Differential Human Factors in User Data

Clemens Stachl

München 2016

Differential Human Factors in User Data

Inauguraldissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig–Maximilians–Universität München

vorgelegt von
Clemens Stachl
aus Leoben, Österreich

München, den 04.10.2016

Erstgutachter: Prof. Dr. Markus Bühner

Zweitgutachter: Prof. Dr. Moritz Heene

Tag der mündlichen Prüfung: 07.02.2017

Acknowledgments

An der Entstehung dieser Dissertation haben neben mir zahlreiche andere Personen und Organisationen wissentlich und unwissentlich Einfluss genommen. Aus diesem Grund möchte ich mich zumindest bei den wichtigsten bedanken und sie hier namentlich erwähnen. Auf fachlicher Ebene gilt mein Dank zuallererst meinem Betreuer, Prof. Dr. Markus Bühner. Besonders möchte ich ihm für die mir ermöglichte kreative Freiheit in meinem Tun und dem schier unendlichen, mir gegenüber geäußerten Vertrauen bedanken. Ebenfalls bedanken möchte ich mich bei Jiew-Quay Au, Prof. Dr. Bernd Bischl, Prof. Dr. Moritz Heene und Dr. Sven Hilbert für die Durchsicht, Korrektur und Verbesserung dieser Arbeit. Ebenfalls bedanken möchte ich mich bei Jiew-Quay Au, Alexandra Dietl, Martin Fröhlich, Victoria Gebhardt, Sarah Huber, Jakob Kaiser, Matthias Lamm, Ramona Schödel und Ilona Weigand, welche als StudentInnen wertvolle Beiträge in Form von Studiendurchführung, Daten, und Ideen geliefert haben. Darüber hinaus möchte ich mich bei der AUDI AG bedanken, ohne deren Finanzierung, die Dissertation in der vorliegenden Form nicht möglich gewesen wäre. Obwohl die vorliegende Arbeit eine rein wissenschaftliche ist, möchte ich mich auch bei einigen Personen aus meinem privatem Umfeld bedanken. Besonderer Dank gilt meinen Eltern, welche mir unter anderem Durchhaltevermögen und Penetranz gelehrt, und es mir durch ihre Unterstützung ermöglicht haben, ein Universitätsstudium zu absolvieren. Darüber hinaus bedanke ich mich auch bei Nadine, die mich im letzten Jahr meiner Promotionszeit sehr unterstüzt und immer wieder aufgebaut hat.

Contents

List of Figures

List of Tables

Nomenclature

ACC	Accuracy
AFA	Act Frequency Approach
AG	Aktiengesellschaft
APA	American Psychological Association
API	Application Programming Interface
AUC	Area Under the Curve
BAC	Balanced Accuracy
BAS	Behavioral Activation System
BFI	Big Five Inventory
BFSI	Big Five Structure Inventory
BIS	Behavioral Inhibition System
CI	Confidence Interval
CPU	Central Processing Unit
CV	Cross-Validation
DSS	Dawid-Sebastian Score
ESM	Experience Sampling Method
EU	European Union
EUR	Euro
FN	False Negative
FP	False Positive
GPS	Global Positioning System
HCI	Human Computer Interaction
INSBAT	Intelligence Structure Battery
iOS	iPhone Operating System
LASSO	Least Absolute Shrinkage and Selection Operator
LMU	Ludwig-Maximilians-Universität München

LOOCV	Leave One Out Cross-Validation
M	Mean
MLR	Machine Learning in R
MMCE	Mean Misclassification Error
MSE	Mean Squared Error
NEO-FFI	NEO-Five Factor Inventory
NEO-PI-R	NEO-Personality Inventory- Revised
NEO-PI3	NEO-Personality Inventory 3
NIR	No-Information Rate
PCA	Principal Component Analysis
PDA	Personal Digital Assistant
PFER	Pairwise Family Error Rate
PLS	Partial Least Squares
ROC	Receiver Operator Curve
SD	Standard Deviation
SMOTE	Synthetic Minority Oversampling Technique
SSL	Secure Sockets Layer
TIPI	Ten Item Personality Inventory
TN	True Negatives
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
UK	United Kingdom
VIF	Variance Inflation Factor
WiFi	Wireless Fidelity

Abstract

This thesis investigates how differential human factors, such as demography and personality, are related to actual individual behavior. Within this broad context, this work addresses the prevailing lack of real behavior in the scientific field of psychology and differential-/social psychology in particular. Furthermore, this work provides an introduction to the practice of data-logging as a promising alternative to self-reports for the collection of behavioral data. Additionally, we introduce new data-analytical concepts from the field of machine learning in order to appropriately handle large and noisy datasets, such as technical logs. To illustrate these concepts we provide three empirical studies, using behavioral logging procedures. In the first study we report on data obtained in a virtual automotive driving simulation. Using these data, we demonstrate how individual driving patterns can be used to predict driver gender with high accuracy from basic automotive driving logs. Additionally, we provide information about the most important variables associated with male and female driving styles. Two additional studies utilize a specially designed Android application, to automatically collect behavioral user data in a privacy protecting manner from participants private smartphones. The second study describes how most stable mobile application usage on smartphones can be predicted from individual personality and demography scores and highlights implications for personality sensitive recommender systems. The third study demonstrates how individual personality can potentially be predicted, using a wide range of user interactions, with a machine learning approach. Finally, we discuss the reported results within the context of previous research and highlight possible implications of technological advancements for psychological science.

Zusammenfassung

Die vorliegende Dissertation beschäftigt sich damit, wie differentielle menschliche Faktoren wie Geschlecht und Persönlichkeit mit tatsächlichem individuellem Verhalten in Verbindung stehen. Innerhalb dieses breiten Kontexts, thematisiert die vorliegende Arbeit das Fehlen von tatsächlichem Verhalten in der wissenschaftlichen Psychologie und der Sozial- und Persönlichkeitspsychologie im Besonderen.

Als vielversprechende Alternative zu Selbstauskünften aus Fragebögen bietet diese Arbeit eine Einführung in die Praxis des Datenloggings zur Erhebung von Verhaltensdaten. Darüber hinaus werden neue Analysemethoden aus dem Bereich des maschinellen Lernens vorgestellt, welche es ermöglichen, große Datensätze, effektiv zu analysieren. Im Hauptteil dieser Arbeit, werden drei empirische Studien mit Datenlogging vorgestellt.

In der ersten Studie wurde das Fahrverhalten von 145 Männern und Frauen in Form von basalen Logging-Daten aufgezeichnet. Alle Teilnehmer fuhren für 20 Minuten auf einer standardisierten Strecke, in einem virtuellen Fahrsimulator der AUDI AG. Aus diesen basalen Fahrdaten wurden anschließend Variablen zu Beschleunigung, Geschwindigkeit, Pedalnutzung und Lenkwinkel extrahiert. Die extrahierten Variablen wurden anschließend verwendet um das Geschlecht neuer FahrerInnen aus Fahrdaten vorherzusagen. Hierbei wurde ein regularisiertes Elastic-Net Klassifikationsmodell mit 10 $\times$ 10 facher Kreuzvalidierung auf 70% der Daten trainiert. Die restlichen Daten (30%) wurden verwendet, um die Vorhersagekraft des Modells zu testen.

Es konnte gezeigt werden, dass sich das Geschlecht von FahrerInnen, in durchschnittlich 81% neuer Fälle, erfolgreich vorhersagen lässt. Zusätzlich konnten für die Prädiktion wichtige Variablen identifiziert werden. Vor allem Variablen mit Bezug zum Beschleunigungsverhalten (Veränderung der Geschwindigkeit über die Zeit, Aktuation des Gaspedals) waren wichtig um beide Geschlechter zu trennen. Diese Studie zeigt, dass bereits basale Fahrparameter Rückschlüsse auf das Geschlecht einer Person ermöglichen. Abschließend werden Implikationen für adaptive, personalisierte Systeme diskutiert. Zwei weitere Studien untersuchen explorativ das Potential einer Smartphone App zur automatischen Aufzeichnung von Verhaltensdaten in der Psychologie.

In der ersten der beiden Studien wurde untersucht, inwiefern die kategorielle Nutzung von Smartphone Applikationen zur Validierung von selbstberichteter Persönlichkeit auf Faktoren- und Facettenniveau verwendet werden kann. Hierzu füllten insgesamt 137 TeilnehmerInnen das Big Five Struktur Inventar (Arendasy, 2009) sowie einen demographischen Fragebogen im Labor des Psychologie Departments aus. In einer anschließenden Feldphase, wurden über 60 Tage hinweg pseudonymisierte Nutzungsdaten auf den Smartphones der Personen aufgezeichnet.

Diese Nutzungsdaten wurden im Anschluss genutzt um die Häufigkeit der Verwendung von Applikationen in 14 unterschiedlichen Kategorien zu berechnen. Die Nutzungshäufigkeiten wurden anschließend als abhängige Variablen in Regressionsmodellen verwendet. Als Prädiktoren dienten die jeweils statistisch wichtigste Persönlichkeits- bzw. demographische Variable einer App-Nutzungskategorie.

Die Ergebnisse deuten darauf hin, dass insbesondere Ausprägungen in den Persönlichkeitsfaktoren Extraversion, Gewissenhaftigkeit und Verträglichkeit, sowie das Alter und Geschlecht, die Nutzung von Apps in mehreren Kategorien vorhersagen. Es zeigten sich jedoch kaum Unterschiede zwischen Persönlichkeitswerten auf Faktoren und Facetteniveau. Diese Studie zeigt wie automatisch generierte Nutzungsdaten von Smartphones potentiell zur Validierung von Selbstauskunftsfragebögen genutzt werden können. Außerdem bieten diese Ergebnisse neue Einblicke in die Manifestation von Big Five Persönlichkeitsdimensionen im alltäglichen Verhalten.

In der anderen der beiden Studien wird exemplarisch demonstriert wie sich Smartphone Nutzungsparameter potentiell zur Erkennung von Persönlichkeitsausprägungen verwenden lassen. Zusätzlich zu den Appnutzungsparametern aus der vorherigen Studie wurde aus der Vielzahl an Nutzungsparametern insgesamt 679 Prädiktorvariablen extrahiert und zur Vorhersage von Big Five Persönlichkeitsausprägungen auf Faktoren und Facettenebene verwendet. Die Prädiktoren umfassten die groben Verhaltensbereiche: Kommunikation, Mobilität, App-Nutzung, Aktivität bei Tag und Nacht, Kameranutzung, Musikkonsum und generelle Smartphonenutzung.

Zur Persönlichkeitsvorhersage wurden Faktoren und Facettenwerte in "hoch" und "nicht hoch" geteilt und in binären Klassifikationsmodellen vom Typ: Random Forest, Gradient Boosting und Elastic Net verwendet. In allen Modellen wurde ein Nested-Resampling Ansatz verwendet. Dies zielte darauf ab, Overfitting im Training des Algorithmus und Überschätzung der Vorhersagegüte zu verhindern. Die Vorhersage von Persönlichkeitsausprägungen erwies sich generell als schwierig und erreichte nur in den beiden Facetten Pflichtbewusstsein und Bedachtsamkeit, der Big Five Dimension Gewissenhaftigkeit, eine überzufällige Vorhersagegenauigkeit.

Korrelationen und Kennzahlen der Variablengüte weisen darauf hin, dass Variablen mit Bezug zur zeitlichen Varianz und Regelmäßigkeit von Events besonders prädiktiv für die beiden Facetten sind. Abschließend werden die gefundenen Ergebnisse im Zusammenhang bisheriger Studien aufgearbeitet und diskutiert.

Im vorliegenden Kontext zeigen diese Studien, dass differentielle menschliche Eigenschaften wie Demographie und Persönlichkeit mit objektiven Verhaltensdaten assoziiert sind. Mit Einschränkungen können diese Assoziationen genutzt werden, um Vorhersagen über Persönlichkeit und Verhalten zu treffen. In diesem Sinne könnte Datenlogging als mögliche Alternative zu Selbstberichten über Verhalten genutzt werden um psychometrische Tests zu validieren, kritische Verhaltensmuster vorherzusagen (z.B. depressive Episoden) und technische Systeme besser an einzelne Personen anzupassen. Zusätzlich könnten Methoden aus dem Bereich des maschinellen Lernens, robustere und praktisch anwendbarere Modelle für psychologische Fragestellungen ermöglichen.

Chapter 1

Introduction

The American Psychological Association (APA) defines psychology as the study of the mind and behavior (APA, 2016). Although aspects of the mind such as feelings, emotions, and motivations are important for psychological science, solely behavioral influences of these aspects become evident and tangible (Furr, 2009). For this reason, the investigation and understanding of behavior is often formulated as the main goal of psychologists (APA, 2016).

However, frequent research practices in the field do not exactly hold up to this definition. In fact, different researchers have repeatedly criticized the absence of real behavior as well as the ambivalent usage of the term "behavior" (Baumeister, Vohs, & Funder, 2007; Fleeson, Gallagher, Carolina, & Gallagher, 2009; Furr, 2009; Lewandowski Jr & Strohmetz, 2009; Poorthuis, Thomaes, Denissen, van Aken, & Orobio de Castro, 2014; Vazire & Mehl, 2008). For a definition of behavior, see (Furr, 2009).

In this work we investigate how psychological science can use modern sensor and network technologies collect data about actual behavior and how these data can be related to individual differences such as gender and personality (G. Miller, 2012; Yarkoni, 2012). In contrast to most literature in psychological science, this work focuses on actual behavior in contrast to self-reported measures of individual behavior.

Initially, we provide an excerpt of relevant literature with focus on the Big Five personality theory and elaborate on research methods for behavioral data in psychological science, such as self-reports and behavioral observation. Additionally, the collection of behavioral data logs from mobile devices will be described as a possible supplement to these two approaches. Additionally, we will provide a brief overview of new, promising methodological tools from the field of *Machine Learning* that could aid psychological research in the prediction of criteria (e.g., behavior).

The main contribution of this dissertation however consists of three empirical studies, focusing on associations between automatically generated traces of behavior and big five personality as well as demography. In the first study we investigate automotive driving behaviors as manifested in data logs, obtained from a virtual driving simulator. We illustrate how prediction modeling techniques, introduced later in Section 1.3, can be used to predict driver's gender from log-data, and report on the importance of the most predictive variables. The second and third study utilize smartphones as gathering tools for behavioral data and illustrate in a similar fashion how behavioral outcomes are related to personality traits and how personality factors can potentially be recognized from usage data. We conclude with a discussion about possible implications this and similar work could have on psychological science.

1.1 The Big Five Personality Theory

The overarching description of people's personality has been a persevering challenge in empirical psychology. For this purpose, many different models of personality have been proposed (John, Naumann, & Soto, 2008). Partially these models were proposing consistent dimensions of personality, partially they were lacking common ground - focusing on different aspects (John et al., 2008).

Though, since its emergence in the late 90s, the *Big Five* personality trait theory (P. T. Costa & McCrae, 1992; Goldberg, 1981) has been established as the most prominent personality theory in psychological science. The Big Five model was created with an psycholexical approach (Allport & Odbert, 1936; Norman, 1963). The basic idea behind this approach is that relevant personality phenotypes are manifested in natural language and that a words prevalence in a language use corresponds with its importance as an attribute (for a detailed overview of the model's history, see DeRaad and Boele (2000)).

The model describes people's tendencies of behavior and attitudes on five broad dimensions that hierarchically consist of several sub-facets. The broad factors describe the dimensions extraversion-introversion, emotional stability-neuroticism, agreeableness, conscientiousness, and openness or intellect. However, naming of these dimensions varies slightly across different personality questionnaires. The model has been intensively studied, replicated and used as basis for many personality questionnaires

(Arendasy, 2009; P. T. Costa & McCrae, 1992; Gosling, Rentfrow, & Swann, 2003; R. R. McCrae, Costa Jr., & Martin, 2005; Rammstedt & John, 2007).

Extraversion The dimension extraversion-introversion corresponds to an individual's outgoing tendency in the form of behavior as well as in its own experience. People with extraverted personality enjoy the interaction with others and experience more positive affect in general (Diener, Sandvik, Pavot, & Fujita, 1992; McNiel & Fleeson, 2006). Furthermore extraverts tent to be enthusiastic and assertive about activities. They also tend to get bored more easily in desolate situations and seek for external stimulation (Butt & Phillips, 2008; Damrad-Frye & Laird, 1989; H. J. Eysenck, 1967).

High levels in extraversion are generally associated with engagement in behavior (Hirsh, Deyoung, & Peterson, 2009) as well as the amount, and duration of positive emotions (Asendorpf & Neyer, 2012). This tendencies can also be related to the *behavioral activation system (BAS)* described in the reinforcement sensitivity theory (Gray & McNaughton, 2003).

Introversion is defined as missing extraversion rather than the opposite of extraversion. Therefore, people high on introversion enjoy spending time alone over spending time with others but also enjoy social situations as much as extraverts (Diener et al., 1992). However, in general they do experience less positive affect than extraverts (Lucas & Baird, 2004). In contrast to shy people, introverts do not necessarily fear social encounters.

Emotional Stability Another important personality dimension related to emotions is the emotional stability-neuroticism dimension. Unlike the extraversion-introversion dimension, it is related to the frequency and duration of negative feelings and emotions (Asendorpf & Neyer, 2012). People with high emotional stability experience less feelings of anxiety, depression. Furthermore emotional stability is associated with higher tolerance for stress, frustrations, temptations, and the mastering of social situations.

Highly neurotic people experience more feelings of this kind. However, emotionally stable people do not necessarily experience more positive emotions, as the prevalence of these is rather related to the independent extraversion-introversion dimension. Neuroticism is also associated with the restraint from behavior (Hirsh et al., 2009). This association could also be related to the *Behavioral Inhibition System (BIS)*, described by Grays biopsychological theory of personality (Gray & McNaughton, 2003). Different levels of BIS activation, describe an individual's response sensitivity to anxiety related stimuli in an given environment. Dependent on an individual's sensitivity to punishment and reward absence, BIS activation leads to the avoidance of unpleasant events. Neuroticism was associated with a higher activity of the BIS (Boksem, Tops, Kostermans, & De Cremer, 2008). However, as described by Gray and McNaughton (2003), this relationship is additionally dependent on the respective position on the extraversion-introversion dimension. Neuroticism is also the big five trait most closely related to psychopathology (Ormel et al., 2013) and instable relationships (Malouff, Thorsteinsson, Schutte, Bhullar, & Rooke, 2010; Ozer & Benet-Martínez, 2006).

Agreeableness The big five factor agreeableness describes how cooperative and socially harmonic a person is. Most five factor model questionnaires include subfacets in relation to trust, genuineness, helpfulness, modesty, and tender-mindedness. Together with the extraversion-introversion dimension, agreeableness is the big five personality factor most important for interpersonal relationships and conflicts (Jensen-Campbell & Graziano, 2001). Agreeable people generally get along better with others as they show more respect to the interests and perspectives of other people (Jensen-Campbell & Graziano, 2001).

Furthermore, they are more motivated to get along better with others and help even without motivation (Graziano, Habashi, Sheese, & Tobin, 2007). People with low scores in agreeableness are less concerned about the welfare of others and are less willing to cooperate. Very low agreeableness scores can even be associated with manipulating personality and dishonesty (Jakobwitz & Egan, 2006).

Conscientiousness As summarized by MacCann (MacCann, Duckworth, & Roberts, 2009), most studies investigating the main components of conscientiousness with the psycholexical approach, discovered three common facets. Orderliness describes a persons tendency to be thorough, careful, organized. Industriousness, a facet that describes how prepared and self-disciplined a person is in relation to the achievement of duties and work related goals. The third often discovered facet of conscientiousness describes how reliable and responsible a person is. Less consistent facets of conscientiousness include the tendency to pursue activities and goals consequently, and whether someone prefers traditional/conventional values and behaviors over alternative and new ones (MacCann et al., 2009). In general conscientious people describe themselves as efficient, organized, and rather not as easy-going and disheveled. A large collection of behaviors associated with conscientiousness was reported by Jackson et al. (2010). Furthermore, conscientiousness is the big five personality factor, most predictive for both professional and academic performance (Barrick & Mount, 1991; Poropat, 2009).

Openness openness/intellect (DeYoung, 2015), openness to experience (R. R. McCrae & Costa, Paul T., 1997) or simply openness is the big five personality dimension associated with the ability and tendency to seek, detect, comprehend, and utilize as well as appreciate complex and abstract novel information (DeYoung, 2015). People scoring high on this factor are often found in creative and artistic professions (Barrick, Mount, & Gupta, 2003). The openness, intellect or culture factor is also the big five dimension that has been subject to major debate in the literature, involving not only its name (Matthews, Deary, & Whiteman, 2009). Like the other four personality dimensions, the factor openness was statistically discovered through factor analysis.

However, researchers argued that the construct is not homogeneous enough and is separable into a factor containing the NEO-PI-R facets *Feelings*, *Aesthetics*, and *Fantasy* as well as additional one or two factors containing the other facets *Ideas*, *Actions*, and *Values* (DeYoung, 2015; DeYoung, Grazioplene, & Peterson, 2012; Jang, Livesley, Angleitner, Riemann, & Vernon, 2002; Mussel, Winter, Gelléri, & Schuler, 2011). This separation suggests that the openness construct consists of one rather stable factor resembling affinity with artistic aspects like feelings fantasy and aesthetics as well as other aspects accumulating facets of intellect.

Despite the popularity of the big five model, controversy remains regarding the factorial structure (Eysenck, Hans J., 1991), the ability of behavioral prediction (Mischel, 2004), as well as the theoretical background (Block, 2010; H. J. Eysenck, 1992). Various studies found different numbers of factors, enumerating one (Saucier, Goldberg, & Institute, 2001), two (DeYoung, 2006; Saucier et al., 2001), three (H. J. Eysenck, 1997; H. Eysenck, 2013; Saucier et al., 2001), six (Deary, 1996), seven (Saucier et al., 2001), eight (Tellegen & Waller, 2008), and 16 factors (Cattell & Mead, 2003), highlighting the disunity in the field. However, it is also important to note that not all of these factor resolutions claim to grasp an exhausting description of human personality.

One big point of criticism in relation to the big five model is its derivation which is purely based on factor analysis. Although this approach is methodologically reasonable (and we do not share this particular point of criticism), it misses an universal solution for model choice in the case of multiple models. Furthermore, the big five model has also been criticized for a lack of grounding in theory as factors were identified based on statistical relationships. In relation to the big five model, debate remains whether the factors of agreeableness and conscientiousness should be better combined to one (Aluja, Garcia, & Garcia, 2002; Eysenck, Hans J., 1991; H. J. Eysenck, 1992).

More recent research has utilized correlations between big five factors in order to extract higher order meta-factors. One group of researcher around Colin G. DeYoung suggested that the factors conscientiousness, emotional stability and agreeableness can be combined to a single factor *Stability* and the factors extraversion and openness/intellect formed a second factor - *Plasticity* (DeYoung, 2006). This higher model does not compete with the classical big five model, as with the exception of emotional stability it does not capture a large amount of big five variance (DeYoung, 2006). Still, both metafactors seem to be predictive for the engagement and restraint of behaviors (Hirsh et al., 2009).

1.2 Collection of Behavioral Data in Psychology

1.2.1 The Questionnaire Approach

In psychology, and especially in personality- and social psychology, the most frequently used approach to the collection of latent variables (e.g., Big Five Personality) is the use of standardized and normed self-report questionnaires (Baumeister et al., 2007; Paulhus & Vazire, 2007; Poorthuis et al., 2014). Dependent on the latent criteria to be measured and the type of test, sentences, short phrases or adjectives are used as items.

Self-report questionnaires offer a series of benefits for researchers. In general they are easy to administer and analyze, efficient, and offer economic advantages over other methods (Furr, 2009; Paulhus & Vazire, 2007). Furthermore, they offer the opportunity to gain insights into people's inner states, attitudes, and motivational aspects of behavior (Paulhus & Vazire, 2007). Therefore, a wide range of thoughts, feelings and behaviors can theoretically be collected within a relatively short period of time. This assumption however only holds if a respondent's answers actually correspond to their true feelings, thoughts, and behaviors and that people consciously or unconsciously provide correct information about themselves.

As this is not the case, the self-report method was also associated with a serious of caveats and problems (Paulhus & Vazire, 2007; Podsakoff, MacKenzie, & Podsakoff, 2012; Vaerenbergh & Thomas, 2013). Response styles such as the social desirable (Paulhus, 1991), acquiescent and extreme responding (Baumgartner & Steenkamp, 2001) have been shown to exert non-trivial influences on results of personality questionnaires. See Vaerenbergh and Thomas (2013) for a review.

Furthermore, these response styles can also be triggered in specific situations such as job interviews. The term *faking* refers to the deliberate action of providing answers that are expected to portray one-self in a most positive or beneficial way (Arendasy, Sommer, Herle, Schützhofer, & Inwanschitz, 2011).

Often self-report methods are also used to question people about previous behaviors. This approach can be problematic as people often simply do not remember their behaviors and seem to be bad at providing estimations about how often they engage in certain activities (Boase & Ling, 2013; Gosling, John, Craik, & Robins, 1998; Kobayashi & Boase, 2012; Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Although it is possible to brief people in order to make them more aware of their behaviors, this might also alter the investigated behaviors in the first place.

In order to tackle memory-related influences on self-reports and to collect data about situation-behavior contingencies, experience sampling methods (ESM) were invented. In ESM studies, participants are required to fill out short questionnaires, surveys or other self-reports after regular time periods or when certain conditions are met (e.g., a certain threshold of environmental noise is exceeded). This provides both finer data granularity as well as information about consistency, and variation of self-reported behavior. Current computerized approaches can also provide information about the location and the time a question was answered. Beyond that, ESM data is affected by the same limitations as classical self-reports (Furr, 2009). See Trull and Ebner-Priemer (2013) for a current review of the methodology.

Self-report questionnaires aiming at the assessment of latent psychological variables must be validated in order to test whether the obtained measures are descriptive of the investigated construct (Funder, 2012). The comparison of self-ratings with others-ratings as well as others-others comparisons provide indications for the validity of a particular self-report measure. However, there is broad consensus in the science community that the ultimate criterion for validity is the prediction of behavior and outcomes (Funder, 2012; Ozer & Benet-Martínez, 2006). If, for example self-rated personality can predict life outcomes typically associated with a particular personality trait, this provides strong support for the validity of the test.

Unfortunately, self-proclaimed validation studies frequently use self- or others- report questionnaires about past behavior instead of actually recorded behavior (Jackson et al., 2010; R. R. McCrae & Costa, 1987). However, this approach is problematic as self-report questionnaires about past behavior, are expected to exert the same potentially biasing influences on obtained measures as on the questionnaire they are being used as an validation method for. With regard to the investigation of behavioral manifestations of personality, this is troubling as no actual behavior is ever recorded at any time in the validation process (Baumeister et al., 2007; Furr, 2009; Poorthuis et al., 2014). Furthermore, some studies show that in fact large measurement errors are present in self-report measures about behavior (Boase & Ling, 2013; Kobayashi & Boase, 2012).

Nevertheless, self-report assessment of latent traits, until now, remains the well-beaten path in personality and social psychology.

1.2.2 Behavioral Observation

In addition to self-report questionnaires, the direct observation of actual individual behavior is the most obvious method for behavioral data collection. Most behavioral observation studies are conducted in a standardized or controlled environment and behaviors are video or audio recorded. Subsequently, recorded behaviors are coded by (ideally) independent raters for analysis (Furr, 2009).

The method of behavioral observation is based on the notion that personality traits are manifested in behavior and that characteristic behaviors can be consistently observed across time and situations (e.g., conscientious people are acting reliable on both Monday and Thursday, at home and at work). Furthermore, it is assumed that people with higher scores in a latent trait (e.g., conscientiousness) should exert typical behaviors more frequently than people with low scores in the trait. This aspect of personality manifestation in aggregated frequencies of relevant behavior has originally been proposed in the *Act-Frequency Approach* (AFA) Buss and Craik, 1983. While Buss himself had to rely on the aggregation of self-reports in his study, he already stated that one day the systematic monitoring of individual behavior over standard periods of time will eventually enable the analysis of manifest dispositions (like personality). However, due to its reliance on retrospective self-reports and the intention to mark behaviors as prototypical for a certain latent trait the AFA has become unpopular in psychological science (Block, 1989; Fleeson et al., 2009; Gosling et al., 1998).

Unlike self-reports, behavioral observations are not as biased by response styles (Furr, 2009) as actual behavior is observed and not reported. Furthermore, behavioral observation offers the possibility to observe behavior in real-time, greatly eliminating the memory bias of self-reports. Behavioral observation also offers the possibility to observe multiple behaviors simultaneously.

Behavioral observation studies, formerly popular in personality and social psychology (Gerard & Mathewson, 1966; Haney, Banks, & Zimbardo, 1973; Milgram & Van den Haag, 1978), are only sparsely found in psychological publications nowadays (Baumeister et al., 2007). Developmental psychology seems to be a lonely exception to this trend, due to the fact that subjects cannot be burdened with self-report questionnaires. Baumeister et al. (2007) argues that this absence of direct behavioral investigations and the embrace of self-report studies have possibly been initiated by the cognitive revolution in the 1980ies and has prevailed ever since.

Although the lack of studies investigating actual behavior has to be highlighted, there also exists a series of major difficulties associated with the method of classical behavioral observation. First of all, behavioral observation is expensive in terms of money, time, and manpower. In classical observation studies, performed in a controlled laboratory situation, usually triggering, recording, and especially behavioral coding and consequent data analysis (e.g videos) can be very demanding. Furthermore, rigorous planning and execution of such a study can take a very long time and require specially trained personnel (Furr & Funder, 2009). Another difficulty in the practice of behavioral observation lies in the standardized identification and categorization of relevant behaviors. Therefore, a suitable coding system for behaviors must be adapted or specifically developed.

Additionally, ethical considerations often make it impossible to conduct behavioral observation of relevant criteria. For example, it would be ethically problematic to investigate cheating behavior of people living in a stable relationship with regard to their personality. Furthermore, former behavioral observation studies (e.g., Milgram and Van den Haag (1978)) would not be possible today for ethical considerations. Beyond that, it is often not possible to observe peoples behavior without them knowing. Therefore, the act of observation itself might alter the observed behaviors by inducing self-presentation effects (e.g., people would probably not show cheating behavior when being observed).

As behavioral observation studies are mostly conducted in controlled lab settings, the generalizability and ecological validity of the obtained results is to be questioned. To sum up, these difficulties make behavioral observation studies often simply unfeasible and do not conform well with currently common research practices of frequent publication. However, both, self-reports and behavioral observation bear methodological difficulties as well as distinct advantages. Therefore, if a most accurate assessment of latent traits is desired, a combination of self and others ratings as well as behavioral observation is desired.

This work shows possibilities of how large amounts of behavior-related data can be gathered and can be related to criteria such as demography and personality. We do not advance the view that self-report measures are not an important and valuable part of research in social sciences. However, we pledge for an increased incorporation of new behavioral measures to complement data obtained via self-reports.

1.2.3 Data Logging

Social science researchers have been using mobile electronic devices for about 20 years (Mehl, Pennebaker, Crow, Dabbs, & Price, 2001; G. Miller, 2012) for data collection purposes. Personal digital assistants (PDAs) as well as electronically activated recorders (EAR) have been utilized in conversation analysis (Mehl & Pennebaker, 2003), experience sampling (Hektner, Schmidt, & Csikszentmihalyi, 2007) and diary studies (Bolger, Davis, & Rafaeli, 2003). However, as these devices are expensive and require special programming, the collective, automatized recording of large data samples as well as the related analysis of data has remained a challenge for many researchers in social sciences.

However, rapid developments in digital technology, such as the rapid miniaturization of electronics (Moore, 2006), the price inflation of electronics, as well as their capabilities in terms of available sensors, processing power, and connectivity could make the collection of research data has much easier (G. Miller, 2012; Yarkoni, 2012). Furthermore, the availability of extremely capable consumer electronics makes it unpractical to use expensive, inconvenient, and specially programmed devices for data collection. Mobile phones for example have rapidly developed from normal phones to phones with additional features to extremely capable mobile computers with the option to place calls. Therefore, the alternative to the use of specifically programmed devices for data collection in small samples is the use of peoples private devices.

Furthermore, modern operating systems incorporate large numbers of *Application Programming Interfaces* (APIs). APIs allow developers to access hardware in a standardized way, making sensors accessible to developers (Google, 2016b). These are capable of grasping a wide range of changes in environmental parameters and can be used to

Figure 1.1: Smartphone Log Data

Tabelle	Timestamp	Wochentag	Datum	Uhrzeit	.ocation_Latitude	ocation_Longitud	Typ	Name	PackageName	Category
Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
SCREEN	1424879265920	Wed	25.02.2015	16:47:45	48.1553399	11.5824795	OFF_LOCKED	NULL	NULL	NULL
SCREEN	1424881569637	Wed	25.02.2015	17:26:09	48.1787544	11.5894373	ON_LOCKED	NULL	NULL	NULL
SCREEN	1424881572190	Wed	25.02.2015	17:26:12	48.1787544	11.5894373	ON_UNLOCKED	NULL	NULL	NULL
APP	1424881575865	Wed	25.02.2015	17:26:15	48.1787544	11.5894373	NULL	Facebook	com.facebook...	Social
APP	1424881584060	Wed	25.02.2015	17:26:24	48.1787544	11.5894373	NULL	TouchWiz-Start	com.sec.andr...	-
WIFI	1424881587948	Wed	25.02.2015	17:26:27	48.1787544	11.5894373	connected	NULL	NULL	NULL
APP	1424881601552	Wed	25.02.2015	17:26:41	48.1787544	11.5894373	NULL	DB Navigator	de.hafas.andr...	Transportation
APP	1424881623297	Wed	25.02.2015	17:27:03	48.1787544	11.5894373	NULL	TouchWiz-Start	com.sec.andr...	-
SCREEN	1424881632540	Wed	25.02.2015	17:27:12	48.1787544	11.5894373	OFF_LOCKED	NULL	NULL	NULL
WIFI	1424881641265	Wed	25.02.2015	17:27:21	48.1787544	11.5894373	disconnected	NULL	NULL	NULL
SCREEN	1424881732997	Wed	25.02.2015	17:28:52	48.1787544	11.5894373	ON_LOCKED	NULL	NULL	NULL
SCREEN	1424881733952	Wed	25.02.2015	17:28:53	48.1787544	11.5894373	ON_UNLOCKED	NULL	NULL	NULL
APP	1424881737844	Wed	25.02.2015	17:28:57	48.1787544	11.5894373	NULL	MVG Fahrinfo	de.swm.mvgf...	Transportation
APP	1424881756284	Wed	25.02.2015	17:29:16	48.1789485	11.5894438	NULL	TouchWiz-Start	com.sec.andr...	-
SCREEN	1424881760719	Wed	25.02.2015	17:29:20	48.1789485	11.5894438	OFF_LOCKED	NULL	NULL	NULL
SCREEN	1424882028896	Wed	25.02.2015	17:33:48	48.1789485	11.5894438	ON_LOCKED	NULL	NULL	NULL
SCREEN	1424882030209	Wed	25.02.2015	17:33:50	48.1789485	11.5894438	ON_UNLOCKED	NULL	NULL	NULL
APP	1424882033841	Wed	25.02.2015	17:33:53	48.1789485	11.5894438	NULL	WhatsApp	com.whatsapp	Communication
APP	1424882127840	Wed	25.02.2015	17:35:27	48.1789485	11.5894438	NULL	TouchWiz-Start	com.sec.andr...	-
SCREEN	1424882128523	Wed	25.02.2015	17:35:28	48.1789485	11.5894438	OFF_LOCKED	NULL	NULL	NULL
SCREEN	1424883104279	Wed	25.02.2015	17:51:44	48.1789485	11.5894438	ON_LOCKED	NULL	NULL	NULL
SCREEN	1424883106360	Wed	25.02.2015	17:51:46	48.1789485	11.5894438	ON_UNLOCKED	NULL	NULL	NULL
APP	1424883107846	Wed	25.02.2015	17:51:47	48.1789485	11.5894438	NULL	WhatsApp	com.whatsapp	Communication
APP	1424883111840	Wed	25.02.2015	17:51:51	48.1789485	11.5894438	NULL	TouchWiz-Start	com.sec.andr...	-
SCREEN	1424883113573	Wed	25.02.2015	17:51:53	48.1789485	11.5894438	OFF_LOCKED	NULL	NULL	NULL
SCREEN	1424884003347	Wed	25.02.2015	18:06:43	48.1789485	11.5894438	ON_LOCKED	NULL	NULL	NULL
SCREEN	1424884013264	Wed	25.02.2015	18:06:53	48.1789485	11.5894438	OFF_LOCKED	NULL	NULL	NULL
SCREEN	1424884070895	Wed	25.02.2015	18:07:50	48.1789485	11.5894438	ON_LOCKED	NULL	NULL	NULL

Figure 1.1: Smartphone Log Data obtained with the Android logging app used in Sections 2.2 and 2.3. Events of phone usage are visible with GPS location and timestamp.

create timestamped event data (logs). Those can in return be used to calculate variables that provide information about an individual's behavior along time and locations. See Figure 1.1 for an example.

Outgoing calls on a mobile can for example be aggregated and correlated with extraversion scores (Montag et al., 2014). The average time of the first log event per day provides an approximation of when a person gets up in the morning. This information could then be used to predict conscientiousness in a new sample. We will elaborate on that in Section 2.3.

Furthermore, these developments make it possible to conduct studies with much larger sample sizes using off-the-shelf consumer technology at low cost. Only the development of a mobile application is necessary in order to retrieve information from and to send content to a personal smartphone. Furthermore, this approach theoretically allows for worldwide, unobtrusive data collection in an ecologically valid form, at little cost in personnel and money.

1.3 Predictive Modeling

In this section, we briefly describe the in psychology not yet commonly used *Predictive Modeling* techniques. The concepts introduced here are helpful in order to better understand data analysis in Section 2. Although the terms *Machine Learning*, *Statistical Learning* and *Predicitve modeling* are used interchangeably, we will use the term *predictive modeling* throughout this chapter for the sake of consistency. As this chapter represents only a very brief introduction, the interested reader is advised to consult (James, Witten, Hastie, & Tibshirani, 2013; Kuhn & Johnson, 2013), or even (Aggarwal, 2015; Hastie, Trevor and Tibshirani, Robert and Friedman, Jerome, 2009) for more detailed information. In general the collected information in this chapter consists of extracts from two introductory books about predictive modeling (James et al., 2013; Kuhn & Johnson, 2013).

1.3.1 Prediction and Inference

Statistical modeling generally follows two main motivations: the gain of information (inference) and the prediction of outcomes (Breiman, 2001). In both cases the association of the vector of input variables $\mathbf{X}$ (independent variables) with the vector of output variables $\mathbf{Y}$ (dependent variables) is investigated. As Breiman describes in his famous article (Breiman, 2001), two different approaches exist in the field of statistical modeling.

On the one hand, the classical data modeling culture in which the relationship between $\mathbf{X}$ and $\mathbf{Y}$ is assumed expressible as a stochastic model (e.g., linear regression model). On the other hand, algorithmic modeling culture, where analyses are not always based on specific distributional assumptions (e.g., gaussian normal distribution). Algorithmic modeling only assumes that the sample is taken from some sort of unknown multivariate distribution and that the real relationship between $\mathbf{X}$ and $\mathbf{Y}$ is assumed to be complex and unknown (Breiman, 2001). Hence, algorithmic modeling culture tries to find a function $f(\mathbf{X})$ that uses $\mathbf{X}$ in order to predict the outcome variables $\mathbf{Y}$.

In psychology the most common motivation for data analysis so far has been statistical inference, for example identification of behavioral underpinnings in the form of human understandable models (e.g., if $\mathbf{X}$ increases $\mathbf{Y}$ increases as well). Alternatively, data can be analyzed in order to achieve a maximum of predictive performance with regard to the criterion variable (e.g., predicting **Y** from **X** using a function $f(\mathbf{X})$). Although both approaches are mutually valuable, usually there exists a tradeoff. Explainable models (e.g., linear regression) often do not represent the reality of the **X** and **Y** relationship, and highly predictive models (e.g., random forest) are often not intuitively understandable to humans.

Predictive modeling deals with the prediction of binary, categorical, or continuous outcomes. Predictive models with binary (e.g., gender) and categorical (e.g., level of education) outcome measures are referred to as classification tasks, models with continuous outcomes (e.g., salary) are referred to as regression tasks.

Furthermore, predictive models can be divided in supervised and unsupervised learning tasks. In a supervised learning task for each instance of x_i, $i = 1, ..., n$, there is an associated response y_i, whereas in unsupervised learning tasks (e.g., cluster analysis) no information about y_i responses is provided. In this section we will focus on supervised learning methods. Predictive modeling mostly follows a relatively fixed sequence of analytical steps beginning with the pre-processing of the data, training of an algorithm and concluding with the evaluation of predictive performance. We will elaborate on these steps in the upcoming sections.

1.3.2 Pre-processing

At the beginning of most data analysis endeavors, the data has to be first pre-processed, so algorithms can be trained on it. This process typically involves several steps. The order of these steps is not fixed and depends on the type of data and the research intentions. Often, data is not in the right format, has missing values or is provided in the form of continuous timestamp data, data logs, text, images and so on.

Data transformation can be useful in order to remove skewness, better describe variance in the data or handle outliers. Centering and scaling of a variable (commonly known as z-transformation) induces a mean of zero and a standard deviation of one to the respective variable[1]. Some models (e.g., regularization techniques, LASSO, ridge regression, elastic net) require predictors to be on the same scale (Friedman, Hastie, & Tibshirani, 2010). Data transformations can also be useful in order to remove significant skewness from variables (e.g., violation of the normal distribution assumption). In order to achieve this, data can be replaced by its log, square root, or inverse. Alternatively, power transformations can be used to increase normality to a given variable.

[1] only if the standard deviation is used for scaling

The Box-Cox transformation is a power transformation family that can help to induce normality to a given variable (Box & Cox, 1964). In a similar fashion, the Yeo-Johnson (Yeo & Johnson, 2000) transformation can also be applied on negative values of **X**. Please also note that although the transformation of variables can be useful in many instances, legitimate criticism has been expressed about unreflected practice of data transformation (O'Hara & Kotze, 2010; Tabachnick & Fidell, 2012). Furthermore, data transformations can also complicate interpretation of single values as the transformed scale of variables does no longer correspond to the original units.

Outliers are defined as data points that are exceptionally different from the mainstream of the remaining values of a given variable. They often induce problems in models (especially with non-robust linear models) and can distort associations between variables. However, care has to be taken not to hastily remove those in small samples (often the case in psychological science) as they might indicate parts of not-yet sampled subpopulations or tails of a not yet visibly skewed distribution. A good approach to this problem is the use of robust statistical methods (Kafadar, 2003) such as the robust variance or the robust mean which are superior to their parametric counterparts in almost all cases (Erceg-Hurn & Mirosevich, 2008). Winsorizing or trimming is one concrete approach to the handling of univariate outliers (Erceg-Hurn & Mirosevich, 2008). In trimming, values that are more extreme than a specified cutoff (e.g., lowest and highest 10% of values or z-transformed values greater than 3), are replaced with the maximum or minimum of the remaining data points.

Another approach to outlier handling, is the spatial sign transformation that projects all cases on a multidimensional sphere with equal distance to the center (Serneels, De Nolf, & Van Espen, 2006). However, as this procedure goes beyond the scope of this chapter we will not elaborate on it. For practices of outlier handling in high dimensional data sets, see (Aggarwal, 2015).

Another problem that is encountered in almost all data sets, at varying degree, are missing values. Initially, it is of interest to understand why data is missing. For example, values could not have been recored in the first place or be related to the criteria (e.g., missing GPS values in a phone-logging study could be related to the personality of the user, a variable one might want to predict based on GPS data). Often missing values are concentrated in single predictors and often this variable can be excluded as a whole. The removal of single cases or even whole variables is not problematic in data sets with many cases and predictors, however can be costly in small samples (as often prevalent in social sciences). As an alternative to the removal of cases with missing values, empty data entries can be imputed. Therefore, the median, the expectation-maximization value, or similar measures can be imputed in order to avoid loss of data (Kuhn & Johnson, 2013; McLachlan & Krishnan, 1997).

In order to train a predictive model on the data, variables (term often interchangeably used with features) often have to be extracted or computed from the raw data set. In text mining for example (Yarkoni, 2010), frequencies of certain words or word categories (e.g., nouns) could be extracted from a text. In natural language processing (Mairesse, Walker, Mehl, & Moore, 2007) linguistic features such as pitch, loudness, word use, speed could be extracted as numerical representations.

In the case of a very large number of initial variables, clusters and components of commonly shared variance can be extracted from a large number of variables and again used in the model. To achieve this, a principal component analysis (PCA), partial least squares (PLS) or cluster analysis can be performed beside others. Once features have been extracted, the data set ideally exists in the form of so called *tidy data*. Each row now represents an observation and each column represents a variable or feature (Wickham, 2014).

In addition to the creation of predictors, uninformative predictors eventually have to be removed from the data set. Highly intercorrelated predictors (*collinearity*, *multicollinearity*) or variables with little or no variation in the containing values are generally referred to as uninformative predictors, as they do not add much new information (variance) to the data. Intercorrelated predictors, share common variance and can especially cause problems with linear regression models.

The presence of collinearity can be identified by calculation of the variance inflation factor (VIF) (Fox & Monette, 1992), however many modern predictive modeling software offer special algorithms for the removal of uninformative predictors (Bischl et al., 2016; Kuhn, 2015). Variables with little or no variance are expected to increase a models complexity and to cause problems during resampling, please see Section 1.3.4. In general, the removal of uninformative predictors often improves fit and or stability of prediction models.

1.3.3 Performance Evaluation & Overfitting

Prediction models are mostly categorized as either classification or regression problems. Dependent on the type of problem, different measures exist in order to evaluate the accuracy of models. In both cases these measures somehow express towards which degree the predicted outcome values $\hat{y}$ differ from the actual outcome values y.

Regression

For regression problems, the most often used accuracy measure is the *mean squared error* (MSE).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2 \tag{1.1}$$

The MSE will be small when the actual values y are very close to the predicted values $\hat{y}$ and very large if predictions and actual outcomes differ significantly. Equation 1.1 shows that the MSE becomes large when the sum of squared differences between predicted and true score become larger. The MSE in Equation 1.1 is calculated on the training data (the part of the dataset an algorithm is fitted on, details in Section 1.3.4) and can be misguiding as an indicator for how well the model will predict new samples. The reason for this is that modern data analysis algorithms can be very flexible and perfectly fit a model to a given dataset with minimal MSE. Problematically, very flexible models often show worse prediction accuracy on new samples in comparison with more general models. This effect is caused as a flexible model tries to catch all variation in the data, including both, variations caused by the true relationship, as well as unsystematic variation. As only the systematic variation in the data provides information of the true relationship between the predictor variables and the criterion, the modeling of noise causes false predictions on new samples. In this case MSE in the training and test set deviates greatly. This effect is also called overfitting. Therefore, the MSE obtained from new, independent test data is important in order to evaluate how well a model generalizes. The test error can be estimated in several ways in order to obtain a more reliable measure for how well an particular algorithm will extrapolate on new data. Some of these approaches will be introduced in Section 1.3.4.

Table 1.1: Confusion-Matrix

	Response	
Truth	high	nothigh
high	292 (TP)	178 (FP)
nothigh	246 (FN)	654 (TN)

Note: Confusion-matrix for the binary prediction task of personality scores from Section 2.3; the high cell numbers were induced by an artificial upsampling procedure.

Classification

In classification problems the outcome variable is categorical (multinomial) or binary (binomial). Therefore, the MSE is not suitable for model performance evaluation. A very common way to describe performance of a classification model is a simple *confusion matrix*. The confusion matrix is a cross-table showing correspondence of real and predicted class labels. The diagonal numbers represent correct classifications, the off-diagonal cells contain misclassified cases. In Table 1.1 a confusion matrix of a binary classification task is presented. In this example, cases are labeled as either *"high"* or *"not high"*. These labels refer to binned personality scores of the conscientiousness facet sense of duty, predicted with a gradient boosting classification algorithm, tuned and cross-validated with a nested resampling approach are visible. For details about this classification task, see 2.3.

$$Accuracy = \frac{\textit{number of correct classifications}}{\textit{number of all classifications}} \tag{1.2}$$

In addition to a confusion matrix, other performance measures can be computed. The most basic measure is the accuracy rate (ACC). Considering the confusion matrix in Table 1.1 we can calculate an accuracy score of 0.69 by summing up the diagonal scores ($292 + 654 = 946$) and dividing them by the total number of cases ($946/1370 = 0.69$). Please note that the number of cases (1370) in this example has been increased artificially by tenfold in order to enable better training of the algorithm.

Despite its intuitiveness, this measure is problematic for several reasons. Accuracy scores in unbalanced datasets (e.g., more cases labeled as "not high") are not really meaningful. In Table 1.1, the *"high"* class (N_{high} = 470) has more cases than the *"not high"* class ($N_{nothigh}$ = 900). Simple classification of all cases to the bigger class would result in an accuracy score of 0.66, without any predictive value. The percentage of the most prevalent class is sometimes referred to *"No-information rate"* (NIR) or classification baseline. For a binary classification task, this score is 0.5 with equal class sizes, but can be considerably higher for unbalanced criteria. Therefore, a classifier with an accuracy score above the NIR can be considered as reasonable when accurate prediction of all classes is equally desirable.

The accuracy score is also problematic if the prediction of one class is more important than the prediction of the other. For example, in cancer screening the consequences can be much more fatal when a cancer case is missed than if a patient is falsely labeled as cancerous and assigned to further examinations. In this case, the *Sensitivity* and *Specificity* measures as well as the *Positive Predictive Value* and the *Negative Predictive Value* can be more informative about the desired performance of a classifier. These measures take into account the prediction accuracy of the specific classes. In general true-positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN) classifications are distinguished (see also Table 1.1). Using this differentiation it is possible to calculate sensitivity, specificity, true positive and true negative predictive values of a classifier. Sensitivity measures the proportion of positive cases that have been classified correctly in the test set (people high in sense of duty and classified as such). Specificity measures the proportion of correctly classified negative cases. Therefore, sensitivity and specificity can be used to calculate a *balanced accuracy* score. The true positive predictive value measures the ratio of TPs in the total number of cases, classified as positive. The true negative predictive value describes the ratio of TNs in the total number of cases, classified as negative.

$$Sensitivity = \frac{TP}{TP + FN} \tag{1.3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{1.4}$$

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2} \tag{1.5}$$

$$\textit{True Positive Predictive Value} = \frac{TP}{TP + FP} \tag{1.6}$$

$$\textit{True Negative Predictive Value} = \frac{TN}{TN + FN} \tag{1.7}$$

Furthermore, is it possible to also quantify predictions in the form of class probabilities. Although generally a case with a class probability of 0.52 as well as another case with a probability for the same class of 0.98 will both be classified in the same category, the first case is classified with less confidence. Prediction probabilities are usually especially interesting in applications where not only the definite classification (e.g., 1 or 0 in a binary task) but also the confidence of a decision is of importance, or wrong decisions are very costly.

Often, it is difficult to train a classifier with both, high sensitivity and specificity. However, a good way to illustrate both errors (and their relationship) is to plot them against each another in a receiver operator characteristic curve (ROC), while varying across hyperparameter settings. In Figure 1.2 the ROC curve of the gender classification from Section 2.1 is visible. In this particular ROC curve, the threshold parameter (of class probabilities) was tuned and the respective pairs of the true and false positive rate are plotted.

A related measure is the area under the curve (AUC). AUC takes values between 0 and 1, with higher values being better. In a balanced binary classification task, a classifier with random performance will have a AUC around 0.5. Therefore, a classifiers performance can be considered as important if it scores well above 0.5.

Figure 1.2: Receiver Operator Characteristic Curve

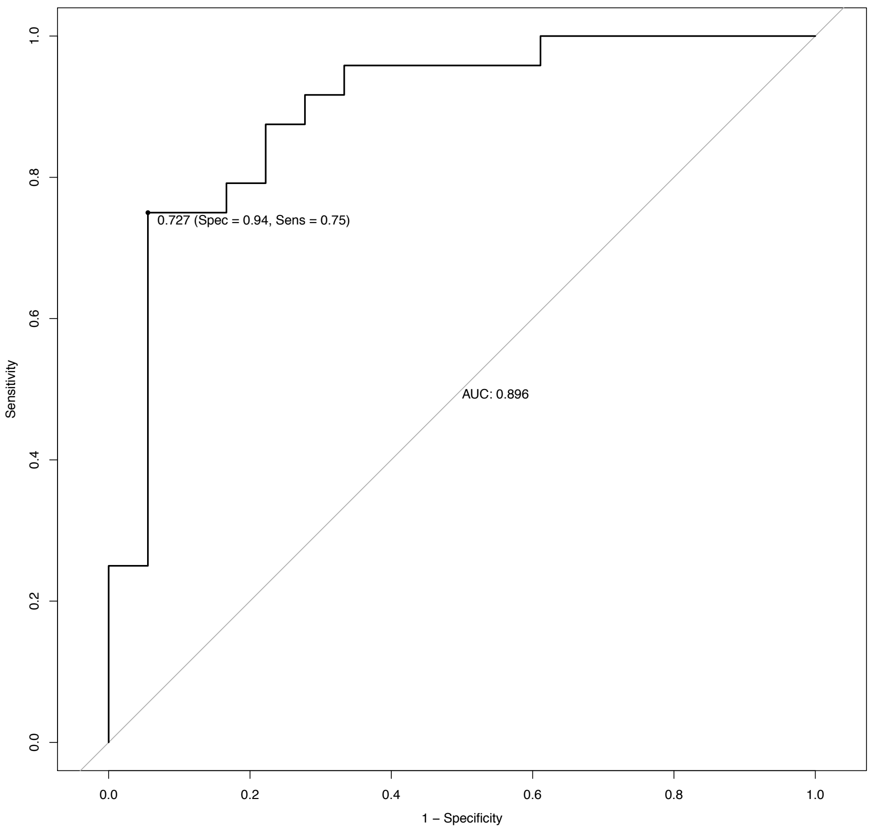

Figure 1.2: ROC curve of the gender classifier in Section 2.1.

The Bias-Variance Tradeoff

The relationship between the MSE in test and training set in relation to model flexibility is the result of the *Bias-Variance Trade-Off*. Precisely the error (MSE) of a model can be separated into three components (James et al., 2013).

$$E(y_i - \hat{f}(x_i))^2 = Var(\hat{f}(x_i)) + [Bias(\hat{f}(x_i))]^2 + Var(\epsilon) \tag{1.8}$$

In Equation 1.8, the term $E(y_i - \hat{f}(x_i))^2$ represents the expected MSE of a method. The part $Var(\hat{f}(x_i))$ refers to the variance of method - how flexible a model is in following single data points. In other words, the variance of a method refers to the degree a model would change when applied to a new set of data. Flexible models are usually high in variance, therefore capable to describe complex non-linear relationships. However, they are also likely to overfit.

Contrarily, simple models (e.g., linear regression) are rather unflexible and are likely to underfit the data as they cannot catch the real relationships (bias). This part of the error is described by the second term $[Bias(\hat{f}(x_i))]^2$ in Equation 1.8. The squared bias of an method refers to how unable a model is to capture the true relationship of **x** and **y**. Finally, the last part $Var(\epsilon)$ refers to the variance of the irreducible error terms.

Therefore, more flexible models have higher variance and less flexible models have more bias. Finding a model both low in bias and variance constitutes the main goal of prediction modeling and will lead to high prediction accuracy in new samples. This relationship is referred to as the *Variance-Bias tradeoff*.

1.3.4 Data Splitting & Resampling

Once the data is pre-processed and available in the right format, the most characteristic part of the supervised learning method starts. A suitable algorithm has to be identified and trained on a set of data in order to make predictions on a new dataset. For example in Section 2.1 we trained an *Elastic Net* algorithm to recognize the drivers gender from basic automotive driving parameters such as the maximum speed or the average acceleration. In order to estimate how well this algorithm will extrapolate on a new population of drivers (predict the gender of new drivers), we had to separate the model fitting process from the prediction process. At this point it cannot be overstated that the fitting of a function must happen on one set of data and the performance evaluation on another. This is important because as mentioned in Section 1.3.4, it is possible to theoretically train an algorithm that produces zero prediction error if training and testing is performed only on one set of data. However, this will lead to massive overfitting.

Generally, it is of importance to provide enough data on which an algorithm can be trained on but also to ensure that the used test-set is large enough in order to give an realistic estimation of predictive performance on new samples. Since Mosier (1951), at the latest, the psychological science community is aware that models that have been fitted on a particular set of data cannot be predicted again on the very same set of data. Fitted models in fact have to be cross-validated on a completely new set of data in order to draw realistic conclusions about how well a particular model will generalize to the population.

In an ideal case our sample consists of many subjects N with only a small number of predictor variables $\mathbf{X}$ that are systematically related to our outcome variables $\mathbf{Y}$. In this case we could split our data in three parts: a training set, a testing set, and a validation set. Furthermore, we could then train algorithms on the training set, obtain estimations about the predictive ability through prediction on the test set and eventually test again on the validation set after all modeling is completed.

However, in many cases (and especially in psychological studies) samples are often not very large and one cannot afford the luxury of single data set splits. Furthermore, there exists mounting evidence that single training-test-set splits with small samples are not necessarily favorable as performance will vary greatly (due to the relatively high probability of a single extreme case to be in either test or training set) (Bischl, Mersmann, Trautmann, & Weihs, 2012; Molinaro, Simon, & Pfeiffer, 2005). Therefore, the use of more sophisticated re-sampling techniques is advisable in order to make use of the available data, in a most effective way and to get a more realistic picture about the performance of a particular model.

Resampling techniques work similar to single training-test-set splits, but repeat this procedure many times in order to train the algorithm on different subsets of the data (recycling). The general idea remains the same: train on a subset of data, predict on another, and aggregate performance estimations across all iterations. Several established resampling techniques have been invented (Bischl et al., 2012) with differences in how the dataset is split and which subsamples of the data are selected. In this section we

Figure 1.3: 3-Fold Cross-Validation.

will introduce *(Repeated) K-Fold Cross-Validation*, *Bootstrapping*, *Subsampling and Nested Resampling*, for more techniques, please refer to (Bischl et al., 2012; Kuhn & Johnson, 2013).

Classical k-fold *Cross-Validation* (CV) refers to the procedure of randomly splitting a given dataset into k folds of roughly equal size, while using $k-1$ folds for training and the remaining fold for testing (prediction). This procedure is than repeated with all other folds being the test set once (see Figure 1.3). Consequently, the k estimates of performance are then summarized.

In the case of unbalanced samples (e.g., less females than males) it is helpful to use *stratified sampling* when performing classification tasks. This ensures that roughly the same ratio of cases with the respective attribute (e.g., male and female), with regard to the original data set, is present in each of the k folds. In repeated cross-validation the k random splits are performed several times and performance is aggregated across all iterations. In general it is usually necessary to find a trade-off between computational efficiency and bias reduction (difference between estimated and true predictive performance of an model). In that sense, larger numbers of k lead to an continuous decrease in bias and an simultaneous increases of computational efforts.

In the extreme case of *leave-one-out cross-validation* (LOOCV) the number of k is equal to the number of cases in the data set. This approach is usually computationally burdensome (Bischl et al., 2012; Kuhn & Johnson, 2013), but can be performed efficiently in some cases (Bischl et al., 2012). As performance between LOOCV and repeated CV is comparable (Molinaro et al., 2005), the latter should be preferred from the perspective of computational efficiency.

Another well-known method for resampling is *the Bootstrap*. Bootstrapping refers

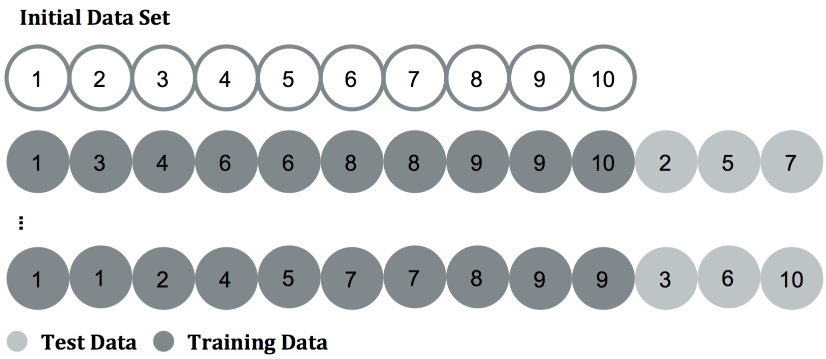

Figure 1.4: Bootstrap resampling.

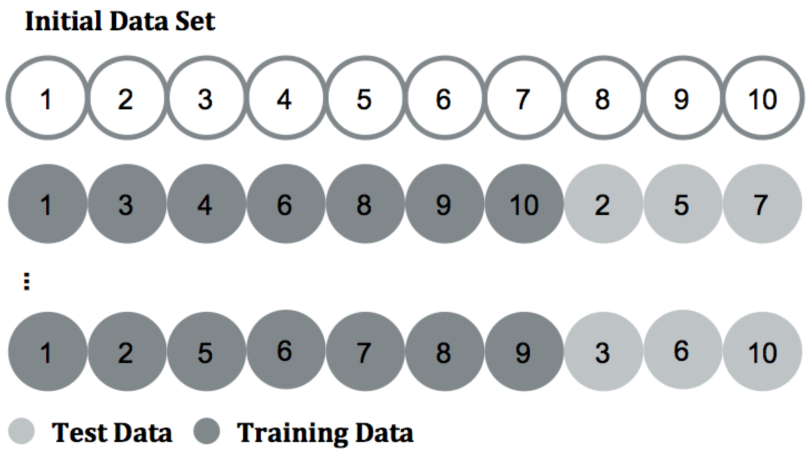

Figure 1.5: Subsampling.

to an equally distributed drawing of a sample from a data set with replacement (Efron & Tibshirani, 1986), see Figure 1.4. Although, similar to CV, boostrap resampling uses much higher numbers of k (e.g., 500) and the training set is of equal size as the complete data set. Most notably, samples can be represented multiple times in the training and test sets. As this can lead to overfitting in small samples, alternative approaches have been proposed but will not be further discussed here (see Efron and Tibshirani (1997) for details).

Subsampling or *Monte-Carlo-Cross-Validation* is very similar to the Boostrap method, however samples are drawn without replacement, see Figure 1.5. As the bootstrap method has been shown to be problematic with repeated measures, subsampling should be preferred (see Bischl et al. (2012) for a discussion).

Figure 1.6: Nested Resampling; hyperparamter tuning is performed in the respective inner resampling iteration (4-fold CV), tuned paramters used for model evaluation in the respectiv outer iteration (3-fold CV), mean predictive performance is calculated across all outer resampling iterations.

In addition, unbiased estimation of prediction performance, resampling techniques should also be used for the selection of important variables the tuning of model parameters and even the selection of suitable models. In these cases *Nested Resampling* designs are necessary (Bischl et al., 2012). This is essential as for the determination of optimal hyperparameter settings (see Section 1.3.4) and variable selection, as these can mostly not be chosen without looking at the available data. When doing so, it is important to keep test and training data separated. The basic idea behind nested resampling is to perform parameter tuning, model selection, and variable selection in an inner resampling loop while using an outer resampling loop for model evaluation. In Figure 1.6 such a nested resampling design is illustrated. Using this approach, the respective training and test parts of the data set remain separated. The design consists of an outer loop with 3-fold cross-validation and an inner loop with 4-fold cross-validation.

Feature Selection

An important part of predictive modeling that should be performed within resampling is the selection of important variables for prediction of the criteria. In many cases, many predictor variables are available in a given data set. Not all of these variables $\mathbf{X}$ might be effectively related to the outcome variables $\mathbf{Y}$. Therefore, a valuable subset of the available predictors must be chosen.

This challenge is continuously gaining importance as the availability of high dimensional data sets is increasing rapidly. Feature selection is not only necessary in order to make models less complex and more intuitively understandable, it is also required by certain models (such as ordinary least squares regression) to have less predictor variables than cases. Furthermore, predictors with no informative value can affect model performance negatively. Some models (such as the Elastic Net or tree based methods) overcome this problem due to coefficient regularization and integrated feature selection.

In addition to models with integrated feature selection (e.g., Least Absolute Shrinkage and Selection Operator (LASSO), Random Forest), separate procedures for feature selection such as wrappers (forward, backward stepwise) or filters but also more sophisticated techniques, such as genetic algorithms or simulated annealing can be applied. See Chapter 19 in Kuhn and Johnson (2013) for a summary.

Model Tuning

In addition to the model fitting process during training of an algorithm, optimal tuning parameters of models (hyperparameter) can be adjusted in order to optimize learning efforts with respect to a criterion (e.g., maximize accuracy). These parameters define the complexity of models and influence performance as well as can mostly not be calculated with a simple formula and must be determined through resampling.

In penalized linear models (e.g., the LASSO) the shrinkage parameter λ can be tuned in order to minimize the MSE during training. In tree-based ensemble methods (Breiman, Friedman, Olshen, & Stone, 1984) (e.g., Random Forest, Gradient Boosting) hyperparameters such as the number of trees, and the amount of variables considered at each split (m_{try}) can be tuned. As stated before in Section 1.3.4 this procedure should ideally happen during resampling.

Model Selection

In algorithmic modeling culture it is a common approach, to search for an algorithm that predicts the outcome variables $\mathbf{Y}$ using the predictor variables $\mathbf{X}$ with a maximum of accuracy (Breiman, 2001). It is not uncommon to compare the performance of complex - *Black-Box* models (e.g., Random Forests, Neural Nets, Support Vector Machines) with simpler models (e.g., logistic regression or linear regression) and investigate whether they produce comparable results. If the same prediction accuracy is achieved it mostly makes sense to prefer a simpler models for the sake of interpretability. Modern statistical software such as the *mlr* R-package (Bischl et al., 2016) provide convenience functions to achieve this.

Chapter 2

Empirical Studies

This chapter reports on three studies investigating associations of automatically generated logs of behavior and individual differences. The first study shows how standard driving parameters from automotive vehicles systematically vary with respect to gender. Furthermore, we demonstrate how basal driving parameters can be used to predict gender with high accuracy with a machine learning approach.

The second study focuses on the use of mobile applications on smartphones and describes how big five personality facets are predictive of app-usage frequencies in several categories. The last study investigates how behavior-related features can be identified from smartphone log data and modeled in a statistical learning setting in order to recognize individual personality scores with a machine learning approach. Study 2 and 3 studies are part of a larger research project, conducted at LMU between September 2014 and August 2015, initiated by myself.

2.1 Study 1: Gender Recognition from Automotive Driving Data

2.1.1 Abstract

The recognition and utilization of user-specific information is of increasing importance in relation to modern recommender systems and adaptive user interfaces. Associated with this trend is the increased need for privacy protecting measures in personalized systems. This work demonstrates the possibility to recognize user-gender from automotive driving data with high accuracy in an identity protecting manner. The analysis

shows that variables in relation to acceleration, gas pedal actuation as well as situation dependent driving speed are especially informative about driver gender. The results and implications are discussed in relation to possible applications in adaptive user interfaces and personalized systems. The following study corresponds to an enriched version of the initially submitted yet published paper *"Show Me How You Drive and I,ll Tell You Who You Are Recognizing Gender Using Automotive Driving Parameters"* (Stachl & Bühner, 2015).

2.1.2 Introduction

The capability to distinguish between both genders is an important ability in order to interpret gender-sensitive social information and develops at an age of approximately four (Martin & Halverson, 1981). Humans utilize a series of cues to identify other peoples gender. Amongst features such as clothing and voice, humans recognize gender from visual features like the face or body structures (Bruce et al., 1993).

Furthermore, these and other features have been intensively studied in order to train statistical classifiers for automatic gender recognition (Abdollahi, Valavi, & Ahmadi Noubari., 2009; Bekios-Calfa, Buenaposada, & Baumela, 2014; Cao, Dikmen, Fu, & Huang, 2008; Hadid & Pietikäinen, 2009). Besides the characteristics described above, people also infer other's gender through observation of natural behavior for which gender differences have been reported in various areas such as risk taking (Byrnes, Miller, & Schafer, 1999), aggression (Knight, Guthrie, Page, & Fabes, 2002), and most frequently in spatial abilities (Coluccia & Louse, 2004). Gender differences in behavior can be partially explained by biological as well as evolutionary and socio-cultural factors. However, the missing consent concerning this topic is reflected in the still ongoing nature-nurture debate (Eagly & Wood, 2013).

Analysis of user behavior for statistical recognition of demographics as well as psychometrics has recently gained popularity, especially with regard to computer and web technology. This development is directly related to great advances in mobile computing technology and human computer interaction. Modern ubiquitous web and sensor technology exists in many every-day objects and makes it possible to unobtrusively collect large amounts of behavioral data. Some researchers even refer to this new approach as Psychoinformatics (Yarkoni, 2012) or Computational Social Science (Cioffi-Revilla, 2010).

In a previous study, Hu, Zeng, Li, Niu, and Chen (2007) used web browsing data to predict gender and age. Others utilized various data from mobile phones to predict a multitude of demographic attributes (Zhong, Tan, Mo, & Yang, 2013). Results of other researchers suggest that certain smartphone user behaviors as well as facebook likes could possibly be used to even infer self reported big five personality traits (De Montjoye, Quoidbach, Robic, & Pentland, 2013; Youyou, Kosinski, & Stillwell, 2015) such as extraversion (Montag et al., 2014).

However, the analysis of behavioral driving data in the automotive context has been largely neglected for the purpose of gender recognition. In relation to driving behavior, previous research showed that traffic related mortality is higher for men than for females in the majority of countries worldwide (Twisk, Bos, Shope, & Kok, 2013; Zhu, Zhao, Coben, & Smith, 2013). These results are supported by other reports showing that although young men describe themselves as better drivers they drive riskier, use less safety equipment, and reported more risky driving behavior in comparison with females (Barr et al., 2015; Fernandes, Hatfield, & Soames Job, 2010; Vardaki & Yannis, 2013). Whereas analysis of automotive driving parameters (speed, acceleration, steering angle etc.) previously focused on aspects like fuel consumption, exhaust emissions, and mobility patterns (Brundell-Freij & Ericsson, 2005; Ericsson, 2000a, 2000b, 2001; Nielsen, Østergaard, Marra, & Træholt, 2010), the implications of individual differences in relation to automotive driving parameters have mostly been investigated as predictors for unsafe or risky driving (Guo & Fang, 2013; Lonczak, Neighbors, & Donovan, 2007; Lucidi, Mallia, Lazuras, & Violani, 2014).

The only data (known to us) related to gender specific driving behavior, recorded at a technical parameter level, was collected in two studies by Ericsson (2000a) and Ericsson (2000b) and an earlier investigation by Redsell, Lucas, and Ashford (1993). All of these studies investigated the associations between several factors (among them gender) with driving parameters especially fuel consumption. Redsell et al. (1993) noted that especially in changing environmental conditions (transition between street types) driver specific factors were associated with changes in fuel consumption. Ericsson (2000b) discovered that the average acceleration was generally higher for men compared to women. This pattern was especially pronounced on a low speed street type. Average speed was not different between both genders, except on one street type where men drove faster in comparison with women. The author interpreted acceleration and velocity interaction effects with different street types as an indication for alternations in the street environment to trigger most gender or driver specific variation in driving parameters. In another study Ericsson (2000a) investigated the variation of automotive driving patterns with regard to human factors. In addition to differences in acceleration patterns, they also found that in average, females drove at lower speed in comparison with men. However, it is difficult to generalize results of these studies due to small sample sizes ($N_{Redsell2013} = 6$, $N_{Ericsson2000b} = 12$, $N_{Ericsson2000a} = 29$ families), as well as in one case the fact that only the effect of a participant with no regard to a specific variable (gender) was investigated (Redsell et al., 1993). Family samples (data from cars labeled as either male or female if more than 75% of the total driving was performed by one gender) were used.

Comprehensively, these findings indicate that behavioral gender differences might be reflected in individual driving parameters and could possibly have an effect on variables like fuel consumption and emission exhaust.

The recording of automotive driving parameters in real world settings is costly and bears financial and actuarial difficulties. Virtual driving simulators are frequently used in industrial and academic settings for research and evaluations. Furthermore, modern driving simulators offer the possibility to record individual driving behaviors in a highly standardized, safe and cost-effective manner (Kaptein, Theeuwes, & Van Der Horst, 1996). Although the external validity of driving simulator results has to be questioned (Mullen, Charlton, Devlin, & Bedard, 2011), large sample studies are almost infeasible without initial leads from simulated driving studies. With this work, we intend to investigate the possibility to infer driver-gender from automotive driving parameters. In relation to previous research describing gender related differences in driving behavior (Ericsson, 2000a, 2000b; Redsell et al., 1993), we expect variables in accordance to acceleration and speed to be good statistical predictors of drivers gender.

Furthermore meaningful information with regard to gender could possibly be extracted in driving situations where the type of driving situation is changing (e.g., change from a rural road to a highway, or at intersections). Therefore, we hypothesize that data related to vehicle acceleration at changing driving situations will be predictive for gender recognition. Nonetheless, a major part of this investigation was to identify possible additional meaningful predictors in an exploratory fashion (see the Method section for details).

Aims of this study were the accurate statistical recognition of driver gender, based on automotive driving parameters as well as the identification of promising gender sensitive parameters beyond those identified in previous research. In addition, we also wanted to describe the data with an interpretable model in order to better understand dependences between gender and driving parameters.

2.1.3 Method

Participants

A total of 182 subjects participated in the virtual driving simulation. All participants were haphazardly recruited from the pool of AUDI employees in Ingolstadt, Germany. Since some participants (N = 37) experienced heavy symptoms of simulator sickness, they had to stop the simulation and their data were excluded from the sample. A final sample of 145 participants remained for statistical analysis. Gender was not totally equally distributed in our sample with 83 men and 62 women. The mean age of all participants was 32 years. Most participants (N = 65/44.8%) were between 18 and 28 years old, 50 participants (34.5%) were between 29 and 39 years old, 25 participants (17.2%) were between 40 and 50 years old and 5 participants (3.4%) were 51 or older. The sample was skewed in terms of education, as 71.7% of all participants had college or university education. Data collection and experimental procedures were coordinated between the AUDI AG workers committee and the Ludwig-Maximilians-Universität München (LMU) in order to be conducted in a most privacy protecting and non-invasive manner.

Apparatus

The driving task took place in a driving simulator of the AUDI AG in Ingolstadt, Germany. The used driving simulator consists of a circular $2.6\text{m}^2 \times 13.3\text{m}^2$ 250° frontal and side projection surface, with 16 million pixels as well as a $6\text{m}^2 \times 3\text{m}^2$ projection surface with 4.6 million pixels, located behind the car mockup. A visual refresh-rate of 60Hz and a data collection rate of 25Hz were used during the experiment. See Figure 2.1 for an overview of the driving simulator and the mockup. A specifically designed test track was used during the experiment. Various sections including straights, crossroads, roundabouts, lane changes and highways were implemented in the track. During these sections, variables in relation to speed, lane departure, braking force, gas pedal pressure, steering angle were collected. The drive along the 23.7km test track took approximately 20 minutes.

Figure 2.1: Driving Simulator and Mockup

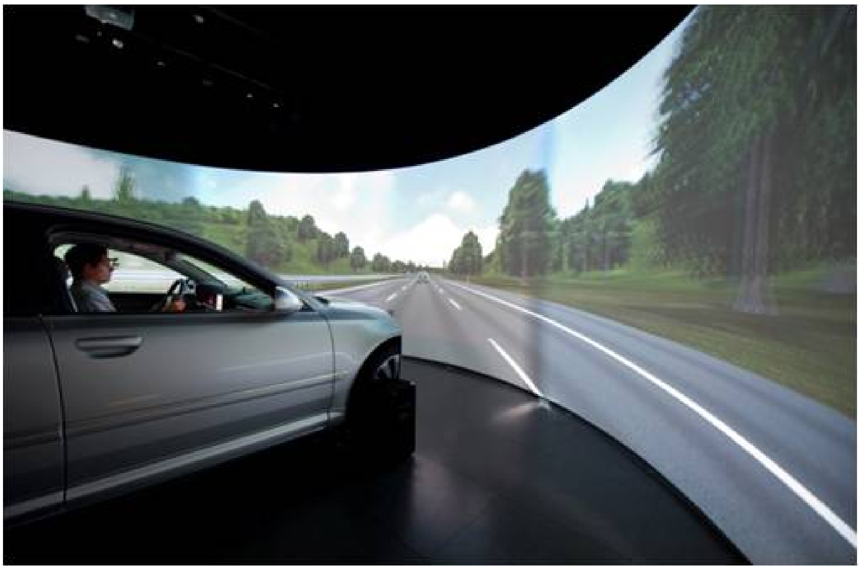

Figure 2.1: Driving simulator apparatus and car mockup at AUDI headquarters in Ingolstadt, Germany. Picture, as courtesy by the AUDI AG.

Procedures

Participants arrived at the laboratory and received a standardized written instruction with general information about the experimental procedures, as well as a short demographic questionnaire. On completion of the questionnaire, participants were guided to the driving simulator mockup. Once in the car, participants were verbally instructed about the interactions they had to perform during the experiment as well as possible effects of simulator sickness. During the drive, participants were verbally navigated along the route. Although participants were alone in the car mockup during the complete duration of the experiment, verbal communication with the experimenters was possible at all times.

Statistical Analysis

To create features for statistical modeling we used combinations of various standard driving parameters with the current driving situation ($p = 370$). Both descriptive measures (mean and standard deviation) and distributional measures (percentage of time in certain value ranges) were used for model creation. An overview of the recorded driving parameters is provided in Table 2.1.

Prior to predictive modeling, we applied a series of data pre-processing procedures.

As linear models are sensitive to predictor noise, near zero variance variables, highly correlated predictors (if $r > .80$) and two cases containing missing values were removed from the data set.

After pre-processing, 190 of the initial 370 predictor variables remained in the data set. For modeling, the data set was randomly split into a training (n = 101/70%) and a testing set (n = 42/30%). Considering the high number of predictors (p = 190) in relation to our sample size we used binomial elastic net regularized regression to statistically classify driver gender using a subset of most contributory predictors. The elastic net model represents a combination of the ridge regression and the LASSO (Least Absolute Shrinkage and Selection Operator), especially suitable for p > n problems (Zou & Hastie, 2005) and is capable to perform both shrinkage of correlated predictors and grouped variable selection. The model was trained with 10 fold 10 times repeated cross validation in order to avoid overfitting. During each resampling iteration the respective sample was centered and scaled. See Section 4 for syntax and data to reproduce this analysis. The reproducible code does not include variable extraction, as this part of the analysis was performed by the AUDI AG.

All data processing as well as statistical analyses in this study were performed with statistical software R 3.3.1 (R Core Team, 2016). Additionally several external packages were used for this purpose. We used the *caret* and *glmnet* packages for predictive modeling (Friedman et al., 2010; Kuhn, 2015), the *doParallel*, and *doMC* packages for computational parallelization (Analytics & Weston, 2015a, 2015b), and the *ggplot2* package for visualization. Furthermore, the *kernlab*, *pROC*, and *lattice* packages were used for miscellaneous purposes (Karatzoglou, Smola, Hornik, & Zeileis, 2004; Robin et al., 2011; Sarkar, 2008).

Table 2.1: Driving Feature Overview

Velocity	Measure
	M/SD
% of time	0-15 km/h
% of time	15-30 km/h
% of time	30-50 km/h
% of time	50-70 km/h
% of time	>70 km/h
Steering Wheel Angle	rad
	M/SD
% of time	< -5
% of time	-5 > < -3
% of time	-3 > < -1
% of time	-1 > < 0
% of time	0 > < 1
% of time	1 > < 3
% of time	3 > < 5
% of time	> 5
Gas Pedal & Break Pedal actuation	
	M/SD
% of time	0-25%
% of time	25-50%
% of time	50-75%
% of time	75-100%
Acceleration/Deceleration	m/s2
	M/SD
% of time	< -2.5
% of time	- 2.5 > < -1.5
% of time	- 1.5 > < -1.0
% of time	- 1.0 > < -0.5
% of time	- 0.5 > < 0.0
% of time	0.0 > < 0.5
% of time	0.5 > < 1.0
% of time	1.0 > < 1.5
% of time	1.5 > < 2.5
% of time	> 2.5

Note: Extracted features related to velocity, steering wheel angle, pedal actuation and deceleration-acceleration. Features were seperately calculated at the beginning of the drive, at straight sections, crossings, roundabouts and highway ramps.

2.1.4 Results

The final elastic net model was trained to maximize area under the curve (AUC), with parameters $\alpha = 0.1$ and $\lambda = 0.28$ in the regression equation. The AUC measure represents the probability of the model to accurately classify two randomly sampled participants based on their driving parameters. The AUC of the final model applied on the test set (N=42) was 0.90 ($CI^{95\%}_{Auc} = [0.79, 1]$). See Table 2.2 for the performance measures. The comparison of the lower specificity 0.67 and high sensitivity 0.96 shows that the model is more successful in classification of males (positive class) in comparison with females. This imbalance was most likely induced due to disproportional ratios of men an women in the sample.

Variable Importance

The final model included 116 non-zero predictors. Variable importance measures of the 40 top-ranked predictors are visible in Figure 2.2 and show that a variety of parameters contributed to the model. In addition to the direction of the effect we also notice that certain types of measures are more often present in the top ranked predictors than others. Most notably, almost half of all variables are associated with acceleration (18), while only two predictors are related to actuation of the braking pedal. Roughly the same amount of variables related to steering wheel angle (8), velocity (7) and gas pedal actuation were ranked among the top predictors.

In Figure 2.3, two top ranked distributional measures in relation to acceleration patterns are plotted against each other with gender indicators based on color and shape. The plot has to be interpreted in the way that both quantitative variables represent percentages of acceleration values in a specified range (eg. Acc > 1.5 < 2.5 Crossing, refers to values between 1.5 and 2.5). With the help of the distributions on the top and the right side it is visible that a larger number of male values tend to be represented more frequently in the lower range of acceleration (0 >< .5), whereas females are more spread out. Although a gender specific pattern is visible, it is also intuitive that both classes are not clearly separable only using these two measures.

Furthermore, simple Welch t-tests show that mean comparisons do not necessarily help to explain gender differences in driving parameters. Whereas the comparison of standard deviations in gas pedal actuations at crossing1 (cross1_gas_SD) are significantly different for both genders ($t(111) = -5.34, p < 0.001, d = 0.91, 1 - \beta = 0.99; M_m = 10.47, SD_m = 1.28, M_f = 11.94, SD_f = 1.6$), the comparison of the standard deviation

Table 2.2: Prediction Performance Measures

Measure	Value
AUC	0.90
95% $CI_{(Auc)}$	0.79, 1
Accuracy (Acc)	0.83
95% $CI_{(Acc)}$	0.69, 0.93
Balanced Accuracy	0.81
Sensitivity	0.96
Specificity	0.67
No Information Rate (NIR)	0.57
P-Value [Acc > NIR]	0.0003
Pos Pred Value	0.79
Neg Pred Value	0.92
Positive Class	Men

Note: Standard performance measures of the elastic net classifier as evaluated on the test set.

in average highway speed (highway_total_v_SD) is not ($t(125) = 1.91, p = 0.058, d = 0.38, 1-\beta = 0.62; M_m = 10.08, SD_m = 2.03, M_f = 9.40, SD_f = 2.14$). Although it would be interesting to describe more of the important variables in our model with further detail, we do not elaborate on this aspect as this would go beyond the scope of this paper.

Figure 2.2: Elastic Net Variable Importance

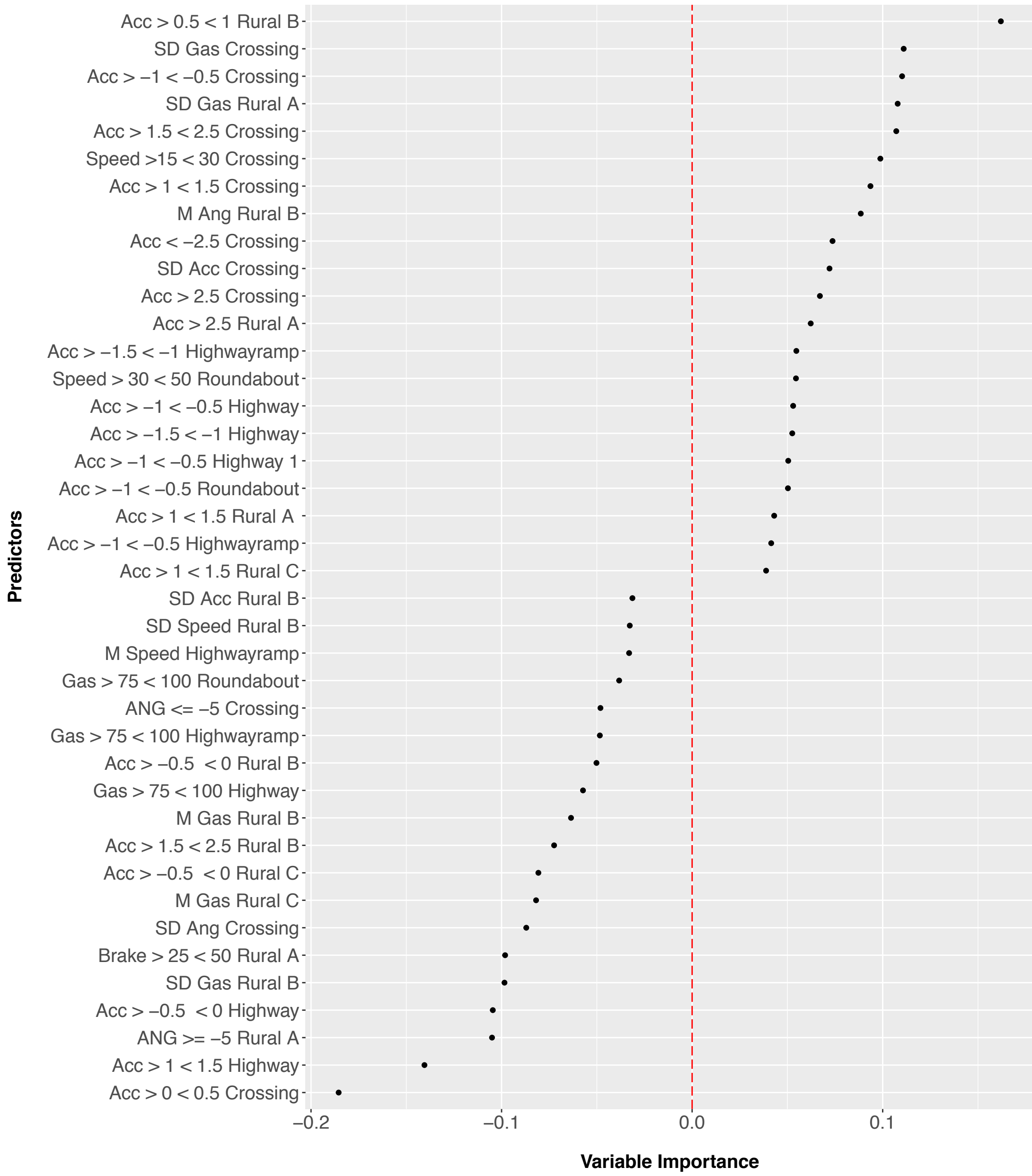

Figure 2.2: Variable importance measures (regularized β coefficients of the final model) for the top 40 predictors of gender. Positive values refer to variables predictive for male, negative values refer to variables predictive of female gender. Abbreviations: M = mean, SD = standard deviation, Gas = gas pedal actuation, Acc = acceleration, Ang = steering wheel angle, Brake = brake pedal actuation.

Figure 2.3: Scatterplot of Distributional Acceleration Features

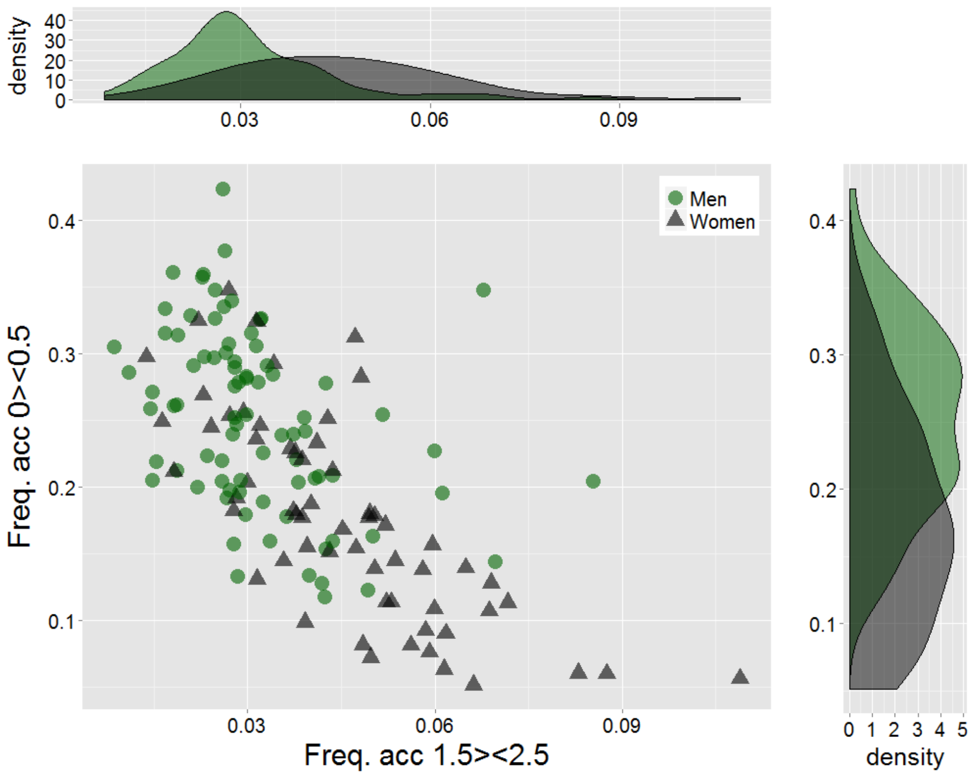

Figure 2.3: Scatterplot and density plots of two top predictors related to gender specific acceleration patterns. Values represent frequencies of values in a certain category (e.g., a value of 0.03 in the variable Freq. 1.5><2.5 indicates that about 3% of all acceleration values ranged from 1.5 to 2.5 at the crossing driving situation.

2.1.5 Discussion

To our knowledge, this is the first paper that investigates gender recognition based on automotive driving parameters. The goal of this paper was to investigate whether driving data based on a 20 minute drive in a simulator is sufficient for accurate prediction of driver-gender using a machine learning approach. Our results show that although we did not use personal information such as text input or video data, it was possible to classify gender well above chance, purely based on technical driving parameters.

As hypothesized, features relating to acceleration in dynamic driving situations were identified as most important predictors in the model. However, variable importance measures also illustrate that additional parameters with relation to individual acceleration behavior were contributing to the final model. Additionally, speed (velocity), gas pedal actuation and measures related to the steering wheel angle turned out to be especially predictive in our model. These gender specific patterns in acceleration, gas pedal actuation and velocity are in accordance with previous research (Ericsson, 2000a, 2000b; Redsell et al., 1993) that indicates possible gender differences in automotive driving parameters. Intuitively, these differences in driving parameters could be closely related to gender differences in spatial orientation, reported in other studies (Zhu et al., 2013).

However, the real reasons for gender differences in driving patterns remain unclear and should be investigated prospectively (Coluccia & Louse, 2004). In consideration of the relatively short virtual test drive (20min), classification accuracy is quite impressive. Furthermore, if these results are reproducible in real life driving situations, alterations or suggestions for adaptations of systems in the car could be made possible after a very short period of time. Gender differences could be used to improve human-machine interaction in adaptive user interfaces. Navigational strategies for example have been reported to be different for both genders in previous studies. Whereas men mostly use Euclidean information when orienting, it was reported that women predominantly rely on landmark information (Dabbs Jr., Chang, Strong, & Milun, 1998; Lawton, 1994).

In addition to gender sensitive interfaces for navigational tasks gender prediction while driving could also be used to account for more individual aesthetical needs in adaptive user interfaces. For example, results of a study suggest design strategies like gamification as differentially appealing to both genders (Koivisto & Hamari, 2014). Furthermore, gender-adaptive systems in vehicles could alter aspects like seat ergonomics, temperature, or even interface characteristics like colors and points of interest in a map.

Especially useful might be the alteration of system adjustments that usually do not justify the installation of a button or menu entry or are not intuitively understandable to the average user. For example, steering effort is a factor that could very well be adjusted to user-specific needs in a subliminal manner (Anand, Terken, & Hogema, 2011). Considering the wide variety of gender differences reported so far, many more adjustments in adaptive user interfaces could be explored. The analysis of behavioral gender differences does not exactly simplify gender recognition in comparison with facial image classification (Bekios-Calfa et al., 2014; Hadid & Pietikäinen, 2009). However, it could prove as useful in situations where neither visual nor linguistic information can be collect (cars without camera). Furthermore, gender recognition via camera might be privacy violating, as in addition to gender, an image could reveal the identity of a person.

Limitations

Even though, driving data obtained in a high-end virtual driving simulation offers high degrees of standardization and does relate to real driving situations, notable discrepancies exist in comparison with data collected in real settings (Mullen et al., 2011). Therefore, care has to be taken generalizing these results to real driving situations. In real driving contexts, different and additional variables might be informative about driver gender. Furthermore, variables like the steering wheel angle as well as actuation of the braking pedal might yield different values once recorded in real life settings, due to the lags in the simulation (Mullen et al., 2011). In addition, factors like virtual distance perception (that have been heavily investigated outside of the automotive context) might cause deviations in virtual in comparison with in real driving behavior. For a review see (Renner, Velichkovsky, & Helmert, 2013).

2.1.6 Conclusions and Future Work

This work demonstrates the possibility to recognize drivers gender with high accuracy based on standard driving parameters obtained in a 20 minute virtual test drive. Furthermore, automated, non-camera based gender recognition from automotive driving parameters opens new possibilities for gender adaptive systems and user interfaces in the car. The present work acts as a starting point for further research in relation to the analysis of driving parameters with regard to the recognition of user specific characteristics.

However, future studies should focus on more complex criteria such as the interaction between gender, age and personality traits. This could be promising as gender and age are known to interact with big five personality factors such as emotional stability, agreeableness (Chapman, Duberstein, Sorensen, & Lyness, 2007; Vecchione, Alessandri, Barbaranelli, & Caprara, 2012) and sub-facets like assertiveness and excitement seeking (J. Costa, Terracciano, & McCrae, 2001) as well as openness to feelings. The latter ones might be reflected in individual driving behavior.

The presented results do not shed light on underlying reasons for gender differences in driving parameters. Therefore, continuative research should further investigate the cause of gender differences in the observed variables by e.g. linking them to biological or cognitive theories. In direct relation to that, the current results should be compared and validated based on driving parameters obtained in real life driving settings, a goal we intend to achieve in the near future.

2.1.7 Author Contributions

Besides myself, Prof. Dr. Markus Bühner contributed to the creation of this article. Prof. Dr. Bühner acted as the supervising author of this article.

2.1.8 Acknowledgements

This research was funded by the AUDI AG, Ingolstadt, Germany. The views and conclusions expressed in this paper are those of the authors and do not necessarily reflect the official policy or position of the AUDI AG. We thank the AUDI AG, for providing the driving simulator facilities as well as technical assistance with the conduct of the study.

2.2 Study 2: Validation of Self-Reported Personality with Mobile Application Usage

2.2.1 Abstract

The present work investigates the potential of behavioral validation of personality self-reports with data of mobile application usage on smartphones. Relationships between personality and app-usage in 14 categories are investigated at factor and facet level. A total of 137 subjects (87w, 50m) with an average age of 24 (SD = 4.72) and above average education level (96% completed a levels) participated in a 90 minutes psychometric lab session as well as in a consequent 60 days passive logging study in the field. Our results suggest that personality is related to the use of mobile applications towards a larger degree than previously reported. Beyond demographics, extraversion, conscientiousness and agreeableness predict application usage at factor and facet level. Furthermore, Big Five factor and facet level scores show comparable predictive performance. This work illustrates how behavioral proxy measures can be used to validate self-reports of personality with actual behavior. Furthermore, this study provides new insights into behavioral manifestations of personality

2.2.2 Introduction

Personality refers to relatively stable individual differences in characteristic patterns of thinking, feeling and behaving across time and situations. Individual differences in personality have been shown to predict important life outcomes and behaviors on individual and inter-individual level (Ozer & Benet-Martínez, 2006; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007).

In psychological science, personality is mostly defined within trait theories. They make two main assumptions: First, they assume a certain stability of personality dispositions (e.g., Sociableness) over time and situations. Second, personality traits are believed to systematically change individual behavior (Matthews et al., 2009) (e.g., more sociable people have more contact and interactions with others). Whereas, pure patterns of thinking and feeling are hard to observe, patterns of behavior should be directly observable and aggregatable across situations. These aggregations can then be used to construct a picture of behaviorally represented personality aspects (Vazire, 2010).

Many different trait models of personality have been proposed with an initially varying number of relatively independent dimensions (John et al., 2008). However, since its development in the late 90s, the *Big Five* personality trait theory (P. T. Costa & McCrae, 1992; Goldberg, 1981) has emerged as the most widely accepted model in psychology. Created by using a psycholexical approach (Allport & Odbert, 1936; Norman, 1963), the Big Five model provides a wide description of human personality for applications within and beyond scientific research

The model describes people's tendencies of behavior and attitudes on five broad dimensions that hierarchically consist of several sub-facets. The five broad factors describe the dimensions extraversion-introversion, emotional stability-neuroticism, agreeableness, conscientiousness and openness or intellect.

Measurement of personality is usually achieved with normed and standardized self-report questionnaires. These are commercially available and exist in a wide range at different length and level of measurement detail - suitable for different applications (Arendasy, 2009; P. T. Costa & McCrae, 1992; Gosling et al., 2003; R. R. McCrae et al., 2005; Rammstedt & John, 2007).

Furthermore, self-reports have to be validated in order to ensure that personality is actually measured. Often this is performed by the comparison with results of other self report measures or measures about the same person obtained by others (Funder, 2012). However, relating self-report personality measures to behavioral criteria is considered the gold standard of validation for latent constructs (Funder, 2012; Vazire, 2010). This is of relevance as personality can only be considered as an important construct if it meaningfully helps to predict individual behavior (Funder, 2012).

As criticized (Baumeister et al., 2007), many apparent validation studies nevertheless rely on self-report measures of typical behaviors (Fleeson et al., 2009; Wu & Clark, 2003). This is problematic as previous studies have shown that self reports of behavior usually include large amounts of measurement errors (Boase & Ling, 2013; Kobayashi & Boase, 2012; Paulhus & Vazire, 2007). The general lack of studies investigating actual behavior with regard to personality psychology as well as the undifferentiated use of the term *"behavior"* has been repeatedly subject to criticism (Baumeister et al., 2007; Fleeson et al., 2009; Furr, 2009; Lewandowski Jr & Strohmetz, 2009; Poorthuis et al., 2014; Vazire & Mehl, 2008). Still, many psychological studies rely on self-report measures, also because the collection of large behavioral samples across many situations can be costly, time consuming and even unfeasible with sufficiently high testing power.

The current work will illustrate how some of these validation difficulties could potentially be overcome with the help of current off-the-shelf consumer technology.

The availability of cheap mobile sensor technology in the form of smartphones enables gathering of large and diverse behavioral samples as validation criteria for personality questionnaires (G. Miller, 2012; Yarkoni, 2012). Although the idea of mobile electronic data collection is not new (Mehl et al., 2001), the potential of smartphones as "*silent observers*", unobtrusively collecting behavioral data, has mostly remained unrecognized in psychological science. This is surprising, since modern smartphones are capable of unobtrusively recording a large variety of behavioral proxy measures over a long period of time, at low cost.

In this regard, the present work focuses on the behavioral validation of personality measures with usage frequencies of mobile application usage.

Mobile applications (apps) are an integral part of current smartphones, tablets and smartwatches as the majority of users' actions are carried out through an app. A rapidly growing number of them caters to users' everyday needs such as communication, information and entertainment. Their wide-spread every-day use and functional specificity made apps an interesting target for research in human-computer-interaction (HCI). Several projects analyzed app usage (Böhmer, Hecht, Schöning, Krüger, & Bauer, 2011), related user behavior (Brown, McGregor, & McMillan, 2014), launching habits (Hang, De Luca, Hartmann, & Hussmann, 2013) and app re-visitation (Jones, Ferreira, Hosio, Goncalves, & Kostakos, 2015). These studies mostly quantified app usage via data logging, including context, such as time and location.

In psychological science, only some previous studies have started to investigate markers of individual behavior manifested in smartphone use (and in actual behavior in general). Some studies collected self-reports of behavior (Butt & Phillips, 2008; Kim, Briley, & Ocepek, 2015; Lane & Manner, 2011). Kim et al. (2015) investigated how sociodemographic variables as well as Big Five personality (measured with the *Ten Item Personality Inventory - TIPI*) are predictive of self-reported categorical app usage. Their results suggest, that demographic variables and especially gender change frequency of general smartphone usage and application use in broad categories. Furthermore, they indicate that *Big Five* personality factors extraversion, openness and conscientiousness predict smartphone use. Specifically, their results suggest that conscientiousness has a negative effect on the use of e-commerce applications (finance and shopping) ($\exp(\hat{\beta}) = 0.89$, CI 95% = $[0.83, 0.96]$) and extraversion on the use of literacy (book, reference management and education) ($\exp(\hat{\beta}) = 0.91$, CI 95%= $[0.83, 0.99]$) as well as relational applications (social network and instant messaging) ($\exp(\hat{\beta}) = 1.04$, CI 95% = $[1.01, 1.08]$). Based on these results, the authors conclude that the impact of personality on app usage frequencies does not extend beyond the effect imposed by demographic variables.

Only a small number of researchers have used an approach similar to ours and directly logged user behavior to examine relationships with personality. These studies mostly focused on communication (Montag et al., 2014, 2015), and showed that call frequencies are related to extraversion (Montag et al., 2014) (e.g., call out count, $r = 0.45$, CI 95% = $[0.19, 0.65]$). Other results of Montag et al. (2015) show small positive correlations of the use of the popular messaging service *WhatsApp* with extraversion ($\rho = 0.18$, CI 95% = $[0.14, 0.22]$) and neuroticism ($\rho = 0.07$, CI 95% = $[0.03, 0.11]$), as well as negative correlations with conscientiousness ($\rho = -0.13$, CI 95% = $[-0.17, -0.09]$).

Another group of researchers has also investigated the relationship of personality and mobile phone usage (Chittaranjan, Blom, & Gatica-Perez, 2013). They examined a large number of correlations of smartphone use and Big-Five personality traits, measured with the TIPI (Gosling et al., 2003) . In addition to call related variables, they considered app usage in eleven broad categories (e.g., office, internet, maps) as predictors for personality.

Compared to Chittaranjan et al. (2013), Kim et al. (2015) as well as (Montag et al., 2015) reported few and rather small associations between personality and app usage, despite the large samples they used in their studies ($N_{Kim2015} = 4154$, $N_{Montag2015} = 2418$). In comparison to the earlier study of (Montag et al., 2014) wherein they reported correlations of up to ($r = 0.45$) in a much smaller sample ($N_{Montag2014} = 49$), the discovered correlations in the later study (Montag et al., 2015) are much smaller (e.g., $r = 0.19$). Montag himself suggested that the lower correlations observed in the second study could be caused by the less reliable personality questionnaire (BFI) (Rammstedt & John, 2007) they used in the second in comparison with the first study, where the NEO-FFI (P. T. Costa & McCrae, 1992) was used. (Kim et al., 2015) used an even shorter questionnaire (TIPI) (Gosling et al., 2003) in combination with self-report data for their statistical analysis.

Although we acknowledge the economic advantage of 10-item questionnaires over longer and more extensive instruments such as the NEO-PI-3 (R. R. McCrae et al., 2005) or the Big Five Structure Inventory (BFSI) (Arendasy, 2009), we do not share the authors' believe (Gosling et al., 2003; Rammstedt & John, 2007) that short instruments enable accurate measurement of the hierarchical *Big Five* dimensions. We rather argue that both reliability and content validity of BFI and TIPI are questionable, based on the fact that the *Big Five* dimensions were lexically derived through dimension reduction procedures and the factor dimensions therefore are converged from several (but 30 at least) lower dimensions, see (DeRaad & Boele, 2000) for an overview. Therefore, the exhaustive measurement of the main personality aspects with ten items (fewer items than facets) seems to be a difficult or at least incomplete task. The validity of short questionnaires has been subject to discussion in previous research (Heene, Bollmann, & Bühner, 2014; M. Ziegler, Kemper, & Kruyen, 2014). However, we will not elaborate on this issue as it goes beyond the scope of this paper.

Furthermore, previous research mostly tried to establish linear relationships between behavioral frequencies and personality (Chittaranjan et al., 2013; Montag et al., 2014, 2015), assuming a Gaussian distribution. However, count data usually rather resembles a (Quasi) Poisson distribution, a fact that could affect results of data analysis (O'Hara & Kotze, 2010). Moreover, in order to predict behavioral criteria from individual personality scores, either broad factor values or sub-facet scores can be used. Some previous research suggests that personality facet measures provide independent prediction value in relation to behavioral criteria in addition to factor level scores (M. Ziegler et al., 2014). However, disagreement is prevalent in current research concerning this topic (Ashton, Paunonen, & Lee, 2014; Salgado, Moscoso, & Berges, 2013). Uncertainty remains with regard to whether factor or facet level scores are better for the prediction of behavioral categories. Previous studies analyzing logging data only used factor level personality scores and related them to app usage behavior (Chittaranjan et al., 2013; Montag et al., 2014, 2015). In the present study we relate both factor and facet personality scores to behavioral criteria and compare the obtained results.

In summary, it is not clear whether inconsistencies in previous results are attributable to the use of less reliable psychometric instruments (Chittaranjan et al., 2013; Kim et al., 2015; Montag et al., 2015), to the use of self-report data (Kim et al., 2015) or improvable analytics. However, the current situation motivates our more intensive investigation of the expected relationships between app usage and *Big Five* personality, in particular by combining behavioral data-logging with fine-grained personality measures. Precisely, we used the BFSI (Arendasy, 2009) a comprehensive, reliable, and detailed personality inventory, which as its authors claim, measures personality in accordance to the *Partial-Credit-Model* (Masters, 1982) at factor and facet level. In relation to the lexical derivation of the *Big Five*, the BFSI uses adjectives and short phrases as items for personality assessment. This could help circumvent previously reported problems regarding the comprehensibility of longer sentence-based items (such as in the NEO-FFI (P. T. Costa & McCrae, 1992; R. R. McCrae et al., 2005)). In the present work we analyzed actual app usage behavior, considering more categories of app usage and consequently more types of behavior in the analysis.

Previous studies have reported some associations between personality and app use on smartphones, however more guidance for the practitioner with regard to which associations are the most stable and promising ones for further investigation are needed. Hence, this work aims to identify the most stable statistical relationships between *Big Five* personality aspects and individual behavior, manifested in mobile application usage in order to provide solid starting points for prospective research.

2.2.3 Method

Data used in this work constitutes a fraction of a larger research project at Ludwig-Maximilians-Universität München (LMU), investigating relationships between psychological variables and a wide range of behavior, logged via smartphones (such as calls, app-usage, messages, geoposition etc.). However, this paper focuses on app usage behavior only, and explores its relationship with personality. Therefore, further descriptions will only include data dimensions related to the present analyses. Data collection took place between September 2014 and August 2015 at Munich, Germany, EU.

Participants

We recruited 137 participants, 87 women and 50 men, via social media, forums, blackboards, flyers, and on campus. The obtained sample was rather young with a mean age of 24 years (SD = 4.72) and 75% of the participants being 26 or younger. The majority of the sample had at least completed high school (96%) and 31% of all participants had completed education at university level. All subjects gave written consent prior to participation and could withdraw participation in the study as well as demand deletion of non-anonymized data at any time. The study was approved by the responsible IRB and data protection officer.

Personality Measures & Demographics

Big five personality was measured with the German version of the *Big Five* Personality Inventory (BFSI) (Arendasy, 2009) in a laboratory setting. The BFSI was selected for personality assessment due to its unambiguous items as well as its favorable psychometric properties. In contrast to more common personality scales such as the new NEO-PI-3 (R. R. McCrae et al., 2005), the authors of the BFSI (Arendasy, 2009) report conformity to the partial credit model. We used the person parameter of the partial credit model instead of sum scores for all analyses.

The BFSI consists of 300 items (adjectives and short phrases) and measures *Big Five* personality dimensions (agreeableness, openness to experience, conscientiousness, extraversion and emotional stability/absence of neuroticism) on factor and facet level with a four-step likert scale ranging from "*untypical for me*" to "*typical for me*".

In addition to the personality scores, age, gender and current level of completed education was collected. Gender was recorded dichotomously with "1" representing male and "2" representing female participants. Level of education was subdivided in five categories from no education to finished university degree. Please take this into consideration when looking at correlations in Tables 2.3, 2.4, as well as regression coefficients in Table 2.6.

Behavioral Measures

User behavior was recorded via an Android logging app (available for Android 4.0 or higher), specifically designed for this purpose. In particular, the app recorded which apps were used when and at which location. It also logged screen activation states. The data was regularly transferred to our server, once participants were connected to WiFi, using SSL encryption. Afterwards, the logged data was further enriched with information retrieved from the Google Play Store (Google, 2016a), such as app category, description, rating and number of downloads.

With regard to the analysis described later, usage frequencies of apps in a certain category (e.g., *Communication*) were aggregated. The categorization as provided by the Google Play Store (Google, 2016a) was used as a basis. However, more suitable labels had to be assigned to some apps with a non-appropriate category label. (See section 2.2.3 and the supplemental files for more information.) The final dataset consisted of event-based, timestamp-sorted data and contains 3,246,821 entries with an average of 23,699 events (SD = 12,165.42) per participant.

Procedure

The study was conducted in two stages. In a lab session, participants gave written consent and completed a personality inventory, subscales of an intelligence test and a demographic questionnaire. As the intelligence subscales were intended for use in another study, we will not elaborate on those. The testings took place at the psychological laboratory at the university's psychology department. Subsequently, the logging app was installed on participant's private smartphone and tested for functionality.

Once operational, the app logged a great variety of anonymous usage related parameters for the consequent 60 days, regularly uploading it to our servers when the respective Android smartphone was connected to WiFi. The app stayed in the background: Participants did not have to complete any tasks or actions to avoid influencing their natural smartphone use. As a sole exception, they were reminded (via a pop-up message) to re-enable location sensor and app history access (Android 5.0 and higher) in case they had turned off these settings (e.g., in order to save battery). After 60 days of logging, participants were invited to receive their compensation (individual personality profile and 30 EUR or course credit for students). During this meeting, an additional manual backup of the collected data was retrieved from the device.

Data Analysis

Prior to modeling we had to pre-process and clean the data. Although in general we used the app categorization provided by the Google Play Store (Google, 2016a), a number of apps were clearly mislabeled and had to be re-labeled in order to perform meaningful data analysis. We share the notion that the labeling of information is always somehow subjective, therefore we provide the full list of relabeled apps as an supplemental file to this article. Furthermore, we had to exclude a large number of bloatware[1] and background apps that showed up as actual usage in the collected logs (see the supplemental files for a complete list).

In order to handle univariate outliers in the data, we first identified robust z-transformed values with values larger than three. Robust z-transformation was done by subtracting values by the median and dividing the result by the median absolute deviation. The median absolute deviation is a robust measure of variability in an univariate data sample. The values were than adjusted to the maximum value of the remaining data points

[1]Bloatware refers to pre-installed, mostly unwanted software that often negatively affects system performance of devices

(winsorizing) (Erceg-Hurn & Mirosevich, 2008). This procedure allowed us to not waste data while limiting the effects of extreme, possibly spurious data points. In total only two values of the variable *App Usage Lifestyle* were adjusted. Furthermore, we only included app usage variables with a median absolute deviation larger than zero, excluding categories with no or almost no variation in the data.

Prior to regression modeling, we investigated descriptive statistics as well as correlations between the Big Five factors and the demographic variables. In order to investigate the relationship of personality and demography on app-usage behavior, we performed a two-step analysis for factor and facet scores respectively. Due to the number of predictors (facet analysis) and because of the expected multi-collinearity between the personality and demographic variables (visible in Table 2.3), we used a conservative stability selection procedure (Hofner, Boccuto, & Göker, 2015; Meinshausen & Bühlmann, 2010) in combination with the popular Least Angular Shrinkage Selection Operator (LASSO) penalization regression (Friedman et al., 2010). This procedure was chosen in order to only select the most reliable predictors while penalizing for correlations between them. The LASSO regressions were modeled under the assumption of a Poisson distribution with each app usage category as the respective criteria.

Stability selection refers to a relatively new concept that adds resampling procedures to variable selection, such as LASSO and therefore making the selection procedure more reliable (Meinshausen & Bühlmann, 2010). This procedure avoids to fit only one model on the data, instead many different models are fitted on subsets of the data. Therefore, variables that repeatedly (above a certain threshold) add predictive value to different models are chosen.

Furthermore, this procedure allows for assessment of the selection stability of variables while controlling for sample error. Hofner et al. (2015) suggests the upper limit of the *pairwise family error rate (PFER)* to be set at $\alpha < PFERmax < m\alpha$, where m represents the number of predictors, and α represents the respective significance level ($m_{\text{factor}}\alpha = 8 \times 0.05 = 0.4$ and $m_{\text{facet}}\alpha = 33 \times 0.05 = 1.65$ in our case). Based on this recommendation, we used a PFER of 0.2. We chose this parameter value since PFER represents the tolerable number of falsely selected noise variables. Therefore, we kept this value well below one, tolerating less than one noise variable. Furthermore, we fixed the number of selected variables to 1, choosing only the most influential predictor for the respective app usage category.

In a second step, the respective predictor selected through stability selection, was again used as predictors in Quasi-Poisson regressions, with app usage categories as the respective criteria. This was performed because regression coefficients of a penalized model are hard to interpret. We chose generalized linear regression over linear regression analysis as count data usually follows a Poisson distribution (O'Hara & Kotze, 2010). In order to account for over-dispersion in our data set, we assumed Quasi-Poisson distributions instead of Poisson distributions for the dependent variables.

Both global *Big Five Personality* and subfacets of the *Big Five Personality* scores, as well as demographics, were used as predictors in the regression models. This procedure was repeated for each app usage category respectively. In order to compare factor with facet models, we used the *Dawid-Sebastian* score as a measure of model fit. This measure is similar to the mean squared error but additionally accounts for overdisperison in count data (Czado, Gneiting, & Held, 2009). For a Quasi-Poisson distributed random variable X, with $E(X) = \mu$ and $Var(X) = \theta\mu$, the *Dawid-Sebastian* score for an observed value x is calculated as follows:

$$DSS(x) = \frac{(x-\mu)^2}{\theta\mu} + 2\log(\theta\mu). \tag{2.1}$$

For example, X could be the usage frequency of *Communication* apps, μ would represent the expected app usage frequency (needs to be estimated) and the variance of app usage is $\theta\mu$ where θ is the overdispersion parameter of assumed Quasi-Poisson distribution. In order to obtain an unbiased estimation of model fit, we used a Monte-Carlo resampling procedure. In particular, we created test (10%) and training set (90%), fitted a Generalized Linear Regression model with Quasi-Poisson distribution on the training set and calculated the mean DSS across all observations. In order to calculate the DSS, μ and θ were estimated from the training set. This procedure was repeated 100 times for each criterion and DSS scores were averaged across all observations in each test set. In comparison analyses, we made sure that equal test and training set splits were used. Please note that for some app-usage categories (visible in Table 2.5) no modeling was performed as not enough data was available, they are therefore not reported in the results section (e.g., *Comics*), see Table 2.6 for all the predicted categories.

All data processing as well as statistical analyses in this study were performed with statistical software R 3.3.1 (R Core Team, 2016). Additionally, several external packages were used for this purpose. We used the *glmnet* package for statistical modeling and the *stabs* package for stability selection (Friedman et al., 2010; Hofner & Hothorn, 2015). Information about syntax and data for reproduction of the presented results can be found in Section 4.

2.2.4 Results

Descriptive Statistics

As visible in Table 2.3 correlations between demographic as well as the Big Five personality factors were present in the data. Due to deviations from Gaussian distributions in all app usage categories, we used Spearman correlations for all correlations in our analysis (Yarkoni, 2010). Age was correlated with Education (ρ =0.42, $p < 0.001$), with older people being more educated. Female gender was positively associated with both extraversion ($\rho = 0.20$, $p = 0.01951$) and agreeableness ($\rho = 0.26$, $p = 0.00252$). Furthermore, several correlations between the Big Five factors were present in the data set. The highest correlation was observed between extraversion and openness ($\rho = 0.58$, $p < 0.001$). This correlation is quite high for allegedly independent personality dimensions. However, as the variance inflation factor (VIF) for both extraversion (VIF = 1.82) and openness (VIF = 1.69) is smaller than 4 (VIF) (Dormann et al., 2013; Fox & Monette, 1992) we proceeded with the analysis. The relationship between extraversion and openness is in accordance with literature and adds to the ongoing debate about the structure of the openness dimension and its relationship with extraversion (DeYoung, 2006). Additionally, extraversion was correlated with emotional stability ($\rho = 0.42$, $p < 0.001$) as well as with agreeableness ($\rho = 0.37$, $p < 0.001$). Surprisingly, no correlation between emotional stability and *Gender* was present in our data (ρ = -0.06, n.s). For all additional correlations see Table 2.3.

Table 2.4 shows pairwise Spearman correlations of psychometrics and demographics with usage of app-categories. Several relationships are visible. As the number of calculated correlations would induce a multiple testing problem we will only elaborate on reliable relationships, after variable selection.

In total, 2,835 different apps were used by the 137 participants in our study with an average of 12.42 different apps used per day. Apps of the *Communication* category were on average used most frequently (37.10 times a day), whereas apps of the *Comics* category were used most infrequently (0.08 times a day). *Game* apps show the longest usage duration on average. More information about app categories as well as the top apps of each category is provided in Table 2.5. Please note: The table is sorted by the average number of app-uses in the respective category.

Table 2.3: Pairwise Spearman Correlations Between Big Five Factor Scores and Demographics

	Predictors	1	2	3	E	A	ES	C
1	Gender	1						
2	Age	-0.10 [-0.26, 0.07]	1					
3	Bildung	-0.04 [-0.20, 0.13]	0.42 [0.27, 0.55]	1				
E	Extraversion	0.20 [0.03, 0.36]	0.00 [-0.16, 0.17]	0.01 [-0.15, 0.18]	1			
A	Agreeableness	0.26 [0.09, 0.41]	0.07 [-0.10, 0.23]	0.07 [-0.10, 0.23]	0.34 [0.18, 0.48]	1		
ES	Emotional Stability	-0.06 [-0.22, 0.11]	0.02 [-0.15, 0.19]	0.06 [-0.11, 0.23]	0.42 [0.27, 0.55]	0.23 [0.07, 0.39]	1	
C	Conscientiousness	0.14 [-0.03, 0.30]	-0.06 [-0.23, 0.11]	0.03 [-0.14, 0.20]	0.19 [0.02, 0.35]	0.17 [0.00, 0.33]	0.29 [0.13, 0.44]	1
O	Openness	0.14 [-0.02, 0.30]	0.04 [-0.13, 0.20]	0.04 [-0.13, 0.21]	0.58 [0.45, 0.68]	0.42 [0.27, 0.55]	0.31 [0.15, 0.45]	0.29 [0.13, 0.44]

Note: Pairwise Spearman correlations between *Big Five* measures, demographic variables and app usage categories. Square brackets contain 95% confidence intervals.

Prediction of App Use – Big Five Factor Level

The feature selection procedure reported stable personality and demography predictors for a total of 13 app usage categories (see Table 2.6). Besides gender, age and education, the three *Big Five* factors extraversion, agreeableness and conscientiousness were chosen as meaningful behavioral predictors in the variable selection. Therefore, emotional stability and openness did not provide enough unique predictive value for the app usage criteria. The highest stabilities in feature selection could be observed for Gender as a predictor for the use of *Music & Audio* (0.99) and for extraversion as a predictor for the use of *Communication* applications (0.94). The lowest acceptable stability could be observed for *Education* as a predictor for *Lifestyle* app usage (0.67) and extraversion as a predictor for *Media & Video* applications (0.69).

Table 2.4: Pairwise Spearman Correlations Between Demographics, Big Five Factor Scores and App Usage.

	Gender	Age	Bildung	E	A	ES	C	O
Tools	0.06 [-0.11, 0.23]	-0.1 [-0.26, 0.07]	-0.08 [-0.24, 0.09]	0.06 [-0.11, 0.23]	0.12 [-0.05, 0.28]	-0.05 [-0.22, 0.12]	-0.04 [-0.21, 0.13]	0 [-0.17, 0.17]
Games	-0.08 [-0.24, 0.09]	-0.29 [-0.44, -0.13]	-0.19 [-0.35, -0.02]	-0.05 [-0.22, 0.12]	0.02 [-0.15, 0.19]	0.02 [-0.15, 0.19]	-0.15 [-0.31, 0.02]	-0.13 [-0.29, 0.04]
Entertainment	-0.15 [-0.31, 0.02]	-0.21 [-0.36, -0.04]	-0.2 [-0.36, -0.03]	0.02 [-0.15, 0.19]	-0.03 [-0.2, 0.14]	-0.01 [-0.18, 0.16]	-0.06 [-0.23, 0.11]	-0.07 [-0.23, 0.1]
Productivity	-0.2 [-0.36, -0.03]	-0.06 [-0.23, 0.11]	0.02 [-0.15, 0.19]	0.05 [-0.12, 0.22]	0.01 [-0.16, 0.18]	0.02 [-0.15, 0.19]	-0.08 [-0.24, 0.09]	-0.05 [-0.22, 0.12]
News & Magazines	-0.17 [-0.33, 0]	0.12 [-0.05, 0.28]	-0.01 [-0.18, 0.16]	-0.03 [-0.2, 0.14]	-0.06 [-0.23, 0.11]	0.02 [-0.15, 0.19]	-0.08 [-0.24, 0.09]	-0.11 [-0.27, 0.06]
Photography	-0.11 [-0.27, 0.06]	-0.05 [-0.22, 0.12]	-0.13 [-0.29, 0.04]	0.12 [-0.05, 0.28]	-0.04 [-0.21, 0.13]	-0.16 [-0.32, 0.01]	0.03 [-0.14, 0.2]	0.03 [-0.14, 0.2]
Shopping	-0.15 [-0.31, 0.02]	-0.14 [-0.3, 0.03]	-0.13 [-0.29, 0.04]	0.12 [-0.05, 0.28]	0.03 [-0.14, 0.2]	0.05 [-0.12, 0.22]	0.08 [-0.09, 0.24]	-0.05 [-0.22, 0.12]
Communication	-0.01 [-0.18, 0.16]	-0.09 [-0.25, 0.08]	-0.12 [-0.28, 0.05]	0.27 [0.11, 0.42]	0.03 [-0.14, 0.2]	-0.05 [-0.22, 0.12]	-0.07 [-0.23, 0.1]	-0.01 [-0.18, 0.16]
Books & Reference	-0.06 [-0.23, 0.11]	-0.2 [-0.36, -0.03]	-0.09 [-0.25, 0.08]	0.03 [-0.14, 0.2]	-0.06 [-0.23, 0.11]	-0.12 [-0.28, 0.05]	-0.11 [-0.27, 0.06]	0.03 [-0.14, 0.2]
Travel & Local	-0.12 [-0.28, 0.05]	0.02 [-0.15, 0.19]	0.06 [-0.11, 0.23]	0.11 [-0.06, 0.27]	-0.13 [-0.29, 0.04]	-0.09 [-0.25, 0.08]	-0.23 [-0.38, -0.06]	-0.08 [-0.24, 0.09]
Music & Audio	-0.34 [-0.48, -0.18]	-0.21 [-0.36, -0.04]	-0.15 [-0.31, 0.02]	0.14 [-0.03, 0.3]	-0.01 [-0.18, 0.16]	0.04 [-0.13, 0.21]	-0.06 [-0.23, 0.11]	-0.03 [-0.2, 0.14]
Business	-0.08 [-0.24, 0.09]	-0.22 [-0.37, -0.05]	-0.14 [-0.3, 0.03]	-0.01 [-0.18, 0.16]	-0.12 [-0.28, 0.05]	-0.13 [-0.29, 0.04]	-0.08 [-0.24, 0.09]	-0.03 [-0.2, 0.14]
Lifestyle	-0.04 [-0.21, 0.13]	-0.07 [-0.23, 0.1]	-0.13 [-0.29, 0.04]	0.15 [-0.02, 0.31]	0.02 [-0.15, 0.19]	-0.03 [-0.2, 0.14]	0.04 [-0.13, 0.21]	-0.06 [-0.23, 0.11]
Transportation	0.04 [-0.13, 0.21]	-0.17 [-0.33, 0]	0.03 [-0.14, 0.2]	0.18 [0.01, 0.34]	0.2 [0.03, 0.36]	0.18 [0.01, 0.34]	0.07 [-0.1, 0.23]	0.08 [-0.09, 0.24]
Weather	0.01 [-0.16, 0.18]	0.05 [-0.12, 0.22]	0.13 [-0.04, 0.29]	0.11 [-0.06, 0.27]	-0.04 [-0.21, 0.13]	0.01 [-0.16, 0.18]	0.07 [-0.1, 0.23]	-0.05 [-0.22, 0.12]
Browser	-0.09 [-0.25, 0.08]	-0.25 [-0.4, -0.09]	-0.14 [-0.3, 0.03]	0.06 [-0.11, 0.23]	0.02 [-0.15, 0.19]	-0.1 [-0.26, 0.07]	-0.12 [-0.28, 0.05]	-0.03 [-0.2, 0.14]
Media & Video	0.16 [-0.01, 0.32]	-0.15 [-0.31, 0.02]	-0.16 [-0.32, 0.01]	0.07 [-0.1, 0.23]	0.09 [-0.08, 0.25]	-0.19 [-0.35, -0.02]	-0.13 [-0.29, 0.04]	-0.04 [-0.21, 0.13]
Social	0.07 [-0.1, 0.23]	-0.28 [-0.43, -0.12]	-0.2 [-0.36, -0.03]	0.09 [-0.08, 0.25]	0.03 [-0.14, 0.2]	-0.05 [-0.22, 0.12]	0.01 [-0.16, 0.18]	-0.11 [-0.27, 0.06]

Note: Pairwise Spearman correlations between *Big Five* measures, demographic variables and app usage categories. Square brackets contain 95% confidence intervals. Abbreviations stand for Entertainment, Productivity, News & Magazines, Photography, Shopping, Communication, Books & Reference, Travel & Local, Music & Audio, Lifestyle, Transportation, and Media & Video from left to right; E = extraversion, A = agreeableness, ES = emotional stability, C = conscientiousness, O = openness.

Table 2.5: Descriptive Statistics - App Usage

Category	$M_{\#uses}$	$SD_{\#uses}$	M_{usage}	SD_{usage}	#Apps	#Users	Top 5 Apps
Communication	37.10	25.00	31.2s	14.2s	184	137	WhatsApp, Contacts, Dialer, Mail, Facebook Messenger
Social	8.10	10.37	47.9s	43.7s	120	126	Facebook, Instagram, Snapchat, Twitter, Weibo
Browser	7.25	5.89	71.6s	32.2s	11	136	Chrome, Internet, Firefox, Opera, Dolphin Browser
Tools	5.46	6.84	18.3s	10.7s	568	137	Google Search, Clock, Google Play Store, Calculator, S Voice
Productivity	4.22	3.87	22.3s	9.8s	297	137	Settings, S Planner, Calendar, ColorNote, Google Drive
Photography	4.11	3.47	24.5s	17s	131	129	Gallery, Camera, SnapApp, Album, PicsArt
Games	3.79	5.48	122.3s	153.5s	326	100	Clash of Clans, Quizduell, Candy Crush Saga, Farm Heroes Saga, Trials Frontier
Music & Audio	2.97	1.89	13.6s	10.2s	172	135	Spotify, Music Player, Google Play Music, MP3-Player, SoundCloud
Travel & Local	2.65	1.11	49.3s	26.2s	150	134	Maps, MVV Companion, TripAdvisor, BlaBlaCar, Airbnb
Entertainment	2.60	1.54	72.7s	82.7s	168	134	YouTube, 9GAG, PlayerPro, appinio, PS4-Magazin
Books & Ref.	2.18	1.83	28.9s	59.1s	115	123	Munpia, dict.cc plus, dict.cc, Wikipedia, LEO
Transportation	2.03	1.50	36.7s	30s	54	110	MVG Fahrinfo, DB Navigator, MeinFernbus, Uber, mytaxi
Media & Video	2.01	1.20	40.8s	146.6s	103	134	Video-Player, Google Play Movies, VLC, Video, ZDF
News & Mag.	2.01	1.24	28s	39.8s	114	126	FOCUS Online, SPIEGEL ONLINE, Flipboard, SZ.de, N24 News
Lifestyle	1.50	2.06	21.6s	36.2s	111	75	Tinder, Sleep, Chefkoch, eBay Kleinanzeigen, PAYBACK
Business	1.27	1.29	30.4s	42.9s	59	118	AnyConnect, POLARIS Office Viewer 5, Polaris Viewer 4.1, XING, Quickoffice
Health & Fitn.	1.15	1.61	16.6s	41.1s	105	61	SleepBot, Strava, Fitbit, Freeletics, MyFitnessPal
Shopping	1.14	1.52	34.5s	103.2s	65	69	eBay, mydealz, Amazon, brands4friends, Shpock
Education	1.10	2.42	27.1s	72.8s	106	54	UnlockYourBrain, AnkiDroid, TUM Campus App, Duolingo, Web Opac
Sports	0.87	3.58	17.2s	75.9s	47	34	kicker, Comunio, Kicktipp, Score!, Sportschau
Weather	0.85	0.88	10s	14.9s	45	74	Weather, wetter.com, WetterOnline, WetterApp, Wetter-Widget
Finance	0.63	1.15	12.8s	30.1s	53	39	Sparkasse, Banking 4A, Wüstenrot, YNAB, Banking
Personalization	0.50	4.56	1.7s	8.8s	95	32	Aviate, Backgrounds, Zedge, Flatastico, HD Widgets
Medical	0.21	0.80	2.1s	8.7s	15	18	Lady Pill Reminder, PillReminder, Pillreminder, iPhysikum, Remember Your Pill
Comics	0.08	0.41	5.7s	38.6s	6	6	xkcd Browser, NICHTLUSTIG, Marvel Unlimited, xkcdViewer, xkcd - Now

Note: $M_{\#uses}$ = avg. usage count across all participants and days, $SD_{\#uses}$ = standard deviation of avg. usage count; M_{usage} = avg. single usage duration across all usages, M_{usage} = standard deviation of avg. single usage duration; *#Apps* total number of apps in the category across all participants in our dataset, *#Users* respective number of users that ever used an app from the respective category during data collection, top five apps for each category; M = mean, SD = standard deviation. Table is sorted in descending order by $M_{\#uses}$.

Table 2.6: Variable selection | Prediction of App Usage

		Criteria											
	FACTOR LEVEL	Games	Product.	Photo.	Comm.	Travel	Music & Audio	Business	Lifestyle	Transp.	Browser	Media & Video	Social
Predictors	Gender (female)	-	0.84 | 0.56	-	-	-	0.99 | 0.41	-	-	-	-	-	-
	Age	-	-	-	-	-	-	0.84 | 0.93	-	-	0.79 | 0.95	-	0.78 | 0.90
	Education	-	-	-	-	-	-	-	0.67 | 0.59	-	-	-	-
	Extraversion (E)	-	-	0.68 | 1.40	0.94 | 1.30	-	-	-	-	-	-	0.69 | 1.34	-
	Agreeableness (A)	-	-	-	-	-	-	-	-	0.73 | 1.36	-	-	-
	Conscientiousness (C)	0.81 | 0.54	-	-	-	0.70 | 0.76	-	-	-	-	-	-	-
	Mean DSS	24.12	24.46	20.82	30.13	17.31	19.55	13.57	18.52	17.36	24.53	19.7	27.18
	FACET LEVEL	Games	Product.	Photo.	Comm.	Travel	Music & Audio	Business	Lifestyle	Transp.	Browser	Media & Video	Social
Predictors	Gender (female)	-	.65 | 0.56	-	-	-	0.96 | 0.41	-	-	-	-	-	-
	Age	-	-	-	-	-	-	0.64 | 0.93	-	-	0.59 | 0.95	-	0.67 | 0.90
	Sociableness (E2)	-	-	0.57 | 1.19	0.92 | 1.15	-	-	-	0.56 | 1.27	-	-	0.63 | 1.18	-
	Sense of duty (C3)	-	-	-	-	0.58 | 0.84	-	-	-	-	-	-	-
	Willingness to trust (A1)	-	-	-	-	-	-	-	-	0.68 | 1.25	-	-	-
	Mean DSS	-	24.46	20.61	29.99	17.22	19.55	13.57	18.28	17.23	24.53	19.52	27.18

Note: The left values are the respective probabilities of variable selection at factor or facet level, obtained with stability selection. Right values represent $\exp(\hat{\beta})$ coefficients from quasi-poisson regression models between *Big Five* factor scores, demographics and app usage variables. Numbers greater than one represent a positive, number smaller one a negative relationship. Empty cells refer to not-selected variables. Interpretation: Coefficients greater than 1 describe the percentage of increase in app usage that go along with an increase of 1 unit in the personality score (e.g., extraversion ~ Comm.: 1.3 ≙ 30 % increase). Scores below 1 indicate a negative relationship (e.g., Games*conscientiousness: 0.54 ≙ 100 - 54 ≙ 46 % decrease) and indicate the percentage of decrease in app usage per one unit increase in the respective predictor score). Note that “1” represents male and “2” represents female labeling for the variable gender when interpreting coefficients. Abbreviations of categories stand for Productivity, Photography, Communication, Travel & Local, Transportation, from left to right; DSS = Dawid Sebastian Score.

The psychometric and demographic variables chosen by stability selection, were modeled as predictors in Quasi-Poisson generalized linear regression models. In Table 2.6, positive as well as negative relationships can be observed. Female gender, age, education and conscientiousness seem to be generally negatively associated with app usage. Women seem to use less apps related to *Productivity* (-44%) and most prominently *Music & Audio* (-59%). Age showed rather small negative relationships with app usage. Hence, one unit increase in age was negatively associated with app use in the categories *Business* (-7%), *Browser* (-5%) and *Social* (-10%). Interestingly, *Education* was not associated with the categorical app use, except for a strong negative relationship with *Lifestyle* app usage (-41%). Beyond demographic variables, extraversion was generally positively associated with app usage. An increase of one unit in extraversion was associated with app usage increase in *Photography* (+40%), *Communication* (+30%) and *Media & Video* (+34%) applications. The factor agreeableness was positively associated with the use of apps related to *Transportation* (+36%). One unit increase in conscientiousness decreased app usage for *Games* (-46%) and *Travel & Local* (-24%) apps.

Prediction of App Use – Big Five Facet Level

In general, the same procedure of analysis was performed with personality predictors on factor and facet level and will therefore not be explained again. Table 2.6 shows the results of the features selection and regression modeling on facet level. For a more intuitive understanding of the presented relationships, results from feature selection and regression modeling are described in a combined form in this section. Furthermore, we elaborate on differences between the factor and facet level analyses.

Although the results of feature selection show similarities with the factor level analysis (section 2.2.4), two differences are apparent. The game application usage, predicted at factor level could not be predicted when facet level scores together with demographics were used. Furthermore, *Education* was not selected as an effective predictor when competing against personality on facet level. Hence, the use of *Lifestyle* apps is more effectively explained by the extraversion subfacet sociableness. An one-unit increase in sociableness therefore is related to an 27% increase in the use of *Lifestyle* applications.
Other relationships are in general very similar to the associations found at *Big Five* factor level, please see Table 2.6. Further comparison between factor and facet level predictors show that associations (exp(β) coefficients) with app usage categories are generally higher for factor level predictors in comparison with facet predictors. This is true for

extraversion, agreeableness and conscientiousness in comparison with the respective facets (sociableness, sense of duty, willingness to trust), please compare factor and facet level in Table 2.6.
Comparing facet-level models to factor models, model fit is higher (lower DSS values) on facet level in six out of eleven comparable models and equal in five models. In other words, model fit was not better for any factor model in comparison with a facet level model. Please see the *Mean DSS* rows in Table 2.6 for factor and facet level.

2.2.5 Discussion

Self-reports of personality can be validated with frequencies of mobile application usage on smartphones. In the present study, we provide such evidence for the dimensions extraversion, conscientiousness and agreeableness. Both, factor and facet models were effective in the prediction of categorical app usage, with tendencially better facet models in some instances.

Furthermore, our results suggest that in addition to demographic factors, personality dispositions can be linked to variations in app usage frequencies towards a larger degree than suggested by previous research (Kim et al., 2015; Montag et al., 2014).

In the following we will discuss the various effects discovered in our data and suggest possible explanations as a starting point for prospective research.

Personality and App Usage

Extraversion and its respective facet sociableness were related to increased application usage in categories related to *Photography*, *Communication*, *Lifestyle* and *Media & Video*. Higher frequency of *Communication* app usage are in line with previous literature reporting higher numbers of call-related activities (Chittaranjan et al., 2013; Kim et al., 2015; Montag et al., 2014), as well as a higher usage frequency of the WhatsApp messenger (Montag et al., 2015) for people with higher scores in extraversion.

Furthermore, our results show a positive relationship of *Photography* app usage with extraversion/sociableness. This result is possibly related to previous work, reporting increased photo uploads and photo sharing for higher values in extraversion (Eftekhar, Fullwood, & Morris, 2014; Hunt & Langstedt, 2014). Interestingly, extraversion- sociableness is also the personality dimension that shows the most positive associations with various categories of app usage in general. This might reflect an aspect of the personality trait that is described in the literature as the need for external stimulation (Butt & Phillips, 2008; H. J. Eysenck, 1967). People might satisfy this need through communication, or other channels such as the consumption of *Lifestyle*, *Media & Video*, *Photography* apps.

Agreeableness and the respective facet willingness to trust seem to be related to the use of *Transportation* apps. Higher *Transportation* app usage in agreeable people could be connected to them being more affine to public transportation than private transportation in comparison with less agreeable individuals. This is backed by the notion that agreeable people tend to be more prosocial in the sense of being tolerable of others, preferring cooperation over competition (Graziano & Tobin, 2009).

Conscientiousness and the facet sense of duty was associated with lower frequencies of *Gaming* as well as the use of *Travel & Local* apps. The lower usage of *Gaming* apps of highly conscientious people supports the notion that conscientious people are more focused on their tasks and less likely to engage in procrastination activities (Lee, Kelly, & Edwards, 2006).

Harder to interpret is the negative relationship with the *Travel & Local* app category. It could be related to the fact that conscientious people are usually also more traditional in their behaviors and attitudes. Closer inspection of the top apps in the *Travel & Local* category backs that idea. Besides public transportation, this category also covers more liberal and modern concepts such as ride-sharing (e.g., BlaBlaCar) or home-sharing (e.g., Airbnb). These activities could very well be less interesting for more traditional or conservative people. However, this hypothesis should be experimentally investigated in prospective studies.

Emotional Stability & Openness were not predictive for any app usage categories on neither factor nor facet level. Emotional stability or Neuroticism is a personality dimension that is defined through feelings and emotions rather than actions (John & Robins, 1993; Vazire, 2010) and has even been associated with behavioral restraint (Hirsh et al., 2009). Emotional stability therefore is a dimension that is not easily observable and evaluable. Symptoms of depression, are hard to detect for that reason (Mehl, 2006). As in this work we only focused on app usage frequencies at categorical level, it is not surprising that no stable associations with emotional stability could be observed. However,

considering the association of Neuroticism and its link to depression (Hodgins & Ellenbogen, 2003; Ormel et al., 2013) as well as the link between reduced activity and social contact associated with depression (Burton et al., 2013), variables in relation to these dimensions could be retrieved from data logs (e.g., GPS). Additionally the analysis of word use e.g. in chats could be a promising approach (Yarkoni, 2010).
The missing predictability of openness for app usage, is harder to explain. As reported in Section 2.2.4, openness shares a lot of variance with extraverison, highlighting the heterogeneity of the construct. Furthermore, previous research indicates that extraversion and openness could be even combined to a single personality dimension related to the engagement in behavior and the incorporation of new environmental information (Hirsh et al., 2009). However, Spearman correlations in Table 2.4 do not show any significant correlations between openness and app usage suggesting other reasons in our case. Openness is also considered to be the most heterogenic Big Five dimension (DeYoung, 2015), related to both intellectual abilities and affinity for exploration. Most likely, it can be only be related to more specific behaviors unlike the broad app usage categories used in this study.

Factors vs. Facets

In general, model fits as well as directions of effects are very similar between factor and facet level models. However, game application usage could only be predicted with conscientiousness scores on factor level. Furthermore, direct comparison shows marginally better DSS scores for facet level models in six categories. Related to that, some previous literature reports higher predictive performance of personality facets over Big Five factor scores on behavioral criteria (Anglim & Grant, 2014; Paunonen & Ashton, 2001). Although our results tendencially support this notion, it cannot be concluded that facet level scores are generally better in the prediction of behavior. Differences are marginal and rather suggest a similar predictive performance of the selected factor and facet level traits. On the one hand, model fit scores suggest at least equivalent predictive performance of selected personality facets in comparison with factors. On the other hand regression coefficients for factor level traits are higher throughout the analysis. However, our results do not support previous claims that broad personality traits are better in the prediction of broad behavioral categories (Hogan & Roberts, 1996). However, it also remains to be seen whether more narrow real behavioral criteria, such as single app usage or even isolated behaviors performed within apps are better predictable by narrow traits (Hogan & Roberts, 1996).

Furthermore, it is important to highlight differences between this study and previous studies comparing factor and facet personality scores. In our case, data logs of actual behavior were used, whereas previous literature mainly focused on the prediction of self-reported behavior (Paunonen & Ashton, 2001; Salgado et al., 2013). Furthermore, the measured personality dimensions as well as the psychometric quality of the tests might change measured personality scores. We used the rasch-scaled BFSI (Arendasy, 2009) in order to measure the latent personality variables, while acknowledging the hierarchical *Big Five* structure. The use of other instruments with no conformity to the *Partial-Credit model* (Masters, 1982) might yield different results.

Finally, one has to consider our conservative modeling approach and the hierarchical structure of the Big Five model. We only selected the best predictor for each app usage criterion. Therefore, the consideration of several personality facets could very well improve prediction performance in comparison with factor level scores. However, this goes beyond the scope of this paper and should be investigated in prospective investigations.

Demographics and App Usage

Gender effects suggest lower application usage frequencies in two categories. Most prominently, our results show that women in average used apps related to *Music & Audio* only half as often as men did. This is in accordance with results of Kim et al. (2015) who also reported lower frequencies of entertainment application use (including music) for women. It is unclear why we observed this effect with such stability. One reason for this could be that in our sample, women and men show different listening behaviors. Our logging method does not include music consumption on secondary devices (such as mp3-players or iPods). Therefore, women could for example rather prefer listening to music on separate devices in comparison with men. Furthermore, this could also be related to technology acceptance (Sherman et al., 2000), as many apps in the *Music & Audio* category were related to novel music streaming services (see Table 2.5). According to these statistics (based on an American and UK sample) on Globalwebindex.com (Globalwebindex & Mander, 2015a), differences between men and women in the adoption of paid music streaming services exist, however not at the magnitude observed in our sample.

Another interesting effect suggests that women use less apps related to productivity in comparison with men. caution has to be taken with interpretation of this effect. Although gender differences in productivity have been investigated in previous research (Leahey, 2006; Reed, Enders, Lindor, McClees, & Lindor, 2011), this effect could also be explained otherwise. Although the lower usage of productivity apps could suggest lower productivity for women, it could also suggest the opposite - women use less productivity apps because they do not need them.

Increasing age was generally associated with small decreases in *Business*, *Browser* and *Social* app usage. These results are related to those of Kim et al. (2015) who also reported lower app usage frequencies for relation apps (messaging and social). In both studies age only poses a small effect on app usage. However, at this point it is important to point out that it is very likely that these small effects in the present study as well in the study of (Kim et al., 2015) are caused by self selection effects. Furthermore, they reported a large negative correlation between age and smartphone ownership (r = -0.56) (Kim et al., 2015). Furthermore, in our study only participants with an compatible Android phone could participate. As this suggests that smartphone ownership drastically declines with age and the variation in the data only describes effects of mostly younger participants, the current results cannot rule out different app usage behavior of older people in general.

In contrast to results of Kim et al. (2015), no big effects of *Education* on app usage were found in the present study. The only observed effect shows drastic decreasing app usage frequency of lifestyle apps for people with higher education. This result is contradictory to results of Kim et al. (2015) who reported an increase in the *Information* app category including lifestyle with an increase in education 11% and age 3%. It is not completely clear which apps they clustered in the reported categories (Kim et al., 2015). However, lifestyle app usage in the present study was topped by the popular dating app *Tinder*, mainly used by people in their twenties, a point where university education mostly has not been completed, and people are unmarried. This combination corresponds to the majority of *Tinder* users (backed by Globalwebindex and Mander (2015b)). Thus, the relationship with education might be confounded with age and marital status as well as the skewed education distribution in our sample. A more useful explanation of *Lifestyle* app usage could be provided by the extraversion facet sociableness.

Limitations

The present study has several strengths as well as limitations. The major strengths of the study are related to the detailed measurement of Big Five personality dispositions as well as the reliance on actual behavior recorded with an app specifically designed for that purpose. Additionally, this study focuses on the most stable effects with regard to the implemented stability feature selection and the consequent regression analyses. In particular, this study highlights how mobile application usage varies with regard to personality factors.

However, there are important limitations to be noted. Our sample was purely collected from the German population in Munich with age and education not being perfectly representative of the general population. However, as smartphone usage is less prevalent with older people (Kim et al., 2015), our sample might not be too different to the normal population of smartphone users in that regard. Moreover, usage patterns might differ when compared to, for example, samples from other cities and countries. For instance, availability and popularity of public transportation impacts the use of apps in the related category. Application usage is can also be different with regard to the cultural background and country of an user. However, although some variation in app usage is to be expected, many popular apps for common tasks are globally available or have popular regional equivalents. Furthermore, associations between app usage and age were similar to previous results although smaller statistical associations were to be expected in a more homogeneous sample.

It also has to be noted that in the present study investigated categorical app usage - a fraction of activities trackable on smartphones. Therefore, it is likely that the inclusion of additional parameters (e.g., GPS locations, calls, single app usage) will make it possible to establish more relationships with personality traits.

We highlight that one has to be careful with drawing post-hoc conclusions based on the observed relationships. While our results indicate interesting avenues for both personality research in academics as well personalization research in industrial settings, we understand the reported relationships as promising starting points for closer investigation, not as established facts.

2.2.6 Conclusions & Outlook

This study is one of few to investigate relationships between self-reported *Big Five* personality on factor and facet level and automatically logged measures of app usage. Our results show that variations in mobile application use can be linked to both demographical factors and personality, beyond previously reported associations (Chittaranjan et al., 2013; Kim et al., 2015; Montag et al., 2015).

Precisely, the personality dimensions of extraversion, agreeableness and conscientiousness on factor and facet level, predict app usage in several categories. The discovered associations fit well into the existing literature and provide validating evidence for self-reported personality dimensions. These findings aid development and validation of personality inventories and promote the use of logging data for the investigation of behavioral personality aspects.

Beyond the field of methodological research in psychological science, discovered associations could also help with the development of personality-adaptive recommender-systems in the field of HCI, facilitating the creation of more individual content. Hence, knowledge about an users dispositions could be used in order to choose better default settings and e.g. app rankings and system language style in technical systems (e.g., smartphones).

Further research is needed in order to cross-validate and extend research efforts beyond the investigated factors. Additionally, specific usage patterns such as single application usage should be investigated more closely. In particular, the specific content of apps is likely to be descriptive of the user's personality. For example, some studies indicated that sending images via instant messaging services is related to the user's personality (Hunt & Langstedt, 2014). However, the actual content of a photography might be additionally interesting, as it serves as an indicator of which information a person is willing to share with others and is considered important.

Validation studies should collectively use different metrics obtainable with smartphones and mobile sensors such as call behaviors, geoposition and frequently used words. Related to that, prospective studies should also collect larger samples in order to identify more potential relationships between personality and mobile behavior (Kim et al., 2015). This approach could than help to associate openness and emotional stability with loggable information.

Conclusively, the present study shows how personality is associated with differences in application usage on smartphones and how such behavioral measures can be used in order to validate self-reports of personality.

2.2.7 Author Contributions

Besides myself, Jiew-Quay Au, Bernd Bischl, Heinrich Hussman, Daniel Buschek, Alexander De Luca and Markus Bühner contributed to the creation of this article. Daniel Buschek and Alexander De Luca greatly assisted with creation of the data collection methodology and proof read of the article. Jiew-Quay Au assisted with data cleaning and statistical analysis. Bernd Bischl, Heinrich Hussmann and Mr. Bühner acted as the supervising authors of this article.

2.2.8 Acknowledgements

We want to thank all the students (*Alexandra Dietl*, *Victoria Gebhardt*, *Sarah Huber*, *Jakob Kaiser*) for supporting us with recruiting, psychometric testing and organization as well as *Martin Fröhlich & Matthias Lamm* for their programming efforts. Furthermore we would like to thank the LMU data protection officer Dr. iur. Marco Wehling for his assistance in legal matters. This project was partially funded by a *Google Research Award*.

2.3 Study 3: Personality Recognition from Smartphone Data

2.3.1 Abstract

This study explores the potential of smartphones to collect a broad variety of behavioral proxy measures and use these measures to predict individual big five personality traits on both factor and facet level. Therefore, a total of 137 subjects (87 female) with an average age of 24 (SD = 4.72) and above average education level (96% completed A-levels) participated in a 90 minutes psychometric lab session as well as in a consequent 60 days passive logging study in the field. Prediction modeling was performed in order to recognize individual personality from user data. Although results suggest several correlations between individual personality scores and behavioral variables, prediction modeling shows only limited success. Variable importance measures of predictable personality facets relate to previous research and highlight the potential of automated data collection with consumer electronics in the field of psychological science. Implications for prospective research as well as society are discussed.

2.3.2 Introduction

Individual differences in personality have been repeatedly shown to exert influence on important life outcomes and behaviors on individual and inter-individual level (Ozer & Benet-Martínez, 2006; Roberts et al., 2007). Therefore, the ability to judge someones personality can have non-trivial implications for life outcomes such as intimate relationships (Malouff et al., 2010) or turnover decisions (Zimmerman, 2008).

People are surprisingly fast and accurate at judging their own as well as others personalities (Ready, Clark, Watson, & Westerhouse, 2000), even with complete strangers (Carney, Colvin, & Hall, 2007). Knowledge of someones personality helps us to predict how others will behave in certain situations and what they will prefer (e.g., partying with the crowd or enjoying the evening with selected friends).

Besides effects on individual decisions, knowledge about systematic differences in behaviors and preferences are very valuable for business applications and adaptive systems. Personality has also been linked to differences in individual preferences for things like movies, websites, brands and products (Kosinski, Bachrach, Kohli, Stillwell, & Graepel, 2014, 2013; Youyou et al., 2015). Therefore, knowledge about these relationships could improve ads as well as suggestions for new and unrelated products. Furthermore, adaptive-personalized systems in general could benefit from the inclusion of individual personality scores. Recommendations for restaurants in digital online maps could for example be adapted to individual preferences in order to not simply display ratings from all users but from users similar to oneself (e.g., restaurant ratings). Cars could (under consideration of additional factors) suggest routes and destinations with regard to individual personality.

However, the assessment of individual personality dispositions with conventional self-report methods for use adaptive systems is not feasible in most applications. Hence, automatic recognition of an users personality is desirable in order to effectively adjust system parameters and recommendations. Nowadays, digital records are produced permanently through usage of everyday consumer technology. Visited websites, "likes" in social networks and performed user actions on smartphones ubiquitously create records of these events.

In the present work, we investigate the potential of Android smartphone logs as predictors for individual self-reported personality scores. In particular we use an predictive modeling approach to quantify relationships between personality and smartphone usage in terms of predictive accuracy.

Two previous studies reported successful prediction of big five personality scores from smartphone usage data above chance (Chittaranjan et al., 2013; De Montjoye et al., 2013). Chittaranjan et al. (2013) used continuous smartphone usage data, collected over a period of 17 month and reported correlations between these variables and self-reported personality, measured with the *Ten-Item-Personality-Inventory (TIPI)* (Gosling et al., 2003). Furthermore, they used support vector machines (SVM) (Cortes & Vapnik, 1995) with a radial basis kernel and predicted median binned personality with variables calculated from phone usage.

In a similar project, De Montjoye et al. (2013) collected self-reports of the big five personality dimensions, measured with the Big Five Inventory (BFI) (Rammstedt & John, 2007) and used call and message logs as well as GPS location data as predictors for personality in a three-class prediction problem. They also used an SVM algorithm for classification and reported prediction accuracies of up to 61% (22% higher than the baseline).

However, in both studies variable selection was performed prior to statistical modeling (Chittaranjan et al., 2013; De Montjoye et al., 2013). This approach indicates that reported prediction accuracies might be too optimistic as models were possibly overfitted on the data. It is important to note that variable selection should be performed during the respective resampling iterations in the training stage of the statistical learning model (Bischl et al., 2012; Simon, 2007).

Furthermore, both studies used relatively short personality questionnaires with questionable psychometric properties. De Montjoye et al. (2013) used a 44 item personality questionnaire and Chittaranjan et al. (2013) used the TIPI, at ten item personality questionnaire. Although shorter questionnaires are quicker to administer and score, reliability and validity of these instruments needs to be questioned (Heene et al., 2014; M. Ziegler et al., 2014). In particular it is uncertain that very short questionnaires (e.g., TIPI) are able to fully record the hierarchical structure of the big five personality model.

Comprehensively, it is unclear whether reported predictive capabilities of smartphone usage data for self-reported personality are too optimistic due to possible overfitting. Furthermore, predictions could even be improved with better self-report questionnaires and more available data. Therefore, in this study we intend to circumvent methodological shortcomings of previous investigations and extend beyond previous attempts with a more reliable measurement of self-reported personality, a larger sample, and more sophisticated resampling strategies.

Besides the usage of smartphone data for the prediction of individual differences from user behavior, several other approaches with regard to behavior exist:

Related to the Big Five personality models' development with an psycho-lexical approach (Allport & Odbert, 1936; Norman, 1963), language use has been investigated as a promising predictor of individual self-reported personality. Word usage in relation to word usage in conversation has been investigated (Fast & Funder, 2008; Yarkoni, 2010). In addition to word count, Mairesse and Walker (2006), Mairesse et al. (2007) used a wide range of linguistic features (e.g., prosody, utterances) to predict individual personality scores with accuracy above the baseline. Others used acoustic features (e.g., pitch and loudness) (Polzehl, Moller, & Metze, 2010).

Newer studies have investigated relationships between activities on social media and personality (also involving language). The investigation of user data from social networks for personality studies are promising as both individual preferences (e.g., likes) and linguistic information can be retrieved. Farnadi et al. (2016) provides a nice overview of this approach.

Most impressively, Youyou et al. (2015) used a sample of 86,220 Facebook users to predict individual self-reports of personality from Facebook likes. Furthermore, they compared accuracy of personality judgment between computer models with personality judgments about the same person made by significant others. Their results indicate that predictive models show superior predictive accuracy (self-other agreement) as well as higher external validity for the prediction of personality related life outcomes in comparison with ratings of facebook friends.

Usage of data from social networks is especially promising for utilization in recommender systems as many people own an account and online data is accessible at ease. More difficult however is to provide personality judgments for people who are relatively inactive on social networks or people who do not own an user account.

In this study we investigate the possibility to infer big five personality scores from the wide range of behavioral proxy measures, extracted from naturalistic smartphone usage. Furthermore we describe a large number of variables, derived from previous literature and show their relationship with big five personality facets. We also intend to show how social science can benefit from the usage of modern statistical learning procedures in order to describe the complex and manifold relationships of self-reported personality and associated behaviors.

2.3.3 Method

Participants

In total 178 participants were recruited from the academic population at LMU, as well as from forums, social media, blackboards, flyers, and direct recruitment in the streets of Munich, Germany, between September 2014 and April 2015.

The data of 41 participants could not be used for analysis as both the data transfer to the server did not work properly and they did not pick up their compensation at the end of the study. The final sample consisted of 50 males and 87 females. In average, participants were 24 years old (SD = 4.72), with 75% of the participants being 26 or younger. The majority of the sample had at least completed high school (96%) and 31% of all participants had completed education at university level. All subjects gave written consent prior to participation and could withdraw participation in the study as well as demand deletion of non-anonymized data at any time. Participants received a document including the results of the completed psychometric tests as well as either course credit (3h), financial compensation (EUR 30) or a combination of both (EUR 15 and 1.5h). All subjects participated voluntarily and signed a written consent form in advance of participation. The conduct of this study was approved by the responsible IRB as well as the responsible data protection officer of the LMU. Please note that the sample used in this case is equal to the sample used in Section 2.2. However, in this study, much more variables (in addition to app usage frequencies) were used in the analysis.

Apparatus

Personality Measures & Demographics Big Five Personality dimensions were measured on facet level via six subscales, respectively. Therefore, the computer-based German version of the Big Five Structure Inventory (BFSI) was used due to its good psychometric properties and short duration (Arendasy, 2009). In contrast with more common personality inventories such as the NEO-PI-R or the more current version the NEO-PI-3 (R. R. McCrae et al., 2005), the BFSI uses adjectives or short statements and has been developed using item response theory rather than classical test theory. The authors report psychometric benefits over other similar questionnaires due to conformity with the partial credit model (Masters, 1982).

The BFSI consists of 300 items (adjectives and short phrases) and measures the *Big Five* personality dimensions (agreeableness, openness to experience, conscientiousness, extraversion and emotional stability/absence of neuroticism) on factor and facet level with a four-step likert scale ranging from "*untypical for me*" to "*rather untypical for me*" to "*rather typical for me*" to "*typical for me*".

In addition to the personality scores, age, gender and current level of completed education was collected. Gender was recorded dichotomously with "0" representing male and "1" representing female participants. Level of education was subdivided in five categories from "*no education*" to "*compulsory education*" to "*vocational training*" to "*A-levels*" to "*finished university degree*".

Additionally, selected subtests of the German version of the Intelligence Structure Battery (INSBAT) were administered. However, intelligence scores were collected for a different study and will therefore not be described in further detail. Relationship status, music listening behavior and the number of personal mobile devices were collected with a demographic questionnaire.

Android Application The majority of data was collected with an *Android* smartphone app, specifically designed to effectively log user data in an anonymous and most unobtrusive way. In Figure 2.4 an image of the app-interface is visible. Collected user data was regularly transferred to a server at LMU using SSL encryption once participants were connected to a wireless network. In contrast with traditional methods for behavioral observation, the usage of a smartphone application enabled us to unobtrusively record a large number of behaviorally related parameters.

The logged measures included but were not limited to, event-related data about calls, messages, longitude/latitude position, app starts/installations, screen activation, flight mode activation, Bluetooth connections, played music, battery charging status, photo and video events (no actual photos or videos) and connection events to wireless networks were collected. Additionally, message lengths, installed apps, technical device characteristics were collected. Irreversibly hash encoded versions of contact names and phone numbers were collected in order to distinguish contacts while preventing possible identification of individuals. Actual contact names, phone numbers and contents of messages, calls etc. were neither logged nor analyzed at any time. Furthermore, the app was automatically disabled after 60 days and did not collect any further data beyond that date.

External Data In addition to data collected via self report measures and the logging application, we enriched the existing data set with additional parameters from online repositories by the use of web-based application program interfaces (APIs). The Echonest API[2] was used to collect additional metadata such as danceability, genre, speechiness, acousticness etc. about songs participants listened to. Additionally we retrieved application related information such as descriptions, ratings and number of downloads from the Google Play Store (Google, 2016a).

Procedure

Data collection was performed in two steps. In an initial lab session, participants had to read and sign a standardized written consent form, explaining details of the study including data collection, storage and processing. Consequently, they had to complete a series of psychometric and demographic tests at the psychological department at LMU (see the Apparatus/Personality section for details).

[2] http://developer.echonest.com/docs/v4/song.html

Figure 2.4: Android Logging Application

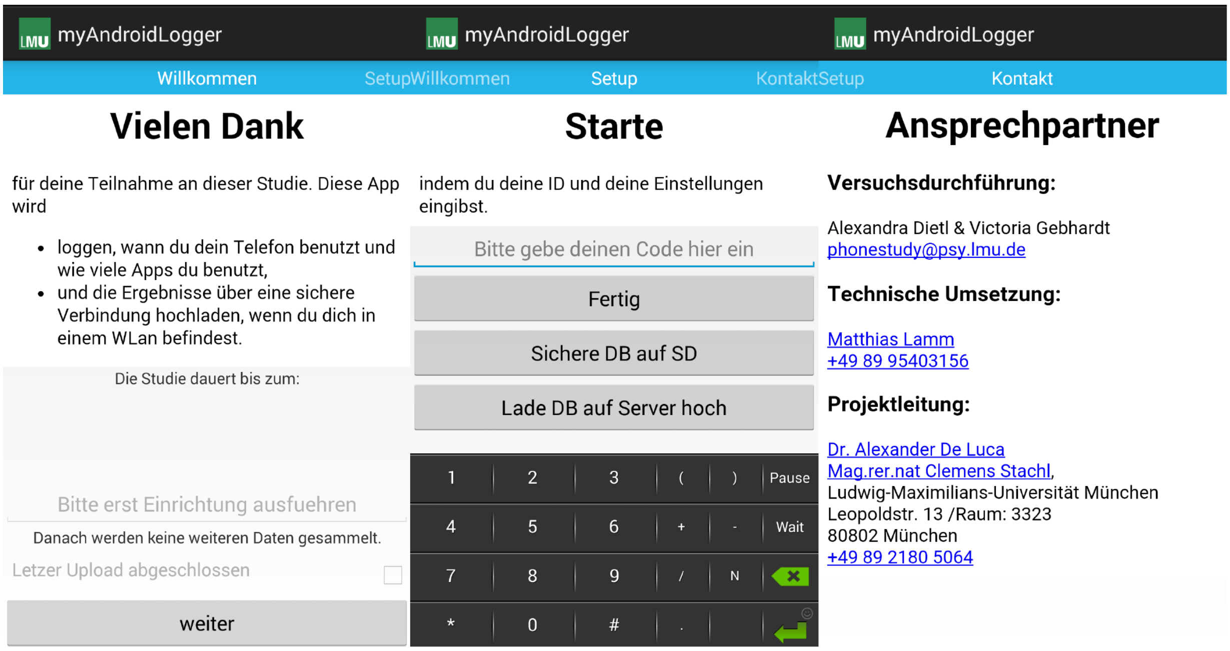

Figure 2.4: Interface of the logging application used in study 2 and 3.

After completion, a specifically developed Android app was installed on each participants private Smartphone (see the Section 2.3.3 for details) and registered to an five-digit number id. The id was only used to related results from the psychometric tests to the phone usage data - at no point it was possible to relate it to a specific name of a participant. For 60 days, this app recorded usage data in an anonymous way and transferred encrypted data to a server at LMU. During this period, no user interaction was necessary and the app did not display information of any kind. An exception was a pop-up message that reminded participants to keep location services turned on, in case this permission was deactivated.

At the end of the data collection period, participants were reminded in a pop-up message to contact the research staff in order to receive their compensation. During the compensation session, a manual backup of the usage data was secured from all devices.

Data Analysis

All data processing as well as statistical analyses in this study were performed with statistical software R 3.3.1 (R Core Team, 2016). Additionally, several external packages were used for this purpose. We used the *mlr* package for predictive modeling (Bischl et al., 2016), the *parallelMap* package for parallel computing (Bischl & Lang, 2015), the *gbm*,*glmnet* and *ranger* packages for model fitting(Friedman et al., 2010; Ridgeway, 2015; Wright & Ziegler, 2015), as well as several other packages (*caret*, *RANN*, *irace*,*kernlab*, *e1071*, *xtable*, *dplyr*, *psych*, *mailR*) for other purposes (Arya, Mount, Kemp, & Jefferis, 2015; Dahl, 2016; Karatzoglou et al., 2004; Kuhn, 2015; López-Ibáñez, Dubois-Lacoste, Stützle, & Birattari, 2011; Premraj, 2015; Revelle, 2016; Wickham & Francois, 2016). The syntax as well as the necessary data files can be found in the appendix. Please note that syntax documentation for this study does not include predictor extraction, as this step was performed by my former master student *Jiew-Quay Au*. Furthermore, binning of the criterion variables is also not documented in the syntax as the used norm values are copyright protected by the Schuhfried GmbH, Mödling, Austria.

Predictor Extraction In the beginning, the raw data consisted in the form of timestamp sorted event data. However, in order to predict personality scores from usage data, we had to extract predictor variables (features) from the raw data. This extraction step was performed under consideration of results from previous studies as well as availability of data and resulted in a total number of 679 predictor variables.

These features related to the behavioral categories of *communication*, *mobility*, *application usage*, *day and nighttime activity*, *camera usage*, *music consumption*, and *general phone usage*. These features quantified frequency, length/duration and response rate (only calls and messages), variance, regularity and entropy of events. Features in relation to communication were extracted because associations with the personality dimensions extraversion and agreeableness were reported in previous research (Chittaranjan et al., 2013; De Montjoye et al., 2013; Montag et al., 2014, 2015). Mobility measures, such as the average radius of gyration per day were expected to be predictive of the personality dimension emotional stability (Burton et al., 2013; De Montjoye et al., 2013; Saeb et al., 2015). However, as the majority of our participants had high frequencies of missing GPS values, we could not use mobility variables for prediction modeling and will therefore not elaborate on these measures. Measures in relation to the usage of applications were calculated because previous research suggested associations with personality dimensions (Chittaranjan et al., 2013; Xu, Frey, Fleisch, & Ilic, 2016), see also Section 2.2. Features in relation to day and nighttime activities, with focus on the regularity of these events were extracted based on previous research (Jackson et al., 2010) especially the body of research concerning circadian typology (Tonetti, Fabbri, &

Natale, 2009). We calculated features quantifying the camera usage due to previously reported associations with big five personality dimensions (Chittaranjan et al., 2013; Eftekhar et al., 2014; Hunt & Langstedt, 2014). Features in relation to music listening behavior we calculated based on previous reports of personality specific music preferences (Greenberg et al., 2016, 2015; Langmeyer, Guglhör-Rudan, & Tarnai, 2012). A full list of all used features (after pre-processing) with summary statistics is provided in the Appendix 4.

To overcome outlier-related problems of classical estimators (such as the mean), we generally used robust estimators for the calculation of predictor variables. For predictor variables indicating the central tendency we used the huber mean (Kafadar, 2003) estimator, for variables indicating the variation of a variable we generally used the robust location-free scale estimate (Rousseeuw & Croux, 1993). As some of the calculated variables in the present data set are rather unfamiliar in the field of psychological science, we provide brief explanations here.

In addition to the frequency of events, the regularity of events can provide information about distinct behavior. Likely, the regularity of which e.g. certain places are visited or behaviors are executed provides information with regard to individual differences and lifestyle (Williams, Whitaker, & Allen, 2012). As previously reported, conscientious people report to follow a more ordered daily routine in comparison with less conscientious individuals (Jackson et al., 2010). Similar to (De Montjoye et al., 2013) we used the method described in Williams et al. (2012) to quantify repeated routine over time. $D(\cdot) = 0$ indicates perfect regularity and higher values indicate more irregularity. For this study we chose the window size to be $\omega = 24$ (hours).

Inspired by De Montjoye et al. (2013), we additionally calculated measures of entropy for various behavioral variables. Entropy as defined in information theory describes the uncertainty in a countable random variable **X** or the retrievable information through inspection of that variable. More concrete elements in a variable **X** (e.g., unique phone numbers in all phone numbers somebody called), can be represented as a probability vector $\mathbf{p} = (p_1, ..., p_d)'$ with $p_1 + ... + p_d = 1$. Therefore, the *Shannon*- entropy is defined as

$$\hat{\mathbf{H}}(\hat{\mathbf{p}}) = -\sum_{j=1}^{d} \hat{p}_j \log \hat{p}_j \tag{2.2}$$

Likely, some contacts will be called more frequently than others, therefore the respective probability $\hat{p}_j$ can be higher or lower. The entropy $\hat{H}(\hat{p})$ describes towards which degree unique contacts are called equally often. Therefore, entropy will be minimal if all contacts are called equally often and large if some contacts are called much more frequently than others. Consequently, this measure will vary between people according to how unequal these probabilities are distributed. In particular we used the bias-corrected *Miller-Madow* estimator of the *Shannon* entropy (G. A. Miller, 1955), where $\hat{d}$ is some estimate of the number of bins with nonzero probability.

$$\hat{\mathbf{H}}_{MM} = \hat{\mathbf{H}} + \frac{\hat{d} - 1}{2n} \tag{2.3}$$

Pre-Processing Prior to modeling we had to pre-process and clean the data. We removed highly-correlated ($r > 0.75$) variables, predictors with no or near zero-variance from further analyses (cutoff 10%) and variables with more than 30% missing values. This resulted in a final dataset of 238 predictor variables and 35 criterion variables, please see Appendix 4 for details on all variables.

Some features provided information about app usage. Although in general we used the app categorization provided by the Google Play Store (Google, 2016a), a number of apps were clearly mislabeled and had to be re-labeled in order to perform meaningful data analysis. Furthermore, we had to exclude a large number of bloatware[3] and background apps that showed up as actual usage in the collected logs. All other pre-processing steps were completed directly during training of the respective algorithm in order to avoid overfitting.

In order to predict self-reported personality scores from log data we performed two steps of analysis. First, we used a random forest (Wright & Ziegler, 2015) binary classifier and predicted gender from all predictor variables. Therefore, we used a nested resampling design with 200 subsampling iteration (95% split rate) as the inner resampling loop and 10×10 repeated cross-validation as the outer resampling loop. In the inner resampling, hyperparamters and threshold tuning as well as artificial upsampling was performed. During each resampling iteration, the respective sample was centered and scaled. Furthermore, missing values were imputed with a k-nearest-neighbors algorithm, in order not to waste any samples (Troyanskaya et al., 2001). Due to class imbalance, we used the synthetic minority oversampling technique (SMOTE) (Chawla,

[3]Bloatware refers to pre-installed, mostly unwanted software that often negatively affects system performance of devices

Bowyer, Hall, & Kegelmeyer, 2002) in order to add more instances in the respective minority class. This was also performed during each inner resampling iteration. Subsequently, we added the mean prediction score of gender (between 0 and 1) to the data set as new predictor for following personality models.

Second, we predicted each personality factor and facet in a with a prediction modeling approach. Specifically, we predicted personality dimensions in binary classification tasks. We chose the classification approach as regression was to hard under consideration of the available sample size. We binned criteria in *"high"* and *"not high"* by using the 70th percentile of the respective norm sample. Hence, cases with personality scores above the 70th percentile of the norm sample were classified as *"high"*, cases below as *"not high"*.

Prediction Modeling We performed prediction modeling for all criteria with three different algorithms. Specifically, we used a random forest (Wright & Ziegler, 2015), a gradient boosting machine (Friedman, 2002) and an elastic net binomial regression (Friedman et al., 2010). All three algorithms perform automated variable selection and can handle many predictors simultaneously. Furthermore, we used nested resampling designs for all algorithms with slightly different settings between them. The random forest used 200 subsampling iterations (95% split) in the inner tuning loop and ten × 20 repeated cross-validation in the outer resampling loop. The gradient boosting models were tuned with five subsampling iterations (95% split) in the inner and ten × three repeated cross-validation in the outer resampling loop. The elastic net used ten × cross-validation in the inner and nine × eight repeated cross-validation in the outer resampling loop. Settings for imputation, threshold tuning and upsampling were equal to those in the gender prediction task before. We repeated this analysis sequence for all criteria. The predicted gender variable was not used for the gradient boosting models.

Consequently, we investigated whether aggregated prediction performance of each individual criteria was above the *no information rate* (NIR). No information rate refers to the accuracy a predictor can achieve if all cases are simply classified in the majority category (e.g., if in a sample 60% of participants are female and 40% are male, classification of gender would therefore have a NIR of 60%).

2.3.4 Results

Descriptive Statistics

The used dataset (after preprocessing) contained 273 variables. Summary statistics for all variables are presented in the Appendix 41. Correlations between the Big Five variables as well as the additionally collected demographic variables are visible in Table 2.3 in Section 2.2.4. As these correlations have been already discussed in Section 2.2.4, they will not be discussed again at this point. In summary, the observed correlations do not induce collinearity issues for the consequent analyses. In Tables 43 to 47, in the Appendix, we provide the all correlations between personality factors, facets and predictor variables.

Prediction of Gender

In a first step the gender was predicted in a binary classification task from smartphone usage variables in order to use the obtained predictions as a new predictor variable for the prediction of personality. We chose this approach as gender showed correlations with personality (see Table 2.2.4) and therefore could help predict personality. Furthermore, it was also important to not use the actual gender levels as those would not be available in an applied setting.

In Tables 2.7 a summary of the classification results is provided. The random forest classifier achieved a classification accuracy of 0.75 roughly 11% above the NIR. Based on the individual resampling results, we calculated the average predicted gender score for each case and fed it back into the data set.

Prediction of Personality

In Tables 2.8, 2.9 and 2.10, results of prediction modeling are visible. None of the used algorithms could predict any personality variables at factor level. Solely the conscientiousness facets sense of duty and caution could be predicted with accuracy above the NIR with at least one algorithm. Furthermore, it is visible that regularized linear models were not successful in the prediction of any criteria whereas non-linear models (random forest and gradient boosting) could predict at least one criterion above chance.

Table 2.7: Random Forest Performance Measures - Gender Classification

Measure	Value
Accuracy (Acc)	0.75
95% $CI_{(Acc)}$	0.72, 0.77
Balanced Accuracy	0.71
Sensitivity	0.59
Specificity	0.83
No Information Rate (NIR)	0.64
P-Value [Acc > NIR]	< 0.0001
Pos Pred Value	0.67
Neg Pred Value	0.78
Positive Class	Men

Note: Standard performance measures of the random forest classifier as evaluated on the test set.

Table 2.8: Random Forest Performance Measures - Personality

	MMCE	ACC	BAC	TPR	TNR	NIR	ACC > NIR
Emotional.Stability	0.56	0.44	0.43	0.41	0.46	0.72	-
Extraversion	0.43	0.57	0.56	0.54	0.58	0.69	-
Openness	0.52	0.48	0.45	0.38	0.53	0.69	-
Conscientiousness	0.43	0.57	0.55	0.50	0.61	0.64	-
Agreeableness	0.73	0.27	0.48	0.87	0.10	0.77	-
Carefreeness	0.62	0.38	0.51	0.75	0.26	0.76	-
Equanimity	0.56	0.44	0.46	0.52	0.41	0.72	-
Positive.mood	0.62	0.38	0.46	0.70	0.22	0.66	-
Self.consciousness	0.46	0.54	0.56	0.61	0.51	0.69	-
Self.control	0.35	0.65	0.61	0.51	0.70	0.69	-
Emotional.robustness	0.68	0.32	0.49	0.92	0.06	0.69	-
Friendliness	0.50	0.50	0.54	0.69	0.39	0.64	-
Sociableness	0.52	0.48	0.50	0.54	0.45	0.70	-
Assertiveness	0.54	0.46	0.46	0.51	0.42	0.59	-
Dynamism	0.41	0.59	0.61	0.68	0.55	0.66	-
Adventurousness.	0.47	0.53	0.56	0.68	0.44	0.61	-
Cheerfulness	0.53	0.47	0.51	0.65	0.36	0.64	-
Openness.imagination	0.48	0.52	0.50	0.41	0.58	0.64	-
Openness.aesthetics	0.37	0.63	0.51	0.32	0.70	0.82	-
Openness.feelings	0.60	0.40	0.47	0.62	0.31	0.72	-
Openness.actions	0.43	0.57	0.55	0.43	0.67	0.61	-
Openness.ideas	0.41	0.59	0.52	0.23	0.81	0.61	-
Openness.value.norms	0.45	0.55	0.54	0.54	0.55	0.68	-
Competence	0.50	0.50	0.50	0.50	0.50	0.69	-
Love.of.order	0.45	0.55	0.56	0.60	0.52	0.69	-
Sense.of.duty	0.37	0.63	0.62	0.52	0.73	0.54	TRUE
Ambition	0.50	0.50	0.49	0.42	0.56	0.64	-
Discipline	0.43	0.57	0.48	0.20	0.76	0.66	-
Caution	0.40	0.60	0.62	0.68	0.55	0.62	-
Willingness.to.trust	0.55	0.45	0.50	0.62	0.38	0.70	-
Genuineness	0.53	0.47	0.45	0.41	0.49	0.75	-
Helpfulness	0.64	0.36	0.48	0.86	0.11	0.66	-
Obligingness	0.65	0.35	0.50	0.89	0.11	0.69	-
Modesty	0.53	0.47	0.51	0.62	0.39	0.68	-
Good.naturedness	0.63	0.37	0.51	0.81	0.22	0.74	-

Note: Mean performance measures of the random forest binary classification task as evaluated on the test set; MMCE = mean misclassification error, ACC = accuracy, BAC = balanced accuracy, TPR = true positive rate, TNR = true negative rate, NIR = no information rate; successfully predicted criteria are bold.

Table 2.9: Gradient Boosting Performance Measures - Personality

	MMCE	ACC	BAC	TPR	TNR	NIR	ACC > NIR
Emotional.Stability	0.55	0.45	0.44	0.46	0.42	0.72	-
Extraversion	0.48	0.52	0.49	0.55	0.42	0.70	-
Openness	0.46	0.54	0.49	0.62	0.35	0.70	-
Conscientiousness	0.39	0.61	0.60	0.62	0.58	0.64	-
Agreeableness	0.44	0.56	0.53	0.58	0.49	0.77	-
Carefreeness	0.65	0.35	0.44	0.29	0.58	0.79	-
Equanimity	0.48	0.52	0.50	0.54	0.46	0.73	-
Positive.mood	0.53	0.47	0.46	0.50	0.42	0.66	-
Self.consciousness	0.52	0.48	0.50	0.43	0.58	0.69	-
Self.control	0.36	0.64	0.62	0.67	0.57	0.69	-
Emotional.robustness	0.62	0.38	0.45	0.29	0.62	0.71	-
Friendliness	0.55	0.45	0.45	0.45	0.45	0.71	-
Sociableness	0.49	0.51	0.49	0.52	0.47	0.73	-
Assertiveness	0.52	0.48	0.49	0.48	0.49	0.64	-
Dynamism	0.46	0.54	0.54	0.54	0.54	0.71	-
Adventurousness.	0.52	0.48	0.48	0.48	0.48	0.65	-
Cheerfulness	0.52	0.48	0.52	0.41	0.63	0.68	-
Openness.imagination	0.53	0.47	0.45	0.51	0.39	0.69	-
Openness.aesthetics	0.43	0.57	0.55	0.59	0.51	0.84	-
Openness.feelings	0.56	0.44	0.47	0.39	0.56	0.72	-
Openness.actions	0.41	0.59	0.54	0.67	0.41	0.67	-
Openness.ideas	0.50	0.50	0.49	0.50	0.48	0.65	-
Openness.value.norms	0.43	0.57	0.54	0.62	0.45	0.68	-
Competence	0.50	0.50	0.49	0.50	0.47	0.74	-
Love.of.order	0.38	0.62	0.61	0.62	0.61	0.72	-
Sense.of.duty	0.31	0.69	0.69	0.71	0.66	0.66	TRUE
Ambition	0.46	0.54	0.53	0.57	0.50	0.64	-
Discipline	0.39	0.61	0.59	0.64	0.54	0.74	-
Caution	0.36	0.64	0.64	0.65	0.64	0.62	TRUE
Willingness.to.trust	0.49	0.51	0.53	0.49	0.58	0.70	-
Genuineness	0.39	0.61	0.57	0.64	0.49	0.82	-
Helpfulness	0.44	0.56	0.55	0.58	0.52	0.70	-
Obligingness	0.52	0.48	0.48	0.48	0.47	0.77	-
Modesty	0.48	0.52	0.52	0.53	0.52	0.75	-
Good.naturedness	0.42	0.58	0.56	0.50	0.61	0.74	-

Note: Mean performance measures of the gradient boosting machine classification task as evaluated on the test set; MMCE = mean misclassification error, ACC = accuracy, BAC = balanced accuracy, TPR = true positive rate, TNR = true negative rate, NIR = no information rate; successfully predicted criteria are bold.

Table 2.10: Elastic Net Performance Measures - Personality

	MMCE	ACC	BAC	TPR	TNR	NIR	ACC > NIR
Emotional.Stability	0.55	0.45	0.47	0.50	0.43	0.72	-
Extraversion	0.48	0.52	0.52	0.51	0.52	0.69	-
Openness	0.49	0.51	0.45	0.28	0.61	0.69	-
Conscientiousness	0.48	0.52	0.50	0.43	0.57	0.64	-
Agreeableness	0.53	0.47	0.44	0.40	0.49	0.77	-
Carefreeness	0.49	0.51	0.49	0.45	0.52	0.76	-
Equanimity	0.46	0.54	0.52	0.47	0.57	0.72	-
Positive.mood	0.53	0.47	0.46	0.44	0.48	0.66	-
Self.consciousness	0.46	0.54	0.54	0.55	0.54	0.69	-
Self.control	0.42	0.58	0.58	0.59	0.58	0.69	-
Emotional.robustness	0.57	0.43	0.46	0.54	0.38	0.69	-
Friendliness	0.49	0.51	0.55	0.66	0.43	0.64	-
Sociableness	0.55	0.45	0.47	0.53	0.42	0.70	-
Assertiveness	0.55	0.45	0.47	0.56	0.38	0.59	-
Dynamism	0.51	0.49	0.51	0.55	0.46	0.66	-
Adventurousness.	0.56	0.44	0.47	0.64	0.31	0.61	-
Cheerfulness	0.58	0.42	0.47	0.62	0.32	0.64	-
Openness.imagination	0.56	0.44	0.45	0.46	0.44	0.64	-
Openness.aesthetics	0.52	0.48	0.52	0.58	0.46	0.82	-
Openness.feelings	0.53	0.47	0.46	0.44	0.48	0.72	-
Openness.actions	0.39	0.61	0.61	0.59	0.62	0.61	-
Openness.ideas	0.46	0.54	0.51	0.37	0.65	0.61	-
Openness.value.norms	0.43	0.57	0.58	0.61	0.55	0.68	-
Competence	0.45	0.55	0.53	0.50	0.57	0.69	-
Love.of.order	0.47	0.53	0.52	0.50	0.54	0.69	-
Sense.of.duty	0.49	0.51	0.51	0.46	0.55	0.54	-
Ambition	0.47	0.53	0.49	0.38	0.61	0.64	-
Discipline	0.57	0.43	0.45	0.51	0.39	0.66	-
Caution	0.43	0.57	0.55	0.48	0.62	0.62	-
Willingness.to.trust	0.53	0.47	0.46	0.45	0.48	0.70	-
Genuineness	0.45	0.55	0.49	0.39	0.60	0.75	-
Helpfulness	0.38	0.62	0.60	0.54	0.65	0.66	-
Obligingness	0.53	0.47	0.45	0.40	0.50	0.69	-
Modesty	0.46	0.54	0.51	0.42	0.60	0.68	-
Good.naturedness	0.50	0.50	0.48	0.44	0.52	0.74	-

Note: Mean performance measures of the elastic net classification task as evaluated on the test set; MMCE = mean misclassification error, ACC = accuracy, BAC = balanced accuracy, TPR = true positive rate, TNR = true negative rate, NIR = no information rate.

Variable Importance

In addition to pure prediction accuracy, it is also informative to closer investigate some of the most important variables for the respective criteria. Although, classification was performed as a binary task, we also provide correlations and plots using the continuous criterion variables in order to provide some information about possible associational directions. In Table 2.11 the ten most important predictors for the gradient boosting classification of the conscientiousness facets sense of duty and caution are visible. In general the variable importance measures here consider interactions between the predictors (up to nine dimensions in this case) and take both linear and non-linear relationships into account. For a more intuitive, yet incomplete illustration, we also provide Spearman correlations between criteria and predictors in 2.11. Some of the top variables are similar for both facets. In general this ranking suggests that variables containing information about the irregularity and stability of day and night time activity were important for the classification of both conscientiousness facets. Furthermore, the Spearman correlations suggest that most associations between both criteria and predictors are negative (e.g., sense of duty*Variance of downtime duration on weekdays $\rho = -0.31$). Hence, less variance in daily events (e.g., morning evening) and higher irregularity of daily events was associated with higher scores in the conscientiousness facet sense of duty. In the upper part of Figure 2.5 sense of duty facet values in binned and continuous form are plotted against the predictor *variance of downtime duration at weekdays*. The plot suggests that the variance of downtime durations is similar to a personality score of approximately 2.8 and then mainly decreases with even higher scores in sense of duty.

Figure 2.5 shows that people with higher scores in sense of duty in average have lower values in the variable *variance in the duration of downtime during weekdays*. However, the fitted polynomial and the data distribution indicate that the relationship between both variables is not strictly linear. Downtime was defined as the longest duration between the last and first event an user performed in the evening and the next morning of the following day. As some events still happen during nighttime (e.g., apps updating, wifi on/off etc.), we calculated the longest usage breaks from both sides of the night and ignored up to eight events during that period.

Therefore, people with high values in this variable had nightly smartphone usage breaks with different lengths (sometimes slept long, on other days slept briefly). Participants with low values had nights without smartphone usage with approximately the

Table 2.11: Variable Importance and Spearmen Correlations

Sense.of.Duty	Importance	ρ	Caution	Importance	ρ
Variance of downtime duration on weekdays	14.75	-0.31	Variance of the first event on weekdays	4.15	-0.27
Number of events during downtime	6.88	-0.31	Average number of days with less than one hour phone usage	3.94	-0.14
Variance of the first event on weekdays	6.55	-0.26	Total number of contacts with one assigned tel. number	3.50	-0.15
Irregularity of all aggregated events	4.96	-0.28	% Entertainment app usage	3.28	0.13
% Travel & Local app usage in the evening	4.58	-0.17	Variance of the last events on weekdays	3.28	-0.26
Variance of the first events on weekends	4.52	-0.26	Number of events during downtime	3.07	-0.27
Average number of Travel & Local app uses on weekdays	2.23	-0.22	Variance of the daily number of incoming calls	3.01	-0.20
Irregularity of the last events on weekends	2.12	-0.17	% of Transportation app usage on mornings	2.71	0.22
% of Productivity app usage on evenings	1.83	-0.05	Variance of downtime duration on weekdays	2.58	-0.22
Ratio of installed and used apps	1.21	0.11	Average battery charge when recharged	2.12	-0.02

Note: Relative importance of the top 10 predictors on the classification for the conscientiousness facets sense of duty and caution; Pearson correlations (ρ) between the respective criteria and the predictor variables; Spearman correlations have been calculated based on the continuous criteria, on the complete data set.

same duration. Another important predictor of sense of duty was the total number of events that were recorded during the time participants were allegedly sleeping. The lower part of Figure 2.5 shows sense of duty plotted against the total number of events (of each participant) that happened during downtime. The data suggests that people with higher scores in sense of duty had less events happening during downtime than participants with low scores.

Furthermore, application usage in the categories *Travel & Local* and *Productivity* was predictive for the facet sense of duty. Spearman correlations suggest negative associations of *Travel & Local* app-usage on weekdays ($\rho = -0.22$) and evenings ($\rho = -0.17$). Similarly, caution was predicted by the percentage of application usage in the categories *Entertainment* in general and *Transportation* in the morning. Those associations are shown as positive by the correlation coefficients ($\rho_E = 0.13$, $\rho_T = 0.22$).

The most important predictor for the facet caution was the temporal variance of the first event during weekdays. This measure indicates how much variation there is across all first events and across all weekday-mornings. Participants with high values in this variable got up at many different times during weekdays, people with low values got up at similar times. The respective Spearman correlation suggests a negative relationship between caution and the variance in first events on weekdays ($\rho = -0.27$).

Unlike for sense of duty, the average amount of days with less than one hour of total phone usage, the number of saved contacts with only one assigned number and the average battery charge of the phone when it was connected to a charger also predicted caution.

Figure 2.5: Scatterplots and box-plots of criteria and important predictor variables

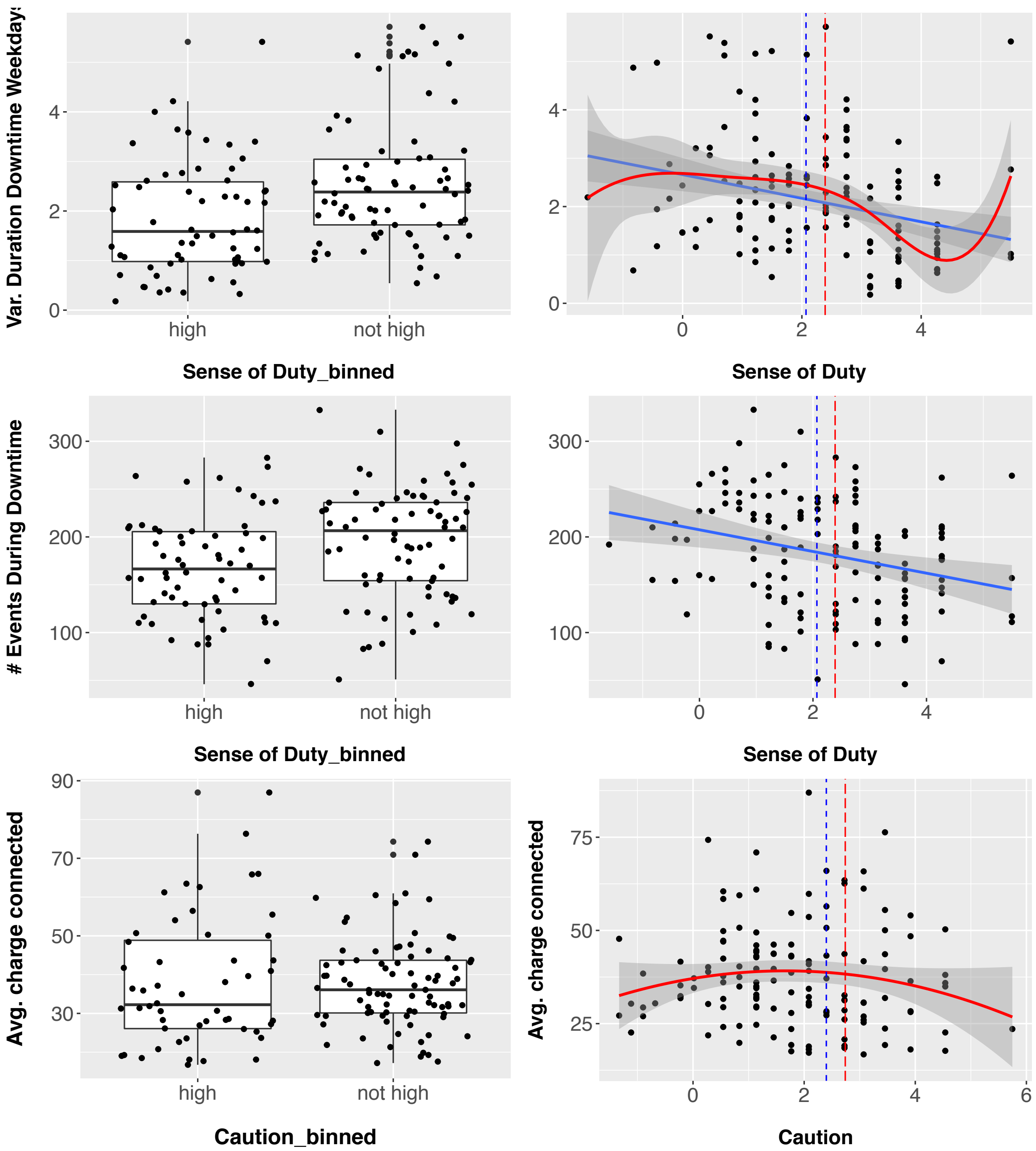

Figure 2.5: The first and second row of boxplots and scatterplots shows associations between the facet sense of duty and the variation of smartphone downtime at weekdays and the number of events during sleep; the lower graphs show the association between the facet caution and the average battery charge when the phone was connected to a charger; the blue solid lines are robust regressions, the red solid line is a 5th order spline with shades indicating 95% confidence intervals; dashed lines represent cutoffs for binning (70% percentile of the respective norm sample) for males (blue) and females (red) respectively; avg = average, # = number, var = variance.

2.3.5 Discussion

Prediction of Conscientiousness

Results of this study show that the prediction of personality from smartphone usage data was in general not successful with the available data. As an exception, it was possible to predict high personality scores in the facets sense of duty and caution of the conscientiousness factor above chance. Variable importance and Spearman correlations indicate that, especially the temporal variance as well as regularity of events was important for successful predictions. However, it is not possible to largely confirm previous reports (Chittaranjan et al., 2013; De Montjoye et al., 2013) reporting classification successes for all personality dimensions. Besides personality, it was possible to predict gender above the NIR.

Although as a whole, the reported results are sobering, the inspection of the top predictors for sense of duty and caution still provide insights for prospective research efforts. Less variance and irregularity in reoccurring events (as suggested by Spearman correlations), indicates that more order in daily activities was associated with high scores in both criteria. This suggests that the predictor variables most closely related to both facets of conscientiousness, picked up on the tendency of conscientious people to organize their lifes in an orderly fashion with regard to activities and hours of the day. This conclusion fits well with results of Jackson et al. (2010), who reported that breaking daily routines was one of the best negative predictors for conscientiousness ($r = -0.40$). Furthermore, the lower nightly usage of Travel & Local apps of people high in sense of duty could indicate that they do not travel much in the evenings and rather stay at home to be fit for the next day at work. People high in caution used apps related to public transportation more often in the morning, this matches well conceptions of conscientious people being careful not to miss appointments or their bus. This is also supported by Jackson et al. (2010) who reported that the behavior related items *"Get to appointment on time", ($r = 0.32$); "Miss appointments", ($r = -0.44$); "Leave for work at the exact time we had planned", ($r = 0.24$) ; "Miss the bus", ($r = -0.32$)* were items associated with conscientiousness. Several other variables (e.g., *Average battery charge when recharged; % of Productivity app usage on evenings*) were predictive in our models. However, as the absence of a linear univariate effect on the criteria suggests interactions and nonlinear associations (see also Figure 2.5), discussion of those would exceed the aim of this paper - we will not elaborate on them any further.

These results highlight associations between behaviors and self-reported levels of conscientiousness. Furthermore, these results also indicate that people with high scores in these dimensions can be potentially identified automatically through analysis of their daily events on their smartphone. Specifically, third party apps could easily log when a person is active on the phone and sell these data to companies interested in how conscientious a specific person is. As conscientiousness is the personality dimension most closely related to job performance and negative turnover decisions (Barrick & Mount, 1991; Zimmerman, 2008), hiring or firing decisions could for example be additionally based on data like this. Furthermore, insurance companies could use these data in order to create more personalized insurance contracts, as people with high scores in conscientiousness are less likely to show risky behaviors and unhealthy habits (Ozer & Benet-Martínez, 2006). Although, accuracy in the present study is not very high for predictable dimensions, it is likely that it will increase with additional variables and larger sample sizes.

Our results show that no other personality dispositions could be successfully predicted above the NIR. This is counterintuitive as several correlations suggest linear associations between most behavioral predictors and personality traits. Especially surprising is that extraversion could not be predicted from user data as several associations were observed with extraversion in the present data set (see Table 4 in the appendix) and previous studies (Chittaranjan et al., 2013; De Montjoye et al., 2013; Montag et al., 2014, 2015). However, the absence of more success with the prediction of personality facets from user data also shows that correlations between predictor and criterion are more easily established than the cross-validated prediction of a criterion with the simultaneous consideration of multiple predictor variables and their intercorrelations. Personality refers to relatively stable individual differences in characteristic patterns of thinking, feeling and behaving across time and situations (Specht, Egloff, & Schmukle, 2011). This study only investigated how behavioral features are predictive of overall self-reported personality scores, largely ignoring personality aspects manifested in cognitive and emotional patterns. Furthermore, this argumentation is also supported by the fact that we could show that self-reported personality scores correlate with measures of actual behavior (see Section 2.3.4) and predict categorical behaviors (see Section 2.2) in several categories. One could argue it is possible to relate personality to behaviors and to predict behaviors. However, prediction of personality purely based on behaviors seems difficult or incomplete.

The predictability of personality dimensions based on behavior must also be related to the number of relevant behaviors grasped with our method of collection. Therefore, it is possible that the behavioral measures we extracted from the raw data, were not equally related to all personality dimensions. Furthermore, the extracted features might be to rough for the complexity of personality dimensions. For example people high in openness might not be different to others in terms of how much they use communication apps, but rather through the content of their communications (Yarkoni, 2010). In the case of openness to aesthetics, the highest correlations were observed for variables in relation to photo app usage (e.g., perc_Photography*openness to aesthetics, $\rho = 0.31$; total_number_shared_photos*openness to aesthetics, $\rho = 0.27$), see Table 4 in the appendix. Although these associations provide interesting insights, the actual content of the shared photos could provide additional predictive power for personality prediction. Although we tried to account for more specific parts of personality related variance through the creation of features with regard to time of the day, one might have to even consider even more fine-grained data in order to predict personality traits.

Furthermore, some previous research states that personality traits are also differently associated with engagement or restraint from behavior (Hirsh et al., 2009), making personality manifestations in behavior easier or harder to observe. In the case of the personality dimension neuroticism (low emotional stability), location data could have provided predictive value (Hodgins & Ellenbogen, 2003; Ormel et al., 2013), unfortunately we could not use those measures due to high numbers of missing values.

Furthermore, previous research suggests that personality is strongly manifested in individual preferences in addition to behaviors (Kosinski et al., 2014, 2013). As impressively shown by Youyou et al. (2015), these measures can also be retrieved from user data and could complement current measures in predictive value.

Limitations

This study comes with several benefits as well as important limitations.

Designed as a classical logging study, we analyzed actual and naturally-occurring behavior, automatically generated through phone usage. This approach allowed us to record large quantities of behavioral proxy measures and highlight associations with personality dimensions. Furthermore, this is the first study that shows (limited for two facets of conscientiousness) that automatically generated metrics of smartphone usage can successfully predict high personality scores while taking precautions not to over-fit the data. (Chittaranjan et al., 2013; De Montjoye et al., 2013).

However, important limitations have to be noted. Our rather small sample was purely collected from the German population in Munich with age and education not being perfectly representative of the general population. However, as smartphone usage is less prevalent with older people (Kim et al., 2015), our sample includes the most representative group. For the assessment of personality traits, we only used self-reports. However, one could criticize that some research indicates that not all personality dimensions are equally rateable by the self (Vazire, 2010). Therefore, others-ratings could have altered the collected personality scores.

Furthermore, we only analyzed data about actual behaviors, variables about for example individual preferences would possibly improve personality predictions (Youyou et al., 2015).

Moreover, app usage patterns might differ when compared to, for example, samples from other cities and countries. For instance, availability and popularity of public transportation impacts the use of apps in the related category. Application usage can also be different with regard to the cultural background and country of an user. However, although some variation in app usage is to be expected, many popular apps for common tasks are globally available or have popular regional equivalents. Furthermore, associations between app usage and age were similar to previous results although smaller statistical associations were to be expected in a more homogeneous sample.

The results with regard to the predictable facets can be related to existing literature (Jackson et al., 2010; MacCann et al., 2009), associating aspects of conscientiousness with the order of life events. However one has to be careful as the provided explanations are partially drawn post-hoc and our data is missing ground truth for the observed behaviors.

2.3.6 Conclusions & Outlook

This work shows how facets of conscientiousness can be predicted from very basic usage data, mainly related to the order and regularity of daily events. Furthermore, this study is one of few that predicted single personality facets based on recordings of actual behaviors. Although we observed several correlations between personality traits and behavioral variables, we were only able to predict two personality facets from these measures. Therefore, we conclude that despite the flexible prediction algorithms we used in this study, more and different variables are needed in order to recognize personality dimensions from automatically generated user data.

This work served as an exploratory study to test this methodology in terms of data collection and the prediction of psychometric measures. However, prospective studies should collect data from larger samples and consider additional parameters, related to behavioral and cognitive preferences. Additionally, more measures of individual activity (e.g location data, physical activity tracking, audio features etc.) could be collected in future studies. Most likely, data from additional participants with more GPS data will improve the prediction of some personality aspects (e.g., emotional stability) As an additional approach, prospective studies could include experience sampling methods in order to understand reasons for behavior and in order to create better behavioral features for prediction purposes.

Despite the relatively limited success in terms of personality prediction of this study, it is likely that further efforts will improve on this aim. Therefore, we see at least two possible developments for the field of psychological science. First, continuing digitalization of our world will eventually make it possible to predict human behaviors and preferences based on previous preferences and behaviors. In this case, psychological science could drastically change due to then available exhaustive methodologies. This change would shift the current theoretical focus of the discipline to a rather data driven - computational social science (Cioffi-Revilla, 2010).

Another possibility however is the failure of the personality prediction endeavor. Further studies would continue to collect more and more behaviors and individual preferences through usage data and still fail to reliably predict personality. In this case, personality theories also will be at doubt as even large collections of individually different behaviors and preferences cannot fully predict personality scores. This would be surprising as personality is expected to be directly manifested in individual behavior. Existing theories will be questioned in relation to their relevance for practical applications and daily life.

However, at this point it remains unclear how logging data research will influence psychological science. Nevertheless, it could hold potential for the improvement of existing theories, through increased consideration of actual behavior. This again could help to better predict outcomes in daily life, and therefore increase practical relevance of the discipline.

2.3.7 Author Contributions

Besides myself, Jiew-Quay Au, Bernd Bischl, and Markus Bühner contributed to the creation of this article. Jiew-Quay Au assisted with data cleaning and predictor extraction. Bernd Bischl and Markus Bühner acted as the supervising authors of this article.

2.3.8 Acknowledgements

We want to thank our student researchers (*Alexandra Dietl*, *Victoria Gebhardt*, *Sarah Huber*, *Jakob Kaiser*) for supporting us with recruiting, psychometric testing and organization as well as *Martin Fröhlich & Matthias Lamm* for their programming efforts. Furthermore, we would like to thank the LMU data protection officer Dr. iur. Marco Wehling for his assistance in legal matters. This project was partially funded by a *Google Research Award*.

Chapter 3

General Discussion

In the present work, we investigated how data logging capabilities of cars and smartphones can be used to effectively collect behavioral data in psychological science. Furthermore, we investigated to which extend differential human factors such as big five personality dimensions and demographics are reflected in individual patterns of user data. We could show that gender is systematically reflected in automotive driving patterns as well as smartphone usage and that gender can be predicted using these data. Beyond the recognition of gender, we showed that big five personality traits predict usage frequencies of categorical smartphone applications. Furthermore, our results indicate that facets of conscientiousness can potentially be recognized from user data on smartphones.

Within the investigated context and limitations, these results support the notion that personality traits, demography, and behavior are associated with another and that personality scores and demography can predict behavior and vice versa. These results are in line with previous research, highlighting the importance of associations between big five personality traits and various life outcomes as well as self-reported behaviors (Fleeson et al., 2009; Ozer & Benet-Martínez, 2006; Roberts et al., 2007). Beyond the implications discussed individually for each single study (see Sections 2.1.5, 2.2.5 and 2.3.5), in this Section we address important aspects of this work in a broader context, including possible influences, this and similar work could have on the field of psychological science and differential psychology in particular.

3.1 Actual Behavior and Prediction

This work diverges from many other studies in two important aspects. First, we used no self-reported measures of behavior, but purely focused on actual behavior, automatically logged in the respective devices. This approach was chosen, as personality is traditionally expected to be manifested in individual behavior (Fleeson et al., 2009). While a missing focus on actual behavior in psychological science has been debated repeatedly (Baumeister et al., 2007; Fleeson et al., 2009; Furr, 2009; Lewandowski Jr & Strohmetz, 2009; Poorthuis et al., 2014; Vazire & Mehl, 2008), current technological advances in mobile sensor and networking technology could radically facilitate reach of this goal (G. Miller, 2012; Yarkoni, 2012). These possibilities are further underlined by the present work, which illustrates how a wide range of daily behaviors and activities can be collected unobtrusively, and relatively easy with the use of off-the shelf smartphones. Furthermore, we show how these data can be used for studies in psychological research.

Second, this work predominately aims at the prediction of outcome variables rather than the testing of hypotheses. On the one hand, explanatory research has traditionally dominated psychological science. On the other hand, it has been criticized that, despite psychological theories do not lack detail, they are able to tell very little about what people actually do in real life (Baumeister et al., 2007; Yarkoni & Westfall, 2016). In this sense, Yarkoni and Westfall (2016) recently suggested that psychological science might benefit from an focus on the prediction of behaviors and outcomes in addition to the explanation of those. Although, thoroughly designed confirmatory experiments were until now considered the gold standard in psychological science, additional focus on prediction could help to relate psychological research more closely to real-world problems and aid understanding of existing concepts (Mozer & Lindsey, 2016; Yarkoni & Westfall, 2016). Using data like ours, researchers could for example help to predict critical behavioral patterns and events, such as episodes of schizophrenia and depression (Ben-Zeev et al., 2014; Saeb et al., 2015).

Although we undertook several prediction efforts in relation to behavioral variables and personality traits, our analysis to predict personality traits from user data show quite limited success (see Section 2.3). This suggests that our predictor variables did not grasp enough trait relevant variance in behavior. However, it remains unclear whether we simply did not collect enough behavioral data, or if the prediction of personality traits requires more detailed information about cognitive and emotional factors. Furthermore, stable constructs, such as conscientiousness might be recognizable through rather general behavioral styles (e.g., an increased regularity and structure in life). Whereas, the prediction of rather heterogeneous personality dimensions (such as openness (DeYoung, 2015)) might require more complex aggregations of behavioral data.

However, this also leads to the problem of insufficiently large sample sizes in psychological science (Holmes, 1983; Marszalek, Barber, Kohlhart, & Holmes, 2011). Although in this work we took uttermost care not to overfit our data, prospective studies should collect larger samples (in the thousands) in order to realize more stable prediction models (Yarkoni & Westfall, 2016).

3.2 Challenges of Data Logging Studies

Data logging methodologies provide a set of advantages over conventional behavioral observation studies. However, some difficulties currently exist with regard to automated data logging on smartphones, that must be taken into consideration. First, rich logging of smartphone usage is currently only possible within the *Android* operating system, as other operating systems (e.g., Apples iOS) do not allow access to many sources of user data. Second, despite Android currently having a worldwide market share of around 82% (Gartner, 2016), researchers could induce some sampling bias as excluded iPhone users might represent a different sociodemographic population (Richter & Statista, 2014). Nevertheless, the larger market penetration of the Android system will eventually enable the recruitment of a more diverse sample from the population of smartphone users.

Third, another problem associated with the Android operating system is its fragmentation. Different versions of the operating system exist with (OpenSignal, 2015) interchanging API availability and options. This situation makes it not only hard for developers to create new applications, it can also cause problems with data logging. Effectively this means that some signals might not be available on phones with a particular version of the operating system.

Unrelated to Android fragmentation, emphasis has to be put on intensive application testing as inefficient logging procedures (e.g., too frequent GPS logs) can dramatically reduce battery life of mobile devices, impair user experience, and potentially influence user behavior (e.g., people use their phone less because the battery drains quickly). Additionally, battery drain might also result in higher drop out rates from studies. Furthermore, it has to be noted that the sensor accuracy of smartphones potentially does not live up to specially designed scientific measurement devices in all dimensions and varies between devices from various manufacturers. Especially, accuracy of logged activity data (e.g., steps) seems to vary significantly between devices and manufacturers (Stisen et al., 2015). Nevertheless, the ongoing development of better sensors will eventually eliminate that concern. Furthermore, in terms of study design considerations, some inaccuracies might be acceptable in exchange for larger samples and the recording of actual behavior.

In addition to technical challenges, the analysis of log-data demands new skills from researchers. Data has to be transformed into an usable format before it can be analyzed. As data is originally recorded in the form of timestamped events, variables have to be calculated from the raw data set. Furthermore, the drastically increasing dimensionality of collected datasets demands new statistical and computational methods for analysis. For example log-files typically contain lots of missing values and extend along millions of rows. Taking this into consideration it is not feasible to manually find and correct errors in the data. These requirements encourage cooperation with disciplines such as computer science and data science.

Furthermore, log data can mostly only be analyzed with regard to usage frequencies, regularities, variation and so on. The ground truth of an observed behavior often remains unclear. Even if sequences of behavior seem logical (e.g., the use of a communication app follows the use of a photography app - the user might have sent a picture), ultimately it is uncertain why an user performed a certain action. For the investigation of a psychological phenomenon it might be interesting to know why someone acted in a certain way in a situation (e.g., using the phone in a social situation to escape personal contact or to connect with significant others). Although, this kind of information can still be gathered by e.g., experience sampling, the automatic recognition of behavioral context remains a challenge.

3.3 Conclusion

Data logging studies and prediction modeling provide new opportunities to link psychological variables to behavior more directly. This will eventually make it possible to build some sort of personality prediction algorithm with sufficiently high accuracy.

However, there is one catch associated with the prediction of latent personality variables from user data, not addressed in previous research (Chittaranjan et al., 2013; De Montjoye et al., 2013). The most accurate prediction of personality from usage data can only serve as an intermediate goal to proof the concept and validity of the logging methodology. In a best case scenario, these efforts would lead to the prediction of the exact likert ratings of the user personality self-report inventory. However, personality traits are only useful to the point where they help to predict individual differences in behavior, preferences and life outcomes. This raises the question whether single variable representations of personality dimensions are most useful for the achievement of this goal. First, self-reports are associated with a series of problems (Paulhus & Vazire, 2007; Podsakoff et al., 2012; Vaerenbergh & Thomas, 2013) and should not necessarily be seen as the perfect numerical representation of the underlying latent traits.

Second, from a practical point of view, it is possible that combinations of high-dimensional digital records about an individual might one day serve better as predictors for future behavior in comparison with self-reported personality traits. Most theories on personality traits are based on indicators derived from responses to self-report questionnaires. The latent personality traits, extracted from these models, are used to describe systematic individual tendencies in behaviors, cognition and emotions. However, as more direct measures of individual behaviors and preferences will become traceable, existing theories of individual differences might be challenged by personality models based on indicators from log data.

For now, it remains to be seen if these models are eventually more successful in the prediction of relevant life outcomes. If so, this could trigger a new era of theory building in personality research. Ultimately, one day individual differences might be better described as complex aggregations of words spoken, places visited, things liked, music listened to, humans befriended, products bought, steps walked, movies watched and so on.

Chapter 4

Supplemental Files

All supplemental files are accessible for reviewers in an open science framework project:

- **Study 1**

  1. ***study1data.csv.***
     The csv file contains the data set used for prediction modeling.
  2. ***study1_analysis.html.***
     The file contains executable R-code in order to reproduce the reported results using the provided data set.

- **Study 2**

  1. ***study2data.csv.***
     The csv file contains the data set used for feature selection and regression modeling.
  2. ***study2_analysis.Rmd.***
     The file contains Markdown code in order to reproduce the reported results using the provided data set.
  3. ***study2appcategories.csv.***
     The file contains an exhaustive list of all logged applications, the respective package name and the labeling used for feature extraction, as well as the original labeling obtained from the Google Play Store (Google, 2016a).

4. ***study2bloatware.csv.***
   The file contains a list of all apps we considered as *Bloatware* and were therefore not included in the analysis of app usage frequencies.

- **Study 3**

  1. ***study3data.RDS.***
     The RDS file contains the data set used for prediction modeling.
  2. ***study3bdata.RDS.***
     The RDS file contains the binned criteria variables used for prediction modeling.
  3. ***study3_analysis.html.***
     The file contains executable R-code in order to reproduce the reported results using the provided data sets.

Appendix

Table 1: Summary Statistics Study 3

Variable	N	M	SD	Min	Max	Range
Gender	137	1.64	0.48	1.00	2.00	1.00
Bildungsgrad	137	4.26	0.57	2.00	5.00	3.00
Age	137	23.58	4.71	18.00	50.00	32.00
Emotional.Stability	137	-0.04	0.70	-2.00	2.52	4.52
Extraversion	137	0.03	0.74	-1.98	1.88	3.85
Openness	137	0.01	0.72	-1.84	2.12	3.96
Conscientiousness	137	0.08	0.77	-1.63	1.81	3.44
Agreeableness	137	-0.16	0.75	-2.11	1.80	3.91
Carefreeness	137	0.03	1.18	-2.58	3.24	5.82
Equanimity	137	0.48	1.03	-2.30	3.27	5.57
Positive.mood	137	0.92	1.44	-4.55	5.59	10.15
Self.consciousness	137	0.72	1.11	-2.42	3.90	6.32
Self.control	137	0.70	1.01	-2.10	3.36	5.46
Emotional.robustness	137	0.68	1.27	-1.75	5.53	7.28
Friendliness	137	1.43	1.33	-1.70	5.41	7.11
Sociableness	137	1.35	1.73	-3.41	5.64	9.05
Assertiveness	137	0.80	1.42	-2.30	5.61	7.91
Dynamism	137	1.37	1.52	-2.02	5.94	7.96
Adventurousness.	137	0.44	1.56	-3.25	5.27	8.52
Cheerfulness	137	1.82	1.66	-3.23	6.09	9.32
Openness.to.imagination	137	1.30	1.45	-2.04	5.33	7.37
Openness.to.aesthetics	137	0.34	1.21	-2.38	4.61	6.99
Openness.to.feelings	137	2.10	2.23	-5.65	6.04	11.69
Openness.to.actions	137	1.50	1.41	-2.75	5.42	8.16
Openness.to.ideas	137	1.88	1.44	-0.85	5.51	6.37
Openness.to.the.value.and.norm.system	137	0.93	1.04	-1.61	4.86	6.47
Competence	137	1.05	1.30	-1.87	4.43	6.31
Love.of.order	137	1.21	1.63	-4.34	5.67	10.01
Sense.of.duty	137	2.20	1.46	-1.59	5.50	7.10

Ambition	137	2.20	1.62	-1.40	5.86	7.25
Discipline	137	1.77	1.53	-1.13	5.75	6.88
Caution	137	1.78	1.42	-1.33	5.75	7.08
Willingness.to.trust	137	0.23	1.32	-3.09	4.21	7.30
Genuineness	137	1.00	0.91	-1.20	4.25	5.45
Helpfulness	137	1.59	1.46	-2.47	6.04	8.52
Obligingness	137	0.89	1.15	-1.86	3.71	5.57
Modesty	137	0.58	1.18	-2.68	3.91	6.59
Good.naturedness	137	1.92	1.73	-2.99	6.40	9.39
total_number_missed_calls	137	25.10	29.74	0.00	176.00	176.00
total_duration_calls	137	26780.43	39109.90	0.00	274926.00	274926.00
total_duration_incoming_calls	137	9830.32	15087.52	0.00	92611.00	92611.00
avg_duration_incoming_calls_weekend	137	192.56	264.41	0.00	1974.50	1974.50
avg_duration_outgoing_calls_weekend	137	174.17	277.44	0.00	2068.00	2068.00
avg_leng_incoming_sms	137	80.88	21.67	0.00	129.84	129.84
avg_leng_outgoing_sms	137	84.23	39.68	0.00	236.91	236.91
var_duration_calls	137	171269.23	293671.00	0.00	1717624.93	1717624.93
var_duration_incoming_calls	137	30623.04	115820.88	0.00	1048486.11	1048486.11
var_incoming_sms_leng	137	1763.31	913.73	0.00	3849.96	3849.96
var_outgoing_sms_leng	137	2798.47	3419.69	0.00	26338.19	26338.19
var_duration_calls_weekend	137	24313.31	105859.88	0.00	1013878.47	1013878.47
ratio_avg_duration_incoming_outgoing_calls	113	3.10	5.59	0.00	47.48	47.48
total_number_contacts_end	137	100.50	172.78	0.00	1537.00	1537.00
total_number_added_contacts	137	17.27	99.38	0.00	1121.00	1121.00
total_number_contacts_with_one_number	137	132.37	90.58	0.00	418.00	418.00
total_number_contacts_with_two_numbers	137	17.55	25.85	0.00	221.00	221.00
total_number_contacts_with_mail	137	66.07	168.75	0.00	1273.00	1273.00
total_number_unique_contacts_who_called	137	2.42	3.65	0.00	20.00	20.00
total_number_unique_contacts_outgoing_sms	137	2.61	3.28	0.00	22.00	22.00
avg_completeness_score_contacts	137	0.20	0.08	0.00	0.47	0.47
entropy_of_contact_missed_calls	137	1.23	0.97	0.00	3.06	3.06
entropy_of_contact_sms_weekday	137	1.33	0.70	0.00	2.94	2.94
entropy_of_contact_missed_calls_weekend	137	0.84	0.84	0.00	2.66	2.66
entropy_of_contact_outgoing_sms_weekday	137	0.98	0.71	0.00	2.49	2.49
entropy_of_contact_outgoing_sms_weekend	137	0.63	0.67	0.00	2.31	2.31
response_rate_sms	137	0.14	0.13	0.00	0.55	0.55
response_rate_missed_call_answer_with_sms	137	0.01	0.04	0.00	0.33	0.33
response_rate_calls_weekend	137	0.24	0.26	0.00	1.00	1.00
response_rate_calls_weekday	137	0.25	0.25	0.00	1.00	1.00
percent_calls_night	137	0.29	0.13	0.00	0.66	0.66
percent_sms_night	137	0.30	0.13	0.00	0.64	0.64
gps_data_available	137	0.58	0.50	0.00	1.00	1.00
avg_time_last_event_weekday	137	24.16	1.23	21.74	27.86	6.12

avg_time_first_event_sunday	136	9.35	1.12	6.17	12.42	6.25
var_first_event_weekday	137	1.67	1.54	0.00	12.12	12.12
var_last_event_weekday	137	1.50	1.26	0.04	9.09	9.05
var_first_event_weekend	137	2.31	1.55	0.00	9.32	9.32
var_last_event_weekend	137	2.56	2.11	0.04	12.56	12.52
number_nights_more_than_7_hours_downtime	137	36.37	12.10	2.00	60.00	58.00
number_nights_less_than_4_hours_downtime	137	2.16	3.52	0.00	21.00	21.00
var_duration_downtime_weekday	137	2.30	1.49	0.18	11.45	11.27
var_duration_downtime_weekend	137	3.65	2.57	0.28	13.04	12.76
regularity_last_event_all	137	0.16	0.03	0.09	0.26	0.18
regularity_first_event_weekday	137	0.10	0.04	0.03	0.25	0.22
regularity_last_event_weekday	137	0.15	0.04	0.06	0.28	0.22
regularity_last_event_weekend	137	0.15	0.04	0.00	0.25	0.25
number_events_during_sleep	137	181.07	60.31	-34.00	333.00	367.00
ratio_number_apps_inst_apps_used	137	1.75	0.92	0.31	10.11	9.80
avg_number_videos_taken_weekdays	137	0.17	0.47	0.00	4.55	4.55
avg_number_videos_taken_weekend	137	0.18	0.40	0.00	2.57	2.57
avg_inter_event_time_weekend	137	0.00	0.00	0.00	0.01	0.01
total_events_airplaine_db	137	48.88	94.10	0.00	529.00	529.00
bluetooth_used	137	0.28	0.45	0.00	1.00	1.00
total_events_boot_db	137	59.23	52.88	2.00	318.00	316.00
avg_charge_connected	137	37.67	13.65	16.75	86.97	70.22
avg_charge_disconnected	137	73.94	13.90	32.61	97.00	64.39
avg_number_charge_connected_per_day	137	2.79	3.44	0.26	30.12	29.86
number_checking_behaviour_events	137	2365.23	1643.65	181.00	8489.00	8308.00
number_songs_listened_per_day	137	6.76	12.88	0.00	73.56	73.56
percentage_of_songs_listened_between_0_6	137	0.05	0.13	0.00	1.00	1.00
percentage_of_songs_listened_between_6_12	137	0.20	0.24	0.00	1.00	1.00
percentage_of_songs_listened_between_12_18	137	0.28	0.27	0.00	1.00	1.00
percentage_of_songs_listened_between_18_24	137	0.25	0.26	0.00	1.00	1.00
entropy_music_genres_morning	137	2.12	1.71	0.00	4.55	4.55
number_music_audio_apps	137	5.59	2.61	0.00	15.00	15.00
number_business_apps	137	1.79	1.13	0.00	7.00	7.00
number_photography_apps	137	2.90	1.93	0.00	12.00	12.00
number_books_and_reference_apps	137	2.25	2.03	0.00	14.00	14.00
number_tools_apps	137	15.55	4.65	3.00	31.00	28.00
number_games_puzzle_apps	137	0.78	1.44	0.00	8.00	8.00
number_weather_apps	137	1.74	1.30	0.00	5.00	5.00
number_finance_apps	137	1.26	1.17	0.00	6.00	6.00
number_education_apps	137	0.98	1.95	0.00	19.00	19.00
number_sports_apps	137	0.47	0.99	0.00	6.00	6.00
number_games_board_apps	137	0.12	0.45	0.00	3.00	3.00
number_games_racing_apps	137	0.14	0.42	0.00	2.00	2.00

number_antivirus_and_security_apps	137	0.74	0.75	0.00	3.00	3.00
number_battery_saver_task_killer_apps	137	0.36	0.55	0.00	2.00	2.00
calendar_apps_used	137	0.43	0.50	0.00	1.00	1.00
avg_plusone_scores	137	2387644.54	1074668.89	404516.17	5544395.75	5139879.58
total_number_shared_photos	137	21.07	26.17	0.00	177.00	177.00
regularity_all_aggr_events	137	1.13	0.27	0.69	2.59	1.90
download_count..1.000.000.000...5.000.000.000.	137	6.37	2.23	2.00	12.00	10.00
download_count..10.000...50.000.	137	0.86	1.31	0.00	6.00	6.00
download_count..5.000...10.000.	137	0.20	0.43	0.00	2.00	2.00
download_count..5.000.000...10.000.000.	137	2.02	2.17	0.00	11.00	11.00
download_count..50.000...100.000.	137	0.42	0.85	0.00	4.00	4.00
number_apps_messenger_used	137	1591.46	1430.52	0.00	6853.00	6853.00
number_apps_searchengine_used	137	87.30	394.72	0.00	3420.00	3420.00
avg_usage_time_1h	131	2.01	4.37	0.03	43.44	43.40
avg_usage_time_2h	122	6.46	39.69	0.03	407.67	407.63
avg_usage_time_5h	122	6.38	28.92	0.07	214.91	214.84
avg_usage_time_6h	131	1.90	3.18	0.03	26.95	26.92
avg_usage_time_7h	136	1.74	3.41	0.28	37.15	36.87
avg_usage_time_8h	136	1.55	1.15	0.30	7.41	7.11
avg_usage_time_10h	136	1.45	0.80	0.42	4.97	4.55
avg_usage_time_19h	136	1.34	0.79	0.41	4.50	4.09
avg_usage_time_0h	134	1.69	1.71	0.19	13.19	13.00
usage_count_4h	136	9.29	16.68	0.00	151.00	151.00
usage_count_6h	136	24.93	27.88	0.00	175.00	175.00
usage_count_7h	136	57.24	38.56	3.00	200.00	197.00
usage_count_0h	136	67.62	56.73	0.00	323.00	323.00
app_usage_Tools_perc_morning	137	0.28	0.12	0.00	0.78	0.78
app_usage_Tools_perc_midday	137	0.26	0.11	0.00	0.55	0.55
app_usage_Tools_perc_evening	137	0.33	0.11	0.00	0.66	0.66
app_usage_Tools_perc_night	137	0.13	0.10	0.00	0.57	0.57
app_usage_Finance_perc_midday	137	0.09	0.21	0.00	1.00	1.00
app_usage_Games_perc_morning	137	0.12	0.14	0.00	0.58	0.58
app_usage_Games_perc_midday	137	0.17	0.18	0.00	0.57	0.57
app_usage_Games_perc_night	137	0.06	0.11	0.00	0.70	0.70
app_usage_Entertainment_perc_morning	137	0.14	0.16	0.00	1.00	1.00
app_usage_Entertainment_perc_midday	137	0.20	0.18	0.00	0.69	0.69
app_usage_Entertainment_perc_evening	137	0.34	0.26	0.00	0.89	0.89
app_usage_Entertainment_perc_night	137	0.07	0.11	0.00	0.53	0.53
app_usage_Productivity_perc_morning	137	0.24	0.10	0.02	0.60	0.58
app_usage_Productivity_perc_midday	137	0.38	0.11	0.10	0.71	0.61
app_usage_Productivity_perc_evening	137	0.32	0.12	0.00	0.68	0.68
app_usage_Productivity_perc_night	137	0.06	0.07	0.00	0.36	0.36
app_usage_Personalization_perc_morning	137	0.02	0.07	0.00	0.37	0.37

app_usage_News...Magazines_perc_morning	137	0.14	0.20	0.00	1.00	1.00
app_usage_News...Magazines_perc_midday	137	0.16	0.22	0.00	1.00	1.00
app_usage_News...Magazines_perc_evening	137	0.14	0.19	0.00	1.00	1.00
app_usage_News...Magazines_perc_night	137	0.03	0.07	0.00	0.34	0.34
app_usage_Unknown_perc_morning	137	0.05	0.12	0.00	0.67	0.67
app_usage_Unknown_perc_midday	137	0.09	0.19	0.00	1.00	1.00
app_usage_Unknown_perc_evening	137	0.12	0.23	0.00	1.00	1.00
app_usage_Unknown_perc_night	137	0.04	0.10	0.00	0.59	0.59
app_usage_Photography_perc_morning	137	0.20	0.14	0.00	0.96	0.96
app_usage_Photography_perc_midday	137	0.37	0.16	0.00	0.71	0.71
app_usage_Photography_perc_evening	137	0.31	0.15	0.00	0.65	0.65
app_usage_Photography_perc_night	137	0.05	0.07	0.00	0.33	0.33
app_usage_Shopping_perc_morning	137	0.08	0.15	0.00	0.82	0.82
app_usage_Shopping_perc_midday	137	0.12	0.21	0.00	1.00	1.00
app_usage_Shopping_perc_evening	137	0.12	0.21	0.00	0.85	0.85
app_usage_Communication_perc_morning	137	0.22	0.07	0.07	0.46	0.39
app_usage_Communication_perc_midday	137	0.38	0.05	0.23	0.55	0.31
app_usage_Communication_perc_evening	137	0.35	0.07	0.11	0.56	0.46
app_usage_Books...Reference_perc_morning	137	0.09	0.15	0.00	0.75	0.75
app_usage_Books...Reference_perc_midday	137	0.17	0.23	0.00	0.92	0.92
app_usage_Books...Reference_perc_evening	137	0.17	0.24	0.00	1.00	1.00
app_usage_Books...Reference_perc_night	137	0.03	0.08	0.00	0.50	0.50
app_usage_Travel...Local_perc_morning	137	0.21	0.15	0.00	0.83	0.83
app_usage_Travel...Local_perc_midday	137	0.35	0.20	0.00	1.00	1.00
app_usage_Travel...Local_perc_evening	137	0.30	0.19	0.00	0.81	0.81
app_usage_Travel...Local_perc_night	137	0.06	0.08	0.00	0.40	0.40
app_usage_Music...Audio_perc_morning	137	0.19	0.18	0.00	0.71	0.71
app_usage_Music...Audio_perc_midday	137	0.27	0.21	0.00	1.00	1.00
app_usage_Music...Audio_perc_evening	137	0.28	0.22	0.00	1.00	1.00
app_usage_Music...Audio_perc_night	137	0.05	0.08	0.00	0.39	0.39
app_usage_Medical_perc_midday	137	0.03	0.12	0.00	0.94	0.94
app_usage_Education_perc_morning	137	0.10	0.21	0.00	1.00	1.00
app_usage_Education_perc_midday	137	0.09	0.20	0.00	0.86	0.86
app_usage_Education_perc_evening	137	0.07	0.16	0.00	0.89	0.89
app_usage_Education_perc_night	137	0.01	0.04	0.00	0.30	0.30
app_usage_Business_perc_midday	137	0.18	0.25	0.00	1.00	1.00
app_usage_Business_perc_evening	137	0.12	0.19	0.00	1.00	1.00
app_usage_Business_perc_night	137	0.04	0.13	0.00	1.00	1.00
app_usage_Lifestyle_perc_morning	137	0.08	0.15	0.00	0.84	0.84
app_usage_Lifestyle_perc_evening	137	0.12	0.21	0.00	0.96	0.96
app_usage_Lifestyle_perc_night	137	0.05	0.12	0.00	0.78	0.78
app_usage_Transportation_perc_morning	137	0.16	0.14	0.00	0.65	0.65
app_usage_Transportation_perc_midday	137	0.26	0.20	0.00	0.71	0.71

app_usage_Transportation_perc_evening	137	0.26	0.21	0.00	0.83	0.83
app_usage_Transportation_perc_night	137	0.07	0.10	0.00	0.43	0.43
app_usage_Weather_perc_morning	137	0.14	0.21	0.00	0.83	0.83
app_usage_Weather_perc_evening	137	0.13	0.18	0.00	1.00	1.00
app_usage_Weather_perc_night	137	0.03	0.07	0.00	0.33	0.33
app_usage_Sports_perc_evening	137	0.09	0.21	0.00	0.92	0.92
app_usage_Sports_perc_night	137	0.02	0.11	0.00	1.00	1.00
app_usage_Browser_perc_morning	137	0.24	0.11	0.00	0.65	0.65
app_usage_Browser_perc_midday	137	0.33	0.13	0.00	0.67	0.67
app_usage_Browser_perc_evening	137	0.32	0.12	0.00	0.71	0.71
app_usage_Browser_perc_night	137	0.06	0.06	0.00	0.34	0.34
app_usage_Health...Fitness_perc_morning	137	0.07	0.17	0.00	0.97	0.97
app_usage_Health...Fitness_perc_midday	137	0.07	0.16	0.00	0.67	0.67
app_usage_Health...Fitness_perc_evening	137	0.09	0.18	0.00	0.75	0.75
app_usage_Health...Fitness_perc_night	137	0.02	0.07	0.00	0.50	0.50
app_usage_Media...Video_perc_morning	137	0.09	0.16	0.00	1.00	1.00
app_usage_Media...Video_perc_midday	137	0.16	0.21	0.00	1.00	1.00
app_usage_Media...Video_perc_evening	137	0.20	0.28	0.00	1.00	1.00
app_usage_Media...Video_perc_night	137	0.06	0.17	0.00	1.00	1.00
app_usage_Social_perc_morning	137	0.17	0.13	0.00	0.65	0.65
app_usage_Social_perc_midday	137	0.25	0.18	0.00	0.79	0.79
app_usage_Social_perc_evening	137	0.25	0.17	0.00	0.61	0.61
app_usage_Social_perc_night	137	0.05	0.07	0.00	0.35	0.35
variance_number_incoming_calls_perday	137	49.00	141.25	0.00	1360.29	1360.29
ratio_avg_number_calls_weekday_weekend	136	1.42	1.59	0.34	12.79	12.45
ratio_avg_number_in_calls_weekday_weekend	105	1.32	1.33	0.08	11.84	11.76
ratio_incoming_outgoing_calls_weekday	127	0.35	0.35	0.00	2.00	2.00
ratio_incoming_outgoing_calls_weekend	124	0.38	0.38	0.00	2.00	2.00
ratio_incoming_outgoing_sms	136	4.13	4.29	0.00	32.00	32.00
usage_News...Magazines_apps	137	332.65	1216.90	0.00	11172.00	11172.00
usage_Weather_apps	137	16.02	29.30	0.00	197.00	197.00
usage_Health...Fitness_apps	137	25.93	75.42	0.00	508.00	508.00
number_radio_usage	137	6.84	34.54	0.00	371.00	371.00

Note: Summary statistics of all predictor and criteria variables used in 2.3.

Table 2: Pairwise Spearman Correlations Between Demographics and Predictors

	Geschlecht	Bildungsgrad	Age
total_number_missed_calls	-0.09	-0.11	0.05
total_duration_calls	-0.09	0.06	0.20
total_duration_incoming_calls	-0.08	0.13	0.10
avg_duration_incoming_calls_weekend	-0.05	0.10	0.00
avg_duration_outgoing_calls_weekend	-0.09	0.00	0.14
avg_leng_incoming_sms	0.08	-0.07	-0.07
avg_leng_outgoing_sms	0.32	-0.01	0.09
var_duration_calls	-0.17	-0.01	0.08
var_duration_incoming_calls	-0.07	0.10	0.03
var_incoming_sms_leng	0.11	0.14	0.01
var_outgoing_sms_leng	0.15	0.01	0.05
var_duration_calls_weekend	-0.13	0.05	0.05
ratio_avg_duration_incoming_outgoing_calls	0.10	0.07	-0.17
total_number_contacts_end	0.01	0.08	0.04
total_number_added_contacts	0.04	0.03	0.01
total_number_contacts_with_one_number	0.17	0.11	0.15
total_number_contacts_with_two_numbers	-0.16	0.18	0.07
total_number_contacts_with_mail	-0.18	0.06	0.13
total_number_unique_contacts_who_called	-0.09	0.09	0.08
total_number_unique_contacts_outgoing_sms	0.02	0.09	0.22
avg_completeness_score_contacts	-0.18	0.15	0.04
entropy_of_contact_missed_calls	-0.14	0.03	0.03
entropy_of_contact_sms_weekday	0.13	0.07	0.13
entropy_of_contact_missed_calls_weekend	-0.07	0.01	0.03
entropy_of_contact_outgoing_sms_weekday	0.21	0.07	0.15
entropy_of_contact_outgoing_sms_weekend	0.07	0.16	0.18
response_rate_sms	0.10	0.10	0.13
response_rate_missed_call_answer_with_sms	-0.09	-0.10	0.04
response_rate_calls_weekend	-0.16	-0.04	0.03
response_rate_calls_weekday	-0.25	-0.02	0.05
percent_calls_night	-0.15	0.09	-0.06
percent_sms_night	-0.07	-0.11	-0.11
gps_data_available	-0.13	-0.03	-0.11
avg_time_last_event_weekday	-0.16	-0.08	-0.11
avg_time_first_event_sunday	-0.09	0.16	0.04
var_first_event_weekday	-0.08	0.01	-0.04
var_last_event_weekday	-0.10	0.09	0.16
var_first_event_weekend	0.03	0.04	0.21
var_last_event_weekend	0.09	-0.01	-0.02
number_nights_more_than_7_hours_downtime	0.03	0.05	-0.01

number_nights_less_than_4_hours_downtime	0.03	-0.16	-0.15
var_duration_downtime_weekday	-0.03	0.10	0.17
var_duration_downtime_weekend	0.10	0.01	0.03
regularity_last_event_all	0.12	0.00	0.01
regularity_first_event_weekday	0.12	0.08	0.11
regularity_last_event_weekday	0.01	-0.05	0.05
regularity_last_event_weekend	0.17	-0.01	0.11
number_events_during_sleep	0.20	0.14	0.09
ratio_number_apps_inst_apps_used	-0.05	0.02	0.05
avg_number_videos_taken_weekdays	-0.16	-0.06	-0.05
avg_number_videos_taken_weekend	-0.12	0.06	0.02
avg_inter_event_time_weekend	0.11	0.03	0.10
total_events_airplaine_db	-0.08	-0.07	0.01
bluetooth_used	-0.17	0.04	-0.08
total_events_boot_db	0.01	-0.13	0.06
avg_charge_connected	-0.20	0.05	0.01
avg_charge_disconnected	-0.05	0.03	0.08
avg_number_charge_connected_per_day	-0.11	-0.07	-0.15
number_checking_behaviour_events	-0.00	-0.04	-0.20
number_songs_listened_per_day	-0.19	-0.13	-0.13
percentage_of_songs_listened_between_0_6	-0.24	-0.02	-0.05
percentage_of_songs_listened_between_6_12	-0.05	-0.07	-0.06
percentage_of_songs_listened_between_12_18	-0.21	0.02	0.05
percentage_of_songs_listened_between_18_24	-0.00	-0.16	-0.20
entropy_music_genres_morning	-0.08	-0.04	0.04
number_music_audio_apps	-0.18	-0.13	-0.07
number_business_apps	-0.05	0.03	0.05
number_photography_apps	-0.14	-0.05	-0.01
number_books_and_reference_apps	-0.17	-0.03	-0.05
number_tools_apps	-0.08	0.10	0.11
number_games_puzzle_apps	0.01	-0.14	-0.22
number_weather_apps	0.18	0.18	0.00
number_finance_apps	-0.06	0.06	-0.07
number_education_apps	-0.25	0.11	-0.10
number_sports_apps	-0.16	-0.07	-0.07
number_games_board_apps	-0.04	0.03	-0.01
number_games_racing_apps	-0.13	-0.14	-0.23
number_antivirus_and_security_apps	-0.11	-0.02	0.10
number_battery_saver_task_killer_apps	-0.07	0.02	0.04
calendar_apps_used	-0.35	-0.13	0.02
avg_plusone_scores	0.39	0.06	-0.04
total_number_shared_photos	0.21	-0.01	0.01
regularity_all_aggr_events	0.06	-0.15	-0.03

download_count..1.000.000.000...5.000.000.000.	-0.12	-0.14	-0.18
download_count..10.000...50.000.	-0.40	0.03	0.07
download_count..5.000...10.000.	-0.16	-0.04	0.05
download_count..5.000.000...10.000.000.	-0.25	-0.08	-0.06
download_count..50.000...100.000.	-0.32	-0.05	0.03
number_apps_messenger_used	0.04	-0.18	-0.30
number_apps_searchengine_used	0.03	-0.20	-0.29
avg_usage_time_1h	0.00	-0.08	0.07
avg_usage_time_2h	-0.09	0.16	0.30
avg_usage_time_5h	-0.18	-0.14	0.10
avg_usage_time_6h	0.02	-0.10	0.16
avg_usage_time_7h	-0.06	-0.13	0.13
avg_usage_time_8h	0.00	-0.25	0.03
avg_usage_time_10h	-0.02	-0.24	-0.07
avg_usage_time_19h	0.01	-0.20	-0.09
avg_usage_time_0h	-0.04	-0.18	-0.08
usage_count_4h	0.08	-0.09	-0.13
usage_count_6h	0.08	-0.07	-0.06
usage_count_7h	0.03	-0.06	-0.05
usage_count_0h	-0.16	-0.04	-0.19
app_usage_Tools_perc_morning	0.03	-0.03	0.09
app_usage_Tools_perc_midday	-0.05	-0.09	-0.09
app_usage_Tools_perc_evening	0.12	0.06	-0.05
app_usage_Tools_perc_night	-0.11	0.00	0.09
app_usage_Finance_perc_midday	-0.25	0.02	-0.09
app_usage_Games_perc_morning	-0.05	-0.10	-0.20
app_usage_Games_perc_midday	-0.06	-0.15	-0.25
app_usage_Games_perc_night	-0.10	-0.17	-0.27
app_usage_Entertainment_perc_morning	-0.04	-0.14	-0.12
app_usage_Entertainment_perc_midday	-0.12	-0.13	-0.22
app_usage_Entertainment_perc_evening	-0.03	-0.10	-0.13
app_usage_Entertainment_perc_night	-0.23	-0.14	-0.19
app_usage_Productivity_perc_morning	0.02	-0.13	-0.03
app_usage_Productivity_perc_midday	-0.03	0.01	-0.06
app_usage_Productivity_perc_evening	0.03	0.05	0.12
app_usage_Productivity_perc_night	-0.17	-0.02	-0.05
app_usage_Personalization_perc_morning	-0.21	-0.05	-0.07
app_usage_News...Magazines_perc_morning	-0.14	0.02	0.01
app_usage_News...Magazines_perc_midday	-0.08	-0.01	-0.01
app_usage_News...Magazines_perc_evening	-0.15	-0.01	0.01
app_usage_News...Magazines_perc_night	-0.09	0.05	0.07
app_usage_Unknown_perc_morning	-0.19	-0.01	0.01
app_usage_Unknown_perc_midday	-0.15	-0.06	0.01

app_usage_Unknown_perc_evening	-0.26	-0.05	0.08
app_usage_Unknown_perc_night	-0.24	-0.03	0.01
app_usage_Photography_perc_morning	0.01	-0.12	-0.03
app_usage_Photography_perc_midday	0.11	0.02	0.11
app_usage_Photography_perc_evening	-0.04	0.14	0.09
app_usage_Photography_perc_night	-0.01	-0.07	-0.10
app_usage_Shopping_perc_morning	-0.07	-0.10	-0.09
app_usage_Shopping_perc_midday	-0.18	-0.03	-0.05
app_usage_Shopping_perc_evening	-0.12	-0.06	-0.10
app_usage_Communication_perc_morning	0.13	0.03	0.05
app_usage_Communication_perc_midday	0.02	-0.04	-0.08
app_usage_Communication_perc_evening	-0.05	0.06	0.05
app_usage_Books...Reference_perc_morning	0.02	-0.05	-0.16
app_usage_Books...Reference_perc_midday	-0.01	-0.08	-0.20
app_usage_Books...Reference_perc_evening	-0.11	-0.09	-0.22
app_usage_Books...Reference_perc_night	-0.15	0.04	-0.13
app_usage_Travel...Local_perc_morning	0.17	-0.17	-0.09
app_usage_Travel...Local_perc_midday	-0.01	0.12	-0.05
app_usage_Travel...Local_perc_evening	-0.15	0.13	0.15
app_usage_Travel...Local_perc_night	-0.03	0.03	-0.02
app_usage_Music...Audio_perc_morning	-0.12	0.03	-0.06
app_usage_Music...Audio_perc_midday	-0.21	-0.05	0.02
app_usage_Music...Audio_perc_evening	-0.14	-0.18	-0.21
app_usage_Music...Audio_perc_night	-0.17	-0.01	-0.00
app_usage_Medical_perc_midday	0.21	-0.15	-0.16
app_usage_Education_perc_morning	-0.26	0.13	-0.04
app_usage_Education_perc_midday	-0.27	0.10	-0.00
app_usage_Education_perc_evening	-0.21	0.17	-0.03
app_usage_Education_perc_night	-0.15	0.05	-0.04
app_usage_Business_perc_midday	0.01	-0.19	-0.17
app_usage_Business_perc_evening	-0.12	-0.11	-0.13
app_usage_Business_perc_night	0.15	0.04	-0.15
app_usage_Lifestyle_perc_morning	-0.07	-0.03	0.01
app_usage_Lifestyle_perc_evening	-0.07	-0.06	-0.11
app_usage_Lifestyle_perc_night	-0.11	0.04	0.01
app_usage_Transportation_perc_morning	0.05	0.11	-0.13
app_usage_Transportation_perc_midday	0.02	0.02	-0.15
app_usage_Transportation_perc_evening	0.08	0.09	-0.14
app_usage_Transportation_perc_night	-0.06	-0.02	-0.12
app_usage_Weather_perc_morning	-0.09	0.11	0.09
app_usage_Weather_perc_evening	-0.03	0.12	0.04
app_usage_Weather_perc_night	-0.04	0.10	-0.02
app_usage_Sports_perc_evening	-0.13	-0.03	-0.12

app_usage_Sports_perc_night	-0.19	-0.11	-0.08
app_usage_Browser_perc_morning	-0.00	-0.13	0.02
app_usage_Browser_perc_midday	0.13	0.04	-0.05
app_usage_Browser_perc_evening	-0.07	-0.01	0.08
app_usage_Browser_perc_night	-0.20	-0.11	-0.01
app_usage_Health...Fitness_perc_morning	-0.13	0.11	0.06
app_usage_Health...Fitness_perc_midday	-0.19	0.12	0.08
app_usage_Health...Fitness_perc_evening	-0.11	0.14	0.08
app_usage_Health...Fitness_perc_night	-0.14	0.16	0.05
app_usage_Media...Video_perc_morning	0.04	-0.12	-0.12
app_usage_Media...Video_perc_midday	-0.03	-0.07	-0.02
app_usage_Media...Video_perc_evening	0.15	-0.12	0.03
app_usage_Media...Video_perc_night	-0.09	-0.05	-0.04
app_usage_Social_perc_morning	0.05	-0.15	-0.14
app_usage_Social_perc_midday	-0.09	-0.19	-0.29
app_usage_Social_perc_evening	0.12	-0.14	-0.13
app_usage_Social_perc_night	-0.08	-0.07	-0.16
variance_number_incoming_calls_perday	-0.14	0.11	0.09
ratio_between_avg_number_calls_perweekday_perweekend	-0.04	-0.01	0.11
ratio_between_avg_number_incoming_calls_perweekday_perweekend	-0.17	-0.05	0.11
ratio_incoming_outgoing_calls_weekday	-0.25	0.07	-0.10
ratio_incoming_outgoing_calls_weekend	-0.14	0.11	-0.12
ratio_incoming_outgoing_sms	-0.04	-0.13	-0.18
usage_News...Magazines_apps	-0.16	-0.01	0.12
usage_Weather_apps	0.02	0.14	0.07
usage_Health...Fitness_apps	-0.19	0.14	0.08
number_radio_usage	0.00	0.07	-0.03
number_shazam_apps_used	-0.18	-0.17	-0.04
avg_uses_perday_week_Tools	-0.23	-0.07	-0.12
avg_uses_perday_week_Games	-0.09	-0.19	-0.28
avg_uses_perday_week_Travel...Local	-0.03	0.07	0.02
avg_uses_perday_week_Education	-0.21	0.19	0.01
avg_uses_perday_week_Business	-0.05	-0.22	-0.21
avg_uses_perday_week_Transportation	0.05	0.08	-0.15
avg_uses_perday_week_Weather	0.00	0.13	0.06
avg_uses_perday_Puzzle	0.04	-0.17	-0.19
avg_uses_perday_Trivia	0.03	-0.08	-0.22
avg_uses_perday_Arcade	-0.17	-0.14	-0.19
avg_uses_perday_Casual	0.21	-0.19	-0.21
avg_uses_perday_Lifestyle	-0.14	-0.08	-0.01
avg_uses_perday_end_Photography	-0.10	-0.10	0.05
avg_uses_perday_end_Education	-0.21	0.10	-0.02
avg_uses_perday_end_Business	-0.01	-0.06	-0.15

avg_uses_perday_end_Transportation	0.03	0.08	-0.15
avg_usage_time_day_Tools	-0.16	-0.07	-0.14
avg_usage_time_day_Entertainment	-0.04	-0.18	-0.11
avg_usage_time_day_Communication	0.10	-0.24	-0.18
avg_usage_time_day_Books...Reference	-0.07	0.02	0.11
avg_usage_time_day_Travel...Local	0.14	0.10	0.04
avg_usage_time_day_Music...Audio	-0.05	-0.03	-0.05
perc_Books...Reference	-0.04	0.02	-0.10
perc_Business	-0.04	-0.18	-0.17
perc_Communication	0.23	0.11	0.18
perc_Entertainment	-0.09	-0.11	-0.10
perc_Lifestyle	-0.08	-0.07	-0.02
perc_Media...Video	0.13	-0.10	0.06
perc_Medical	0.20	-0.17	-0.21
perc_Music...Audio	-0.31	-0.08	-0.11
perc_News...Magazines	-0.04	0.14	0.19
perc_Photography	0.21	-0.03	0.05
perc_Productivity	-0.16	0.21	0.21
perc_Shopping	-0.15	-0.11	-0.10
perc_Social	0.12	-0.18	-0.26
perc_Sports	-0.16	-0.06	-0.09
perc_Tools	-0.07	0.15	0.13
perc_Transportation	0.09	-0.07	-0.20
perc_Browser	0.04	0.01	-0.05
perc_Unknown	-0.14	-0.02	-0.03

Note: Pairwise Spearman correlations between demographic and predictor variables from Section 2.3.

Table 3: Pairwise Spearman Correlations Between Openness and Predictors Study 3

Predictors	Openness	O-I	O-A	O-F	O-A	O-ID	O-VN
app_usage_Entertainment_perc_evening	-0.23	-0.14	-0.07	-0.03	-0.34	-0.17	-0.19
entropy_of_contact_outgoing_sms_weekend	0.21	0.17	0.18	0.10	0.25	0.13	0.06
number_battery_saver_task_killer_apps	-0.20	-0.10	-0.16	-0.11	-0.09	-0.21	-0.17
regularity_last_event_all	-0.19	-0.12	-0.05	-0.04	-0.12	-0.25	-0.16
total_number_unique_contacts_who_called	0.18	0.15	0.03	-0.06	0.24	0.20	0.10
app_usage_Health...Fitness_perc_night	-0.18	-0.08	-0.13	0.02	-0.12	-0.20	-0.09
regularity_last_event_weekday	-0.17	-0.08	-0.03	-0.09	-0.11	-0.23	-0.12
app_usage_Communication_perc_midday	0.17	0.12	0.18	-0.06	0.18	0.09	0.17
ratio_incoming_outgoing_sms	-0.17	-0.20	-0.14	-0.11	-0.17	-0.09	-0.05
number_sports_apps	-0.16	-0.05	-0.20	-0.04	-0.14	-0.13	-0.25
app_usage_Tools_perc_evening	-0.16	-0.20	-0.05	-0.11	-0.16	-0.05	-0.12
app_usage_Productivity_perc_midday	0.16	0.07	0.07	0.05	0.26	0.10	0.13
app_usage_Productivity_perc_evening	-0.16	-0.06	-0.08	-0.06	-0.22	-0.12	-0.05
app_usage_Transportation_perc_midday	0.16	0.04	0.09	-0.07	0.05	0.24	0.26
avg_uses_perday_Trivia	-0.16	-0.14	0.00	-0.19	-0.24	-0.05	-0.15
perc_Medical	-0.16	-0.21	-0.06	-0.03	-0.14	-0.13	-0.05
perc_Sports	-0.16	-0.05	-0.21	-0.04	-0.16	-0.12	-0.21
number_books_and_reference_apps	0.15	0.18	0.08	-0.06	0.09	0.15	0.16
app_usage_Sports_perc_evening	-0.15	-0.03	-0.25	-0.03	-0.11	-0.12	-0.26
app_usage_Browser_perc_evening	-0.15	-0.06	-0.12	-0.05	-0.19	-0.14	-0.15
avg_uses_perday_week_Games	-0.15	-0.03	-0.08	-0.18	-0.17	-0.09	-0.15
percent_sms_night	-0.14	0.00	0.01	-0.07	-0.23	-0.08	-0.11
avg_charge_disconnected	-0.14	-0.19	-0.04	-0.21	-0.06	-0.10	-0.07
avg_usage_time_day_Books...Reference	-0.14	-0.08	-0.04	-0.11	-0.12	-0.02	-0.08
number_photography_apps	-0.13	-0.10	-0.10	-0.04	-0.14	-0.03	-0.11
avg_usage_time_5h	0.13	0.13	0.14	0.01	0.13	0.03	0.09
app_usage_News...Magazines_perc_night	-0.13	-0.18	-0.14	-0.12	-0.10	-0.04	-0.00
var_last_event_weekend	0.12	0.06	0.06	-0.04	0.18	0.09	0.15
percentage_of_songs_listened_between_12_18	0.12	0.11	-0.02	0.03	0.08	0.05	0.10
number_games_racing_apps	-0.12	-0.13	-0.11	-0.04	-0.06	-0.09	-0.05
app_usage_Personalization_perc_morning	-0.12	-0.05	-0.22	0.01	-0.07	-0.16	0.02
app_usage_Photography_perc_midday	0.12	-0.03	0.05	0.15	0.10	0.11	0.12
app_usage_Transportation_perc_morning	0.12	0.08	0.09	0.04	0.04	0.11	0.13
app_usage_Sports_perc_night	-0.12	0.00	-0.13	-0.06	-0.11	-0.12	-0.22
perc_Transportation	0.12	0.04	0.13	-0.04	0.04	0.13	0.13
avg_leng_outgoing_sms	0.11	0.12	0.26	0.17	-0.00	0.03	0.01
total_number_contacts_with_two_numbers	0.11	0.06	-0.11	-0.05	0.23	0.11	-0.00
response_rate_missed_call_answer_with_sms	-0.11	-0.04	-0.08	-0.03	-0.01	-0.17	-0.07
usage_count_6h	-0.11	-0.12	-0.09	0.07	-0.09	-0.12	-0.09
usage_News...Magazines_apps	-0.11	-0.15	-0.11	-0.13	-0.08	-0.09	-0.00

perc_Photography	0.11	0.10	0.31	0.22	0.02	0.05	-0.01
total_number_contacts_with_one_number	0.10	-0.05	0.02	0.11	0.24	0.05	0.06
total_events_airplaine_db	0.10	0.05	0.01	0.06	0.09	0.06	0.14
bluetooth_used	0.10	0.13	0.01	0.13	0.11	0.04	-0.03
number_business_apps	-0.10	-0.12	-0.04	0.04	-0.13	-0.06	-0.11
app_usage_Finance_perc_midday	-0.10	-0.14	-0.14	-0.12	-0.04	-0.02	-0.05
app_usage_Communication_perc_morning	-0.10	-0.14	-0.04	0.03	-0.07	-0.09	0.04
app_usage_Music...Audio_perc_midday	0.10	0.12	-0.01	-0.05	0.08	0.06	0.04
app_usage_Music...Audio_perc_night	0.10	0.17	0.01	0.09	0.09	0.03	-0.03
app_usage_Weather_perc_morning	-0.10	-0.10	0.12	-0.05	-0.13	-0.08	-0.09
app_usage_Social_perc_midday	-0.10	-0.08	-0.11	-0.06	-0.10	-0.03	-0.17
avg_usage_time_day_Entertainment	0.10	0.19	0.14	0.03	0.01	0.14	-0.02
perc_Social	-0.10	-0.09	-0.05	-0.05	-0.09	-0.04	-0.10
avg_leng_incoming_sms	-0.09	-0.13	-0.17	0.07	-0.11	-0.00	-0.02
entropy_of_contact_outgoing_sms_weekday	0.09	0.06	0.15	0.03	0.09	0.12	-0.01
number_nights_less_than_4_hours_downtime	-0.09	0.05	-0.03	-0.01	-0.12	-0.14	-0.14
number_events_during_sleep	-0.09	-0.15	0.02	-0.03	0.01	-0.09	-0.10
avg_charge_connected	-0.09	-0.01	-0.13	-0.18	-0.06	-0.08	-0.03
avg_plusone_scores	0.09	-0.02	0.18	0.09	0.03	0.05	0.01
download_count..5.000...10.000.	-0.09	0.03	0.01	-0.05	-0.21	-0.05	-0.06
usage_count_4h	0.09	0.12	0.07	0.02	0.03	0.04	0.13
app_usage_Shopping_perc_midday	-0.09	-0.05	-0.18	-0.09	-0.07	-0.03	-0.12
app_usage_Books...Reference_perc_evening	0.09	0.19	0.06	-0.01	0.03	0.07	0.04
app_usage_Travel...Local_perc_midday	0.09	0.02	0.03	-0.04	0.02	0.16	0.12
app_usage_Music...Audio_perc_evening	-0.09	0.05	-0.06	-0.03	-0.13	-0.10	-0.12
app_usage_Weather_perc_evening	-0.09	-0.10	0.07	-0.04	-0.09	-0.04	-0.17
app_usage_Media...Video_perc_evening	-0.09	-0.07	0.07	0.14	-0.12	-0.20	-0.03
app_usage_Social_perc_evening	-0.09	-0.07	-0.12	-0.04	-0.14	-0.01	-0.09
avg_uses_perday_Arcade	-0.09	-0.05	-0.13	0.09	-0.03	-0.18	-0.04
avg_usage_time_day_Music...Audio	-0.09	-0.09	-0.03	-0.03	0.00	-0.21	-0.07
total_number_unique_contacts_outgoing_sms	-0.08	-0.03	0.03	-0.07	-0.07	-0.03	-0.12
response_rate_calls_weekend	0.08	0.07	-0.04	-0.08	0.12	0.10	0.02
gps_data_available	-0.08	-0.00	-0.07	-0.12	-0.11	-0.08	0.01
number_education_apps	-0.08	-0.01	-0.12	-0.11	-0.03	-0.01	-0.02
total_number_shared_photos	0.08	0.11	0.27	0.31	-0.04	-0.02	-0.03
app_usage_Social_perc_morning	-0.08	0.01	-0.02	-0.02	-0.13	-0.06	-0.14
avg_uses_perday_end_Education	-0.08	-0.05	-0.07	-0.09	-0.07	0.00	-0.09
perc_Communication	0.08	-0.05	0.07	0.13	0.05	-0.01	0.11
perc_Unknown	-0.08	-0.00	-0.01	-0.07	-0.12	-0.05	-0.04
avg_duration_outgoing_calls_weekend	0.07	0.08	-0.04	-0.01	0.05	-0.01	-0.02
percentage_of_songs_listened_between_18_24	0.07	0.12	0.11	0.14	-0.05	0.05	0.01
download_count..1.000.000.000...5.000.000.000.	-0.07	-0.02	-0.06	-0.04	-0.12	-0.05	-0.08
app_usage_Tools_perc_night	0.07	0.17	0.00	0.07	0.03	0.01	0.04

app_usage_News...Magazines_perc_evening	-0.07	-0.14	-0.08	-0.18	-0.04	0.02	0.09
app_usage_Photography_perc_night	0.07	0.13	0.12	0.09	0.03	-0.00	0.00
app_usage_Travel...Local_perc_morning	-0.07	-0.05	0.11	-0.01	-0.10	-0.09	-0.08
app_usage_Weather_perc_night	0.07	0.02	0.13	0.10	0.06	0.06	0.01
app_usage_Browser_perc_midday	0.07	-0.04	0.11	0.10	0.09	0.01	0.05
app_usage_Media...Video_perc_morning	-0.07	-0.08	0.02	0.08	-0.07	-0.14	0.00
perc_Media...Video	-0.07	-0.03	0.07	0.09	-0.10	-0.13	-0.08
perc_Tools	0.07	0.12	0.02	0.04	0.03	0.13	-0.05
total_number_contacts_with_mail	-0.06	-0.00	-0.08	-0.03	-0.02	-0.06	-0.03
avg_time_first_event_sunday	-0.06	-0.09	-0.01	-0.10	0.04	-0.06	0.03
avg_number_videos_taken_weekdays	0.06	0.09	0.12	0.12	0.03	-0.02	-0.02
total_events_boot_db	0.06	0.15	-0.00	0.13	-0.03	-0.04	0.06
percentage_of_songs_listened_between_6_12	0.06	0.07	0.09	0.02	0.05	0.06	0.06
entropy_music_genres_morning	0.06	0.09	-0.01	0.03	0.00	0.03	0.04
number_games_board_apps	0.06	0.16	0.09	0.15	-0.03	-0.03	-0.05
download_count..10.000...50.000.	0.06	0.17	-0.06	-0.10	-0.01	0.09	0.11
avg_usage_time_6h	0.06	0.12	0.07	0.08	0.05	-0.04	-0.01
avg_usage_time_10h	0.06	0.02	0.07	-0.14	0.11	0.09	0.02
avg_usage_time_19h	0.06	0.12	0.19	-0.01	-0.02	0.02	0.06
app_usage_Unknown_perc_evening	0.06	0.16	-0.06	-0.04	0.00	0.13	-0.01
app_usage_Travel...Local_perc_night	0.06	0.03	0.07	0.05	0.10	0.07	0.01
avg_uses_perday_week_Business	0.06	0.09	0.10	-0.03	-0.01	0.05	0.03
avg_uses_perday_week_Transportation	0.06	0.04	0.07	0.01	-0.01	0.05	0.09
avg_uses_perday_end_Transportation	0.06	0.03	-0.00	-0.04	-0.04	0.11	0.08
var_duration_calls	-0.05	-0.01	-0.09	-0.11	-0.08	0.00	-0.04
entropy_of_contact_missed_calls_weekend	0.05	0.03	-0.11	-0.04	0.09	0.06	-0.01
var_first_event_weekday	-0.05	0.02	-0.00	-0.03	-0.07	-0.05	-0.07
var_duration_downtime_weekday	-0.05	-0.00	-0.07	-0.09	0.01	-0.05	-0.04
number_songs_listened_per_day	0.05	0.11	-0.01	0.03	0.01	0.01	0.03
number_weather_apps	0.05	-0.16	0.03	0.05	0.04	0.10	0.04
number_finance_apps	-0.05	-0.11	-0.05	-0.07	-0.02	-0.00	0.04
calendar_apps_used	-0.05	0.11	-0.05	-0.12	-0.09	-0.00	-0.08
regularity_all_aggr_events	-0.05	0.03	0.02	0.26	-0.03	-0.25	-0.07
avg_usage_time_1h	0.05	0.06	0.10	0.04	-0.00	-0.02	0.04
usage_count_7h	-0.05	-0.03	-0.08	0.15	-0.08	-0.08	-0.06
app_usage_Games_perc_midday	-0.05	0.03	-0.05	-0.05	-0.12	-0.01	-0.12
app_usage_News...Magazines_perc_midday	0.05	-0.06	-0.04	-0.10	0.08	0.07	0.11
app_usage_Unknown_perc_night	0.05	0.15	0.00	-0.10	0.03	0.10	0.02
app_usage_Books...Reference_perc_night	0.05	0.14	0.05	-0.04	0.01	0.05	0.06
app_usage_Medical_perc_midday	-0.05	-0.13	0.00	0.02	-0.03	-0.08	0.03
app_usage_Education_perc_morning	-0.05	-0.07	-0.09	-0.02	-0.00	0.02	-0.05
app_usage_Business_perc_evening	0.05	0.09	0.02	0.01	-0.02	0.05	0.03
app_usage_Health...Fitness_perc_midday	-0.05	0.04	-0.04	0.13	-0.03	-0.12	-0.07

ratio_betw._avg_num_calls_perweek_d_e	0.05	0.00	0.06	0.07	0.12	-0.03	0.06
usage_Weather_apps	-0.05	-0.11	0.10	-0.02	-0.05	-0.05	-0.03
number_radio_usage	0.05	-0.07	-0.03	-0.03	0.09	0.06	-0.00
avg_uses_perday_week_Travel...Local	0.05	0.13	0.03	0.04	0.01	-0.04	0.03
avg_uses_perday_Lifestyle	-0.05	0.03	-0.10	0.00	-0.06	-0.07	-0.04
perc_Lifestyle	-0.05	0.03	-0.09	0.01	-0.10	-0.08	-0.05
perc_News...Magazines	-0.05	-0.14	-0.05	-0.06	-0.05	-0.00	0.05
perc_Shopping	-0.05	-0.06	-0.16	-0.09	-0.05	-0.00	-0.07
entropy_of_contact_missed_calls	0.04	-0.01	-0.13	-0.09	0.08	0.10	0.02
response_rate_sms	0.04	0.05	-0.06	0.01	0.06	0.06	-0.06
percent_calls_night	-0.04	-0.02	-0.10	-0.01	-0.05	-0.01	-0.05
avg_number_videos_taken_weekend	-0.04	0.01	-0.04	0.12	-0.06	-0.07	-0.11
download_count..5.000.000...10.000.000.	-0.04	0.06	-0.09	-0.05	-0.10	-0.03	-0.06
number_apps_searchengine_used	0.04	0.01	-0.05	-0.07	0.02	0.08	0.04
avg_usage_time_7h	-0.04	-0.04	0.05	-0.10	-0.06	-0.02	0.03
app_usage_Tools_perc_midday	-0.04	-0.05	-0.04	-0.01	0.05	-0.09	-0.06
app_usage_Shopping_perc_morning	-0.04	0.00	-0.05	-0.04	-0.06	-0.05	-0.13
app_usage_Music...Audio_perc_morning	-0.04	0.03	-0.09	-0.10	-0.03	0.02	0.05
app_usage_Education_perc_evening	-0.04	-0.03	-0.04	-0.03	-0.03	-0.00	-0.05
app_usage_Lifestyle_perc_morning	-0.04	-0.02	-0.16	-0.03	-0.03	-0.04	-0.01
app_usage_Browser_perc_morning	-0.04	-0.07	0.10	0.04	-0.05	-0.05	-0.03
app_usage_Health...Fitness_perc_morning	-0.04	-0.02	-0.07	0.10	0.07	-0.08	-0.03
app_usage_Media...Video_perc_midday	-0.04	-0.02	-0.03	0.14	-0.02	-0.11	-0.07
variance_number_incoming_calls_perday	0.04	0.01	-0.06	-0.05	0.11	0.03	0.02
ratio_betw._avg_number_in_calls_perweek_d_e	0.04	0.14	0.09	0.00	0.09	-0.07	-0.08
avg_uses_perday_week_Education	-0.04	-0.00	-0.10	-0.02	-0.01	-0.01	-0.07
perc_Books...Reference	0.04	0.07	0.15	-0.04	-0.01	0.04	0.06
var_incoming_sms_leng	-0.03	-0.06	0.08	-0.12	-0.03	0.09	0.02
avg_completeness_score_contacts	-0.03	0.02	-0.13	-0.14	0.04	0.04	-0.14
response_rate_calls_weekday	0.03	0.06	-0.13	-0.20	0.12	0.04	0.08
var_last_event_weekday	-0.03	-0.03	-0.03	-0.04	0.01	-0.08	0.04
ratio_number_apps_inst_apps_used	0.03	-0.04	-0.00	-0.01	0.03	0.07	-0.02
avg_number_charge_connected_per_day	0.03	0.14	-0.09	0.06	-0.04	-0.00	-0.01
number_checking_behaviour_events	0.03	0.10	-0.03	0.15	-0.05	-0.03	-0.08
number_tools_apps	-0.03	-0.08	0.02	0.14	-0.08	-0.05	-0.10
avg_usage_time_8h	0.03	0.01	0.07	-0.03	-0.01	0.01	0.09
app_usage_Games_perc_night	-0.03	0.10	-0.06	-0.05	-0.10	0.00	-0.10
app_usage_News...Magazines_perc_morning	0.03	-0.02	0.04	-0.09	0.01	0.08	0.06
app_usage_Lifestyle_perc_night	-0.03	-0.01	-0.12	0.07	0.03	-0.08	0.00
app_usage_Transportation_perc_night	0.03	-0.03	0.04	-0.14	0.02	0.06	0.10
ratio_incoming_outgoing_calls_weekday	0.03	-0.01	-0.08	-0.16	0.08	0.15	0.06
avg_uses_perday_Puzzle	-0.03	0.06	-0.12	-0.01	-0.12	-0.00	-0.05
avg_usage_time_day_Communication	0.03	0.05	0.04	0.18	0.05	-0.12	-0.01

perc_Business	-0.03	0.02	0.05	-0.02	-0.07	-0.04	-0.06
perc_Productivity	0.03	0.02	0.05	-0.11	0.01	0.09	0.07
total_number_missed_calls	0.02	-0.00	-0.10	0.01	0.09	-0.03	-0.07
total_duration_calls	-0.02	-0.01	-0.07	-0.00	0.01	-0.07	-0.03
var_duration_incoming_calls	0.02	-0.01	-0.07	-0.07	-0.01	0.08	-0.01
var_outgoing_sms_leng	0.02	0.08	0.11	0.06	-0.05	0.01	-0.09
total_number_added_contacts	-0.02	0.04	0.00	0.05	0.03	-0.06	-0.19
entropy_of_contact_sms_weekday	0.02	-0.01	0.06	-0.02	0.08	0.04	-0.03
var_first_event_weekend	0.02	0.06	0.08	0.08	0.07	-0.09	-0.02
var_duration_downtime_weekend	-0.02	0.01	0.04	-0.04	0.05	-0.09	-0.14
number_antivirus_and_security_apps	-0.02	-0.01	0.01	-0.12	-0.05	0.04	-0.02
app_usage_Unknown_perc_morning	-0.02	0.10	-0.07	-0.07	-0.06	0.06	-0.06
app_usage_Unknown_perc_midday	0.02	0.11	-0.03	-0.05	-0.03	0.08	0.04
app_usage_Photography_perc_morning	-0.02	-0.06	0.01	0.11	0.02	-0.09	-0.00
app_usage_Photography_perc_evening	-0.02	-0.01	-0.10	-0.08	0.03	0.05	-0.06
app_usage_Shopping_perc_evening	-0.02	-0.00	-0.08	-0.06	-0.02	-0.03	-0.05
app_usage_Education_perc_midday	-0.02	0.01	-0.05	-0.05	0.01	0.02	-0.04
app_usage_Education_perc_night	0.02	-0.01	-0.04	0.01	0.03	0.06	0.06
app_usage_Business_perc_night	0.02	0.01	0.02	0.19	0.04	-0.02	-0.02
app_usage_Transportation_perc_evening	-0.02	-0.04	-0.08	-0.03	-0.13	0.07	-0.03
app_usage_Social_perc_night	-0.02	0.03	-0.03	-0.02	-0.04	-0.04	-0.04
ratio_incoming_outgoing_calls_weekend	0.02	-0.04	-0.05	-0.18	0.05	0.14	0.13
usage_Health...Fitness_apps	-0.02	0.04	-0.02	0.11	0.03	-0.06	-0.06
avg_uses_perday_week_Weather	-0.02	-0.09	0.10	-0.05	-0.01	0.00	-0.02
avg_uses_perday_Casual	-0.02	-0.02	0.06	0.10	-0.05	-0.11	-0.03
avg_uses_perday_end_Business	-0.02	0.08	0.02	0.01	-0.08	-0.07	0.00
avg_usage_time_day_Tools	0.02	0.06	-0.03	-0.05	0.04	0.03	0.00
perc_Browser	0.02	0.05	0.12	0.06	-0.08	0.04	0.10
ratio_avg_duration_incoming_outgoing_calls	-0.01	-0.10	-0.01	-0.01	-0.09	0.11	0.01
total_number_contacts_end	-0.01	0.03	-0.03	0.03	0.10	-0.05	-0.19
number_nights_more_than_7_hours_downtime	-0.01	-0.12	0.00	-0.07	-0.01	0.05	0.07
regularity_first_event_weekday	-0.01	-0.11	0.02	-0.09	0.05	-0.01	0.05
regularity_last_event_weekend	0.01	0.06	0.09	0.13	0.06	-0.12	-0.09
avg_inter_event_time_weekend	0.01	0.01	0.12	-0.03	-0.01	-0.04	0.01
percentage_of_songs_listened_between_0_6	0.01	0.11	-0.06	-0.01	-0.01	0.02	-0.05
number_games_puzzle_apps	0.01	0.05	-0.08	0.01	-0.03	0.04	-0.05
app_usage_Games_perc_morning	-0.01	0.04	-0.02	-0.05	-0.05	0.07	-0.09
app_usage_Entertainment_perc_morning	0.01	0.11	0.10	0.04	-0.12	-0.01	0.00
app_usage_Entertainment_perc_midday	0.01	0.08	0.03	-0.05	-0.02	-0.01	0.00
app_usage_Entertainment_perc_night	0.01	0.13	-0.08	-0.06	-0.02	0.05	-0.02
app_usage_Productivity_perc_morning	-0.01	-0.07	0.01	-0.03	-0.00	0.01	-0.08
app_usage_Productivity_perc_night	-0.01	0.14	-0.11	-0.03	-0.02	-0.05	0.02
app_usage_Communication_perc_evening	-0.01	0.05	-0.13	-0.02	-0.05	0.07	-0.14

app_usage_Books...Reference_perc_morning	0.01	0.07	0.11	-0.01	-0.04	0.00	0.06
app_usage_Books...Reference_perc_midday	0.01	0.02	0.05	-0.04	0.00	0.02	0.11
app_usage_Business_perc_midday	0.01	-0.04	0.07	-0.02	0.02	-0.00	0.04
app_usage_Lifestyle_perc_evening	-0.01	0.08	-0.09	0.06	-0.04	-0.05	-0.04
app_usage_Browser_perc_night	-0.01	0.12	-0.01	0.01	-0.03	-0.03	-0.04
app_usage_Media...Video_perc_night	0.01	0.03	-0.02	0.06	0.05	-0.03	0.03
avg_uses_perday_week_Tools	-0.01	0.08	-0.10	0.04	0.00	-0.00	-0.05
avg_uses_perday_end_Photography	0.01	0.01	0.06	0.10	-0.02	-0.02	-0.02
perc_Entertainment	-0.01	0.13	0.08	-0.05	-0.17	0.03	-0.06
perc_Music...Audio	0.01	0.06	-0.11	-0.03	0.05	-0.00	-0.08
total_duration_incoming_calls	0.00	-0.01	-0.03	-0.03	-0.00	0.02	0.00
avg_duration_incoming_calls_weekend	0.00	-0.05	-0.02	-0.08	-0.04	0.06	0.03
var_duration_calls_weekend	-0.00	-0.00	-0.15	-0.12	0.02	0.07	-0.07
avg_time_last_event_weekday	-0.00	0.14	-0.00	0.05	-0.05	-0.04	-0.06
number_music_audio_apps	-0.00	-0.08	0.09	-0.02	-0.00	-0.04	0.00
download_count..50.000...100.000.	-0.00	0.04	-0.19	-0.06	0.04	0.03	0.06
number_apps_messenger_used	0.00	0.04	-0.02	0.11	-0.02	-0.06	-0.05
avg_usage_time_2h	0.00	0.10	0.03	-0.11	-0.01	0.00	-0.04
avg_usage_time_0h	-0.00	0.09	0.06	0.03	-0.04	-0.06	-0.13
usage_count_0h	0.00	0.12	-0.02	0.09	-0.05	-0.05	-0.03
app_usage_Tools_perc_morning	-0.00	0.06	0.01	0.13	-0.06	-0.00	-0.02
app_usage_Travel...Local_perc_evening	-0.00	0.02	-0.06	-0.02	0.05	0.01	-0.08
app_usage_Health...Fitness_perc_evening	0.00	0.03	-0.04	0.15	0.04	-0.07	0.02
number_shazam_apps_used	0.00	0.03	-0.09	0.06	0.10	-0.08	-0.06
avg_usage_time_day_Travel...Local	-0.00	0.07	0.04	0.07	-0.01	-0.10	-0.02

Note: Pairwise Spearman correlations between Openness (factor, facets) and predictor variables from Section 2.3; table is sorted by absolute ρ values of Openness, in decreasing order. Abbreviations: O-I = Openness to Imagination, O-A = Openness to Aesthetics, O-F = Openness to Feelings, O-A = Openness to Actions, O-ID = Openness to Ideas, O-VN = Openness to the Value and Norm System.

Table 4: Pairwise Spearman Correlations Between Conscientiousness and Predictors Study 3

Predictors	Conscientiousness	C1	C2	C3	C4	C5	C6
regularity_last_event_weekday	-0.31	-0.22	-0.25	-0.25	-0.31	-0.24	-0.25
avg_time_last_event_weekday	-0.28	-0.20	-0.25	-0.27	-0.23	-0.23	-0.21
var_last_event_weekday	-0.28	-0.17	-0.25	-0.25	-0.30	-0.20	-0.25
var_duration_downtime_weekday	-0.26	-0.14	-0.22	-0.31	-0.26	-0.20	-0.21
var_first_event_weekday	-0.24	-0.16	-0.17	-0.26	-0.18	-0.18	-0.26
avg_usage_time_5h	-0.23	-0.12	-0.24	-0.22	-0.21	-0.25	-0.22
app_usage_Transportation_perc_morning	0.23	0.12	0.10	0.23	0.19	0.19	0.24
percentage_of_songs_listened_between_0_6	-0.22	-0.12	-0.15	-0.17	-0.18	-0.23	-0.16
usage_count_0h	-0.22	-0.13	-0.19	-0.21	-0.18	-0.18	-0.14
app_usage_Productivity_perc_morning	0.22	0.17	0.24	0.19	0.11	0.20	0.20
app_usage_Productivity_perc_night	-0.22	-0.21	-0.23	-0.15	-0.16	-0.16	-0.19
app_usage_Communication_perc_morning	0.22	0.18	0.17	0.24	0.15	0.17	0.23
avg_uses_perday_week_Travel...Local	-0.22	-0.07	-0.19	-0.22	-0.14	-0.19	-0.20
regularity_all_aggr_events	-0.21	-0.12	-0.25	-0.28	-0.12	-0.17	-0.15
number_events_during_sleep	-0.20	-0.09	-0.21	-0.31	-0.15	-0.04	-0.24
avg_usage_time_0h	-0.19	-0.16	-0.23	-0.13	-0.10	-0.17	-0.18
avg_uses_perday_Arcade	-0.19	-0.15	-0.15	-0.11	-0.17	-0.22	-0.14
avg_time_first_event_sunday	-0.18	-0.22	-0.16	-0.17	-0.20	-0.09	-0.10
regularity_last_event_all	-0.18	-0.16	-0.18	-0.15	-0.17	-0.10	-0.13
app_usage_Books...Reference_perc_night	-0.18	-0.12	-0.16	-0.12	-0.15	-0.22	-0.09
avg_usage_time_1h	-0.17	-0.21	-0.18	-0.14	-0.10	-0.15	-0.14
avg_usage_time_2h	-0.17	-0.29	-0.12	-0.11	-0.24	-0.13	-0.07
percent_sms_night	-0.16	-0.15	-0.14	-0.18	-0.08	-0.08	-0.20
app_usage_Tools_perc_night	-0.16	-0.14	-0.20	-0.10	-0.13	-0.10	-0.08
app_usage_Photography_perc_night	-0.16	-0.16	-0.13	-0.11	-0.13	-0.08	-0.13
avg_usage_time_day_Music...Audio	-0.16	-0.09	-0.19	-0.11	-0.18	-0.12	-0.11
percent_calls_night	-0.15	-0.12	-0.17	-0.20	-0.04	-0.07	-0.18
var_first_event_weekend	-0.15	-0.02	-0.15	-0.26	-0.14	-0.02	-0.11
app_usage_Tools_perc_morning	0.15	0.19	0.05	0.10	0.11	0.18	0.14
app_usage_Tools_perc_midday	-0.15	-0.09	-0.04	-0.20	-0.14	-0.15	-0.16
app_usage_Transportation_perc_evening	0.15	0.02	0.09	0.21	0.16	0.16	0.05
app_usage_Health...Fitness_perc_midday	-0.15	-0.08	-0.13	-0.16	-0.18	-0.21	-0.08
avg_usage_time_day_Tools	-0.15	-0.08	-0.09	-0.19	-0.11	-0.13	-0.17
gps_data_available	-0.14	-0.18	-0.12	-0.05	-0.13	-0.16	-0.16
number_nights_more_than_7_hours_downtime	0.14	0.09	0.20	0.10	0.12	0.10	0.05
number_nights_less_than_4_hours_downtime	-0.14	-0.07	-0.19	-0.15	-0.11	-0.13	-0.06
number_weather_apps	0.14	0.10	0.14	0.09	0.17	0.10	0.04
avg_plusone_scores	0.14	0.03	0.07	0.15	0.17	0.15	0.05
app_usage_Games_perc_night	-0.14	-0.11	-0.13	-0.17	-0.07	-0.08	-0.16

app_usage_Photography_perc_midday	0.14	0.14	0.08	0.13	0.12	0.12	0.10
app_usage_Books...Reference_perc_evening	-0.14	-0.05	-0.08	-0.17	-0.10	-0.16	-0.11
app_usage_Browser_perc_night	-0.14	-0.15	-0.11	-0.07	-0.13	-0.16	-0.08
app_usage_Health...Fitness_perc_morning	-0.14	-0.02	-0.14	-0.14	-0.14	-0.20	-0.09
app_usage_Health...Fitness_perc_night	-0.14	-0.09	-0.14	-0.19	-0.14	-0.14	-0.09
avg_uses_perday_week_Games	-0.14	-0.15	-0.06	-0.11	-0.11	-0.10	-0.16
usage_count_4h	-0.13	-0.06	-0.19	-0.08	-0.07	-0.09	-0.14
app_usage_Personalization_perc_morning	-0.13	-0.16	-0.17	-0.06	-0.09	-0.11	-0.07
perc_Transportation	0.13	0.10	0.05	0.13	0.13	0.13	0.05
avg_duration_outgoing_calls_weekend	0.12	0.07	0.16	0.08	0.13	0.04	0.06
var_last_event_weekend	-0.12	-0.02	-0.13	-0.17	-0.02	-0.05	-0.17
number_games_board_apps	-0.12	-0.13	-0.16	-0.15	-0.08	-0.10	-0.05
app_usage_Entertainment_perc_night	-0.12	-0.04	-0.11	-0.04	-0.10	-0.19	-0.05
app_usage_Productivity_perc_evening	-0.12	-0.15	-0.06	-0.05	-0.06	-0.11	-0.14
app_usage_Sports_perc_night	-0.12	-0.07	-0.01	-0.12	-0.19	-0.17	-0.06
app_usage_Social_perc_evening	0.12	0.05	0.07	0.15	0.14	0.10	0.05
avg_uses_perday_week_Tools	-0.12	-0.02	-0.06	-0.13	-0.11	-0.16	-0.08
avg_uses_perday_end_Transportation	0.12	0.04	0.05	0.13	0.13	0.14	0.04
avg_usage_time_day_Travel...Local	-0.12	-0.01	-0.09	-0.11	-0.10	-0.13	-0.07
regularity_first_event_weekday	0.11	0.06	0.13	0.11	0.04	0.10	0.08
regularity_last_event_weekend	-0.11	-0.02	-0.15	-0.17	-0.05	0.01	-0.10
ratio_number_apps_inst_apps_used	0.11	0.05	0.13	0.11	0.16	0.06	0.06
number_battery_saver_task_killer_apps	-0.11	-0.11	-0.11	-0.10	-0.06	-0.10	-0.07
number_apps_searchengine_used	0.11	0.09	0.13	0.03	0.12	0.10	-0.00
avg_usage_time_10h	-0.11	-0.01	-0.14	-0.00	-0.08	-0.07	-0.18
app_usage_Tools_perc_evening	0.11	-0.00	0.13	0.13	0.14	0.05	0.03
app_usage_Books...Reference_perc_midday	-0.11	-0.03	-0.06	-0.08	-0.07	-0.15	-0.07
app_usage_Music...Audio_perc_night	-0.11	-0.05	-0.10	-0.11	-0.08	-0.08	-0.09
app_usage_Media...Video_perc_midday	-0.11	-0.06	-0.14	-0.09	-0.11	-0.08	-0.05
app_usage_Media...Video_perc_night	-0.11	-0.03	-0.10	-0.13	-0.12	-0.05	-0.05
number_radio_usage	0.11	0.20	0.08	0.11	0.10	0.09	-0.04
perc_Business	-0.11	-0.08	-0.08	-0.14	-0.12	-0.03	-0.14
var_duration_calls_weekend	0.10	0.06	0.14	0.11	0.17	0.05	-0.03
app_usage_Communication_perc_midday	-0.10	0.03	-0.08	-0.13	-0.11	-0.06	-0.15
app_usage_Travel...Local_perc_evening	-0.10	-0.03	-0.11	-0.17	-0.04	-0.06	-0.09
app_usage_Travel...Local_perc_night	-0.10	0.02	-0.01	-0.14	-0.03	-0.14	-0.13
app_usage_Transportation_perc_midday	0.10	0.08	0.04	0.12	0.08	0.08	0.03
app_usage_Browser_perc_morning	0.10	0.10	0.15	0.09	0.05	0.09	0.09
usage_Health...Fitness_apps	-0.10	0.00	-0.09	-0.18	-0.16	-0.16	-0.04
perc_Shopping	0.10	0.13	0.04	0.05	0.05	0.02	0.05
entropy_of_contact_outgoing_sms_weekday	0.09	0.23	-0.04	0.03	0.10	0.15	0.03
avg_inter_event_time_weekend	-0.09	-0.16	-0.08	-0.03	-0.11	-0.05	-0.05
avg_charge_connected	-0.09	-0.07	-0.13	-0.03	-0.06	-0.03	-0.08

number_antivirus_and_security_apps	0.09	0.09	0.07	0.14	0.04	0.02	0.06
app_usage_Shopping_perc_morning	0.09	0.11	0.08	0.04	0.07	0.01	0.03
app_usage_Music...Audio_perc_midday	-0.09	0.08	-0.02	-0.17	-0.13	-0.13	-0.07
app_usage_Health...Fitness_perc_evening	-0.09	-0.03	-0.08	-0.12	-0.14	-0.16	-0.05
usage_News...Magazines_apps	-0.09	-0.04	-0.04	-0.11	-0.09	-0.10	-0.04
number_shazam_apps_used	-0.09	0.04	0.01	-0.16	-0.09	-0.11	-0.13
avg_uses_perday_end_Business	-0.09	-0.08	-0.05	-0.07	-0.16	-0.05	-0.04
perc_Books...Reference	-0.09	-0.06	-0.07	-0.06	-0.11	-0.11	0.01
perc_News...Magazines	0.09	0.02	0.12	0.03	0.08	0.05	0.12
var_duration_incoming_calls	0.08	0.07	0.05	0.10	0.16	0.03	-0.01
response_rate_missed_call_answer_with_sms	-0.08	-0.02	-0.06	-0.05	-0.09	-0.05	-0.16
var_duration_downtime_weekend	-0.08	-0.01	-0.09	-0.14	-0.04	-0.02	-0.08
number_checking_behaviour_events	-0.08	0.06	-0.12	-0.21	0.04	-0.08	-0.07
percentage_of_songs_listened_between_12_18	0.08	0.17	0.15	0.03	0.04	-0.02	0.10
avg_usage_time_19h	-0.08	-0.11	-0.04	0.02	-0.11	-0.10	-0.05
app_usage_Entertainment_perc_midday	-0.08	0.04	-0.13	-0.10	-0.04	-0.13	-0.08
app_usage_Entertainment_perc_evening	0.08	0.03	0.14	0.14	0.07	-0.05	0.09
app_usage_Business_perc_evening	-0.08	0.01	-0.10	-0.08	-0.11	-0.05	-0.12
app_usage_Media...Video_perc_evening	-0.08	-0.09	-0.15	-0.10	-0.07	-0.01	-0.01
app_usage_Social_perc_night	-0.08	-0.10	-0.14	-0.01	-0.05	-0.06	-0.09
avg_uses_perday_week_Weather	0.08	0.10	0.07	-0.00	0.05	0.16	0.04
avg_uses_perday_Trivia	0.08	-0.05	0.08	0.07	0.15	0.16	-0.05
avg_usage_time_day_Communication	-0.08	0.03	-0.12	-0.13	0.08	-0.10	-0.15
total_number_contacts_end	-0.07	0.04	-0.07	-0.09	-0.07	-0.03	-0.07
total_number_contacts_with_one_number	-0.07	0.08	-0.12	-0.08	0.03	-0.00	-0.18
avg_completeness_score_contacts	0.07	0.08	0.13	0.06	0.06	0.01	0.01
entropy_of_contact_missed_calls	0.07	0.14	0.00	-0.02	0.19	0.09	-0.10
avg_number_charge_connected_per_day	-0.07	0.02	-0.11	-0.12	0.02	-0.05	-0.09
percentage_of_songs_listened_between_6_12	0.07	0.13	0.13	0.12	-0.04	0.01	0.07
entropy_music_genres_morning	0.07	0.08	0.13	0.06	0.10	0.01	0.05
number_books_and_reference_apps	-0.07	0.06	-0.04	-0.08	-0.10	-0.13	-0.00
number_sports_apps	-0.07	-0.07	0.07	-0.06	-0.09	-0.17	-0.03
download_count..5.000.000...10.000.000.	-0.07	-0.04	-0.01	-0.06	-0.05	-0.11	-0.08
avg_usage_time_6h	-0.07	-0.08	-0.08	-0.03	-0.03	-0.15	-0.01
usage_count_7h	0.07	0.12	0.05	0.08	0.08	0.05	0.07
app_usage_News...Magazines_perc_morning	0.07	0.10	0.15	-0.06	0.06	0.08	0.05
app_usage_Shopping_perc_midday	0.07	0.09	0.04	0.05	0.02	-0.00	0.04
usage_Weather_apps	0.07	0.08	0.04	-0.00	0.06	0.13	0.05
avg_uses_perday_end_Photography	0.07	0.13	0.02	0.05	0.04	0.07	0.07
avg_leng_incoming_sms	0.06	-0.04	0.03	0.06	0.09	0.03	0.04
avg_leng_outgoing_sms	0.06	-0.07	0.03	0.11	0.05	0.08	0.11
response_rate_calls_weekday	-0.06	0.07	-0.04	-0.07	0.00	-0.06	-0.17
total_events_boot_db	-0.06	-0.10	-0.04	-0.11	-0.06	-0.01	0.03

number_songs_listened_per_day	-0.06	0.04	0.03	-0.04	-0.06	-0.11	-0.06
number_photography_apps	0.06	0.05	0.02	0.03	0.04	0.08	0.09
calendar_apps_used	-0.06	0.03	-0.05	-0.04	-0.10	-0.12	-0.01
download_count..50.000...100.000.	-0.06	0.07	-0.06	-0.06	-0.07	-0.09	-0.02
number_apps_messenger_used	-0.06	0.01	-0.08	-0.13	0.08	-0.07	-0.09
app_usage_Education_perc_night	0.06	-0.01	0.06	0.14	-0.06	0.06	0.13
app_usage_Lifestyle_perc_night	-0.06	-0.03	-0.08	-0.10	0.03	-0.07	-0.08
app_usage_Browser_perc_midday	-0.06	-0.02	-0.14	-0.11	0.00	0.03	-0.07
app_usage_Media...Video_perc_morning	-0.06	-0.09	-0.10	-0.06	-0.02	0.03	-0.05
avg_usage_time_day_Entertainment	0.06	0.19	0.02	0.03	-0.03	-0.02	0.07
perc_Communication	0.06	0.10	-0.00	-0.03	0.16	0.09	-0.01
perc_Unknown	-0.06	-0.03	-0.06	-0.03	-0.04	-0.08	-0.01
var_incoming_sms_leng	0.05	0.05	-0.01	0.03	0.05	0.09	0.05
var_outgoing_sms_leng	0.05	0.00	0.02	0.05	0.07	0.07	0.02
bluetooth_used	-0.05	0.00	-0.07	-0.04	-0.07	-0.06	-0.05
number_music_audio_apps	-0.05	0.02	-0.02	-0.08	-0.04	-0.04	-0.06
app_usage_Unknown_perc_morning	-0.05	0.06	-0.02	-0.06	-0.04	-0.04	-0.03
app_usage_Business_perc_midday	-0.05	0.03	-0.07	-0.07	-0.10	-0.04	-0.05
app_usage_Transportation_perc_night	-0.05	-0.08	-0.08	-0.01	-0.00	0.02	-0.14
variance_number_incoming_calls_perday	-0.05	0.05	-0.08	-0.05	0.08	-0.01	-0.21
ratio_betw._avg_num_calls_perweek_d_e	-0.05	0.00	-0.07	-0.04	-0.03	-0.03	-0.10
avg_uses_perday_week_Transportation	0.05	0.02	-0.01	0.04	0.06	0.04	0.02
avg_uses_perday_Casual	-0.05	0.00	-0.08	0.03	-0.03	-0.07	-0.04
perc_Entertainment	0.05	0.08	0.04	0.11	-0.06	-0.03	0.14
total_number_missed_calls	-0.04	0.03	-0.05	-0.07	0.10	-0.02	-0.17
total_number_contacts_with_mail	-0.04	-0.02	-0.01	-0.02	-0.05	-0.08	-0.04
entropy_of_contact_sms_weekday	0.04	0.17	-0.06	-0.04	0.07	0.11	-0.07
avg_number_videos_taken_weekdays	-0.04	0.02	-0.07	-0.03	-0.07	-0.01	0.01
avg_number_videos_taken_weekend	0.04	0.07	0.02	-0.08	0.06	0.07	0.04
total_events_airplaine_db	0.04	0.07	0.12	0.09	-0.01	-0.05	0.07
percentage_of_songs_listened_between_18_24	-0.04	0.02	-0.03	-0.08	0.03	0.01	-0.13
number_education_apps	-0.04	-0.04	0.01	0.01	-0.12	-0.09	0.02
number_games_racing_apps	-0.04	-0.10	0.01	-0.02	-0.05	-0.05	0.03
download_count..1.000.000.000...5.000.000.000.	-0.04	-0.00	-0.09	-0.05	-0.07	-0.06	-0.02
avg_usage_time_7h	-0.04	-0.06	0.04	0.05	-0.04	-0.12	-0.01
app_usage_Unknown_perc_night	-0.04	0.05	-0.02	-0.04	-0.07	-0.05	0.03
app_usage_Shopping_perc_evening	0.04	0.01	0.04	0.04	0.01	-0.03	-0.01
app_usage_Communication_perc_evening	-0.04	-0.08	-0.02	-0.02	0.04	-0.06	-0.04
app_usage_Music...Audio_perc_morning	0.04	0.11	0.07	0.04	0.04	-0.01	0.02
app_usage_Education_perc_midday	-0.04	-0.05	0.02	-0.04	-0.10	-0.06	0.02
app_usage_Lifestyle_perc_morning	0.04	0.04	-0.02	-0.08	0.14	0.05	-0.01
app_usage_Weather_perc_morning	0.04	0.00	0.06	0.02	0.03	0.07	0.03
app_usage_Weather_perc_night	0.04	0.12	-0.01	-0.03	0.04	0.09	0.03

app_usage_Sports_perc_evening	-0.04	-0.08	0.07	-0.02	-0.07	-0.13	0.02
app_usage_Browser_perc_evening	-0.04	-0.18	-0.01	0.02	0.05	-0.05	-0.11
ratio_betw._avg_number_in_calls_perweek_d_e	-0.04	0.00	0.04	-0.07	-0.06	-0.01	-0.10
avg_uses_perday_week_Business	-0.04	0.04	0.00	-0.10	-0.05	-0.01	-0.07
avg_uses_perday_Puzzle	-0.04	-0.10	0.01	-0.00	-0.04	-0.02	-0.04
perc_Media...Video	-0.04	-0.04	-0.08	-0.05	-0.05	0.03	0.06
total_number_added_contacts	-0.03	0.04	-0.03	-0.06	-0.05	0.00	-0.02
entropy_of_contact_outgoing_sms_weekend	0.03	0.10	-0.04	0.03	0.07	0.03	-0.03
response_rate_calls_weekend	-0.03	0.05	-0.01	-0.02	0.00	-0.04	-0.12
avg_charge_disconnected	0.03	-0.11	0.03	0.14	-0.04	0.04	0.01
avg_usage_time_8h	-0.03	-0.06	-0.02	0.06	0.01	-0.09	-0.05
app_usage_Finance_perc_midday	-0.03	0.06	-0.02	-0.04	-0.01	0.03	-0.15
app_usage_Games_perc_midday	0.03	0.03	0.07	-0.05	0.04	0.04	-0.04
app_usage_Entertainment_perc_morning	-0.03	0.06	-0.07	0.01	-0.04	-0.09	0.03
app_usage_News...Magazines_perc_evening	0.03	0.06	0.09	-0.01	0.03	0.01	0.03
app_usage_Unknown_perc_midday	-0.03	0.09	-0.04	-0.05	-0.00	-0.01	-0.01
app_usage_Medical_perc_midday	-0.03	0.08	-0.05	-0.08	0.06	-0.03	-0.13
app_usage_Lifestyle_perc_evening	0.03	0.04	0.00	-0.05	0.13	0.02	-0.01
app_usage_Social_perc_midday	-0.03	0.07	-0.06	-0.02	-0.01	-0.07	-0.11
ratio_incoming_outgoing_sms	-0.03	-0.10	0.01	0.01	-0.00	-0.06	-0.08
perc_Lifestyle	0.03	0.03	0.03	-0.08	0.10	-0.01	0.02
perc_Productivity	0.03	0.01	0.12	0.06	-0.08	0.06	0.07
perc_Social	0.03	0.04	-0.05	0.09	0.09	0.02	-0.06
total_duration_incoming_calls	-0.02	0.02	-0.06	0.01	0.11	-0.02	-0.13
avg_duration_incoming_calls_weekend	0.02	0.02	-0.05	0.06	0.14	0.01	-0.07
ratio_avg_duration_incoming_outgoing_calls	-0.02	-0.02	-0.05	0.01	0.07	0.01	-0.08
entropy_of_contact_missed_calls_weekend	-0.02	0.01	-0.04	-0.05	0.11	-0.00	-0.11
number_games_puzzle_apps	0.02	-0.05	0.07	0.05	0.03	0.05	-0.01
download_count..5.000...10.000.	-0.02	-0.08	0.02	-0.03	-0.07	-0.03	0.06
usage_count_6h	-0.02	-0.01	-0.02	0.09	-0.03	-0.02	-0.05
app_usage_Games_perc_morning	0.02	0.03	0.02	0.00	0.07	0.05	-0.03
app_usage_Productivity_perc_midday	-0.02	0.15	-0.07	-0.09	-0.03	-0.04	0.00
app_usage_Photography_perc_morning	0.02	-0.03	-0.02	0.09	0.06	0.02	0.05
app_usage_Travel...Local_perc_midday	-0.02	0.05	-0.08	-0.08	0.02	0.01	-0.09
app_usage_Music...Audio_perc_evening	-0.02	-0.05	0.08	-0.05	0.00	-0.05	-0.07
app_usage_Business_perc_night	0.02	0.01	0.01	0.06	0.01	0.02	0.05
app_usage_Weather_perc_evening	0.02	0.02	0.02	-0.02	0.02	0.08	-0.02
ratio_incoming_outgoing_calls_weekday	0.02	0.04	0.05	0.09	0.02	0.02	-0.02
avg_uses_perday_week_Education	0.02	0.00	0.01	0.02	-0.04	0.00	0.07
perc_Photography	0.02	0.04	0.00	0.01	-0.02	0.06	0.08
perc_Sports	-0.02	-0.09	0.13	-0.02	-0.08	-0.12	0.02
perc_Tools	-0.02	-0.03	0.02	-0.01	-0.09	-0.01	-0.02
perc_Browser	-0.02	0.02	-0.05	0.01	-0.03	-0.06	0.03

total_duration_calls	-0.01	0.03	-0.02	-0.05	0.10	-0.02	-0.11
var_duration_calls	0.01	0.01	0.04	0.05	0.08	-0.04	-0.06
total_number_contacts_with_two_numbers	0.01	0.10	-0.02	0.07	0.04	-0.05	-0.10
total_number_unique_contacts_who_called	-0.01	0.12	-0.06	-0.03	0.09	-0.00	-0.10
total_number_unique_contacts_outgoing_sms	-0.01	0.03	0.04	-0.05	-0.02	0.01	0.02
response_rate_sms	-0.01	0.12	-0.07	-0.11	0.06	0.06	-0.06
number_business_apps	0.01	-0.02	0.04	-0.04	0.01	0.04	0.00
number_tools_apps	0.01	-0.07	0.02	0.01	0.07	0.02	0.01
number_finance_apps	0.01	0.13	-0.16	-0.07	0.05	0.07	-0.09
total_number_shared_photos	0.01	-0.02	0.03	-0.03	0.07	-0.02	0.02
app_usage_News...Magazines_perc_midday	0.01	0.10	0.06	-0.09	0.03	-0.02	-0.03
app_usage_News...Magazines_perc_night	0.01	-0.04	0.01	0.05	0.02	0.01	0.02
app_usage_Unknown_perc_evening	-0.01	0.12	0.05	-0.08	-0.02	-0.04	0.01
app_usage_Photography_perc_evening	-0.01	0.05	-0.02	-0.03	-0.00	-0.03	-0.07
app_usage_Books...Reference_perc_morning	0.01	0.04	-0.03	0.02	0.06	0.00	0.02
app_usage_Travel...Local_perc_morning	0.01	0.03	0.00	0.07	-0.08	-0.04	0.10
app_usage_Education_perc_evening	-0.01	-0.07	-0.02	0.03	-0.06	-0.04	0.07
ratio_incoming_outgoing_calls_weekend	0.01	0.02	-0.04	0.09	0.04	0.04	-0.05
avg_uses_perday_Lifestyle	-0.01	-0.01	0.00	-0.12	0.06	-0.02	-0.01
avg_uses_perday_end_Education	-0.01	-0.06	0.07	0.05	-0.09	-0.05	0.09
avg_usage_time_day_Books...Reference	0.01	0.03	0.02	-0.00	-0.08	0.06	-0.03
perc_Music...Audio	-0.01	0.04	0.01	-0.01	-0.05	-0.00	-0.00
download_count..10.000...50.000.	-0.00	0.02	0.04	-0.02	-0.06	-0.06	0.04
app_usage_Education_perc_morning	0.00	-0.01	0.01	0.00	0.00	-0.02	0.04
app_usage_Social_perc_morning	0.00	0.00	-0.02	0.09	-0.00	-0.04	-0.02
perc_Medical	-0.00	0.03	-0.03	-0.03	0.10	0.01	-0.10

Note: Pairwise Spearman correlations between Conscientiousness (factor, facets) and predictor variables from Section 2.3; table is sorted by absolute ρ values of Conscientiousness, in decreasing order. Abbreviations: C1 = Competence, C2 = Love of Order, C3 = Sense of Duty, C4 = Ambition, C5 =Discipline, C6 = Caution.

Table 5: Pairwise Spearman Correlations Between Extraversion and Predictors Study 3

Predictors	Extraversion	E1	E2	E3	E4	E5	E6
total_number_contacts_with_one_number	0.30	0.24	0.32	0.20	0.28	0.21	0.16
avg_usage_time_day_Communication	0.26	0.25	0.35	0.17	0.18	0.20	0.12
total_number_contacts_with_two_numbers	0.21	0.20	0.22	0.12	0.14	0.17	0.17
avg_charge_disconnected	-0.21	-0.11	-0.16	-0.08	-0.18	-0.10	-0.17
number_apps_messenger_used	0.21	0.19	0.28	0.11	0.13	0.17	0.11
percentage_of_songs_listened_between_18_24	0.20	0.09	0.10	0.14	0.18	0.20	0.14
app_usage_Productivity_perc_midday	0.20	0.16	0.11	0.11	0.12	0.16	0.17
app_usage_Music...Audio_perc_night	0.20	0.17	0.18	0.12	0.13	0.22	0.12
app_usage_Weather_perc_night	0.20	0.15	0.19	0.13	0.12	0.18	0.06
response_rate_calls_weekend	0.19	0.08	0.17	0.19	0.16	0.19	0.10
var_last_event_weekend	0.19	0.13	0.24	0.09	0.11	0.14	0.14
app_usage_Medical_perc_midday	0.19	0.19	0.24	0.17	0.18	0.09	0.04
number_radio_usage	0.19	0.15	0.19	0.18	0.21	0.11	0.10
avg_uses_perday_week_Transportation	0.19	0.23	0.18	0.07	0.12	-0.04	0.23
total_number_contacts_end	0.18	0.15	0.14	0.12	0.08	0.18	0.14
total_number_unique_contacts_who_called	0.18	0.10	0.11	0.13	0.20	0.23	0.08
entropy_of_contact_outgoing_sms_weekend	0.18	0.18	0.15	0.10	0.21	0.16	0.06
ratio_incoming_outgoing_sms	-0.18	-0.24	-0.12	-0.12	-0.14	-0.07	-0.11
number_checking_behaviour_events	0.17	0.12	0.20	0.09	0.20	0.07	0.07
app_usage_Entertainment_perc_midday	0.17	0.04	0.15	0.13	0.13	0.09	0.15
app_usage_Lifestyle_perc_morning	0.17	0.04	0.11	0.14	0.22	0.02	0.13
app_usage_Browser_perc_midday	0.17	0.15	0.13	0.06	0.12	0.05	0.07
number_shazam_apps_used	0.17	0.12	0.14	0.14	0.14	0.21	0.09
entropy_of_contact_missed_calls_weekend	0.16	0.09	0.21	0.12	0.19	0.13	0.03
response_rate_sms	0.16	0.11	0.13	0.16	0.18	0.03	0.08
regularity_last_event_weekend	0.16	0.16	0.14	0.02	0.08	0.08	0.14
entropy_music_genres_morning	0.16	0.08	0.13	0.20	0.20	0.14	0.05
regularity_all_aggr_events	0.16	0.18	0.21	0.04	0.12	0.09	0.07
variance_number_incoming_calls_perday	0.16	0.12	0.18	0.09	0.17	0.17	0.06
avg_uses_perday_week_Travel...Local	0.16	0.08	0.10	0.06	0.19	0.17	0.09
avg_uses_perday_Trivia	-0.16	-0.16	-0.14	-0.07	-0.08	-0.11	-0.15
avg_uses_perday_end_Transportation	0.16	0.22	0.17	0.03	0.10	0.01	0.20
total_number_added_contacts	0.15	0.11	0.12	0.13	0.07	0.12	0.09
entropy_of_contact_missed_calls	0.15	0.08	0.16	0.21	0.17	0.14	0.04
regularity_last_event_weekday	-0.15	-0.13	-0.13	-0.01	-0.15	0.01	-0.21
app_usage_Travel...Local_perc_night	0.15	0.07	0.12	0.04	0.17	0.14	0.08
app_usage_Lifestyle_perc_evening	0.15	0.11	0.12	0.09	0.18	0.04	0.12
app_usage_Media...Video_perc_night	0.15	0.06	0.14	0.07	0.16	0.12	0.06
entropy_of_contact_outgoing_sms_weekday	0.14	0.13	0.07	0.20	0.17	-0.00	0.05
total_events_airplaine_db	0.14	0.08	0.13	0.07	0.10	0.14	0.09

bluetooth_used	0.14	0.15	0.09	0.02	0.07	0.13	0.17
app_usage_News...Magazines_perc_midday	0.14	0.00	0.19	0.18	0.18	0.10	0.07
app_usage_Communication_perc_midday	0.14	0.05	0.13	0.14	0.06	0.13	0.16
avg_uses_perday_Lifestyle	0.14	0.06	0.12	0.07	0.18	0.04	0.13
avg_uses_perday_end_Education	-0.14	-0.12	-0.15	-0.12	-0.15	-0.02	-0.07
avg_usage_time_day_Books...Reference	-0.14	-0.18	-0.16	0.03	-0.16	-0.12	-0.07
perc_Communication	0.14	0.17	0.16	0.18	0.09	0.03	-0.00
total_number_missed_calls	0.13	0.07	0.17	0.10	0.15	0.19	-0.02
response_rate_calls_weekday	0.13	0.02	0.14	0.15	0.13	0.17	0.12
app_usage_Entertainment_perc_evening	-0.13	-0.09	-0.08	-0.05	-0.07	-0.13	-0.05
app_usage_Travel...Local_perc_evening	0.13	0.05	0.09	0.06	0.13	0.12	0.11
ratio_incoming_outgoing_calls_weekend	-0.13	-0.10	-0.15	-0.11	-0.09	0.02	-0.08
avg_uses_perday_end_Photography	0.13	0.11	0.04	0.19	0.15	0.12	0.01
avg_usage_time_day_Travel...Local	0.13	0.14	0.09	-0.06	0.11	0.04	0.14
perc_Medical	0.13	0.14	0.21	0.15	0.10	-0.00	0.03
perc_Sports	-0.13	-0.13	-0.12	-0.07	-0.07	-0.11	-0.05
total_duration_calls	0.12	0.12	0.15	0.11	0.14	0.08	-0.02
avg_number_charge_connected_per_day	0.12	0.03	0.11	0.08	0.12	0.05	0.11
number_songs_listened_per_day	0.12	0.03	0.05	0.06	0.14	0.17	0.10
number_sports_apps	-0.12	-0.13	-0.16	-0.11	-0.05	-0.10	-0.00
app_usage_Music...Audio_perc_midday	0.12	0.02	0.02	0.14	0.13	0.09	0.11
app_usage_Lifestyle_perc_night	0.12	0.07	0.12	0.07	0.13	0.06	0.03
ratio_incoming_outgoing_calls_weekday	-0.12	-0.10	-0.08	-0.10	-0.13	0.02	-0.07
perc_Lifestyle	0.12	0.05	0.07	0.11	0.18	-0.02	0.09
entropy_of_contact_sms_weekday	0.11	0.09	0.15	0.17	0.08	-0.00	0.04
var_first_event_weekday	-0.11	-0.09	-0.05	-0.01	-0.14	0.01	-0.15
avg_number_videos_taken_weekend	0.11	0.10	0.10	0.12	0.13	0.01	0.11
percentage_of_songs_listened_between_12_18	0.11	0.01	0.04	0.16	0.18	0.11	0.04
number_games_puzzle_apps	0.11	0.07	-0.00	0.01	0.15	0.06	0.17
total_number_shared_photos	0.11	0.27	0.19	0.02	0.00	0.04	0.06
usage_count_4h	0.11	0.10	0.18	0.01	0.02	0.18	0.05
app_usage_Tools_perc_evening	-0.11	-0.02	-0.07	-0.14	-0.06	-0.16	-0.08
app_usage_Media...Video_perc_evening	0.11	0.08	0.14	0.06	0.16	-0.04	-0.04
avg_uses_perday_week_Tools	0.11	0.00	0.15	-0.04	0.09	0.11	0.09
avg_duration_outgoing_calls_weekend	0.10	0.10	0.09	0.12	0.13	0.11	0.01
gps_data_available	-0.10	-0.02	0.03	-0.09	-0.11	-0.09	-0.13
number_finance_apps	0.10	-0.01	0.07	0.16	0.13	0.08	0.03
number_games_racing_apps	-0.10	-0.08	-0.10	-0.07	-0.17	-0.01	-0.01
app_usage_Tools_perc_midday	0.10	0.13	0.18	0.05	0.05	0.03	0.03
app_usage_Shopping_perc_evening	0.10	0.05	0.08	0.10	0.12	0.11	0.07
app_usage_Transportation_perc_morning	0.10	0.13	0.02	0.08	0.09	0.02	0.16
app_usage_Browser_perc_night	-0.10	-0.08	-0.07	-0.08	-0.11	0.06	-0.12
usage_Weather_apps	0.10	0.06	0.09	0.10	0.10	0.11	-0.04

avg_uses_perday_week_Weather	0.10	0.03	0.09	0.11	0.10	0.12	-0.03
perc_News...Magazines	-0.10	-0.15	-0.03	0.03	-0.01	-0.11	-0.11
avg_time_first_event_sunday	-0.09	-0.06	-0.01	-0.09	-0.16	0.08	-0.10
var_duration_downtime_weekday	-0.09	-0.11	-0.07	0.03	-0.07	-0.01	-0.14
regularity_last_event_all	-0.09	-0.04	-0.09	-0.01	-0.13	0.03	-0.12
avg_usage_time_7h	-0.09	-0.14	-0.08	0.01	0.01	-0.06	-0.13
app_usage_Shopping_perc_morning	0.09	0.04	0.03	0.18	0.08	0.05	0.07
app_usage_Books...Reference_perc_morning	0.09	0.08	0.09	0.02	0.06	0.02	0.03
app_usage_Music...Audio_perc_morning	0.09	-0.01	0.03	0.05	0.08	0.05	0.09
app_usage_Transportation_perc_midday	0.09	0.12	0.10	0.07	0.07	-0.07	0.05
app_usage_Sports_perc_night	-0.09	-0.10	-0.09	-0.05	-0.07	0.00	-0.04
perc_Shopping	0.09	-0.00	0.07	0.15	0.12	0.09	0.06
perc_Transportation	0.09	0.15	0.05	0.05	0.08	-0.00	0.14
total_duration_incoming_calls	0.08	0.10	0.14	0.05	0.09	0.07	-0.02
percent_sms_night	-0.08	-0.10	0.05	-0.01	-0.13	-0.08	-0.14
avg_number_videos_taken_weekdays	0.08	0.03	0.07	0.07	0.08	0.15	0.01
avg_inter_event_time_weekend	-0.08	-0.02	-0.07	-0.10	-0.10	-0.09	-0.02
percentage_of_songs_listened_between_6_12	0.08	0.04	0.00	0.02	0.04	0.06	0.11
number_books_and_reference_apps	0.08	-0.03	-0.00	0.03	0.08	0.12	0.13
download_count..50.000...100.000.	0.08	0.02	-0.04	0.07	0.11	0.07	0.07
avg_usage_time_2h	-0.08	-0.07	0.04	-0.13	-0.05	0.02	-0.07
avg_usage_time_10h	0.08	-0.04	0.01	0.08	0.13	0.12	0.03
usage_count_7h	0.08	0.05	0.11	-0.01	0.09	0.01	0.08
app_usage_Productivity_perc_evening	-0.08	-0.10	0.01	-0.10	-0.01	-0.11	-0.08
app_usage_Photography_perc_night	0.08	0.05	0.15	-0.01	0.09	0.10	-0.01
app_usage_Education_perc_evening	-0.08	-0.06	-0.12	-0.08	-0.10	0.04	-0.06
app_usage_Sports_perc_evening	-0.08	-0.12	-0.06	-0.05	-0.01	-0.02	-0.02
app_usage_Social_perc_morning	-0.08	-0.08	-0.08	-0.04	-0.04	-0.07	-0.04
app_usage_Social_perc_evening	0.08	0.02	0.13	0.10	0.13	0.05	0.04
usage_Health...Fitness_apps	0.08	0.19	0.10	0.10	0.01	0.05	-0.01
avg_uses_perday_week_Business	0.08	-0.05	-0.01	0.11	0.16	0.02	0.07
perc_Music...Audio	0.08	0.05	-0.00	0.08	0.06	0.15	0.06
perc_Browser	-0.08	-0.08	-0.11	-0.10	0.01	-0.20	-0.05
percent_calls_night	0.07	0.03	0.15	0.03	0.03	0.11	-0.05
var_first_event_weekend	0.07	0.06	0.04	0.02	0.02	0.12	0.04
avg_charge_connected	-0.07	-0.13	-0.04	-0.09	-0.06	-0.06	-0.00
number_business_apps	-0.07	-0.10	-0.08	-0.02	0.04	-0.11	-0.10
download_count..1.000.000.000...5.000.000.000.	-0.07	-0.12	-0.10	0.04	-0.08	0.01	-0.02
avg_usage_time_8h	-0.07	-0.07	-0.11	-0.06	0.02	0.00	-0.06
usage_count_0h	0.07	0.05	0.16	-0.03	0.00	0.14	0.02
app_usage_News...Magazines_perc_morning	0.07	-0.11	0.06	0.17	0.12	0.06	0.04
app_usage_Unknown_perc_morning	-0.07	-0.08	-0.13	0.08	-0.07	-0.01	-0.08
app_usage_Photography_perc_morning	0.07	0.09	0.03	0.02	0.13	-0.07	0.01

app_usage_Transportation_perc_night	0.07	0.06	0.14	-0.06	0.00	0.07	0.12
app_usage_Browser_perc_morning	-0.07	-0.12	-0.15	0.02	-0.02	-0.00	-0.10
app_usage_Browser_perc_evening	-0.07	-0.05	0.04	-0.13	-0.05	-0.09	0.00
avg_uses_perday_week_Games	-0.07	-0.11	-0.08	-0.12	-0.07	0.06	0.03
avg_uses_perday_Arcade	-0.07	0.03	-0.09	-0.23	-0.07	0.07	-0.01
avg_leng_incoming_sms	-0.06	-0.02	0.06	-0.00	-0.04	-0.18	-0.03
ratio_avg_duration_incoming_outgoing_calls	-0.06	-0.02	0.03	-0.08	-0.02	-0.09	-0.11
percentage_of_songs_listened_between_0_6	0.06	-0.01	0.08	0.01	0.08	0.08	0.04
number_music_audio_apps	0.06	0.05	0.03	0.09	0.07	0.15	-0.05
number_photography_apps	0.06	-0.02	0.03	0.13	0.08	0.02	-0.02
number_games_board_apps	0.06	0.12	0.03	-0.11	-0.02	-0.01	0.16
download_count..10.000...50.000.	0.06	-0.08	-0.01	0.10	0.14	0.02	0.14
download_count..5.000...10.000.	-0.06	0.00	-0.13	-0.04	-0.04	-0.14	-0.01
number_apps_searchengine_used	0.06	-0.08	0.04	0.04	0.08	0.02	0.08
avg_usage_time_5h	0.06	0.05	0.11	0.00	0.04	0.16	-0.01
app_usage_Tools_perc_morning	0.06	-0.07	-0.04	0.18	0.11	0.01	0.03
app_usage_News...Magazines_perc_evening	0.06	-0.11	0.10	0.10	0.10	0.03	0.02
app_usage_Communication_perc_morning	-0.06	-0.07	-0.06	0.05	0.00	-0.17	-0.04
app_usage_Travel...Local_perc_morning	-0.06	-0.02	-0.10	-0.06	-0.03	-0.03	-0.03
app_usage_Music...Audio_perc_evening	0.06	0.06	0.05	0.04	0.03	0.05	-0.00
app_usage_Weather_perc_evening	0.06	0.03	0.04	0.07	0.10	0.07	-0.08
app_usage_Media...Video_perc_midday	0.06	0.05	0.05	-0.03	0.11	0.06	-0.06
avg_uses_perday_Casual	0.06	0.14	0.06	-0.06	0.02	0.07	0.09
avg_usage_time_day_Tools	0.06	-0.02	0.13	-0.03	0.03	0.11	0.04
avg_usage_time_day_Music...Audio	0.06	0.05	0.11	0.05	-0.02	0.07	-0.00
perc_Entertainment	-0.06	-0.07	-0.13	0.08	-0.01	-0.09	0.01
perc_Tools	-0.06	-0.03	-0.06	-0.17	-0.09	-0.04	0.01
avg_leng_outgoing_sms	-0.05	0.11	0.03	-0.17	-0.09	-0.12	-0.06
number_weather_apps	0.05	0.01	0.14	0.04	0.09	0.02	-0.07
avg_usage_time_19h	-0.05	-0.04	-0.06	-0.11	-0.00	-0.03	0.03
app_usage_Books...Reference_perc_midday	0.05	0.07	0.04	-0.07	0.03	0.03	0.04
app_usage_Books...Reference_perc_evening	0.05	0.06	0.10	-0.13	0.05	0.03	0.03
app_usage_Business_perc_night	0.05	0.15	0.15	-0.03	0.01	0.02	-0.06
app_usage_Health...Fitness_perc_morning	0.05	0.18	0.10	0.03	-0.03	0.04	-0.07
ratio_betw._avg_number_in_calls_perweek_d_e	0.05	0.00	0.01	0.05	0.04	-0.01	0.16
avg_uses_perday_Puzzle	0.05	0.04	-0.01	-0.02	0.05	-0.00	0.11
perc_Unknown	-0.05	0.00	-0.13	0.00	-0.02	0.04	-0.03
avg_completeness_score_contacts	0.04	0.06	0.00	0.08	-0.04	0.08	0.06
number_nights_less_than_4_hours_downtime	0.04	0.03	0.07	0.05	-0.00	0.07	-0.02
total_events_boot_db	0.04	0.08	0.03	-0.02	0.07	-0.03	0.02
number_battery_saver_task_killer_apps	-0.04	-0.10	-0.03	-0.02	-0.04	-0.08	0.04
usage_count_6h	-0.04	-0.04	0.02	-0.08	-0.03	0.08	-0.09
app_usage_Games_perc_night	0.04	-0.00	-0.01	-0.04	-0.01	0.12	0.11

app_usage_Personalization_perc_morning	-0.04	0.04	-0.05	-0.13	-0.10	-0.02	0.07
app_usage_Photography_perc_midday	0.04	0.14	0.01	-0.01	-0.01	-0.01	0.05
app_usage_Photography_perc_evening	0.04	-0.01	0.02	0.13	0.06	0.11	-0.09
app_usage_Education_perc_morning	-0.04	-0.04	-0.05	0.03	-0.04	0.03	-0.02
app_usage_Health...Fitness_perc_night	-0.04	0.03	-0.03	0.00	-0.06	-0.00	-0.12
avg_usage_time_day_Entertainment	0.04	0.06	0.08	0.10	-0.00	-0.01	0.07
perc_Business	-0.04	-0.08	-0.08	-0.01	0.01	-0.07	-0.01
perc_Productivity	-0.04	-0.06	-0.14	0.04	-0.01	-0.09	0.01
var_duration_calls_weekend	0.03	-0.01	0.06	0.10	0.09	0.08	-0.10
var_last_event_weekday	-0.03	-0.07	0.06	0.01	-0.05	0.01	-0.12
var_duration_downtime_weekend	0.03	-0.02	0.01	0.04	0.04	0.13	-0.01
regularity_first_event_weekday	-0.03	-0.05	-0.05	0.05	-0.02	-0.04	-0.05
ratio_number_apps_inst_apps_used	0.03	-0.02	0.01	0.06	0.07	0.03	-0.05
number_antivirus_and_security_apps	0.03	-0.07	-0.03	0.10	0.07	-0.02	0.06
app_usage_Tools_perc_night	0.03	0.06	0.05	-0.03	-0.04	0.06	-0.01
app_usage_Games_perc_morning	-0.03	-0.05	-0.10	-0.03	-0.03	0.08	0.04
app_usage_Entertainment_perc_night	0.03	-0.02	0.09	-0.01	-0.00	0.13	0.05
app_usage_Unknown_perc_evening	0.03	-0.02	-0.06	0.13	0.05	0.09	-0.04
app_usage_Unknown_perc_night	-0.03	-0.09	-0.09	0.08	-0.01	0.07	-0.03
app_usage_Business_perc_evening	0.03	-0.00	0.03	-0.05	0.06	0.05	0.05
app_usage_Social_perc_midday	0.03	-0.09	0.06	0.10	0.03	0.01	-0.00
usage_News...Magazines_apps	-0.03	-0.10	0.00	0.04	0.02	-0.01	-0.02
perc_Books...Reference	-0.03	0.00	-0.03	-0.08	-0.04	-0.05	-0.03
perc_Photography	0.03	0.13	-0.00	-0.01	-0.01	0.02	0.00
var_duration_calls	-0.02	-0.04	0.01	0.02	0.02	0.01	-0.07
total_number_contacts_with_mail	0.02	0.07	0.05	0.05	-0.06	-0.02	0.03
total_number_unique_contacts_outgoing_sms	-0.02	-0.01	-0.04	0.13	0.05	-0.11	-0.15
avg_time_last_event_weekday	0.02	0.01	0.09	-0.02	-0.02	0.09	-0.03
download_count..5.000.000...10.000.000.	0.02	0.03	-0.03	0.01	0.03	0.08	0.07
avg_usage_time_0h	-0.02	-0.08	-0.04	0.01	0.05	0.11	-0.07
app_usage_Finance_perc_midday	0.02	-0.07	-0.02	0.16	0.05	0.05	-0.02
app_usage_Unknown_perc_midday	0.02	0.02	0.00	0.13	-0.01	0.02	-0.03
app_usage_Communication_perc_evening	0.02	0.05	0.01	-0.07	0.04	0.04	-0.02
app_usage_Books...Reference_perc_night	-0.02	0.03	0.04	-0.17	-0.06	0.01	-0.00
app_usage_Education_perc_midday	-0.02	-0.03	-0.03	-0.02	-0.08	0.09	-0.01
app_usage_Health...Fitness_perc_evening	0.02	0.20	0.11	0.01	-0.04	-0.02	-0.07
avg_uses_perday_end_Business	-0.02	-0.02	0.03	-0.07	-0.02	-0.03	-0.04
avg_duration_incoming_calls_weekend	0.01	0.05	0.11	-0.02	0.01	0.04	-0.07
var_outgoing_sms_leng	0.01	0.07	0.11	-0.05	-0.00	-0.15	0.00
response_rate_missed_call_answer_with_sms	0.01	0.03	0.04	0.06	0.01	0.04	-0.08
number_nights_more_than_7_hours_downtime	-0.01	-0.03	-0.03	0.03	0.03	-0.14	0.01
number_education_apps	0.01	-0.01	0.00	-0.02	0.00	-0.04	0.09
calendar_apps_used	-0.01	-0.10	-0.06	0.10	0.04	0.03	0.03

avg_plusone_scores	-0.01	0.02	0.00	-0.00	-0.02	0.03	-0.07
avg_usage_time_1h	-0.01	0.03	0.01	-0.02	0.02	0.00	-0.03
app_usage_Games_perc_midday	0.01	0.03	-0.05	0.01	-0.01	0.06	0.09
app_usage_Entertainment_perc_morning	-0.01	0.00	-0.05	0.05	0.02	-0.09	0.01
app_usage_Productivity_perc_morning	-0.01	-0.07	-0.11	0.14	0.02	-0.02	0.02
app_usage_Shopping_perc_midday	0.01	-0.07	-0.03	0.15	0.09	0.02	-0.03
app_usage_Travel...Local_perc_midday	-0.01	0.02	0.06	0.00	-0.02	-0.11	0.02
app_usage_Education_perc_night	-0.01	0.01	-0.07	-0.02	-0.04	-0.01	0.08
app_usage_Business_perc_midday	0.01	-0.04	-0.06	0.03	0.08	-0.00	0.01
app_usage_Health...Fitness_perc_midday	0.01	0.18	0.09	0.00	-0.03	-0.01	-0.08
app_usage_Media...Video_perc_morning	-0.01	-0.03	0.01	-0.00	0.03	-0.02	-0.11
app_usage_Social_perc_night	0.01	-0.03	-0.01	0.02	0.00	0.18	0.01
avg_uses_perday_week_Education	-0.01	0.00	-0.06	0.02	-0.01	-0.00	0.03
perc_Media...Video	-0.01	-0.02	-0.02	0.06	0.04	-0.14	-0.09
perc_Social	0.01	0.01	0.04	0.01	0.04	0.02	-0.02
var_duration_incoming_calls	0.00	0.01	0.10	0.02	0.05	0.02	-0.09
var_incoming_sms_leng	0.00	0.01	0.04	0.00	0.01	-0.07	0.09
number_events_during_sleep	-0.00	-0.02	0.10	0.09	-0.04	-0.01	-0.12
number_tools_apps	0.00	0.01	0.04	-0.01	0.05	-0.01	-0.06
avg_usage_time_6h	-0.00	0.04	-0.00	-0.09	0.05	-0.01	0.05
app_usage_Productivity_perc_night	-0.00	0.02	0.09	-0.05	-0.04	0.02	-0.02
app_usage_News...Magazines_perc_night	0.00	-0.07	0.03	0.04	0.05	-0.03	-0.05
app_usage_Transportation_perc_evening	0.00	0.10	0.08	-0.09	-0.04	-0.16	0.12
app_usage_Weather_perc_morning	0.00	-0.02	0.02	-0.01	0.02	0.02	-0.11
ratio_betw._avg_num_calls_perweek_d_e	-0.00	-0.02	-0.10	-0.05	0.04	-0.02	0.09

Note: Pairwise Spearman correlations between Extraversion (factor, facets) and predictor variables from Section 2.3; table is sorted by absolute ρ values of Extraversion, in decreasing order. Abbreviations: E1 = Friendliness, E2 = Sociableness, E3 = Assertiveness, E4 = Dynamism, E5 = Adventurousness, E6 = Cheerfulness.

Table 6: Pairwise Spearman Correlations Between Agreeableness and Predictors Study 3

Predictors	Agreeableness	A1	A2	A3	A4	A5	A6
percent_sms_night	-0.29	-0.12	-0.20	-0.27	-0.21	-0.16	-0.22
number_events_during_sleep	-0.21	-0.07	-0.16	-0.22	-0.26	-0.06	-0.15
avg_uses_perday_week_Transportation	0.21	0.22	0.16	0.19	0.16	-0.02	0.18
app_usage_Transportation_perc_evening	0.20	0.18	0.10	0.21	0.16	0.11	0.13
app_usage_Transportation_perc_morning	0.19	0.14	0.09	0.14	0.17	0.08	0.17
total_number_shared_photos	0.18	0.11	0.07	0.22	0.01	0.04	0.22
avg_uses_perday_Casual	0.18	0.17	0.14	0.12	0.13	0.07	0.17
app_usage_Unknown_perc_morning	-0.17	-0.03	-0.15	-0.11	-0.19	-0.17	-0.15
avg_usage_time_day_Books...Reference	-0.17	-0.17	-0.13	-0.15	-0.13	-0.09	-0.11
avg_leng_outgoing_sms	0.16	0.07	0.06	0.22	0.00	0.08	0.20
percent_calls_night	-0.16	-0.13	-0.09	-0.16	-0.11	-0.11	-0.10
ratio_betw._avg_num_calls_perweek_d_e	0.16	0.24	0.17	0.15	0.08	0.13	0.06
avg_uses_perday_end_Transportation	0.16	0.17	0.10	0.13	0.15	0.03	0.09
response_rate_calls_weekend	-0.15	-0.14	0.02	-0.14	-0.13	-0.12	-0.17
number_business_apps	-0.15	-0.13	-0.17	-0.09	-0.05	-0.14	-0.07
app_usage_Photography_perc_midday	0.15	0.09	0.14	0.12	0.16	0.08	0.15
var_duration_downtime_weekend	-0.14	-0.03	-0.11	-0.19	-0.16	-0.09	-0.11
regularity_last_event_weekday	-0.14	-0.03	-0.08	-0.13	-0.22	0.05	-0.16
bluetooth_used	0.14	0.12	0.16	0.15	0.10	0.05	0.11
total_events_boot_db	0.14	0.09	0.08	0.17	0.11	0.05	0.09
total_number_unique_contacts_outgoing_sms	-0.13	0.02	-0.02	-0.07	-0.17	-0.21	-0.14
number_weather_apps	0.13	0.08	0.15	0.10	0.11	0.04	0.09
number_battery_saver_task_killer_apps	-0.13	-0.05	-0.04	-0.16	-0.04	-0.06	-0.13
calendar_apps_used	-0.13	-0.09	-0.06	-0.08	-0.06	-0.16	-0.12
app_usage_News...Magazines_perc_evening	-0.13	-0.10	0.02	-0.13	-0.04	-0.15	-0.14
app_usage_Unknown_perc_night	-0.13	-0.01	-0.11	-0.12	-0.11	-0.15	-0.15
app_usage_Business_perc_night	0.13	0.10	0.08	0.16	0.06	0.02	0.17
app_usage_Transportation_perc_night	0.13	0.11	0.12	0.11	0.04	0.10	0.06
perc_Transportation	0.13	0.19	0.10	0.09	0.04	0.03	0.09
var_first_event_weekend	0.12	0.13	0.12	0.10	0.04	0.08	0.06
var_duration_downtime_weekday	-0.12	-0.05	-0.07	-0.15	-0.12	-0.01	-0.11
avg_usage_time_6h	0.12	0.19	0.07	0.09	0.11	0.02	0.11
app_usage_Finance_perc_midday	-0.12	-0.12	-0.13	-0.07	-0.14	-0.18	-0.07
app_usage_Transportation_perc_midday	0.12	0.15	0.07	0.07	0.13	0.04	0.10
app_usage_Health...Fitness_perc_evening	0.12	0.01	0.08	0.11	0.17	0.03	0.16
perc_Business	-0.12	-0.02	-0.02	-0.09	-0.06	-0.19	-0.06
download_count..1.000.000.000...5.000.000.000.	-0.11	-0.13	-0.10	-0.07	-0.10	-0.14	0.02
usage_count_4h	0.11	0.04	0.10	0.11	0.03	0.11	0.09
app_usage_News...Magazines_perc_morning	-0.11	-0.12	0.00	-0.10	-0.04	-0.11	-0.09

app_usage_News...Magazines_perc_night	-0.11	-0.06	-0.03	-0.06	-0.02	-0.05	-0.14
avg_uses_perday_week_Business	-0.11	-0.04	0.06	-0.07	-0.05	-0.27	-0.06
avg_uses_perday_Puzzle	0.11	0.07	0.13	0.08	0.05	0.10	0.03
avg_uses_perday_Trivia	-0.11	-0.04	-0.15	-0.07	-0.06	-0.07	-0.11
avg_uses_perday_Arcade	0.11	0.04	0.07	0.03	0.09	0.12	0.11
perc_Media...Video	-0.11	-0.10	-0.09	-0.02	-0.07	-0.03	-0.08
var_incoming_sms_leng	-0.10	-0.00	-0.05	-0.07	-0.10	-0.10	-0.13
gps_data_available	-0.10	-0.02	-0.16	-0.07	-0.02	-0.13	-0.13
avg_charge_connected	-0.10	-0.08	-0.05	-0.08	0.02	-0.11	-0.08
number_games_puzzle_apps	0.10	0.08	0.17	0.08	0.05	0.04	0.03
usage_count_7h	0.10	0.08	0.02	0.13	0.02	0.02	0.16
app_usage_Entertainment_perc_midday	-0.10	-0.10	-0.07	-0.02	-0.08	-0.18	-0.03
app_usage_Photography_perc_evening	-0.10	-0.00	-0.04	-0.08	-0.11	-0.13	-0.08
ratio_incoming_outgoing_sms	-0.10	-0.16	-0.10	-0.16	0.01	0.06	-0.09
total_number_contacts_end	0.09	-0.09	0.06	0.15	0.01	0.00	0.11
total_number_contacts_with_two_numbers	0.09	0.05	0.12	0.05	0.18	-0.02	0.03
number_games_board_apps	0.09	-0.00	0.10	0.06	0.05	-0.01	0.09
number_antivirus_and_security_apps	-0.09	-0.17	0.04	-0.04	-0.02	-0.12	-0.08
app_usage_Unknown_perc_midday	-0.09	-0.01	-0.13	-0.08	-0.11	-0.12	-0.05
app_usage_Music...Audio_perc_evening	-0.09	0.02	-0.07	-0.04	-0.10	-0.03	-0.07
app_usage_Weather_perc_evening	-0.09	-0.09	-0.14	-0.03	-0.10	-0.04	-0.02
app_usage_Media...Video_perc_morning	-0.09	-0.02	-0.07	-0.03	-0.12	-0.02	-0.05
app_usage_Social_perc_midday	-0.09	-0.03	-0.08	-0.07	-0.06	-0.30	-0.02
perc_Books...Reference	-0.09	-0.08	0.02	-0.09	-0.04	-0.13	-0.04
var_duration_calls	-0.08	-0.09	0.04	-0.05	-0.03	-0.08	-0.13
regularity_last_event_weekend	0.08	0.12	0.07	0.06	-0.02	0.04	0.11
avg_number_videos_taken_weekend	0.08	-0.11	0.01	0.11	0.06	0.00	0.11
download_count..5.000.000...10.000.000.	0.08	0.01	0.07	0.10	0.11	-0.05	0.07
avg_usage_time_10h	-0.08	0.00	0.09	-0.06	-0.03	-0.08	-0.16
app_usage_Tools_perc_morning	-0.08	-0.14	-0.05	-0.05	-0.14	0.01	-0.01
app_usage_Games_perc_midday	0.08	0.05	-0.07	0.07	0.06	-0.06	0.13
app_usage_Unknown_perc_evening	-0.08	-0.06	-0.08	-0.07	-0.10	-0.07	-0.04
perc_Photography	0.08	-0.03	0.15	0.08	-0.10	0.10	0.09
total_number_contacts_with_one_number	0.07	0.07	0.10	0.08	-0.01	-0.01	0.09
entropy_of_contact_missed_calls	-0.07	-0.01	-0.01	-0.05	-0.05	-0.16	-0.11
entropy_of_contact_missed_calls_weekend	-0.07	-0.00	-0.04	-0.06	-0.06	-0.13	-0.09
entropy_of_contact_outgoing_sms_weekday	-0.07	0.10	0.01	0.01	-0.19	-0.25	-0.07
response_rate_calls_weekday	-0.07	-0.01	0.04	-0.08	-0.03	-0.04	-0.12
var_last_event_weekday	-0.07	-0.07	-0.01	-0.12	-0.03	0.02	-0.10
number_nights_less_than_4_hours_downtime	-0.07	-0.09	-0.13	-0.04	-0.09	-0.03	-0.01
avg_number_videos_taken_weekdays	-0.07	-0.22	-0.06	-0.04	-0.05	0.06	-0.02
avg_charge_disconnected	-0.07	-0.01	-0.07	-0.09	0.02	0.04	-0.11
number_music_audio_apps	-0.07	-0.06	-0.02	-0.06	0.02	-0.13	-0.02

number_education_apps	-0.07	0.06	-0.06	-0.10	-0.01	-0.18	-0.07
avg_usage_time_19h	0.07	0.14	0.09	0.07	0.09	0.03	0.01
usage_count_6h	0.07	0.05	0.03	0.07	-0.07	0.16	0.12
app_usage_Business_perc_midday	-0.07	0.07	-0.00	-0.08	-0.02	-0.14	-0.01
app_usage_Weather_perc_morning	-0.07	-0.12	-0.19	-0.01	-0.02	-0.01	-0.04
avg_uses_perday_week_Weather	-0.07	-0.13	-0.13	-0.02	-0.06	0.01	-0.01
avg_leng_incoming_sms	0.06	-0.00	-0.01	0.06	0.04	0.11	0.06
var_duration_calls_weekend	-0.06	-0.08	0.02	-0.05	-0.04	-0.07	-0.11
entropy_of_contact_sms_weekday	-0.06	0.05	-0.02	-0.00	-0.15	-0.17	-0.05
var_first_event_weekday	-0.06	-0.04	-0.05	-0.06	-0.11	0.01	-0.06
regularity_last_event_all	-0.06	0.08	-0.02	-0.06	-0.20	0.06	-0.07
regularity_first_event_weekday	-0.06	-0.07	0.00	-0.11	-0.06	0.06	-0.04
app_usage_Tools_perc_night	0.06	-0.05	0.07	0.11	-0.01	0.08	0.00
app_usage_Entertainment_perc_morning	-0.06	-0.10	-0.14	-0.03	-0.01	-0.13	0.03
app_usage_Entertainment_perc_night	0.06	0.01	-0.04	0.10	0.10	0.02	0.00
app_usage_Personalization_perc_morning	0.06	0.08	0.01	0.03	0.09	0.11	0.06
app_usage_Shopping_perc_midday	-0.06	-0.06	-0.12	-0.01	-0.05	-0.22	-0.03
app_usage_Books...Reference_perc_morning	-0.06	-0.08	-0.01	-0.00	-0.07	-0.17	0.03
app_usage_Medical_perc_midday	0.06	0.18	0.12	0.01	0.09	-0.09	0.09
app_usage_Education_perc_night	0.06	-0.08	0.01	0.08	0.10	-0.02	0.04
app_usage_Lifestyle_perc_evening	0.06	-0.00	0.02	0.08	0.06	-0.04	0.08
app_usage_Health...Fitness_perc_morning	0.06	0.00	-0.00	0.04	0.09	-0.05	0.12
app_usage_Media...Video_perc_midday	0.06	0.05	-0.03	0.12	0.04	-0.02	0.08
ratio_incoming_outgoing_calls_weekend	-0.06	-0.07	0.06	-0.10	-0.03	0.08	-0.16
usage_News...Magazines_apps	-0.06	-0.04	0.04	-0.06	0.03	-0.13	-0.09
usage_Health...Fitness_apps	0.06	-0.00	-0.04	0.06	0.10	-0.07	0.12
avg_uses_perday_end_Education	-0.06	-0.03	-0.12	-0.08	-0.06	-0.04	-0.04
avg_usage_time_day_Tools	-0.06	0.07	-0.03	-0.10	-0.00	-0.14	0.02
avg_usage_time_day_Travel...Local	0.06	0.10	0.12	-0.03	0.11	-0.03	0.05
var_outgoing_sms_leng	0.05	0.09	-0.01	0.10	-0.02	-0.06	0.07
response_rate_sms	-0.05	0.11	0.03	-0.00	-0.14	-0.15	-0.03
percentage_of_songs_listened_between_0_6	-0.05	-0.05	-0.03	-0.00	-0.03	-0.08	-0.02
download_count..10.000...50.000.	-0.05	-0.08	-0.03	-0.06	0.08	-0.16	-0.02
avg_usage_time_1h	0.05	0.14	0.17	0.05	0.03	-0.04	-0.00
avg_usage_time_2h	0.05	0.05	0.01	0.04	0.06	0.08	-0.04
avg_usage_time_7h	-0.05	0.01	0.07	-0.05	0.04	-0.09	-0.09
app_usage_Productivity_perc_morning	-0.05	-0.06	0.07	-0.08	-0.02	-0.11	-0.04
app_usage_Productivity_perc_midday	0.05	0.05	-0.12	0.05	0.08	0.03	0.12
app_usage_Shopping_perc_morning	-0.05	-0.07	-0.11	0.01	-0.03	-0.24	0.04
app_usage_Books...Reference_perc_night	0.05	0.01	0.05	0.02	0.05	-0.04	0.11
app_usage_Travel...Local_perc_night	-0.05	-0.10	0.04	-0.02	-0.09	-0.17	-0.02
app_usage_Education_perc_morning	-0.05	0.04	-0.08	-0.03	-0.13	-0.16	-0.02
app_usage_Lifestyle_perc_night	0.05	-0.05	-0.01	0.07	0.08	0.01	0.10

app_usage_Browser_perc_midday	0.05	0.06	0.02	0.01	0.02	-0.01	0.12
avg_uses_perday_Lifestyle	0.05	-0.04	0.01	0.07	0.08	-0.03	0.06
perc_Unknown	-0.05	-0.05	-0.06	-0.01	-0.09	-0.10	-0.01
total_number_missed_calls	-0.04	0.07	-0.01	-0.03	-0.08	-0.04	-0.04
var_duration_incoming_calls	-0.04	-0.04	0.05	-0.03	-0.03	-0.06	-0.10
ratio_avg_duration_incoming_outgoing_calls	-0.04	-0.07	0.04	-0.04	-0.04	0.02	-0.07
percentage_of_songs_listened_between_12_18	-0.04	-0.06	-0.05	-0.01	0.09	-0.21	-0.02
number_photography_apps	-0.04	0.01	-0.03	0.00	-0.05	-0.14	-0.00
number_tools_apps	0.04	0.01	0.03	0.08	0.04	-0.01	0.06
number_finance_apps	-0.04	-0.03	-0.04	-0.02	-0.01	-0.09	-0.04
regularity_all_aggr_events	0.04	0.01	-0.05	0.06	-0.02	0.01	0.14
download_count..5.000...10.000.	-0.04	-0.06	-0.07	-0.04	0.07	-0.08	-0.07
number_apps_searchengine_used	0.04	0.01	0.10	0.00	0.08	-0.00	0.07
app_usage_Productivity_perc_evening	-0.04	0.05	0.05	0.02	-0.18	0.01	-0.10
app_usage_News...Magazines_perc_midday	-0.04	-0.03	0.05	-0.04	0.00	-0.11	-0.04
app_usage_Photography_perc_night	0.04	-0.07	0.01	0.05	-0.05	-0.02	0.06
app_usage_Communication_perc_evening	-0.04	0.08	-0.05	0.01	-0.08	-0.08	-0.04
app_usage_Music...Audio_perc_night	-0.04	-0.08	-0.04	0.02	-0.06	-0.08	0.06
app_usage_Browser_perc_night	0.04	-0.08	-0.01	0.08	0.03	0.02	0.03
app_usage_Health...Fitness_perc_midday	0.04	-0.03	-0.05	0.05	0.12	-0.02	0.10
app_usage_Media...Video_perc_night	0.04	-0.06	0.09	0.10	0.01	0.07	0.03
variance_number_incoming_calls_perday	0.04	0.03	0.09	0.06	-0.01	-0.00	-0.01
ratio_incoming_outgoing_calls_weekday	-0.04	-0.08	0.09	-0.08	0.00	0.08	-0.10
usage_Weather_apps	-0.04	-0.07	-0.11	0.02	-0.05	0.01	0.03
avg_uses_perday_week_Travel...Local	-0.04	-0.03	0.01	-0.05	0.04	-0.15	-0.03
avg_usage_time_day_Communication	0.04	0.10	0.03	0.11	-0.05	-0.11	0.08
perc_News...Magazines	-0.04	-0.04	0.01	-0.04	0.04	-0.05	-0.06
perc_Tools	0.04	-0.01	-0.01	0.01	0.08	0.07	0.09
perc_Browser	0.04	-0.11	-0.01	0.06	0.06	0.14	0.02
total_duration_calls	0.03	0.03	0.08	0.05	-0.03	-0.06	0.01
total_number_added_contacts	0.03	-0.13	0.01	0.10	-0.05	-0.03	0.08
total_number_unique_contacts_who_called	0.03	0.02	0.12	-0.01	0.07	0.04	-0.04
response_rate_missed_call_answer_with_sms	-0.03	0.11	0.02	0.01	-0.07	-0.08	-0.03
avg_time_first_event_sunday	-0.03	-0.00	0.04	-0.04	-0.09	0.01	-0.08
avg_number_charge_connected_per_day	0.03	0.03	0.04	0.06	-0.01	-0.17	0.07
number_checking_behaviour_events	0.03	0.02	-0.05	0.07	-0.06	-0.10	0.13
number_songs_listened_per_day	0.03	0.02	-0.00	0.04	0.09	-0.02	0.04
percentage_of_songs_listened_between_18_24	-0.03	0.01	-0.04	0.05	-0.07	-0.02	0.01
number_sports_apps	-0.03	0.01	-0.11	-0.07	-0.00	-0.02	-0.01
number_apps_messenger_used	0.03	-0.00	-0.03	0.10	-0.09	-0.08	0.10
avg_usage_time_5h	0.03	0.09	0.10	0.02	-0.06	0.05	0.04
app_usage_Photography_perc_morning	0.03	0.02	-0.08	-0.00	0.05	-0.05	0.09
app_usage_Shopping_perc_evening	0.03	0.02	-0.03	0.06	0.00	-0.08	0.04

app_usage_Books...Reference_perc_evening	0.03	0.07	0.07	0.02	-0.01	-0.06	0.09
app_usage_Travel...Local_perc_midday	0.03	0.08	0.09	-0.00	0.05	-0.03	0.00
app_usage_Lifestyle_perc_morning	0.03	-0.04	0.03	0.04	0.06	-0.03	0.04
app_usage_Browser_perc_morning	-0.03	-0.11	-0.10	-0.01	-0.00	0.10	0.00
ratio_betw._avg_number_in_calls_perweek_e_e	0.03	0.05	0.07	0.02	0.10	-0.07	0.02
number_radio_usage	-0.03	-0.09	-0.01	-0.01	-0.01	-0.07	-0.05
avg_uses_perday_end_Business	-0.03	-0.03	-0.01	0.02	-0.02	-0.16	0.03
perc_Communication	-0.03	0.07	-0.00	-0.01	-0.06	-0.05	-0.05
perc_Entertainment	-0.03	-0.09	-0.09	0.02	0.10	-0.15	-0.01
total_duration_incoming_calls	0.02	0.01	0.09	0.05	-0.03	-0.00	-0.02
avg_duration_incoming_calls_weekend	0.02	0.01	0.11	-0.01	0.00	0.01	-0.05
avg_duration_outgoing_calls_weekend	0.02	0.02	0.07	-0.02	0.02	-0.05	-0.01
entropy_of_contact_outgoing_sms_weekend	0.02	0.10	0.08	0.02	-0.06	-0.14	0.02
ratio_number_apps_inst_apps_used	-0.02	0.12	0.06	-0.05	0.01	-0.01	-0.09
total_events_airplaine_db	0.02	-0.03	0.06	0.02	0.04	-0.02	0.02
entropy_music_genres_morning	-0.02	-0.04	-0.09	0.01	0.06	-0.14	0.04
number_books_and_reference_apps	0.02	0.07	0.13	0.01	0.05	-0.07	0.01
avg_plusone_scores	0.02	0.04	0.01	0.00	-0.09	0.05	0.04
app_usage_Tools_perc_midday	0.02	0.07	-0.00	0.00	0.06	-0.14	0.06
app_usage_Communication_perc_midday	0.02	0.01	0.12	-0.04	0.05	-0.03	0.03
app_usage_Music...Audio_perc_morning	-0.02	-0.08	-0.15	0.05	0.11	-0.01	-0.01
app_usage_Sports_perc_evening	0.02	0.00	-0.12	0.02	0.00	0.03	0.04
app_usage_Sports_perc_night	-0.02	-0.02	-0.09	-0.03	-0.02	-0.04	0.00
app_usage_Media...Video_perc_evening	-0.02	0.03	-0.03	0.08	-0.10	0.01	-0.02
app_usage_Social_perc_evening	-0.02	-0.00	0.09	0.02	-0.08	-0.15	0.03
avg_uses_perday_week_Tools	0.02	0.06	-0.01	0.05	0.02	-0.07	0.10
avg_uses_perday_week_Education	0.02	0.11	-0.03	0.02	-0.02	-0.12	0.02
perc_Productivity	0.02	-0.10	0.10	0.02	0.10	-0.01	-0.03
perc_Shopping	0.02	0.01	-0.04	0.06	0.03	-0.08	0.03
avg_completeness_score_contacts	0.01	-0.01	0.05	0.03	0.07	-0.09	-0.02
avg_time_last_event_weekday	-0.01	-0.02	-0.05	0.03	0.00	-0.08	-0.02
var_last_event_weekend	0.01	0.06	0.12	-0.06	0.06	-0.02	-0.04
download_count..50.000...100.000.	0.01	0.11	-0.01	0.03	0.01	-0.10	0.03
avg_usage_time_8h	0.01	0.11	0.07	-0.00	0.10	0.00	-0.08
avg_usage_time_0h	-0.01	0.05	0.04	0.05	-0.08	-0.07	-0.02
app_usage_Tools_perc_evening	-0.01	0.12	-0.06	0.01	-0.02	0.03	-0.03
app_usage_Entertainment_perc_evening	-0.01	-0.08	-0.07	0.04	-0.02	-0.03	0.05
app_usage_Productivity_perc_night	0.01	0.01	0.05	0.00	0.02	-0.05	-0.02
app_usage_Books...Reference_perc_midday	0.01	0.07	0.09	-0.03	0.01	-0.07	0.05
app_usage_Travel...Local_perc_morning	0.01	-0.05	0.01	-0.02	-0.00	-0.02	0.03
app_usage_Travel...Local_perc_evening	-0.01	0.06	-0.05	0.02	0.02	-0.07	0.00
app_usage_Education_perc_midday	0.01	0.01	-0.06	-0.00	0.03	-0.07	0.03
app_usage_Education_perc_evening	-0.01	0.00	-0.10	-0.01	-0.00	-0.04	0.02

app_usage_Weather_perc_night	-0.01	-0.13	-0.01	0.06	-0.02	-0.00	0.03
app_usage_Browser_perc_evening	-0.01	0.09	-0.00	0.01	-0.01	0.04	-0.09
app_usage_Health...Fitness_perc_night	0.01	-0.01	-0.04	0.01	0.04	-0.07	0.03
number_shazam_apps_used	0.01	-0.16	-0.05	-0.00	0.09	-0.05	0.09
avg_uses_perday_week_Games	0.01	0.05	-0.03	-0.00	0.02	0.02	-0.02
avg_uses_perday_end_Photography	0.01	-0.08	0.06	0.06	-0.07	-0.11	0.07
avg_usage_time_day_Music...Audio	-0.01	-0.12	0.03	0.01	-0.00	-0.09	0.03
perc_Lifestyle	-0.01	-0.11	-0.00	0.01	0.05	-0.08	0.04
perc_Medical	-0.01	0.12	0.04	-0.05	0.06	-0.10	0.01
perc_Music...Audio	-0.01	-0.09	-0.09	0.02	0.07	-0.07	0.02
perc_Social	0.01	-0.01	0.07	0.03	-0.10	-0.11	0.09
perc_Sports	0.01	0.02	-0.13	-0.00	0.03	0.00	0.02
total_number_contacts_with_mail	0.00	-0.03	0.09	0.01	0.02	-0.06	0.05
number_nights_more_than_7_hours_downtime	-0.00	0.00	0.07	-0.03	0.06	-0.05	-0.05
avg_inter_event_time_weekend	0.00	0.01	-0.05	-0.01	0.06	0.07	-0.04
percentage_of_songs_listened_between_6_12	0.00	0.01	-0.08	-0.02	0.12	-0.00	0.07
number_games_racing_apps	0.00	-0.01	-0.03	-0.01	0.04	0.04	0.02
usage_count_0h	0.00	-0.06	-0.07	0.05	-0.02	-0.07	0.05
app_usage_Games_perc_morning	-0.00	0.01	-0.06	0.02	-0.06	0.01	0.04
app_usage_Games_perc_night	0.00	0.07	-0.02	0.01	0.01	-0.06	0.03
app_usage_Communication_perc_morning	0.00	-0.03	-0.03	-0.01	0.03	0.05	0.01
app_usage_Music...Audio_perc_midday	0.00	0.03	-0.05	0.01	0.10	-0.14	0.01
app_usage_Business_perc_evening	-0.00	0.05	-0.01	0.05	0.02	-0.06	0.03
app_usage_Social_perc_morning	-0.00	-0.06	0.06	0.01	-0.01	-0.09	0.03
app_usage_Social_perc_night	-0.00	-0.06	0.09	0.02	-0.04	-0.03	-0.01
avg_usage_time_day_Entertainment	-0.00	-0.03	-0.08	0.00	0.11	-0.13	0.09

Note: Pairwise Spearman correlations between Agreeableness (factor, facets) and predictor variables from Section 2.3; table is sorted by absolute ρ values of Agreeableness, in decreasing order. Abbreviations: A1 = Willingness to trust, A2 = Genuineness, A3 = Helpfulness, A4 = Obligingness, A5 = Modesty, A6 = Good Naturedness.

Table 7: Pairwise Spearman Correlations Between Emotional Stability and Predictors Study 3

Predictors	Emotional Stability	ES1	ES2	ES3	ES4	ES5	ES6
app_usage_Photography_perc_night	-0.22	-0.17	-0.21	-0.22	0.00	-0.24	-0.14
ratio_avg_duration_incoming_outgoing_calls	-0.21	-0.22	-0.15	-0.22	-0.14	-0.11	-0.15
response_rate_calls_weekday	0.20	0.22	0.14	0.23	0.15	-0.05	0.11
app_usage_Communication_perc_midday	0.20	0.23	0.14	0.21	0.15	0.01	0.07
app_usage_Media...Video_perc_midday	-0.20	-0.18	-0.23	-0.16	-0.04	-0.13	-0.06
app_usage_Media...Video_perc_evening	-0.20	-0.21	-0.28	-0.11	0.00	-0.18	-0.08
app_usage_Transportation_perc_midday	0.19	0.21	0.12	0.09	0.14	0.16	0.12
app_usage_Transportation_perc_night	0.19	0.22	0.15	0.11	0.13	-0.01	0.12
app_usage_Browser_perc_night	-0.19	-0.08	-0.13	-0.20	-0.15	-0.15	-0.15
perc_Media...Video	-0.19	-0.16	-0.19	-0.21	-0.09	-0.11	-0.03
avg_usage_time_0h	-0.18	-0.11	-0.22	-0.14	-0.03	-0.28	-0.09
app_usage_Business_perc_night	-0.18	-0.17	-0.14	-0.21	-0.06	0.03	-0.19
avg_uses_perday_end_Transportation	0.18	0.15	0.10	0.11	0.17	0.04	0.14
avg_usage_time_1h	-0.17	-0.12	-0.19	-0.16	-0.01	-0.17	-0.11
percent_sms_night	-0.16	-0.12	-0.18	-0.14	0.02	-0.18	-0.18
number_business_apps	-0.16	-0.11	-0.17	-0.14	-0.05	-0.11	-0.05
total_number_shared_photos	-0.16	-0.17	-0.19	-0.10	-0.07	0.00	-0.17
total_number_contacts_with_two_numbers	0.15	0.18	0.10	0.13	0.23	-0.16	0.13
avg_completeness_score_contacts	0.15	0.17	0.19	0.10	0.00	-0.07	0.18
app_usage_Productivity_perc_midday	0.15	0.13	0.18	0.12	0.18	0.03	0.07
app_usage_Transportation_perc_morning	0.15	0.10	0.13	0.05	0.09	0.23	0.17
number_radio_usage	0.15	0.15	0.05	0.05	0.18	0.03	0.18
perc_Transportation	0.15	0.17	0.09	0.05	0.13	0.12	0.08
avg_leng_outgoing_sms	-0.14	-0.12	-0.14	-0.15	-0.15	0.09	-0.11
avg_time_last_event_weekday	-0.14	-0.02	-0.15	-0.09	0.01	-0.29	-0.08
percentage_of_songs_listened_between_6_12	0.14	0.16	0.13	0.12	0.10	-0.08	0.10
number_apps_searchengine_used	0.14	0.21	0.09	0.13	0.13	0.09	0.02
usage_count_6h	-0.14	-0.14	-0.14	-0.16	-0.07	-0.02	-0.14
app_usage_Media...Video_perc_morning	-0.14	-0.11	-0.16	-0.19	-0.08	-0.07	0.01
number_checking_behaviour_events	-0.13	-0.09	-0.21	-0.02	0.03	-0.11	-0.15
regularity_all_aggr_events	-0.13	-0.07	-0.20	-0.06	0.11	-0.24	-0.15
avg_usage_time_2h	-0.13	-0.01	-0.16	-0.12	-0.06	-0.19	-0.05
app_usage_Unknown_perc_morning	-0.13	-0.14	-0.13	-0.08	-0.07	-0.02	-0.03
app_usage_Travel...Local_perc_evening	0.13	0.18	0.12	0.06	0.15	-0.01	0.04
app_usage_Transportation_perc_evening	0.13	0.11	0.11	0.11	0.03	0.08	0.05
app_usage_Browser_perc_evening	-0.13	-0.11	-0.12	-0.10	-0.05	-0.06	-0.14
perc_Photography	-0.13	-0.10	-0.14	-0.13	-0.11	0.01	-0.12
total_number_unique_contacts_who_called	0.12	0.13	0.10	0.16	0.11	-0.06	0.06
regularity_last_event_weekday	-0.12	-0.08	-0.06	-0.20	-0.11	-0.17	-0.09

number_games_puzzle_apps	0.12	0.09	0.11	0.10	0.05	-0.06	0.12
usage_count_0h	-0.12	-0.06	-0.15	-0.07	0.04	-0.26	-0.07
app_usage_Tools_perc_night	-0.12	-0.04	-0.08	-0.12	-0.03	-0.22	-0.07
app_usage_Communication_perc_evening	-0.12	-0.06	-0.10	-0.12	-0.03	-0.09	-0.08
avg_uses_perday_week_Transportation	0.12	0.10	0.08	0.09	0.15	0.03	0.03
avg_uses_perday_end_Business	-0.12	-0.08	-0.17	-0.15	-0.03	-0.04	-0.07
avg_usage_time_day_Entertainment	0.12	0.13	0.09	0.08	0.18	0.05	0.02
avg_duration_outgoing_calls_weekend	0.11	0.13	-0.01	0.15	0.13	0.04	0.07
percent_calls_night	-0.11	-0.06	-0.10	-0.09	0.10	-0.24	-0.10
number_nights_less_than_4_hours_downtime	-0.11	-0.05	-0.18	-0.04	0.00	-0.20	0.01
avg_number_videos_taken_weekend	-0.11	-0.10	-0.08	-0.14	-0.03	-0.11	-0.05
number_songs_listened_per_day	0.11	0.15	0.02	0.14	0.13	-0.15	0.04
number_books_and_reference_apps	0.11	0.18	0.08	0.13	0.08	-0.10	0.05
app_usage_Personalization_perc_morning	0.11	0.13	0.20	0.15	0.03	-0.19	0.12
app_usage_Travel...Local_perc_morning	-0.11	-0.13	-0.11	-0.07	-0.08	-0.04	-0.03
app_usage_Travel...Local_perc_night	-0.11	-0.08	-0.13	-0.08	0.01	-0.10	-0.14
app_usage_Music...Audio_perc_morning	0.11	0.10	0.09	0.12	0.08	-0.07	0.10
app_usage_Browser_perc_midday	0.11	0.11	0.15	0.12	0.13	-0.05	0.03
app_usage_Health...Fitness_perc_midday	-0.11	-0.06	-0.12	-0.08	-0.00	-0.19	-0.06
entropy_of_contact_missed_calls_weekend	-0.10	-0.08	-0.14	-0.04	0.09	-0.19	-0.04
gps_data_available	-0.10	-0.04	-0.15	-0.06	-0.01	-0.14	-0.04
var_first_event_weekday	-0.10	-0.05	-0.04	-0.19	-0.12	-0.11	-0.08
number_finance_apps	0.10	0.11	0.04	0.10	0.13	-0.09	0.14
app_usage_Productivity_perc_evening	-0.10	-0.07	-0.05	-0.10	-0.03	-0.10	-0.11
app_usage_Books...Reference_perc_night	-0.10	-0.04	-0.09	-0.01	-0.05	-0.26	-0.09
app_usage_Weather_perc_evening	-0.10	-0.04	-0.12	-0.19	-0.08	0.05	-0.01
avg_uses_perday_Trivia	0.10	0.12	0.09	0.08	-0.04	0.15	0.11
perc_Business	-0.10	-0.07	-0.13	-0.08	0.02	-0.11	-0.04
perc_Productivity	0.10	0.05	0.16	0.03	-0.06	-0.01	0.19
total_number_contacts_with_one_number	0.09	0.11	-0.05	0.06	0.31	-0.08	0.06
total_number_unique_contacts_outgoing_sms	-0.09	-0.08	-0.08	-0.12	-0.04	-0.08	0.00
entropy_of_contact_missed_calls	0.09	0.05	0.04	0.12	0.15	-0.02	0.08
regularity_first_event_weekday	0.09	0.05	0.09	0.04	0.02	0.15	0.06
avg_number_videos_taken_weekdays	-0.09	-0.07	-0.05	-0.10	-0.02	-0.16	-0.03
entropy_music_genres_morning	0.09	0.13	0.02	0.05	0.18	-0.10	0.12
download_count..50.000...100.000.	0.09	0.15	0.06	0.13	0.00	-0.16	0.19
app_usage_Productivity_perc_night	-0.09	-0.00	-0.09	-0.07	0.04	-0.29	-0.02
app_usage_News...Magazines_perc_morning	0.09	0.10	0.02	0.05	0.04	0.05	0.15
app_usage_News...Magazines_perc_midday	0.09	0.13	0.06	0.04	0.09	-0.03	0.09
app_usage_Unknown_perc_night	-0.09	-0.07	-0.09	-0.04	-0.04	-0.07	-0.01
app_usage_Books...Reference_perc_evening	-0.09	-0.09	-0.01	-0.01	-0.05	-0.17	-0.12
ratio_betw._avg_num_in_calls_perweek_d_e	0.09	0.07	0.04	0.15	0.14	0.03	0.01
avg_uses_perday_end_Education	-0.09	-0.02	-0.05	-0.13	-0.17	-0.04	-0.07

perc_Books...Reference	-0.09	-0.05	-0.05	-0.07	-0.05	-0.09	-0.07
total_number_missed_calls	-0.08	-0.05	-0.13	0.01	0.04	-0.16	-0.09
number_nights_more_than_7_hours_downtime	0.08	0.03	0.16	0.04	-0.03	0.18	0.02
number_battery_saver_task_killer_apps	0.08	0.04	0.11	0.11	0.07	-0.12	0.08
avg_plusone_scores	-0.08	-0.13	-0.09	-0.08	-0.09	0.19	-0.11
download_count..10.000...50.000.	0.08	0.10	0.01	0.08	0.15	-0.13	0.08
number_apps_messenger_used	-0.08	-0.05	-0.17	-0.06	0.11	-0.15	-0.05
avg_usage_time_5h	-0.08	0.04	-0.12	-0.04	0.08	-0.26	-0.08
avg_usage_time_7h	-0.08	-0.11	-0.08	-0.07	-0.01	-0.03	-0.07
app_usage_Games_perc_midday	0.08	0.06	0.10	0.05	0.01	-0.03	0.11
app_usage_Photography_perc_midday	0.08	0.03	0.05	0.06	0.07	0.20	0.07
app_usage_Education_perc_morning	-0.08	-0.06	-0.10	-0.08	-0.07	-0.08	-0.02
app_usage_Education_perc_evening	-0.08	-0.02	-0.05	-0.07	-0.16	-0.03	-0.04
app_usage_Health...Fitness_perc_night	-0.08	-0.03	-0.08	-0.04	-0.04	-0.16	0.00
app_usage_Social_perc_evening	-0.08	-0.08	-0.13	-0.07	0.03	-0.10	0.04
app_usage_Social_perc_night	-0.08	-0.02	-0.11	-0.09	-0.02	-0.17	0.05
ratio_incoming_outgoing_sms	-0.08	-0.06	-0.04	-0.05	-0.08	0.02	-0.08
perc_Music...Audio	0.08	0.16	0.03	0.03	0.00	-0.11	0.18
avg_leng_incoming_sms	-0.07	-0.05	-0.04	-0.13	-0.04	0.06	-0.09
total_number_added_contacts	-0.07	-0.11	-0.03	-0.11	-0.00	-0.12	0.02
entropy_of_contact_sms_weekday	0.07	0.09	0.06	0.07	0.08	-0.06	0.07
response_rate_missed_call_answer_with_sms	0.07	0.09	0.06	0.09	0.06	-0.11	0.04
var_first_event_weekend	0.07	0.10	-0.00	0.09	0.13	-0.14	0.04
var_last_event_weekend	0.07	0.11	0.00	0.03	0.26	-0.07	0.00
total_events_boot_db	-0.07	-0.07	-0.11	0.00	-0.01	-0.04	-0.13
percentage_of_songs_listened_between_0_6	-0.07	-0.04	-0.10	0.02	0.07	-0.26	-0.05
number_tools_apps	-0.07	-0.04	-0.10	-0.08	-0.03	0.01	-0.06
download_count..1.000.000.000...5.000.000.000.	-0.07	0.02	-0.05	-0.07	-0.01	-0.20	-0.00
app_usage_Music...Audio_perc_night	-0.07	-0.04	-0.14	-0.02	0.10	-0.21	0.01
app_usage_Weather_perc_morning	-0.07	-0.02	-0.13	-0.11	-0.09	0.10	0.02
app_usage_Sports_perc_evening	-0.07	-0.08	-0.05	-0.09	-0.06	-0.04	-0.03
avg_uses_perday_week_Travel...Local	-0.07	-0.02	-0.12	-0.00	0.16	-0.26	-0.03
avg_uses_perday_end_Photography	-0.07	-0.00	-0.18	-0.09	0.00	-0.07	0.08
avg_usage_time_day_Travel...Local	0.07	0.05	0.03	0.16	0.14	-0.06	-0.03
total_number_contacts_with_mail	0.06	0.06	0.11	0.02	0.09	-0.07	0.08
avg_time_first_event_sunday	-0.06	-0.03	0.06	-0.14	-0.04	-0.03	-0.07
regularity_last_event_all	-0.06	-0.05	-0.06	-0.14	-0.04	-0.10	-0.04
number_photography_apps	-0.06	0.00	-0.07	-0.10	-0.02	-0.08	0.04
number_games_board_apps	-0.06	-0.03	-0.02	-0.04	0.05	-0.11	-0.16
usage_count_7h	-0.06	-0.11	-0.15	-0.01	-0.02	-0.00	-0.03
app_usage_Tools_perc_midday	0.06	0.07	0.02	0.07	0.16	-0.01	-0.03
app_usage_Entertainment_perc_evening	-0.06	-0.11	0.02	-0.05	-0.04	0.05	-0.11
app_usage_News...Magazines_perc_evening	0.06	0.11	0.02	0.04	0.07	-0.05	0.11

app_usage_Music...Audio_perc_midday	0.06	0.08	0.05	0.09	0.07	-0.16	0.08
app_usage_Music...Audio_perc_evening	0.06	0.10	-0.01	0.06	0.12	-0.06	0.05
app_usage_Business_perc_evening	-0.06	0.02	-0.08	-0.07	0.05	-0.07	-0.09
variance_number_incoming_calls_perday	0.06	0.10	-0.01	0.13	0.14	-0.14	0.01
number_shazam_apps_used	0.06	0.05	0.10	0.08	0.13	-0.13	0.02
avg_uses_perday_week_Games	0.06	0.08	0.08	0.09	-0.01	-0.12	0.10
avg_uses_perday_Arcade	-0.06	-0.02	-0.01	0.03	-0.02	-0.27	-0.11
avg_uses_perday_Casual	0.06	0.05	0.04	0.12	0.08	-0.17	0.02
perc_Shopping	0.06	0.12	-0.02	0.04	0.12	-0.13	0.12
perc_Sports	-0.06	-0.05	-0.04	-0.10	-0.05	0.03	-0.05
perc_Browser	-0.06	-0.12	-0.01	-0.06	-0.08	0.03	-0.04
var_incoming_sms_leng	0.05	0.02	0.04	0.03	0.02	0.00	0.05
bluetooth_used	0.05	0.07	0.04	0.01	0.05	-0.08	0.10
download_count..5.000...10.000.	-0.05	-0.05	-0.01	-0.06	-0.07	-0.06	0.02
avg_usage_time_6h	-0.05	-0.07	-0.01	-0.02	0.01	-0.12	-0.10
app_usage_Entertainment_perc_morning	-0.05	-0.03	-0.00	-0.08	0.04	-0.07	-0.07
app_usage_Photography_perc_morning	-0.05	-0.13	-0.05	-0.01	0.06	-0.05	-0.04
app_usage_Photography_perc_evening	-0.05	0.04	-0.10	-0.02	-0.01	-0.20	0.00
app_usage_Travel...Local_perc_midday	0.05	0.04	0.07	0.09	0.01	-0.05	0.02
app_usage_Lifestyle_perc_morning	0.05	-0.00	-0.01	0.06	0.17	-0.08	0.10
avg_uses_perday_Puzzle	0.05	0.02	0.05	0.07	-0.01	-0.10	0.06
avg_usage_time_day_Communication	-0.05	-0.07	-0.14	0.01	0.22	-0.22	-0.06
perc_Entertainment	0.05	0.09	0.12	-0.02	-0.01	-0.02	0.04
perc_Unknown	-0.05	0.01	-0.05	-0.03	-0.03	-0.15	0.03
var_duration_incoming_calls	-0.04	-0.02	-0.12	0.02	-0.02	-0.04	-0.01
var_outgoing_sms_leng	-0.04	-0.01	-0.04	-0.07	0.01	0.01	-0.05
entropy_of_contact_outgoing_sms_weekend	-0.04	-0.11	-0.06	-0.02	0.09	-0.08	-0.04
regularity_last_event_weekend	0.04	0.06	-0.06	0.05	0.16	-0.15	0.04
number_music_audio_apps	0.04	0.14	0.01	0.00	-0.01	-0.05	0.12
number_sports_apps	-0.04	-0.03	-0.01	-0.04	-0.01	-0.03	-0.07
calendar_apps_used	-0.04	-0.01	-0.01	-0.01	-0.04	-0.22	0.07
download_count..5.000.000...10.000.000.	0.04	0.06	0.07	0.05	0.08	-0.10	0.01
avg_usage_time_8h	-0.04	-0.02	-0.06	0.01	0.03	-0.08	-0.05
app_usage_Entertainment_perc_night	-0.04	0.04	-0.05	-0.05	0.04	-0.20	0.02
app_usage_Productivity_perc_morning	0.04	-0.03	-0.07	0.07	-0.05	0.15	0.10
app_usage_News...Magazines_perc_night	-0.04	0.03	-0.07	-0.06	-0.04	-0.11	0.11
app_usage_Shopping_perc_morning	0.04	0.07	-0.05	0.04	0.10	-0.10	0.11
app_usage_Shopping_perc_evening	0.04	0.11	-0.03	0.01	0.10	-0.11	0.10
app_usage_Books...Reference_perc_morning	-0.04	-0.04	-0.07	0.02	-0.00	-0.07	-0.04
app_usage_Medical_perc_midday	0.04	0.05	-0.02	0.10	0.16	-0.08	-0.02
app_usage_Media...Video_perc_night	-0.04	-0.04	-0.07	-0.05	0.10	-0.15	0.02
ratio_betw._avg_num_calls_perweek_d_e	0.04	0.01	0.04	0.11	-0.01	0.08	-0.03
avg_uses_perday_week_Tools	-0.04	-0.01	-0.06	0.01	0.10	-0.13	-0.07

perc_Social	-0.04	-0.02	-0.05	-0.03	0.01	-0.03	0.01
total_duration_calls	0.03	0.06	-0.05	0.12	0.13	-0.08	-0.03
response_rate_sms	0.03	-0.05	-0.03	0.02	0.10	-0.03	0.02
ratio_number_apps_inst_apps_used	0.03	0.07	0.01	0.03	0.01	0.10	0.02
avg_charge_connected	0.03	0.06	0.05	-0.01	0.00	-0.04	0.12
number_weather_apps	0.03	0.03	0.01	-0.01	0.07	0.19	-0.10
number_games_racing_apps	0.03	0.03	0.07	-0.01	-0.05	0.02	0.11
avg_usage_time_19h	-0.03	-0.03	-0.01	-0.01	0.00	-0.16	-0.00
app_usage_Finance_perc_midday	0.03	0.09	-0.05	0.03	0.06	-0.11	0.13
app_usage_Unknown_perc_midday	-0.03	-0.05	-0.06	0.00	0.04	-0.01	0.01
app_usage_Education_perc_midday	-0.03	0.03	-0.03	0.00	-0.05	-0.08	-0.01
app_usage_Business_perc_midday	-0.03	0.04	-0.07	-0.03	0.06	-0.06	-0.04
app_usage_Sports_perc_night	-0.03	-0.04	-0.01	-0.01	-0.03	-0.11	0.01
app_usage_Health...Fitness_perc_morning	-0.03	0.02	-0.03	0.01	0.03	-0.19	-0.04
app_usage_Social_perc_morning	-0.03	-0.03	-0.02	0.02	-0.05	-0.07	-0.03
ratio_incoming_outgoing_calls_weekday	0.03	0.05	0.08	-0.00	-0.04	-0.01	0.06
ratio_incoming_outgoing_calls_weekend	0.03	0.04	0.08	0.02	-0.11	-0.00	0.08
avg_uses_perday_week_Education	-0.03	-0.02	-0.07	-0.02	-0.09	-0.03	0.02
perc_Communication	0.03	-0.03	0.01	0.10	0.09	0.05	-0.08
total_number_contacts_end	-0.02	-0.04	0.01	-0.04	0.03	-0.09	0.04
entropy_of_contact_outgoing_sms_weekday	0.02	0.01	-0.01	0.01	0.06	-0.07	0.09
response_rate_calls_weekend	0.02	0.05	-0.04	-0.02	0.20	-0.12	0.02
var_duration_downtime_weekday	0.02	0.11	0.02	-0.02	0.05	-0.14	0.01
number_events_during_sleep	-0.02	-0.03	-0.03	-0.07	0.13	-0.02	-0.05
avg_inter_event_time_weekend	-0.02	-0.01	0.01	-0.01	0.05	-0.03	-0.13
percentage_of_songs_listened_between_12_18	0.02	0.04	-0.02	0.04	0.06	-0.11	0.08
usage_count_4h	-0.02	0.01	-0.09	-0.03	0.08	-0.08	-0.03
app_usage_Tools_perc_morning	-0.02	-0.09	-0.11	0.02	0.02	0.00	0.07
app_usage_Tools_perc_evening	-0.02	-0.03	0.06	-0.09	-0.10	0.11	-0.03
app_usage_Games_perc_night	-0.02	0.02	0.00	0.02	0.01	-0.27	0.03
app_usage_Unknown_perc_evening	-0.02	-0.01	-0.03	0.02	0.05	-0.01	-0.02
app_usage_Weather_perc_night	-0.02	0.01	-0.06	-0.07	0.08	0.02	0.02
app_usage_Browser_perc_morning	-0.02	-0.09	-0.01	-0.05	-0.04	0.10	0.01
app_usage_Health...Fitness_perc_evening	-0.02	0.03	-0.02	0.02	0.03	-0.12	-0.03
usage_News...Magazines_apps	0.02	0.11	-0.03	0.01	0.00	-0.10	0.12
usage_Health...Fitness_apps	-0.02	0.04	-0.03	-0.03	0.03	-0.19	0.05
perc_News...Magazines	0.02	0.05	-0.01	-0.00	-0.03	0.07	0.07
avg_duration_incoming_calls_weekend	-0.01	0.04	-0.04	0.04	0.05	-0.08	-0.03
var_duration_calls	0.01	0.01	-0.04	0.07	0.02	-0.05	-0.03
var_duration_calls_weekend	0.01	0.02	-0.06	0.03	0.02	0.05	-0.02
var_last_event_weekday	-0.01	0.08	0.01	0.00	0.03	-0.25	0.01
var_duration_downtime_weekend	0.01	0.02	-0.04	-0.01	0.11	-0.08	-0.00
total_events_airplaine_db	-0.01	-0.04	-0.07	0.02	0.06	-0.06	0.02

avg_number_charge_connected_per_day	-0.01	-0.01	-0.08	0.05	0.09	-0.10	-0.01
number_education_apps	-0.01	0.03	-0.08	-0.03	0.02	-0.11	0.04
number_antivirus_and_security_apps	0.01	-0.01	-0.00	0.03	0.04	0.01	0.03
avg_usage_time_10h	0.01	-0.00	0.01	0.00	0.16	-0.21	0.05
app_usage_Games_perc_morning	0.01	-0.01	0.04	-0.01	-0.03	-0.06	0.06
app_usage_Communication_perc_morning	0.01	-0.11	0.02	0.00	-0.05	0.16	0.07
app_usage_Books...Reference_perc_midday	0.01	0.01	0.02	0.09	-0.02	-0.11	-0.02
app_usage_Education_perc_night	0.01	0.04	0.07	-0.03	-0.10	-0.02	0.07
app_usage_Lifestyle_perc_night	-0.01	-0.02	-0.04	-0.03	0.16	-0.11	-0.00
app_usage_Social_perc_midday	-0.01	0.03	0.03	0.03	0.03	-0.11	-0.04
usage_Weather_apps	0.01	0.03	-0.01	-0.09	-0.01	0.09	0.09
avg_uses_perday_week_Business	-0.01	0.01	-0.11	0.01	0.12	-0.13	0.03
avg_uses_perday_week_Weather	0.01	0.05	-0.02	-0.10	-0.01	0.11	0.09
avg_usage_time_day_Books...Reference	-0.01	-0.01	0.01	-0.08	-0.02	0.04	0.06
avg_usage_time_day_Music...Audio	0.01	0.03	-0.02	-0.01	0.10	-0.19	0.08
perc_Medical	0.01	0.03	-0.06	0.07	0.13	-0.12	-0.02
perc_Tools	0.01	0.03	0.09	-0.05	-0.01	0.10	-0.11
total_duration_incoming_calls	-0.00	0.03	-0.06	0.07	0.07	-0.13	-0.04
avg_charge_disconnected	-0.00	0.02	0.06	-0.13	-0.12	0.10	0.15
percentage_of_songs_listened_between_18_24	-0.00	0.04	-0.11	0.03	0.12	-0.09	-0.03
app_usage_Entertainment_perc_midday	0.00	0.06	-0.02	0.02	0.11	-0.14	-0.01
app_usage_Shopping_perc_midday	-0.00	0.05	-0.08	-0.02	0.04	-0.11	0.10
app_usage_Lifestyle_perc_evening	-0.00	-0.02	-0.04	0.00	0.11	-0.07	0.04
avg_uses_perday_Lifestyle	0.00	0.02	-0.02	-0.01	0.10	-0.09	0.03
avg_usage_time_day_Tools	-0.00	-0.01	-0.01	-0.00	0.13	-0.06	-0.07
perc_Lifestyle	-0.00	-0.01	-0.02	-0.01	0.12	-0.08	0.02

Note: Pairwise Spearman correlations between Emotional stability (factor, facets) and predictor variables from Section 2.3; table is sorted by absolute ρ values of Emotional stability, in decreasing order. Abbreviations: ES1 = Carefreeness, ES2 = Equanimity, ES3 = Positive Mood, ES4 = Self Consciousness, ES5 = Self Control, ES6 = Emotional Robustness.

Bibliography

Abdollahi, M., Valavi, E., & Ahmadi Noubari., H. (2009). Voice-based gender identification via multiresolution frame classification of spectro-temporal maps. In *Proceedings of the International Joint Conference on Neural Networks* (pp. 1–4). doi:10.1109/IJCNN.2009.5178984

Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer International Publishing. doi:10.1007/978-3-319-14142-8

Allport, G. W. & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), i–171. doi:10.1037/h0093360

Aluja, A., Garcia, O., & Garcia, L. F. (2002). A comparative study of Zuckerman's three structural models for personality through the NEO-PI-R, ZKPQ-III-R, EPQ-RS and Goldberg's 50-bipolar adjectives. *Personality and Individual Differences*, *33*(5), 713–725. doi:10.1016/S0191-8869(01)00186-6

Analytics, R. & Weston, S. (2015a). *doMC: Foreach Parallel Adaptor for 'parallel'*. R package version 1.3.4. Retrieved from https://CRAN.R-project.org/package=doMC

Analytics, R. & Weston, S. (2015b). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.10. Retrieved from https://CRAN.R-project.org/package=doParallel

Anand, S., Terken, J., & Hogema, J. (2011). Individual differences in preferred steering effort for steer-by-wire systems. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '11* (p. 55). New York, New York, USA: ACM Press. doi:10.1145/2381416.2381425

Anglim, J. & Grant, S. L. (2014). Incremental criterion prediction of personality facets over factors: Obtaining unbiased estimates and confidence intervals. *Journal of Research in Personality*, *53*, 148–157. doi:10.1016/j.jrp.2014.10.005

APA. (2016). Definition Psychology. Retrieved January 14, 2016, from http://www.apa.org/support/about-apa.aspx?item=7

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

Arendasy, M., Sommer, M., Herle, M., Schützhofer, B., & Inwanschitz, D. (2011, January). Modeling Effects of Faking on an Objective Personality Test. *Journal of Individual Differences*, *32*(4), 210–218. doi:10.1027/1614-0001/a000053

Arya, S., Mount, D., Kemp, S. E., & Jefferis, G. (2015). *RANN: Fast Nearest Neighbour Search (Wraps Arya and Mount's ANN Library)*. R package version 2.5. Retrieved from https://CRAN.R-project.org/package=RANN

Asendorpf, J. B. & Neyer, F. J. (2012). *Psychologie der Persönlichkeit*. Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer.

Ashton, M. C., Paunonen, S. V., & Lee, K. (2014). On the validity of narrow and broad personality traits: A response to Salgado, Moscoso, and Berges (2013). *Personality and Individual Differences*, *56*, 24–28. doi:10.1016/j.paid.2013.08.019

Barr, G. C., Kane, K. E., Barraco, R. D., Rayburg, T., Demers, L., Kraus, C. K., … Kane, B. G. (2015, March). Gender differences in perceptions and self-reported driving behaviors among teenagers. *The Journal of Emergency Medicine*, *48*(3), 366–70.e3. doi:10.1016/j.jemermed.2014.09.055

Barrick, M. R. & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26. doi:10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., Mount, M. K., & Gupta, R. (2003). Meta-Analysis of the Relationship Between the Five-Factor Model of Personality and Holland's Occupational Types. *Personnel Psychology*, *56*(1), 45–74. doi:10.1111/j.1744-6570.2003.tb00143.x

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. doi:10.1111/j.1745-6916.2007.00051.x

Baumgartner, H. & Steenkamp, J.-B. E. (2001). Response Styles in Marketing Research: A Cross-National Investigation. *Journal of Marketing Research*, *38*(2), 143–156. doi:10.1509/jmkr.38.2.143.18840

Bekios-Calfa, J., Buenaposada, J. M., & Baumela, L. (2014). Robust gender recognition by exploiting facial attributes dependencies. *Pattern Recognition Letters*, *36*, 228–234. doi:10.1016/j.patrec.2013.04.028

Ben-Zeev, D., Brenner, C. J., Begale, M., Duffecy, J., Mohr, D. C., & Mueser, K. T. (2014). Feasibility, acceptability, and preliminary efficacy of a smartphone intervention for schizophrenia. *Schizophrenia Bulletin*, *40*(6), 1244–1253. doi:10.1093/schbul/sbu033

Bischl, B. & Lang, M. (2015). *parallelMap: Unified Interface to Parallelization Back-Ends*. R package version 1.3. Retrieved from https://CRAN.R-project.org/package=parallelMap

Bischl, B., Lang, M., Richter, J., Bossek, J., Judt, L., Kuehn, T., ... Kotthoff, L. (2016). *Mlr: machine learning in r*. R package version 2.5. Retrieved from https://github.com/mlr-org/mlr

Bischl, B., Mersmann, O., Trautmann, H., & Weihs, C. (2012). Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation. *Evolutionary Computation*, *20*(2), 249–275. doi:10.1162/EVCO_a_00069

Block, J. (1989). Critique of the act frequency approach to personality. *Journal of Personality and Social Psychology*, *56*(2), 234–45. doi:10.1037/0022-3514.56.2.234

Block, J. (2010). The Five-Factor Framing of Personality and Beyond: Some Ruminations. *Psychological Inquiry*, *21*(1), 2–25. doi:10.1080/10478401003596626

Boase, J. & Ling, R. (2013). Measuring Mobile Phone Use: Self-Report Versus Log Data. *Journal of Computer-Mediated Communication*, *18*(4), 508–519. doi:10.1111/jcc4.12021

Böhmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G. (2011). Falling asleep with Angry Birds, Facebook and Kindle - A Large Scale Study on Mobile Application Usage. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*, 47. doi:10.1145/2037373.2037383

Boksem, M. A., Tops, M., Kostermans, E., & De Cremer, D. (2008). Sensitivity to punishment and reward omission: Evidence from error-related ERP components. *Biological Psychology*, *79*(2), 185–192. doi:10.1016/j.biopsycho.2008.04.010

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–616. doi:10.1146/annurev.psych.54.101601.145030

Box, G. E. P. & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, *26*(2), 211–252. doi:10.2307/2287791

Breiman, L. (2001). Statistical Modeling: The Two Cultures. *Statistical Science*, *16*(3), 199–215. doi:10.2307/2676681. eprint: 0010

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. CRC press.

Brown, B., McGregor, M., & McMillan, D. (2014). 100 days of iphone use: understanding the details of mobile device use. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services* (pp. 223–232). MobileHCI '14. Toronto, ON, Canada: ACM. doi:10.1145/2628363.2628377

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., ... Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces? *Perception*, *22*(2), 131–152.

Brundell-Freij, K. & Ericsson, E. (2005). Influence of street characteristics, driver category and car performance on urban driving patterns. *Transportation Research Part D: Transport and Environment*, *10*(3), 213–229. doi:10.1016/j.trd.2005.01.001

Burton, C., McKinstry, B., Szentagotai Tătar, A., Serrano-Blanco, A., Pagliari, C., & Wolters, M. (2013). Activity monitoring in patients with depression: A systematic review. *Journal of Affective Disorders*, *145*(1), 21–28. doi:10.1016/j.jad.2012.07.001

Buss, D. M. & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review*, *90*(2), 105–126. doi:10.1037/0033-295X.90.2.105

Butt, S. & Phillips, J. G. (2008). Personality and self reported mobile phone use. *Computers in Human Behavior*, *24*(2), 346–360. doi:10.1016/j.chb.2007.01.019

Byrnes, J. P., Miller, D. C., & Schafer, W. D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological Bulletin*, *125*(3), 367–383. doi:10.1037/0033-2909.125.3.367

Cao, L., Dikmen, M., Fu, Y., & Huang, T. S. (2008). Gender recognition from body. In *Proceedings of the 16th ACM International Conference on Multimedia* (pp. 725–728). MM '08. Vancouver, British Columbia, Canada: ACM. doi:10.1145/1459359.1459470

Carney, D. R., Colvin, C. R., & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, *41*(5), 1054–1072. doi:10.1016/j.jrp.2007.01.004

Cattell, H. E. P. & Mead, A. D. (2003). The Sixteen Personality Factor Questionnaire (16PF). In *The SAGE Handbook of Personality Theory and Assessment* (pp. 135–159). Thousand Oaks, CA, US: SAGE Publications Ltd. doi:10.4135/9781849200479

Chapman, B. P., Duberstein, P. R., Sorensen, S., & Lyness, J. M. (2007). Gender Differences in Five Factor Model Personality Traits in an Elderly Cohort: Extension of Robust and Surprising Findings to an Older Generation. *Personality and Individual Differences*, *43*(06), 1594–1603. doi:10.1016/j.paid.2007.04.028

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. doi:10.1613/jair.953. eprint: 1106.1813

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. *Personal Ubiquitous Computing*, *17*(3), 433–450. doi:10.1007/s00779-011-0490-1

Cioffi-Revilla, C. (2010). Computational social science. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(3), 259–271. doi:10.1002/wics.95

Coluccia, E. & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, *24*(3), 329–340. doi:10.1016/j.jenvp.2004.08.006

Cortes, C. & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. doi:10.1007/BF00994018

Costa, J., Paul, Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322–331. doi:10.1037/0022-3514.81.2.322

Costa, P. T. & McCrae, R. R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, *13*(6), 667–673. doi:10.1016/0191-8869(92)90237-J

Costa, P. T. & McCrae, R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Psychological Assessment Resources Inc. doi:10.1037//1040-3590.4.1.5

Czado, C., Gneiting, T., & Held, L. (2009). Predictive model assessment for count data. *Biometrics*, *65*(4), 1254–1261. doi:10.1111/j.1541-0420.2009.01191.x

Dabbs Jr., J. M., Chang, E.-L., Strong, R. A., & Milun, R. (1998). Spatial Ability, Navigation Strategy, and Geographic Knowledge Among Men and Women. *Evolution and Human Behavior*, *19*(2), 89–98. doi:10.1016/S1090-5138(97)00107-4

Dahl, D. B. (2016). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-2. Retrieved from https://CRAN.R-project.org/package=xtable

Damrad-Frye, R. & Laird, J. D. (1989). The experience of boredom: The role of the self-perception of attention. *Journal of Personality and Social Psychology*, *57*(2), 315–320. doi:10.1037/0022-3514.57.2.315

De Montjoye, Y.-A., Quoidbach, J., Robic, F., & Pentland, A. (2013). Predicting personality using novel mobile phone-based metrics. In *Proceedings of the 6th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction* (pp. 48–55). SBP'13. Washington, DC: Springer-Verlag. doi:10.1007/978-3-642-37210-0_6

Deary, I. J. (1996). A (latent) Big Five Personality Model in 1915? A reanalysis of Webb's data. *Journal of Personality and Social Psychology*, *71*(5), 992–1005. doi:10.1037/0022-3514.71.5.992

DeRaad & Boele. (2000). *The Big Five Personality Factors: The Psycholexical Approach to Personality*. Hogrefe & Huber Publishers.

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology*, *91*(6), 1138–51. doi:10.1037/0022-3514.91.6.1138

DeYoung, C. G. (2015). Openness/intellect: A dimension of personality reflecting cognitive exploration. In M. Mikulincer, P. R. Shaver, M. L. Cooper, & R. J. Larsen (Eds.), *APA Handbook of Personality and Social Psychology, Volume 4: Personality Processes and Individual Differences* (pp. 369–399). APA handbooks in psychology. Washington, DC, US: American Psychological Association. doi:10.1037/14343-017

DeYoung, C. G., Grazioplene, R. G., & Peterson, J. B. (2012). From madness to genius: The Openness/Intellect trait domain as a paradoxical simplex. *Journal of Research in Personality*, *46*, 63–78. doi:10.1016/j.jrp.2011.12.003

Diener, E., Sandvik, E., Pavot, W., & Fujita, F. (1992). Extraversion and subjective well-being in a U.S. national probability sample. *Journal of Research in Personality*, *26*(3), 205–215. doi:10.1016/0092-6566(92)90039-7

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., . . . Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. doi:10.1111/j.1600-0587.2012.07348.x

Eagly, A. H. & Wood, W. (2013). The Nature-Nurture Debates: 25 Years of Challenges in Understanding the Psychology of Gender. *Perspectives on Psychological Science*, *8*(3), 340–357. doi:10.1177/1745691613484767

Efron, B. & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, *1*(1), 54–75.

Efron, B. & Tibshirani, R. (1997). Improvements on Cross-Validation: The .632 + Bootstrap Method. *Journal of the American Statistical Association*, *92*(438), 548–560. doi:10.1080/01621459.1997.10474007

Eftekhar, A., Fullwood, C., & Morris, N. (2014). Capturing personality from Facebook photos and photo-related activities: How much exposure do you need? *Computers in Human Behavior*, *37*, 162–170. doi:10.1016/j.chb.2014.04.048

Erceg-Hurn, D. M. & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, *63*(7), 591–601. doi:10.1037/0003-066X.63.7.591

Ericsson, E. (2000a). Driving pattern in urban areas-descriptive analysis and initial prediction model. *Bulletin 185/3000 of Lunds University*, 18. Retrieved from http://lup.lub.lu.se/record/627123

Ericsson, E. (2000b). Variability in urban driving patterns. *Transportation Research Part D: Transport and Environment*, *5*(5), 337–354. doi:10.1016/S1361-9209(00)00003-1

Ericsson, E. (2001). Independent driving pattern factors and their influence on fuel-use and exhaust emission factors. *Transportation Research Part D: Transport and Environment*, *6*(5), 325–345. doi:10.1016/S1361-9209(01)00003-7

Eysenck, Hans J. (1991). Dimensions of personality: 16, 5 or 3? - Criteria for a taxonomic paradigm. *Personality and Individual Differences*, *12*(8), 773–790. doi:10.1207/s15327752jpa4303_23

Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, IL: Thomas Publishing.

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences*, *13*(6), 667–673.

Eysenck, H. J. (1997). Personality and experimental psychology: The unification of psychology and the possibility of a paradigm. *Journal of Personality and Social Psychology*, *73*(6), 1224–1237. doi:10.1037/0022-3514.73.6.1224

Eysenck, H. (2013). *The Structure of Human Personality (Psychology Revivals)*. Psychology Revivals. Taylor & Francis.

Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., Stillwell, D., ... De Cock, M. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, *26*(2), 1–34. doi:10.1007/s11257-016-9171-0

Fast, L. A. & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology*, *94*(2), 334–346. doi:10.1037/0022-3514.94.2.334

Fernandes, R., Hatfield, J., & Soames Job, R. F. (2010). A systematic investigation of the differential predictors for speeding, drink-driving, driving while fatigued, and not wearing a seat belt, among young drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *13*(3), 179–196. doi:10.1016/j.trf.2010.04.007

Fleeson, W., Gallagher, P., Carolina, N., & Gallagher, M. P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*(6), 1097–1114. doi:10.1037/a0016786

Fox, J. & Monette, G. (1992). Generalized Collinearity Diagnostics. *Journal of the American Statistical Association*, *87*(417), 178–183. doi:10.1080/01621459.1992.10475190

Friedman, J. (2002). Stochastic gradient boosting. *Computational Statistics and Data Analysis*, *38*(4), 367–378. doi:10.1016/S0167-9473(01)00065-2

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1–22. doi:10.1359/JBMR.0301229. eprint: NIHMS201118

Funder, D. C. (2012). Accurate Personality Judgment. *Current Directions in Psychological Science*, *21*(3), 177–182. doi:10.1177/0963721412445309

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, *23*(5), 369–401. doi:10.1002/per.724

Furr, R. M. & Funder, D. C. (2009). Behavioral Observation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (Vol. 62, *1*, pp. 189–193). New York, NY, USA: Guilford Press.

Gartner. (2016). Gartner Says Worldwide Smartphone Sales Grew 9.7 Percent in Fourth Quarter of 2015. Retrieved from http://www.gartner.com/newsroom/id/3215217

Gerard, H. B. & Mathewson, G. C. (1966). The effects of severity of initiation on liking for a group: A replication. *Journal of Experimental Social Psychology*, *2*(3), 278–287. doi:10.1016/0022-1031(66)90084-9

Globalwebindex & Mander, J. (2015a). 1 in 4 Spotify Users Pay for the Service. Retrieved March 18, 2015, from http://www.globalwebindex.net/blog/1-in-4-spotify-users-pay-for-the-service

Globalwebindex & Mander, J. (2015b). Why Tinder Has a Lot of Positives to Shout About. Retrieved 22 5, 2015, from http://www.globalwebindex.net/blog/why-tinder-has-a-lot-of-positives-to-shout-about

Goldberg, L. (1981). Language and individual differences: The search for universals in personality lexicons. In L. Wheeler (Ed.), *Review of Personality and Social Psychology* (Vol. 2, pp. 141–165). Beverly Hills, CA: SAGE Publications Ltd.

Google, I. (2016a). Android Apps on Google Play. Retrieved June 16, 2015, from https://play.google.com/store/apps?hl=en

Google, I. (2016b). Sensors Overview | Android Developers. Retrieved August 10, 2016, from https://developer.android.com/guide/topics/sensors/sensors%7B%5C_%7Doverview.html

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

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. doi:10.1016/S0092-6566(03)00046-1

Gray, J. A. & McNaughton, N. (2003, June). *The Neuropsychology of Anxiety*. Oxford University Press. doi:10.1093/acprof:oso/9780198522713.001.0001

Graziano, W. G., Habashi, M. M., Sheese, B. E., & Tobin, R. M. (2007). Agreeableness, empathy, and helping: a person x situation perspective. *Journal of Personality and Social Psychology*, *93*(4), 583–599. doi:10.1037/0022-3514.93.4.583

Graziano, W. G. & Tobin, R. M. (2009). Agreeableness. In M. R. L. R. H. Hoyle (Ed.), *Handbook of Individual Differences in Social Behavior* (pp. 46–61). New York, NY, US: Guilford Press.

Greenberg, D. M., Kosinski, M., Stillwell, D. J., Monteiro, B. L., Levitin, D. J., & Rentfrow, P. J. (2016). The Song Is You: Preferences for Musical Attribute Dimensions Reflect Personality. *Social Psychological and Personality Science*, *7*(6), 597–605. doi:10.1177/1948550616641473

Greenberg, D. M., Müllensiefen, D., Lamb, M. E., & Rentfrow, P. J. (2015). Personality predicts musical sophistication. *Journal of Research in Personality*, *58*, 154–158. doi:10.1016/j.jrp.2015.06.002

Guo, F. & Fang, Y. (2013). Individual driver risk assessment using naturalistic driving data. *Accident Analysis & Prevention*. Emerging Research Methods and Their Application to Road Safety Emerging Issues in Safe and Sustainable Mobility for Older Persons The Candrive/Ozcandrive Prospective Older Driver Study: Methodology and Early Study Findings, *61*, 3–9. doi:10.1016/j.aap.2012.06.014

Hadid, A. & Pietikäinen, M. (2009). Combining appearance and motion for face and gender recognition from videos. *Pattern Recognition*, *42*(11), 2818–2827. doi:10.1016/j.patcog.2009.02.011

Haney, C., Banks, C., & Zimbardo, P. (1973). Interpersonal dynamics in a simulated prison. *International Journal of Criminology & Penology*, *1*(1), 69–97.

Hang, A., De Luca, A., Hartmann, J., & Hussmann, H. (2013). Oh app, where art thou?: on app launching habits of smartphone users. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services* (pp. 392–395). MobileHCI '13. Munich, Germany: ACM. doi:10.1145/2493190.2493219

Hastie, Trevor and Tibshirani, Robert and Friedman, Jerome. (2009). *The Elements of Statistical Learning Data Mining, Inference, and Prediction, Second Edition*. doi:10.1007/978-0-387-84858-7

Heene, M., Bollmann, S., & Bühner, M. (2014). Much ado About Nothing, or Much to do About Something? *Journal of Individual Differences*, *35*(4), 245–249. doi:10.1027/1614-0001/a000146

Hektner, J. M., Schmidt, J. A. (A., & Csikszentmihalyi, M. (2007). *Experience Sampling Method: Measuring the Quality of Everyday Life*. Sage Publications.

Hirsh, J. B., Deyoung, C. G., & Peterson, J. B. (2009, August). Metatraits of the Big Five differentially predict engagement and restraint of behavior. *Journal of Personality*, *77*(4), 1085–102. doi:10.1111/j.1467-6494.2009.00575.x

Hodgins, S. & Ellenbogen, M. (2003). Neuroticism and depression. *The British Journal of Psychiatry*, *182*(1), 79–80. doi:10.1192/bjp.182.1.79

Hofner, B., Boccuto, L., & Göker, M. (2015). Controlling false discoveries in high-dimensional situations: boosting with stability selection. *BMC Bioinformatics*, *16*(1), 144. doi:10.1186/s12859-015-0575-3

Hofner, B. & Hothorn, T. (2015). *stabs: Stability Selection with Error Control*. R package version R package version 0.5-1. Retrieved from http://CRAN.R-project.org/package=stabs

Hogan, J. & Roberts, B. W. (1996). Issues and non-issues in the fidelity-bandwidth trade-off. *Journal of Organizational Behavior*, *17*(6), 627–637. doi:10.1002/(SICI)1099-1379(199611)17:6<627::AID-JOB2828>3.0.CO;2-F

Holmes, C. B. (1983). Sample size in four areas of psychological research. *Transactions of the Kansas Academy of Science (1903-)*, *86*(2/3), 76–80.

Hu, J., Zeng, H. H.-J., Li, H., Niu, C., & Chen, Z. (2007). Demographic prediction based on user's browsing behavior. In *Proceedings of the 16th International Conference on World Wide Web - WWW '07* (p. 151). WWW '07. New York, New York, USA: ACM Press. doi:10.1145/1242572.1242594

Hunt, D. S. & Langstedt, E. (2014). The influence of personality factors and motives on photographic communication. *The Journal of Social Media in Society*, *3*(2). Retrieved from http://www.thejsms.org/tsmri/index.php/TSMRI/article/view/68

Jackson, J. J., Wood, D., Bogg, T., Walton, K. E., Harms, P. D., & Roberts, B. W. (2010). What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness (BIC). *Journal of Research in Personality*, *44*(4), 501–511. doi:10.1016/j.jrp.2010.06.005

Jakobwitz, S. & Egan, V. (2006). The dark triad and normal personality traits. *Personality and Individual Differences*, *40*(2), 331–339. doi:10.1016/j.paid.2005.07.006

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Inc. doi:10.1007/978-1-4614-7138-7

Jang, K. L., Livesley, W. J., Angleitner, A., Riemann, R., & Vernon, P. a. (2002). Genetic and environmental influences on the covariance of facets defining the domains of the five-factor model of personality. *Personality and Individual Differences*, *33*(1), 83–101. doi:10.1016/S0191-8869(01)00137-4

Jensen-Campbell, L. a. & Graziano, W. G. (2001). Agreeableness as a moderator of interpersonal conflict. *Journal of Personality*, *69*(2), 323–361. doi:10.1111/1467-6494.00148

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative big five trait taxonomy: history, measurement, and conceptual issues. In O. P. John,

R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (3rd ed., pp. 114–158). New York, NY, USA: Guilford Press.

John, O. P. & Robins, R. W. (1993). Determinants of interjudge agreement on personality traits: the big five domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality*, *61*(4), 521–551. doi:10.1111/j.1467-6494.1993.tb00781.x

Jones, S. L., Ferreira, D., Hosio, S., Goncalves, J., & Kostakos, V. (2015). Revisitation analysis of smartphone app use. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 1197–1208). UbiComp '15. Osaka, Japan: ACM. doi:10.1145/2750858.2807542

Kafadar, K. (2003). John Tukey and Robustness. *Statistical Science*, *18*(3), 319–331. Retrieved from http://www.jstor.org/stable/3182749

Kaptein, N., Theeuwes, J., & Van Der Horst, R. (1996). Driving Simulator Validity: Some Considerations. *Transportation Research Record: Journal of the Transportation Research Board*, *1550*, 30–36. doi:10.3141/1550-05

Karatzoglou, A., Smola, A., Hornik, K., & Zeileis, A. (2004). Kernlab – an S4 package for kernel methods in R. *Journal of Statistical Software*, *11*(9), 1–20. Retrieved from http://www.jstatsoft.org/v11/i09/

Kim, Y., Briley, D. A., & Ocepek, M. G. (2015). Differential innovation of smartphone and application use by sociodemographics and personality. *Computers in Human Behavior*, *44*, 141–147. doi:10.1016/j.chb.2014.11.059

Knight, G. P., Guthrie, I. K., Page, M. C., & Fabes, R. A. (2002). Emotional arousal and gender differences in aggression: A meta-analysis. *Aggressive Behavior*, *28*(5), 366–393. doi:10.1002/ab.80011

Kobayashi, T. & Boase, J. (2012). No Such Effect? The Implications of Measurement Error in Self-Report Measures of Mobile Communication Use. *Communication Methods and Measures*, *6*(2), 126–143. doi:10.1080/19312458.2012.679243

Koivisto, J. & Hamari, J. (2014). Demographic differences in perceived benefits from gamification. *Computers in Human Behavior*, *35*, 179–188. doi:10.1016/j.chb.2014.03.007

Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, *95*(3), 357–380. doi:10.1007/s10994-013-5415-y

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5802–5. doi:10.1073/pnas.1218772110

Kuhn, M. (2015). *Caret: classification and regression training*. R package version 6.0-58. Retrieved from https://CRAN.R-project.org/package=caret

Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling*. Springer Publishing Company, Inc. doi:10.1007/978-1-4614-6849-3

Lane, W. & Manner, C. (2011). The Impact of Personality Traits on Smartphone Ownership and Use. *International Journal of Business & Social Science*, *2*(17), 22–28.

Langmeyer, A., Guglhör-Rudan, A., & Tarnai, C. (2012). What Do Music Preferences Reveal About Personality?: A Cross-Cultural Replication Using Self-Ratings and Ratings of Music Samples. *Journal of Individual Differences*, *33*(2), 119–130. doi:10.1027/1614-0001/a000082

Lawton, C. A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles*, *30*(11-12), 765–779. doi:10.1007/BF01544230

Leahey, E. (2006). Gender Differences in Productivity: Research Specialization as a Missing Link. *Gender & Society*, *20*(6), 754–780. doi:10.1177/0891243206293030

Lee, D. G., Kelly, K. R., & Edwards, J. K. (2006). A closer look at the relationships among trait procrastination, neuroticism, and conscientiousness. *Personality and Individual Differences*, *40*(1), 27–37. doi:10.1016/j.paid.2005.05.010

Lewandowski Jr, G. W. & Strohmetz, D. B. (2009). Actions Can Speak as Loud as Words: Measuring Behavior in Psychological Science. *Social and Personality Psychology Compass*, *3*(6), 992–1002. doi:10.1111/j.1751-9004.2009.00229.x

Lonczak, H. S., Neighbors, C., & Donovan, D. M. (2007). Predicting risky and angry driving as a function of gender. *Accident Analysis & Prevention*, *39*(3), 536–545. doi:10.1016/j.aap.2006.09.010

López-Ibáñez, M., Dubois-Lacoste, J., Stützle, T., & Birattari, M. (2011). *The irace package, Iterated Race for Automatic Algorithm Configuration*. IRIDIA, Université Libre de Bruxelles, Belgium.

Lucas, R. E. & Baird, B. M. (2004). Extraversion and emotional reactivity. *Journal of Personality and Social Psychology*, *86*(3), 473–485. doi:10.1037/0022-3514.86.3.473

Lucidi, F., Mallia, L., Lazuras, L., & Violani, C. (2014). Personality and attitudes as predictors of risky driving among older drivers. *Accident Analysis & Prevention*, *72*, 318–324. doi:10.1016/j.aap.2014.07.022

MacCann, C., Duckworth, A. L., & Roberts, R. D. (2009). Empirical identification of the major facets of Conscientiousness. *Learning and Individual Differences*, *19*(4), 451–458. doi:10.1016/j.lindif.2009.03.007

Mairesse, F. & Walker, M. (2006). Automatic recognition of personality in conversation. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers* (pp. 85–88). NAACL-Short '06. New York, NY: Association for Computational Linguistics. Retrieved from http://dl.acm.org/citation.cfm?id=1614049.1614071

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*(1), 457–500. doi:10.1613/jair.2349

Malouff, J. M., Thorsteinsson, E. B., Schutte, N. S., Bhullar, N., & Rooke, S. E. (2010). The Five-Factor Model of personality and relationship satisfaction of intimate partners: A meta-analysis. *Journal of Research in Personality*, *44*(1), 124–127. doi:10.1016/j.jrp.2009.09.004

Marszalek, J. M., Barber, C., Kohlhart, J., & Holmes, C. B. (2011, April). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills*, *112*(2), 331–48. doi:10.2466/03.11.PMS.112.2.331-348

Martin, C. L. & Halverson, C. F. (1981). A Schematic Processing Model of Sex Typing and Stereotyping in Children. *Child Development*, *52*(4), 1119–1134. doi:10.2307/1129498

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, *47*(2), 149–174. doi:10.1007/BF02296272

Matthews, G., Deary, I. J., & Whiteman, M. C. (2009). *Personality Traits*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511812736

McCrae, R. R., Costa Jr., T., Paul, & Martin, A., Thomas. (2005). The NEO–PI–3: A More Readable Revised NEO Personality Inventory. *Journal of Personality Assessment*, *84*(3), 261–270. doi:10.1207/s15327752jpa8403_05

McCrae, R. R. & Costa, Paul T., J. (1997). Conceptions and Correlates of Openness to Experience. In *Handbook of Personality Psychology* (pp. 825–847). doi:10.1080/01425690701737481

McCrae, R. R. & Costa, P. T. (1987). Validation of the Five-Factor Model of Personality Across - Instruments and Observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. doi:10.1037/0022-3514.52.1.81

McLachlan, G. J. & Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley.

McNiel, J. M. & Fleeson, W. (2006). The causal effects of extraversion on positive affect and neuroticism on negative affect: Manipulating state extraversion and state neuroticism in an experimental approach. *Journal of Research in Personality*, *40*(5), 529–550. doi:10.1016/j.jrp.2005.05.003

Mehl, M. R. (2006). The lay assessment of subclinical depression in daily life. *Psychological Assessment*, *18*(3), 340–345. doi:10.1037/1040-3590.18.3.340

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): a device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, & Computers*, *33*(4), 517–523. doi:10.3758/BF03195410

Mehl, M. R. & Pennebaker, J. W. (2003). The sounds of social life: a psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*(4), 857–870.

Meinshausen, N. & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(4), 417–473. doi:10.1111/j.1467-9868.2010.00740.x

Milgram, S. & Van den Haag, E. (1978). Obedience to authority. Ziff-Davis Publishing Company.

Miller, G. (2012). The Smartphone Psychology Manifesto. *Perspectives on Psychological Science*, *7*(3), 221–237. doi:10.1177/1745691612441215

Miller, G. A. (1955). Note on the bias of information estimates. Glencoe, IL: Free Press.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, *55*, 1–22. doi:10.1146/annurev.psych.55.042902.130709

Molinaro, A. M., Simon, R., & Pfeiffer, R. M. (2005). Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, *21*(15), 3301–7. doi:10.1093/bioinformatics/bti499

Montag, C., Blaszkiewicz, K., Lachmann, B., Andone, I., Sariyska, R., Trendafilov, B., … Markowetz, A. (2014). Correlating Personality and Actual Phone Usage: Evidence From Psychoinformatics. *Journal of Individual Differences*, *35*(3), 158–165. doi:10.1027/1614-0001/a000139

Montag, C., Blaszkiewicz, K., Sariyska, R., Lachmann, B., Andone, I., Trendafilov, B., … Markowetz, A. (2015). Smartphone usage in the 21st century: who is active on WhatsApp? *BMC Research Notes*, *8*(1), 331. doi:10.1186/s13104-015-1280-z

Moore, G. E. (2006). Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Newsletter*, *20*(3), 33–35. doi:10.1109/N-SSC.2006.4785860

Mosier, C. I. (1951). The need and means of cross validation. I. Problems and designs of cross-validation. *Educational and Psychological Measurement*, *11*, 5–11. doi:10.1177/001316445101100101

Mozer, M. C. & Lindsey, R. V. (2016). Predicting and improving memory retention: psychological theory matters in the big data era. Retrieved from https://www.semanticscholar.org/paper/Predicting-and-Improving-Memory-Retention-Mozer-Lindsey/eee5594f8e88e51f4fb86b43c9f4cb54d689f73c

Mullen, N., Charlton, J., Devlin, A., & Bedard, M. (2011). Simulator Validity: Behaviors Observed on the Simulator and on the Road. Retrieved from http://trid.trb.org/view.aspx?id=1114738

Mussel, P., Winter, C., Gelléri, P., & Schuler, H. (2011). Explicating the openness to experience construct and its subdimensions and facets in a work setting. *International Journal of Selection and Assessment*, *19*(2), 145–156. doi:10.1111/j.1468-2389.2011.00542.x

Nielsen, A. H., Østergaard, J., Marra, F., & Træholt, C. (2010). Driving Pattern Analysis for Electric Vehicle (EV) Grid Integration Study. In *2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)* (pp. 1–6). IEEE. doi:10.1109/ISGTEUROPE.2010.5751581

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: replicated factors structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*(6), 574–583. doi:10.1037/h0040291

O'Hara, R. B. & Kotze, D. J. (2010). Do not log-transform count data. *Methods in Ecology and Evolution*, *1*(2), 118–122. doi:10.1111/j.2041-210X.2010.00021.x

OpenSignal. (2015). Android Fragmentation Visualized. Retrieved from http://opensignal.com/reports/2015/08/android-fragmentation/

Ormel, J., Jeronimus, B. F., Kotov, R., Riese, H., Bos, E. H., Hankin, B., … Oldehinkel, A. J. (2013). Neuroticism and common mental disorders: meaning and utility of a complex relationship. *Clinical psychology review*, *33*(5), 686–97. doi:10.1016/j.cpr.2013.04.003

Ozer, D. J. & Benet-Martínez, V. (2006). Personality and the Prediction of Consequential Outcomes. *Annual Review of Psychology*, *57*(1), 401–421. doi:10.1146/annurev.psych.57.102904.190127

Paulhus, D. L. (1991). Measurement and Control of Response Bias. In *Measures of Personality and Social Psychological Attitudes* (pp. 17–59). Elsevier Inc. doi:10.1016/B978-0-12-590241-0.50006-X

Paulhus, D. L. & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 224–239). Guilford Press.

Paunonen, S. V. & Ashton, M. C. (2001). Big five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, *81*(3), 524–539. doi:10.1037/0022-3514.81.3.524

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *The Journal of Applied Psychology*, *88*(5), 879–903. doi:10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology*, *63*(1), 539–569. doi:10.1146/annurev-psych-120710-100452

Polzehl, T., Moller, S., & Metze, F. (2010). Automatically Assessing Personality from Speech. In *2010 IEEE Fourth International Conference on Semantic Computing*. IEEE. doi:10.1109/ICSC.2010.41

Poorthuis, A. M. G., Thomaes, S., Denissen, J. J. A., van Aken, M. A. G., & Orobio de Castro, B. (2014, January). Personality in Action. *European Journal of Psychological Assessment*, *30*(3), 169–177. doi:10.1027/1015-5759/a000186

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*(2), 322–338. doi:10.1037/a0014996

Premraj, R. (2015). *mailR: A Utility to Send Emails from R*. R package version 0.4.1. Retrieved from https://CRAN.R-project.org/package=mailR

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*(1), 203–212. doi:10.1016/j.jrp.2006.02.001

Ready, R. E., Clark, L. A., Watson, D., & Westerhouse, K. (2000). Self- and Peer-Reported Personality: Agreement, Trait Ratability, and the "Self-Based Heuristic". *Journal of Research in Personality*, *34*(2), 208–224. doi:10.1006/jrpe.1999.2280

Redsell, M., Lucas, G. G., & Ashford, N. J. (1993). Factors affecting car fuel consumption. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering 1989-1996 (vols 203-210)*, *207*(14), 1–22. doi:10.1243/PIME_PROC_1993_207_155_02

Reed, D. a., Enders, F., Lindor, R., McClees, M., & Lindor, K. D. (2011). Gender differences in academic productivity and leadership appointments of physicians throughout academic careers. *Academic Medicine*, *86*(1), 43–47.

Renner, R. S., Velichkovsky, B. M., & Helmert, J. R. (2013). The Perception of Egocentric Distances in Virtual Environments - A Review. *ACM Computing Surveys*, *46*(2), 23:1–23:40. doi:10.1145/2543581.2543590

Revelle, W. (2016). *psych: Procedures for Psychological, Psychometric, and Personality Research*. R package version 1.6.4. Northwestern University. Evanston, Illinois. Retrieved from http://CRAN.R-project.org/package=psych

Richter, F. & Statista. (2014). The Price Gap Between iOS and Android Is Widening. Retrieved from https://www.statista.com/chart/1903/average-selling-price-of-android-and-ios-smartphones/

Ridgeway, G. (2015). *gbm: Generalized Boosted Regression Models*. R package version 2.1.1. Retrieved from https://CRAN.R-project.org/package=gbm

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes. *Perspectives on Psychological Science*, *2*(4), 313–45. doi:10.1111/j.1745-6916.2007.00047.x

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). Proc: an open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinformatics*, *12*, 77.

Rousseeuw, P. J. & Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*, *88*(424), 1273–1283. doi:10.1080/01621459.1993.10476408

Saeb, S., Zhang, M., Karr, C. J., Schueller, S. M., Corden, M. E., Kording, K. P., & Mohr, D. C. (2015). Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. *Journal of Medical Internet Research*, *17*(7). doi:10.2196/jmir.4273

Salgado, J. F., Moscoso, S., & Berges, A. (2013). Conscientiousness, Its Facets, and the Prediction of Job Performance Ratings: Evidence against the narrow measures. *International Journal of Selection and Assessment*, *21*(1), 74–84. doi:10.1111/ijsa.12018

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R*. ISBN 978-0-387-75968-5. New York: Springer. Retrieved from http://lmdvr.r-forge.r-project.org

Saucier, G., Goldberg, L. R., & Institute, O. R. (2001). Lexical studies of indigenous personality factors: premises, products, and prospects. *Journal of Personality*, *69*, 847–879. doi:10.1111/1467-6494.696167

Serneels, S., De Nolf, E., & Van Espen, P. J. (2006). Spatial sign preprocessing: a simple way to impart moderate robustness to multivariate estimators. *Journal of Chemical Information and Modeling*, *46*(3), 1402–9. doi:10.1021/ci050498u

Sherman, R. C., End, C., Kraan, E., Cole, A., Campbell, J., Birchmeier, Z., & Klausner, J. (2000). The Internet Gender Gap Among College Students: Forgotten But Not Gone? *CyberPsychology & Behavior*, *3*(5), 885–94. doi:10.1089/10949310050191854

Simon, R. (2007). Resampling strategies for model assessment and selection. In *Fundamentals of Data Mining in Genomics and Proteomics* (pp. 173–186). Springer.

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: the impact of age and major life events on mean-level and rank-order stability of the Big Five. *Journal of Personality and Social Psychology*, *101*(4), 862–882. doi:10.1037/a0024950

Stachl, C. & Bühner, M. (2015). Show me how you Drive and I'll Tell you who you are Recognizing Gender Using Automotive Driving Parameters. *Procedia Manufacturing*, *3*, 5587–5594. doi:10.1016/j.promfg.2015.07.743

Stisen, A., Blunck, H., Bhattacharya, S., Prentow, T. S., Kjærgaard, M. B., Dey, A., ... Jensen, M. M. (2015). Smart Devices are Different: Assessing and Mitigating Mobile Sensing Heterogeneities for Activity Recognition. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems - SenSys '15* (pp. 127–140). doi:10.1145/2809695.2809718

Tabachnick, B. G. & Fidell, L. S. (2012). *Using Multivariate Statistics (6th Edition)* (6th ed.). Pearson.

Tellegen, A. & Waller, N. G. (2008). Exploring Personality Through Test Construction: Development of the Multidimensional Personality Questionnaire. In *The SAGE Handbook of Personality Theory and Assessment: Volume 2 — Personality Measurement and Testing* (pp. 261–292). London: SAGE Publications Ltd. doi:10.4135/9781849200479.n13

Tonetti, L., Fabbri, M., & Natale, V. (2009). Relationship between Circadian Typology and Big Five Personality Domains. *Chronobiology International*, *26*(2), 337–347. doi:10.1080/07420520902750995

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., … Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, *17*(6), 520–525. doi:10.1093/bioinformatics/17.6.520

Trull, T. J. & Ebner-Priemer, U. (2013). Ambulatory assessment. *Annual Review of Clinical Psychology*, *9*, 151–76. doi:10.1146/annurev-clinpsy-050212-185510

Twisk, D., Bos, N., Shope, J. T., & Kok, G. (2013). Changing mobility patterns and road mortality among pre-license teens in a late licensing country: an epidemiological study. *BMC Public Health*, *13*(333). doi:10.1186/1471-2458-13-333

Vaerenbergh, Y. V. & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195–217. doi:10.1093/ijpor/eds021

Vardaki, S. & Yannis, G. (2013). Investigating the self-reported behavior of drivers and their attitudes to traffic violations. *Journal of Safety Research*, *46*, 1–11. doi:10.1016/j.jsr.2013.03.001

Vazire, S. (2010). Who knows what about a person? The self–other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. doi:10.1037/a0017908

Vazire, S. & Mehl, M. R. (2008). Knowing me, knowing you: the accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology*, *95*(5), 1202–1216. doi:10.1037/a0013314

Vecchione, M., Alessandri, G., Barbaranelli, C., & Caprara, G. (2012). Gender differences in the Big Five personality development: A longitudinal investigation from late adolescence to emerging adulthood. *Personality and Individual Differences*, *53*(6), 740–746. doi:10.1016/j.paid.2012.05.033

Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(1), 1–23. doi:10.18637/jss.v059.i10. eprint: arXiv:1501.0228

Wickham, H. & Francois, R. (2016). *dplyr: A Grammar of Data Manipulation*. R package version 0.5.0. Retrieved from https://CRAN.R-project.org/package=dplyr

Williams, M. M. J., Whitaker, R. M., & Allen, S. M. (2012). Measuring individual regularity in human visiting patterns. In *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust* (pp. 117–122). doi:10.1109/SocialCom-PASSAT.2012.93

Wright, M. N. & Ziegler, A. (2015). Ranger: a fast implementation of random forests for high dimensional data in c++ and r. *arXiv preprint arXiv:1508.04409*.

Wu, K. D. & Clark, L. A. (2003). Relations between personality traits and self-reports of daily behavior. *Journal of Research in Personality*, *37*(4), 231–256. doi:10.1016/S0092-6566(02)00539-1

Xu, R., Frey, R. M., Fleisch, E., & Ilic, A. (2016). Understanding the impact of personality traits on mobile app adoption - Insights from a large-scale field study. *Computers in Human Behavior*, *62*, 244–256. doi:10.1016/j.chb.2016.04.011

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality*, *44*(3), 363–373. doi:10.1016/j.jrp.2010.04.001

Yarkoni, T. (2012). Psychoinformatics: New Horizons at the Interface of the Psychological and Computing Sciences. *Current Directions in Psychological Science*, *21*(6), 391–397. doi:10.1177/0963721412457362

Yarkoni, T. & Westfall, J. (2016). Choosing prediction over explanation in psychology: Lessons from machine learning. doi:10.6084/m9.figshare.2441878.v1

Yeo, I.-K. & Johnson, R. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959. doi:10.1093/biomet/87.4.954

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 201418680. doi:10.1073/pnas.1418680112

Zhong, E., Tan, B., Mo, K., & Yang, Q. (2013). User Demographics Prediction Based on Mobile Data. *Pervasive Mobile Computing*, *9*(6), 823–837. doi:10.1016/j.pmcj.2013.07.009

Zhu, M., Zhao, S., Coben, J. H., & Smith, G. S. (2013). Why more male pedestrians die in vehicle-pedestrian collisions than female pedestrians: a decompositional analysis.

Injury Prevention: Journal of the International Society for Child and Adolescent Injury Prevention, *19*(4), 227–231. doi:10.1136/injuryprev-2012-040594

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short Scales - Five Misunderstandings and Ways to Overcome Them. *Journal of Individual Differences*, *35*(4), 185–189. doi:10.1027/1614-0001/a000148

Zimmerman, R. D. (2008). Understanding the impact of personality traits on individuals' turnover decisions: A meta-analytic path model. *Personnel Psychology*, *61*(2), 309–348. doi:10.1111/j.1744-6570.2008.00115.x

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *67*, 301–320. doi:10.1111/j.1467-9868.2005.00503.x