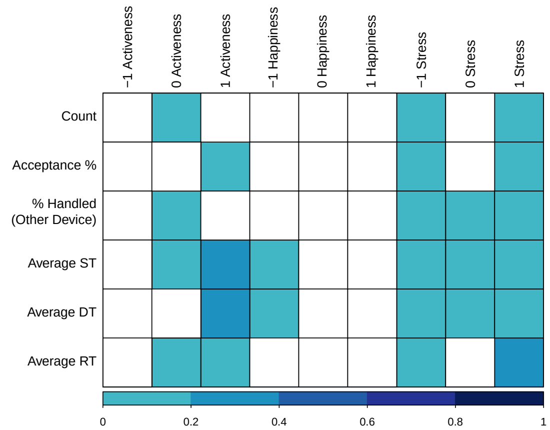


Fig. 3. Results for correlation between emotional states and notification metrics.

	Mean Difference		
Metric	Activeness	Happiness	Stress
Average ST	23.18	NP	NP
Average DT	11.97	NP	NP
Average RT	NP	NP	-58.06**

** refers to the p-values <0.001.
NP refers to not present.

Table 3. Results for the causal effect of emotional states on notification metrics.

or location) than drives both activeness as well as the average ST and DT. For instance, people's activeness might vary with their location and so does the ST and DT of notifications.

However, we observe that there is a statistically significant and negative causal impact of stress on the average RT (i.e., response time for notifications). The negative value indicates that the average response time to notifications is lower for participants with higher stress score (i.e., more stressed). This suggests that people become more attentive to notifications when they are stressed. In another recent study [29] we have also found that user's attentiveness increases as the complexity of an ongoing task increases. Results from this study confirm this and, at the same time, highlight the presence of a causal link.

Note that we do not assess causality between other variables as they do not show significant and moderate association. Indeed, a necessary condition for causality between two variables is the presence of correlation in the first place. Therefore, we indicate *NP* (Not Present) for these relationships in the table.

On the other hand, we also do not see any notification metric for the *preceding hour* that has a significant and moderate relationship with an emotional state. Therefore, no causality analysis is performed for assessing the impact of these metrics on emotional states.

4.3 Emotional States and Phone Usage

In this section we present the results of the correlation and causality analysis for the reported emotional states and phone usage metrics.

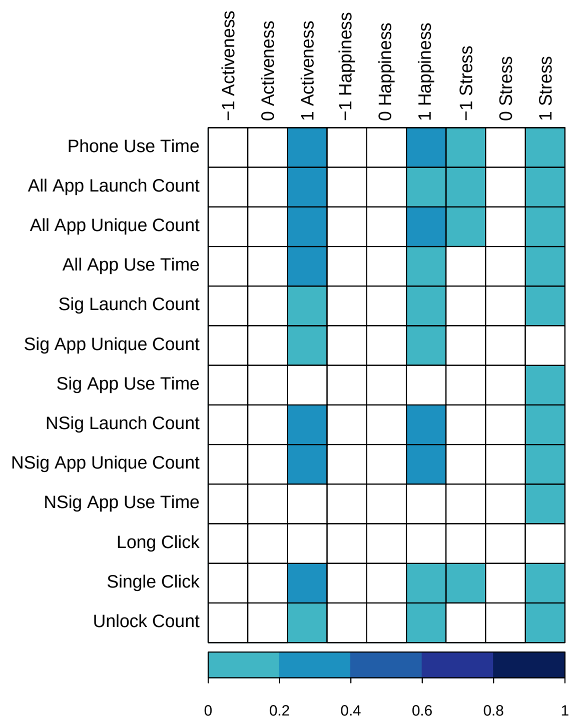


Fig. 4. Results for correlation between emotional states and phone usage metrics.

The key findings of this section are:

- Increase in activeness level positively impacts users' phone usage, number of apps launched and clicks on the screen.
- The happier people report to be, the less inclined they are to use their phone.

We first quantify the association between users' emotional states and their phone usage metrics. As presented in Figure 4 the results show that the activeness level of users has a significantly moderate correlation with seven of the next hour's phone usage metrics. These metrics include usage time of phone and all apps, total number of times (all and non-significant) apps are launched, number of unique (all and non-significant) apps used, and the single click count.

Moreover, the happiness level of users has a significantly moderate correlation with four of the next hour's phone usage metrics. These metrics include phone usage time, number of unique (all and non-significant) apps used, and total number of times non-significant apps are launched.

We then investigate the impact of emotional states (in this case only for activeness and happiness levels as stress level does not correlate moderately with phone usage metrics). The results (presented in Table 4) demonstrate that users' activeness level significantly and strongly impacts the app launch counts, overall usage time, and number of clicks. At the same time, users' happiness level has a significantly strong and negative impact on the

Metric	Mean Difference		
	Activeness	Happiness	Stress
Phone use time	18.51	-16.62**	NP
All app launch count	1.92**	NP	NP
All app unique count	0.18	-0.09	NP
All app use time	39.66**	NP	NP
Non-sig launch count	0.19**	0.08	NP
Non-sig unique app count	0.08**	0.03	NP
Single click count	4.64**	NP	NP

** refers to the p-values <0.001.
NP refers to not present.

Table 4. Results for the causal effect of emotional states on phone usage metrics.

phone usage time. *These results suggest that people tend to spend more time with their phone when they are active. However, increase in the people's happiness makes them less inclined to use the phone.*

Moreover, the activeness level has a weak impact on the number of clicks as well as the unique (all and non-significant) apps that are used in the next hour. This causality relationship is in a sense expected because we already observed that the phone usage increases when people are very active, thus the increase in phone usage increases the likelihood of using more apps and increased number of clicks.

On the other hand, the happiness level has a weak positive impact on the number of all unique apps used, but a weak negative impact on the number of unique non-significant apps used. This suggests that people tend to use their phones and apps less when they are happy; at the same time, there is evidence that they tend to launch non-significant apps. Finally, we do not observe any causal link between users' activeness and happiness levels and other phone usage metrics.

Furthermore, we do not quantify the causal impact of the phone usage metric on the emotional states because there is no moderate correlation between the emotional states and phone usage metrics of previous hour.

4.4 Emotional States and Usage of Specific Applications

As discussed in the previous section, users' emotional states impact on their application usage. In this section, we refine our analysis by performing correlation and causality analysis between emotional states and usage of specific applications (i.e., launch count and usage time for nine types of applications).

The key findings of this section are:

- Users' activeness level has a positive impact on the music app usage.
- Increase in the stress level of users significantly reduces the usage of communication apps.
- People reported being less stressed when their usage of travel apps increases.

Figure 5 presents the results of the correlation analysis between users' emotional states and the usage of specific apps. The results show that the user's activeness has a significantly strong association with the next hour's usage of communication, music and travel apps. We want now to go beyond association and investigate potential casual links between the variables. As a first step, we quantify the impact of users' activeness level on the launch count of communication and travel apps, and usage time of communication and music apps. We select these variables since we observe association between them as discussed in the previous section. Our results

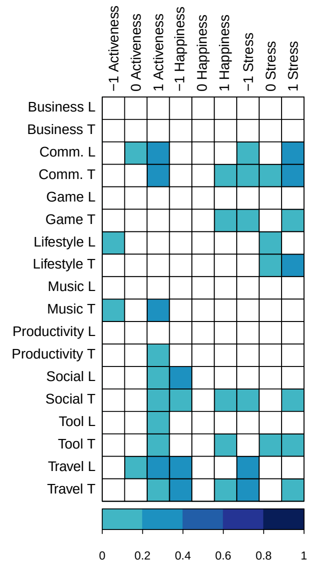


Fig. 5. Results for correlation between emotional states and application metrics. Here, *L* refers to launch count and *T* refers to usage time.

	Mean Difference		
Metric	Activeness	Happiness	Stress
Communication L	0.13	NP	-0.03
Communication T	8.15	NP	-13.09**
Lifestyle T	NP	NP	3.66
Music T	4.58**	NP	NP
Travel L	0.01	NP	NP

** refers to the p-values <0.001.
NP refers to not present.

Table 5. Results for the causal effect of emotional states on application metrics. Here, *L* refers to launch count and *T* refers to usage time.

(presented in Table 5) indicate that the user's activeness level significantly and negatively influences the usage of music apps. This suggests that users listen to music when they are very active, as opposed to when they are less active. This can potentially be due to the fact that people listen to music (thus, use music apps) while travelling, which is when they are active. Moreover, there is no significant causal effect of activeness on the usage of communication and travel apps. Therefore, the observed correlation (between activeness level and the usage of communication and travel apps) is due to other factors that influence both examined variables. For instance, it could be explained by the fact that people generally travel to a new place (by using travel app) during the day time when they are active but not when they are feeling sleepy.

At the same time, the results show that users' stress level has a significantly moderate correlation with the next hour's usage of communication and lifestyle apps. Therefore, we perform causality analysis to quantify the effect of stress on the usage of these apps. The results show that the usage time of communication apps is

	Mean Difference		
Metric	Activeness	Happiness	Stress
Social L	NP	0.05	NP
Travel L	NP	0.13	-0.27**
Travel T	NP	0.12	-0.16

** refers to the p-values <0.005.
NP refers to not present.

Table 6. Results for the causal effect of application metrics on mood. Here, *L* refers to launch count and *T* refers to usage time.

significantly and negatively influenced by the user's stress level. *This indicates that the increase in the stress level makes people reduce their usage of communication apps.* However, there is not causal impact of stress on the usage of lifestyle apps.

On the other hand, the results also show that users' happiness level has a statistically-significantly and moderate correlation with the previous hour's usage of social and travel apps. Additionally, the stress level also moderately correlates with the previous hour's usage of travel apps. Therefore, we perform a causality analysis to check if the users' happiness and stress levels are influenced by their social and travel apps usage patterns. The results presented in Table 6 show that the use of social apps has no causal link with emotional states. However, the use of travel apps has a significantly negative impact on stress level of users. *In other words, the use of travel apps, which is probably linked to the fact that a person is traveling or is going to travel, has a negative impact on the stress level of users (i.e., it decreases their stress level).* Indeed, it is worth stressing that the resulting mood modification is not caused by the application, but by the underlying intention or need of the user to interact with the application itself. For instance, the use of travel apps indicates that the user is likely to travel that is the actual reason of happiness and stress reduction. Therefore, the use of these apps is an indirect signal of user behavior. Quite interestingly, this finding is also supported by the study [11] that shows that (leisure) traveling has been widely regarded as a pursuit to relaxation and mental wellbeing.

4.5 Emotional States and Communication

In this section we present the results of the correlation and causality analysis for the reported emotional states and communication metrics.

We first perform the correlation analysis to quantify the association between users' emotional states and their communication patterns. We present the results in Figure 6. We found that none of the emotional states has a significant or moderate association with the previous and next hours' communication metrics. For this reason we do not perform any further analysis to quantify causality between emotional states and communication metrics.

5 IMPLICATIONS AND LIMITATIONS

In this work we have studied the association between user mood and phone interaction. We have initially conducted a correlation study in order to detect links between user mood and phone interaction. Then, we have attempted to understand the causal impact of users' mood on their interaction with their phone and vice versa.

The findings of this work can be used as a basis for designing more effective computing systems that rely on the analysis of users' emotional states. Examples include personalized services that reflect the actual emotional state of the users, considering not only the association between mood and behavior and smartphone interaction but also the casual links between them.

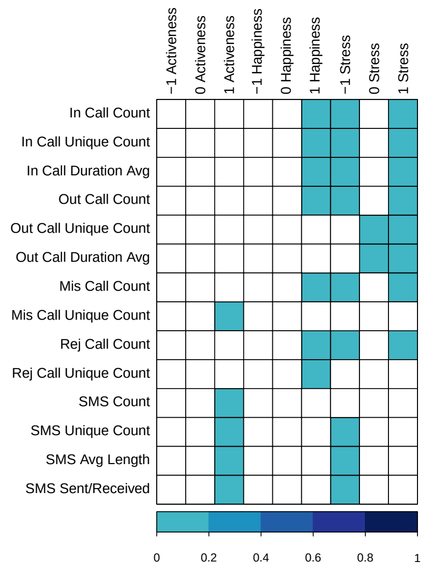


Fig. 6. Results for correlation between emotional states and communication metrics.

This is fundamental especially for marketing [2] and digital health [5, 6] applications. Indeed, one of the most promising areas is digital behavior intervention. Although this study does not focus on investigating behavior change interventions, understanding the causal links between mobile phone interaction and users' cognitive contextual information is of paramount importance for designing more effective systems to support them [15, 33]. In general, we believe that quantifying causality and not only correlation is critical for designing digital behavior intervention tools for a variety of applications from physical activity to psychological support.

Our study utilizes raw sensor data in order to derive high-level information such as location labels and activity, and, consequently, it is subject to limitations and inaccuracies of the inference method. Moreover, self-reported emotion states may be in themselves inaccurate, given the known problems related to biased self-representation in questionnaires [14], or not answered very frequently [30]. Considering also that some users may not be willing to answer any questionnaires when they are very stressed, unhappy or sleepy, our study may fail to capture such extreme cases. This is a common limitation of all smartphone-based studies and, for this reason, we believe it is fundamental to perform additional experiments to reproduce and re-validate these results in different settings in the future.

We are also aware of the limited ecological validity of this study. Thus, although our results indicate a link between mood and phone interaction, it is difficult to make any strong claim in terms of the generalization of our approach. At the same time, we believe that the proposed methodology can be applied to large-scale studies with different population samples. We believe that this study should be replicated in order to test its validity, for example, on different demographics.

Finally, a key limitation of this study stems from our decision to collect data in a real-world scenario with the minimum amount of intervention from our users, i.e., our causality analysis is based on observational data automatically captured from the sensors embedded in users' phones. Using this data we derived a variety of metrics that are considered as confounding variables to control their effects on the treatment and outcome variables of the causal analysis. Moreover, we also considered the effect of autocorrelation, and the current and previous values of emotional state and phone usage to minimize any bias. However, it is not practically possible to capture and control all confounding variables in such observational studies [47]. We cannot exclude that some unobserved variables that have a direct influence on both the treatment and the outcome variables are not included in the analysis. Some example of variables that influence emotional states might be face-to-face interaction [17], diet [20], and amount of sleep [53]. However, it is worth observing that there there is no evidence to demonstrate that these factors also influence phone usage patterns. Therefore, it might be reasonable to assume that that our causal analysis is not biased for such unobserved variables.

It might be possible to monitor these variables, but this will imply the use of ESM techniques in order to collect users' responses. However, this will increase the number of prompts, and, consequently, cause potential annoyance for users, who might leave the study after a short time due to this reason. Alternatively, such variables could be controlled in a lab-based experiment, which will then lack of realism. In fact, our goal was to carry out an real-world *in-the-wild* experiment.

6 RELATED WORK

Today's mobile phones come with a plethora of embedded sensors allowing us to collect information about the user's day-to-day activities [12, 56], mobility [41, 52], the surrounding environment [25], emotional states [10, 21, 24, 36, 37] and much more. This contextual information has been used to model users' interaction with their mobile phones [49]. Previous studies have proposed different approaches to extract the user's application usage [18, 48, 49, 57], receptivity to information [28] and communication patterns [42] for building intelligent systems.

In [49], Srinivasan et al. proposed MobileMiner, a system that performs on-device mining of mobile user's frequent co-occurrence patterns to predict future contextual events. The authors evaluated their mechanism with the data of 106 users collected over 1-3 months. Their results show that MobileMiner could predict the next app to be launched by users with a precision of 80% and a recall of 68%. Another recent study proposed a system that relies on machine learning algorithms for the automatic extraction of rules that reflect user's preferences for receiving notifications in different situations [27]. In [32] Pejovic and Musolesi discussed a mechanism that relies on the contextual information (including activity, location and time of day) to predict opportune moments for delivering notifications.

However, the interaction of users with mobile phones does not solely depend on their physical context, instead it is also associated with numerous aspects of their cognitive context. In recent years, researchers have been trying to uncover the relationship between users' cognitive context and patterns of interaction with mobile phones [7, 22, 29, 34]. In [22] LiKamWa et al. show that application usage and communication patterns are strong indicators of a user's mood, which can be used to infer a user's daily average mood with an accuracy of 66%. In [46] Servia et al. present a longitudinal study, based on data collected by means of a smartphone application, investigating the relationship between user's activity and sociability and a variety of psychological dimensions, such as perception of health, life satisfaction, and connectedness. The authors demonstrate that mobile sensing can be used to predict users' mood with an accuracy of about 70%.

Alvarez-Lozano et al. [7] investigated the changes in the application usage pattern of patients affected by bipolar disorder. The authors show that users' application usage patterns have a strong correlation with different aspects of their self-reported depressive state, sleep and irritability. Mehrotra et al. [29] investigated the effect of both cognitive and physical factors on the user's receptivity to notifications. The authors show that the response time and the perceived disruption from a notification can be influenced by the type, completion level and complexity of the task in which the user is engaged. At the same time, Pielot et al. [34] demonstrated that boredom influences the user's receptivity to information delivered via mobile notifications. Their results show that users are more likely to engage with suggested content on their phones when they are bored.

To the best of our knowledge, only the authors of [54] have performed a causal analysis using sensor data, but they focused exclusively on the causal impact of physical activities (such as walking and running) on the user's stress level.

7 CONCLUSIONS AND FUTURE WORK

In this paper, for the first time, we have performed a causality analysis between users' behavior and mood and mobile phone interaction in terms of notification response, application usage and communication patterns. We collected 5,118 responses to questionnaires for logging users' emotional states (namely, activeness, happiness and stress) from 28 users over a period of 20 days.

First of all, using a non-parametric correlation test (Kendall's Rank), we have shown that users' emotional states moderately correlates with different aspects of notification, phone and application usage, and communication patterns. Then, we have conducted an in-depth causality analysis considering a variety of contextual variables and mood indicators. Moreover, we have investigated whether there is a causal link between mood and interaction with the phone as well as the direction of the link (i.e., whether mood has a causal impact on users' interaction with the phone or interaction with the phone influences users' mood). We have shown that in stressful situations people become more attentive: this results in a lower notification response time. We have also found that people use their phone more when they are active and less when they are happy. With respect to the analysis concerning the causal links between the usage of specific apps and emotional states, we have shown that the increase in users' activeness level reduces the usage of music app. We have also observed that as the stress level increases the usage of communication and lifestyle apps decreases. On the other hand, the causal analysis related to the impact on emotional states indirectly suggest that socializing makes people happier and traveling reduces their stress. However, we have found no association between emotional states and communication metrics.

The potential applications of this work are several, from the design of enhanced search and marketing tools to the development and deployment of more effective mobile systems for behavior intervention. Furthermore, we believe that this work offers new insights into the way people interact with smartphones. More in general, it provides a *quantitative* basis for the development of new methodologies for the design of innovative emotion-aware systems. Understanding causality and not simply correlation is of fundamental importance in the design of systems that affect not only human activities but also emotional states.

As a part of our future research agenda, we plan to go a step further by investigating the causal links between human-smartphone interaction and other conditions such as mood and sleeping disorders.

8 ACKNOWLEDGEMENTS

This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1 and at UCL through the EPSRC grants EP/L018829/2 and EP/L006340/1.

REFERENCES

[1] 2017. Android's Notification Listener Service. (2017). http://developer.android.com/reference/android/service/notification/NotificationListenerService.html.
[2] 2017. Google Now. (2017). http://www.google.com/landing/now/.
[3] 2017. Google Play Store. (2017). https://play.google.com/store/apps.
[4] 2017. Google's Activity Recognition Application. http://developer.android.com/training/location/activity-recognition.htmll. (2017).
[5] 2017. HeadSpace. (2017). https://www.headspace.com.
[6] 2017. HelloMind. (2017). http://www.hellomind.com.
[7] Jorge Alvarez-Lozano, Venet Osmani, Oscar Mayora, Mads Frost, Jakob Bardram, Maria Faurholt-Jepsen, and Lars Vedel Kessing. 2014. Tell me your apps and I will tell you your mood: correlation of apps usage with bipolar disorder state. In *PETRA'14*.
[8] C Daniel Batson, Laura L Shaw, and Kathryn C Oleson. 1992. *Differentiating affect, mood, and emotion: toward functionally based conceptual distinctions.* Sage Publications.

- [9] Christopher Beedie, Peter Terry, and Andrew Lane. 2005. Distinctions between emotion and mood. *Cognition & Emotion* 19, 6 (2005), 847–878.
- [10] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *UbiComp'15*.
- [11] Chun-Chu Chen and James F Petrick. 2013. Health and wellness benefits of travel experiences a literature review. *Journal of Travel Research* 52, 6 (2013), 709–719.
- [12] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, and others. 2008. Activity Sensing in the Wild: a Field Trial of UbiFit Garden. In *CHI'08*.
- [13] Mihaly Csikszentmihalyi and Reed Larson. 1983. The experience sampling method. *New Directions for Methodology of Social and Behavioral Science* 15 (1983), 41–56.
- [14] Mihaly Csikszentmihalyi and Reed Larson. 2014. Validity and reliability of the experience-sampling method. In *Flow and the Foundations of Positive Psychology*. 35–54.
- [15] Martin Fishbein. 1995. Developing effective behavior change interventions: some lessons learned from behavioral research. *NIDA Research Monograph* 155 (1995), 246–261.
- [16] Ronald Aylmer Fisher and Frank Yates. 1938. *Statistical tables for biological, agricultural and medical research*. Longman.
- [17] Joseph P Forgas, Gordon H Bower, and Susan E Krantz. 1984. The influence of mood on perceptions of social interactions. *Journal of Experimental Social Psychology* 20, 6 (1984), 497–513.
- [18] Ke Huang, Chunhui Zhang, Xiaoxiao Ma, and Guanling Chen. 2012. Predicting mobile application usage using contextual information. In *UbiComp'12*.
- [19] Neal Lathia, Veljko Pejovic, Kiran K. Rachuri, Cecilia Mascolo, Mirco Musolesi, and Peter J. Rentfrow. 2013. Smartphones for Large-Scale Behaviour Change Intervention. *Pervasive Computing* 12, 12 (July 2013), 66–73.
- [20] Peter D Leathwood and Patricia Pollet. 1983. Diet-induced mood changes in normal populations. *Journal of Psychiatric Research* 17, 2 (1983), 147–154.
- [21] Hosub Lee, Young Sang Choi, Sunjae Lee, and I.P. Park. 2012. Towards unobtrusive emotion recognition for affective social communication. In *CCNC'12*.
- [22] Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. 2013. Moodscope: Building a mood sensor from smartphone usage patterns. In *MobiSys'13*.
- [23] Bo Lu, Elaine Zanutto, Robert Hornik, and Paul R Rosenbaum. 2001. Matching with doses in an observational study of a media campaign against drug abuse. *J. Amer. Statist. Assoc.* 96, 456 (2001), 1245–1253.
- [24] Hong Lu, Gokul T. Chittaranjan Mashfiqui Rabbi, Denise Frauendorfer, Marianne Schmid Mast, Andrew T. Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. 2012. StressSense: Detecting Stress in Unconstrained Acoustic Environments using Smartphones. In *UbiComp'12*.
- [25] Hong Lu, Wei Pan, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. 2009. SoundSense: Scalable Sound Sensing for People-centric Applications on Mobile Phones. In *MobiSys'09*.
- [26] Prasanta Chandra Mahalanobis. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)* 2 (1936), 49–55.
- [27] Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2016. PrefMiner: Mining User's Preferences for Intelligent Mobile Notification Management. In *Proceedings of UbiComp'16*. Heidelberg, Germany.
- [28] Abhinav Mehrotra, Mirco Musolesi, Robert Hendley, and Veljko Pejovic. 2015. Designing Content-driven Intelligent Notification Mechanisms for Mobile Applications. In *UbiComp'15*.
- [29] Abhinav Mehrotra, Veljko Pejovic, Jo Vermeulen, Robert Hendley, and Mirco Musolesi. 2016. My Phone and Me: Understanding User's Receptivity to Mobile Notifications. In *CHI'16*.
- [30] Abhinav Mehrotra, Jo Vermeulen, Veljko Pejovic, and Mirco Musolesi. 2015. Ask, But Don't Interrupt: The Case for Interruptibility-Aware Mobile Experience Sampling. In *UbiComp'15 Adjunct*.
- [31] Kewei Ming and Paul R Rosenbaum. 2001. A note on optimal matching with variable controls using the assignment algorithm. *Journal of Computational and Graphical Statistics* 10, 3 (2001), 455–463.
- [32] Veljko Pejovic and Mirco Musolesi. 2014. InterruptMe: designing intelligent prompting mechanisms for pervasive applications. In *UbiComp'14*.
- [33] Richard E Petty, Joseph R Priester, and Duane T Wegener. 1994. Cognitive processes in attitude change. *Handbook of Social Cognition* 2 (1994), 69–142.
- [34] Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When attention is not scarce-detecting boredom from mobile phone usage. In *UbiComp'15*.
- [35] Karl H Pribram and Diane McGuinness. 1975. Arousal, activation, and effort in the control of attention. *Psychological Review* 82, 2 (1975), 116–149.

[36] Mashfiqui Rabbi, Shahid Ali, Tanzeem Choudhury, and Ethan Berke. 2011. Passive and In-situ Assessment of Mental and Physical Well-being using Mobile Sensors. In *UbiComp'11*.

[37] Kiran K. Rachuri, Mirco Musolesi, Cecilia Mascolo, Jason Rentfrow, Chris Longworth, and Andrius Aucinas. 2010. EmotionSense: A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research. In *UbiComp'10*.

[38] JA Ressel. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39 (1980), 1161–78.

[39] Donald B Rubin. 1973. Matching to remove bias in observational studies. *Biometrics* 29, 1 (1973), 159–183.

[40] Donald B Rubin. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66, 5 (1974), 688–701.

[41] Adam Sadilek and John Krumm. 2012. Far Out: Predicting Long-Term Human Mobility. In *AAAI'12*.

[42] Antti Salovaara, Antti Lindqvist, Tero Hasu, and Jonna Häkkilä. 2011. The phone rings but the user doesn't answer: unavailability in mobile communication. In *MobileHCI'11*.

[43] Ulrich Schimmack and Reisenzein Rainer. 2002. Experiencing activation: energetic arousal and tense arousal are not mixtures of valence and activation. *Emotion* 2, 4 (2002), 412–417.

[44] Martin EP Seligman. 2004. Can happiness be taught? *Daedalus* 133, 2 (2004), 80–87.

[45] Hans Selye. 1956. *The stress of life.* McGraw-Hill.

[46] Sandra Servia-Rodríguez, Kiran K Rachuri, Cecilia Mascolo, Peter J Rentfrow, Neal Lathia, and Gillian M Sandstrom. 2017. Mobile Sensing at the Service of Mental Well-being: a Large-scale Longitudinal Study. In *WWW'17*.

[47] William R Shadish, Thomas D Cook, and Donald Thomas Campbell. 2002. *Experimental and quasi-experimental designs for generalized causal inference.* Houghton, Mifflin and Company.

[48] Choonsung Shin, Jin-Hyuk Hong, and Anind K Dey. 2012. Understanding and prediction of mobile application usage for smart phones. In *UbiComp'12*.

[49] Vijay Srinivasan, Saeed Moghaddam, Abhishek Mukherji, Kiran K Rachuri, Chenren Xu, and Emmanuel Munguia Tapia. 2014. MobileMiner: Mining your frequent patterns on your phone. In *UbiComp'14*.

[50] Elizabeth A Stuart. 2010. Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics* 25, 1 (2010), 1–21.

[51] Robert E Thayer. 1989. *The biopsychology of mood and arousal.* Oxford University Press.

[52] Arvind Thiagarajan, Lenin Ravindranath, Katrina LaCurts, Samuel Madden, Hari Balakrishnan, Sivan Toledo, and Jakob Eriksson. 2009. VTrack: Accurate, Energy-Aware Road Traffic Delay Estimation Using Mobile Phones. In *SenSys'09*.

[53] Dorthe Kirkegaard Thomsen, Mimi Yung Mehlsen, Søren Christensen, and Robert Zachariae. 2003. Rumination–relationship with negative mood and sleep quality. *Personality and Individual Differences* 34, 7 (2003), 1293–1301.

[54] Fani Tsapeli and Mirco Musolesi. 2015. Investigating causality in human behavior from smartphone sensor data: a quasi-experimental approach. *EPJ Data Science* 4, 1 (2015), 1–15.

[55] Fani Tsapeli, Mirco Musolesi, and Peter Tino. 2017. Model-free Causality Detection: An Application to Social Media and Financial Data. *Physica A* 183, 1 (2017), 139–155.

[56] Tianyu Wang, Giuseppe Cardone, Antonio Corradi, Lorenzo Torresani, and Andrew T Campbell. 2012. WalkSafe: A Pedestrian Safety App for Mobile Phone Users Who Walk and Talk While Crossing Roads. In *HotMobile'12*.

[57] Tingxin Yan, David Chu, Deepak Ganesan, Aman Kansal, and Jie Liu. 2012. Fast app launching for mobile devices using predictive user context. In *MobiSys'12*.