Psychological Research in the Digital Age

Leveraging Smartphones to Predict Psychological Constructs in Daily Life

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
an der Ludwig-Maximilians-Universität München

vorgelegt von
Fiona Alessa Kunz
aus Heilbronn

München 2023

Erstgutachter: Prof. Dr. Markus Bühner

Zweitgutachter: Prof. Dr. David Goretzko

Tag der Abgabe: 12.06.2023

Tag der mündlichen Prüfung: 16.11.2023

Acknowledgments

I would like to thank my supervisor, Professor Markus Bühner, for giving me the opportunity to participate in the PhoneStudy project as an external Ph.D. student and for his trust and guidance throughout my academic journey. I would also like to express my sincere gratitude to Dr. Ramona Schödel for her extraordinary professional and personal support from day one of this project. I am eternally grateful for her belief in me and my dissertation project, her open ear, and her invaluable advice - I could not have asked for a better mentor. Thank you also for challenging and proofreading this dissertation. I would also like to thank all those who helped me along the way and contributed to this research project. The empirical studies in this dissertation would not have been possible without the interdisciplinary collaboration with my fellow researchers in the PhoneStudy team, who were instrumental in making this empirical work a reality. Special thanks go to Florian Bemmann for his valuable technical support and advice.

This dissertation is dedicated to my parents, in gratitude for providing me with all the means to pursue my academic career and for always serving as role models. Finally, I would like to extend a very special thank you to Domi, who has provided me with unwavering support and encouragement over the past years and has been there for me every single day. Without his selfless and unconditional support, this thesis would never have been possible. Thank you for helping me to realize my dream.

Table of Contents

List of Tables

List of Figures

Abstract

The smartphone has become an important personal companion in our daily lives. Each time we use the device, we generate data that provides information about ourselves. This data, in turn, is valuable to science because it objectively reflects our everyday behavior and experiences. In this way, smartphones enable research that is closer to everyday life than traditional laboratory experiments and questionnaire-based methods. While data collected with smartphones are increasingly being used in the field of personality psychology, new digital technologies can also be leveraged to collect and analyze large-scale unobtrusively sensed data in other areas of psychological research.

This dissertation, therefore, explores the insights that smartphone sensing reveals for psychological research using two examples, situation and affect research, making a twofold-research contribution. First, in two empirical studies, different data types of smartphone-sensed data, such as GPS or phone data, were combined with experience-sampled self-report, and classical questionnaire data to gain valuable insights into individual behavior, thinking, and feeling in everyday life. Second, predictive modeling techniques were applied to analyze the large, high-dimensional data sets collected by smartphones. To gain a deeper understanding of the smartphone data, interpretable variables were extracted from the raw sensing data, and the predictive performance of various machine learning algorithms was compared.

In summary, the empirical findings suggest that smartphone data can effectively capture certain situational and behavioral indicators of psychological phenomena in everyday life. However, in certain research areas such as affect research, smartphone data should only complement, but not completely replace, traditional questionnaire-based data as well as other data sources such as neurophysiological indicators. The dissertation also concludes that the use of smartphone sensor data introduces new difficulties and challenges for psychological research and that traditional methods and perspectives are reaching their limits. The complexity of data collection, processing, and analysis requires established guidelines for study design, interdisciplinary collaboration, and theory-driven research that integrates explanatory and predictive approaches. Accordingly, further research is needed on how machine learning models and other big data methods in psychology can be reconciled with traditional theoretical approaches. Only in this way can we move closer to the ultimate goal of psychology to better understand, explain, and predict human behavior and experiences and their interplay with everyday situations.

Zusammenfassung

Das Smartphone ist zu einem wichtigen persönlichen Begleiter in unserem täglichen Leben geworden. Jedes Mal, wenn wir das Gerät benutzen, erzeugen wir Daten, die Informationen über uns selbst liefern. Diese Daten sind für die Wissenschaft wertvoll, weil sie unser Alltagsverhalten und unsere Erfahrungen objektiv widerspiegeln. So ermöglichen Smartphones eine Forschung, die näher am Alltag ist als herkömmliche Laborexperimente und fragebogenbasierte Erhebungsmethoden. Mit Smartphones erfasste Daten werden zunehmend im Bereich der Persönlichkeitspsychologie genutzt, aber auch in anderen Bereichen der psychologischen Forschung können neue digitale Technologien zur Erhebung und Analyse großer Datenmengen eingesetzt werden.

Diese Dissertation untersucht daher die Möglichkeiten der Smartphone Sensing-Methode für die psychologische Forschung am Beispiel der Situations- und Affektforschung und leistet damit einen zweifachen Forschungsbeitrag. Zum einen wurden in zwei empirischen Studien verschiedene Sensor-Datentypen, wie beispielsweise GPS oder Telefondaten, mit klassischen Fragebogendaten kombiniert, um wertvolle Einblicke in das individuelle Verhalten, Denken und Fühlen im täglichen Leben zu gewinnen. Zum anderen wurden prädiktive Modellierungstechniken angewandt, um die großen, hochdimensionalen Datensätze zu analysieren, die durch die Smartphones gesammelt wurden. Um ein tieferes Verständnis der Smartphone-Daten zu erlangen, wurden interpretierbare Variablen aus den Rohdaten extrahiert und die Vorhersageleistung verschiedener automatisierter Machine Learning-Algorithmen systematisch verglichen.

Die Ergebnisse deuten darauf hin, dass mit dem Smartphone erhobene Daten bestimmte Situations- und Verhaltensindikatoren für psychologische Phänomene im täglichen Leben effektiv erfassen können. In bestimmten Forschungsbereichen wie der Affektforschung sollten Smartphone-Daten jedoch herkömmliche fragebogenbasierte Daten sowie andere Datenquellen wie neurophysiologische Indikatoren lediglich ergänzen, aber nicht vollständig ersetzen. Die Dissertation kommt zudem zu dem Schluss, dass die Verwendung von Smartphone-Sensordaten neue Schwierigkeiten und Herausforderungen für die psychologische Forschung mit sich bringt und die traditionellen Methoden und Perspektiven an ihre Grenzen stoßen. Die Komplexität der Datenerhebung, -verarbeitung und -analyse erfordert etablierte Richtlinien für das Studiendesign, interdisziplinäre Zusammenarbeit sowie theoriegeleitete Forschung, die erklärende und prädiktive Ansätze integriert. Dementsprechend muss weiter untersucht werden, wie Machine Learning-Modelle und andere Big Data-Methoden in der Psychologie mit

traditionellen theoretischen Ansätzen in Einklang gebracht werden können. Nur so können wir dem finalen Ziel der Psychologie näherkommen, menschliches Verhalten und Erleben und deren Zusammenspiel mit alltäglichen Situationen besser verstehen, erklären und vorhersagen zu können.

1. General Introduction

One of the foundations of contemporary psychological research is Kurt Lewin's heuristic equation, which states that an individual's behavior is a function of both their personal characteristics and their environment (B = *f*(P, E)) (Lewin, 1936). While Lewin explicitly declared that either the person or the environment can be more influential in determining behavior in certain situations, an intense person-situation debate about the relative importance of the two factors has dominated the research discourse for several decades (e.g., Epstein & O'Brien, 1985; Kenrick & Funder, 1988; Pervin, 1989). This important scientific controversy has led to a growing body of research being framed around the question of whether the person or the situation characteristics matter more (e.g., Fleeson, 2001; Fleeson & Noftle, 2008; Funder, 2009). As most scholars agree that both factors matter, researchers in psychology, from personality and social psychologists to clinical psychologists, continue to focus on understanding person-related features and their interaction with the environment to better explain and predict human behavior (e.g., Beck & Jackson, 2022; Fleeson & Law, 2015; Sherman et al., 2015).

For example, at the person level, previous research has linked dispositional factors, such as personality and affect traits, to specific behavioral patterns, such as pro-social behavior (e.g., Thielmann et al., 2020) or important long-term life outcomes such as quality of life and longevity (Aknin et al., 2018; Ozer & Benet-Martínez, 2006; Roberts et al., 2007; Steptoe et al., 2009). At the situational level, psychologists broadly agree that the environment influences cognition, emotion, and behavior, but there is less consensus on which specific aspects of the environment exert psychological effects (Fleeson & Noftle, 2008; Rauthmann et al., 2014; Sandstrom et al., 2017). Besides the physical environment itself, such as the location visited (Rauthmann et al., 2014; Rauthmann & Shermann, 2020; Saucier et al., 2007), more subjectively experienced characteristics such as the pleasantness or sociability of a situation can influence personality-related behaviors (e.g., Rauthmann et al., 2014; Rauthmann et al., 2015; Rauthmann & Sherman, 2020).

Therefore, building on such previous studies, this dissertation addresses both 'lenses' of psychological research, the person and the environment. Comprising two empirical studies with different research focuses in psychology, this thesis aims to provide new exploratory insights into the interplay between person- and situation-related characteristics. For this purpose, state-of-the-art digital technologies, that have opened up new possibilities for psychological research, are leveraged.

1.1. Psychological Research in the Digital Age

Psychologists can employ various assessment techniques, both in laboratory and real-life settings to better understand and predict human behavior and the interplay of individuals with their environment. Historically, self-assessments have been a staple in psychological research, but there has been a growing interest in more objective measurements of naturalistic behavior, such as behavioral observations in real-life situations, to improve the accuracy of psychological assessments (Baumeister et al., 2007; Yarkoni, 2012). Advances in digital technology have made it more feasible to collect and analyze vast amounts of objective data in addition to subjective questionnaire data (Yarkoni, 2012). This had led to the rapid development of a new field of research called *ambulatory assessment*, which refers to the study of natural behavior in real-life contexts (Conner & Mehl, 2015). This broader framework comprises passive logging through *mobile sensing* (using the built-in sensors of devices such as smartphones also known as *smartphone sensing*), as well as active logging through smartphone-based questionnaires. With their growing popularity for research purposes, several terms have been established, with *digital phenotyping* being used mainly in the field of mental health research (e.g., Onnela & Rauch, 2016; Raballo, 2018). Although focusing on different fields of applied psychology, these research streams have in common that they integrate computer science (or computational) methods into psychology, which is often referred to as *psychoinformatics* (Montag et al., 2016; Yarkoni, 2012). These new approaches for psychological diagnosis as well as intervention use digital data sources to predict for example a person's personality traits (e.g., Chittaranjan et al., 2011; Stachl et al., 2021) or identify changes in a person's emotional state (e.g., Pal et al., 2021; Tzafilkou et al., 2022).

Especially personality psychologists interested in the manifestation of personality traits in daily life have recently pushed for a new focus on more objective data in psychological assessments (Harari, Vaid, et al., 2020; Montag & Elhai, 2019; Renner et al., 2020; Stachl et al., 2020; 2021). However, new digital technologies also hold great potential for other research areas, ranging from affective sciences to situation research. For example, to measure a person's perception of the sociality level of a situation, the person could either be asked to rate the psychological characteristics of a situation or the occurring communication patterns could be observed using unobtrusively sensed behavioral data, such as in-phone communication or social media activity (e.g., Harari et al., 2016; Harari, Müller, et al., 2020; Lane et al., 2014; Montag et al., 2014). Moreover, mobile sensing holds great potential as a clinical tool for monitoring at-risk populations and advancing mobile health (or mHealth) interventions (Mohr et al., 2017). For instance, the accurate measurement of affective experiences can offer the opportunity to augment mental health care by providing personalized, precise, and preemptive interventions that support insights into patterns of health-related behavior and self-management (e.g., Marzano et al., 2015; Naslund, et al., 2015).

Therefore, the present dissertation investigates the opportunities provided by digital technology in psychological research. Concretely, the research leverages smartphones as a data source, utilizing two smartphone-based data collection methods (experience sampling and smartphone sensing), along with conventional survey measures.

1.1.1. Smartphones as Data Collectors

Although smartphones were not intended to be designed for psychological research, they can be repurposed to collect large amounts of ecologically valid data from global samples. When participants download apps specifically developed for research purposes, smartphones can record where they are, what they are doing, and what they can see and hear, as well as run interactive surveys, tests, and experiments (Harari et al., 2017; Miller, 2012; Montag et al., 2016). Integrating headsets, biosensors, or other peripherals through wireless connections can reveal insightful information on a person's thoughts, feelings, and behaviors. Additionally, the growing use of technological devices like smartphones has also enabled a real-time collection of self-reports in psychology (van Berkel et al., 2017).

Experience Sampling

The *experience sampling* (or *ecological momentary assessment* (EMA)) method enables multiple in situ psychological assessments to be taken throughout the day. Overcoming the limitations of classical laboratory or questionnaire-based studies, there has been a surge of psychological research using EMA approaches over the past decades, ranging from personality psychology (e.g., Matz & Harari, 2021; Müller et al., 2020) and affective sciences (e.g., Kuppens et al., 2022) to clinical psychology (e.g., Ebner-Priemer & Trull, 2009).

Smartphone-based experience samplings offer various benefits for psychological assessments to enhance our understanding of psychological experiences in everyday life (see de Vries et al. (2021) for a comprehensive review). First, real-time assessments can increase data accuracy by reducing memory or recall biases (e.g., Ben-Zeev et al., 2009; Ellison et al., 2020; Neubauer et al., 2020). Moreover, the data is collected in real-world environments, improving the ecological validity and generalizability of research findings (e.g., Scollon et al., 2003; Shiffman et al., 2008). Third, repeated assessments enable to capture the dynamic processes such as the intra-individual changes in well-being (e.g., van der Krieke et al., 2017; Wichers et al., 2016). Lastly, experience samplings can be applied in conjunction with other methods, including psychological, physiological, and behavioral data (Scollon et al., 2003). Such multi-method (or multi-modal) assessments can help identify context-specific relationships between personality, affective experience, and behavior (e.g., Goldstein et al., 2021; Hoemann et al., 2020; Huckins et al., 2019).

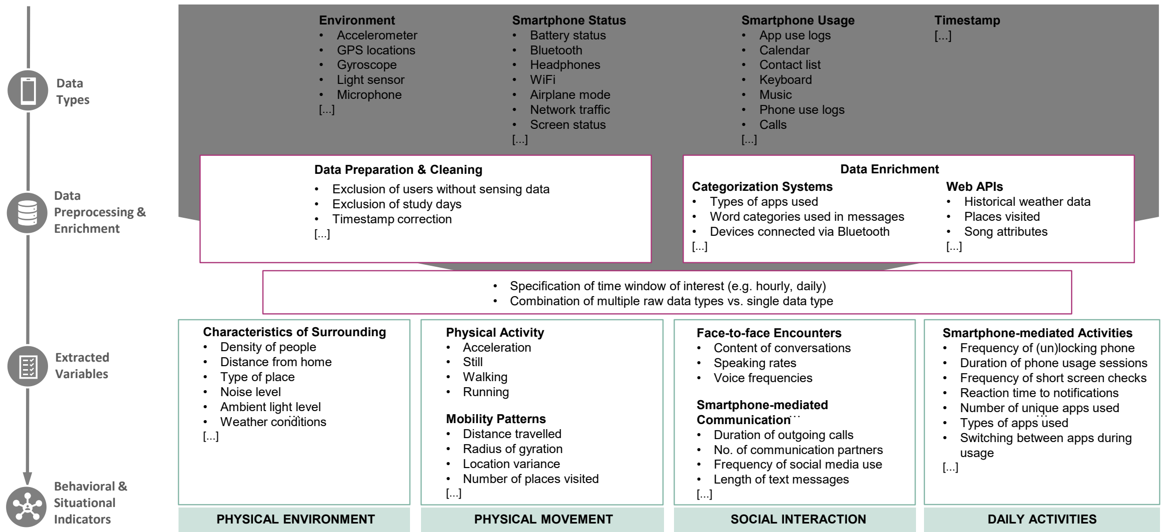
Apart from the tremendous potential, experience sampling also brings new conceptual, methodological, and technological challenges (see Scollon et al. (2003) for a comprehensive review). For example, multiple assessments can result in a high daily burden for participants, risking increased dropouts and low response rates (Shiffman et al., 2008). Accordingly, new psychometric measurements and scales suitable for repeated experience samplings must be developed and validated (de Vries et al., 2021). Furthermore, the technological implementation of the data collection procedure, including a suitable randomized sampling design, can be challenging (van Berkel et al., 2017). Technological challenges include for example strict battery optimization techniques or an appropriate presentation of questionnaires on different devices.

Summing up, although experience sampling approaches hold great new potential, they are still based on self-reporting procedures, which, in turn, are associated with well-known response biases, such as the tendency for people to present themselves in a favorable light or to forget certain details (e.g., Gosling, et al., 1998; Paulhus & Vazire, 2007). Accordingly, self-report measures may not accurately reflect a person's true thoughts, feelings, or behaviors. Therefore, there have been tremendous efforts to collect more objective, behavioral measures, for example by using smartphone sensing data.

Smartphone Sensing

Having established as daily companions, smartphones and their in-built sensors hold great promise to collect a large amount of granular data about their users in everyday life (Harari et al., 2015, 2016; Miller, 2012). Such data can include information about the person's environment (e.g., location, environmental illumination), social interactions (e.g., calls, messages), daily activities (e.g., phone usage, social media activity), or mobility patterns (e.g., distance traveled) (Harari et al., 2016). Collected automatically and unobtrusively in real-world settings, smartphone-generated behavioral and situational data is ecologically valid. The combination of self-reported data and data collected through a smartphone's in-built sensors (e.g., GPS, accelerometer) and log files (e.g., number of calls, app usage) has been used to study various topics. These range from a person's day-night behavior patterns (Schoedel et al., 2020) to the effect of surroundings and activity on mental health (e.g., Ben-Zeev et al., 2015; Canzian & Musolesi, 2015; Lathia et al., 2017; Müller et al., 2020).

Figure 1.1: Data Types, Preprocessing, and Extracted Variables in Smartphone Sensing



Note. This exemplary illustration of data types, extracted variables, and related indicators is not comprehensive. The clustering of smartphone data types and indicators was inspired by Harari et al. (2017) and Schoedel (2020).

Figure 1.1 depicts the types of smartphone sensing data, preprocessing methods, and variables that can be extracted to represent behavioral and situational indicators. GPS data logged by a smartphone, for example, provides information about the environment. Before analysis, data with poor quality (e.g., study days without GPS data due to connection problems) must be removed. The remaining data can be enriched with other data sources using web-based application interfaces (so-called APIs). For example, additional information about the type of a place or the population density of a district can be retrieved for specific points of interest (Mehrotra et al., 2017; Oliveira et al., 2016; Rojas et al., 2016). From this data, variables like the number of restaurant visits can be extracted, providing insights into an individual's mobility patterns. These mobility behaviors can, for example, be used as an indicator of a person's mental health or well-being (Saeb et al., 2017; Sandstrom et al., 2017; R. Wang et al., 2014).

Requiring new skills in app development and data analysis, as well as raising new technical and ethical issues, smartphone sensing is said to transform psychology even more profoundly than previous technological advances (Miller, 2012). Tracking 800 participants for six months can result in one billion activity logs, which correspond to a total data volume of around one terabyte. Accordingly, researchers need to develop new ways of recording, organizing, analyzing, interpreting, and protecting the huge volumes of data that are produced by smartphone sensing apps (Harari et al., 2016; Lazer et al., 2009; Miller, 2012; Montag et al., 2016; 2019). Moreover, there are also some technical limitations. For instance, limited battery power may constrain how many hours per day a mobile sensing app can gather data, especially for energy-demanding GPS data. Connectivity or battery problems can also result in irregular logging intervals and a significant amount of missing data. Sampling rates and logging intervals must be carefully chosen to ensure an appropriate balance between data accuracy and energy efficiency (Miller, 2012; Shepard, et al., 2011; Y. Wang et al., 2009). The complexity of cleaning and analyzing smartphone data due to its high velocity, variety, and volume (the three Vs of big data) requires psychologists to expand their scientific methods by incorporating informatics and statistical methods (Montag et al., 2016; Montag & Elhai, 2019).

1.1.2. Predictive Modeling Method

Therefore, another focus of this work is the application of *predictive modeling* methods to leverage unobtrusively collected smartphone sensing data for psychological research.

The potential of explorative predictive modeling approaches has been highlighted for psychological research for some years now (e.g., Orrù et al., 2020; Yarkoni & Westfall, 2017). While explanatory research aims to understand the underlying mechanisms of psychological phenomena, predictive research focuses on accurately predicting future outcomes (Hofman et al., 2021; Shmueli, 2010; Yarkoni & Westfall, 2017). Predictive modeling and machine learning go hand in hand, as predictive models typically include a machine learning algorithm. Thus, predictive modeling largely overlaps with the field of machine learning and the terms are often used interchangeably (Kuhn & Johnson, 2013, p.1). However, predictive modeling encompasses much more than the tools and techniques for uncovering patterns within data but rather describes the process of developing and training a machine learning model to predict future, unseen data (Kuhn & Johnson, 2013, p.vii). Machine learning methods can efficiently handle large amounts of high-dimensional data by applying automated algorithms. Moreover, comparing the performance of different algorithms can provide valuable insights into the nature of the data being analyzed, such as the linearity of the effects. Besides a clear specification of the practical and/or theoretical utility of accurately predicting the respective outcome, transparency on how the predictive accuracy is evaluated is crucial (Yarkoni & Westfall, 2017). Moreover, the high complexity and low interpretability of machine learning models have often led to warnings for 'black box' models (Pargent & Albert-von der Gönna, 2018; Yarkoni &

Westfall, 2017). Thus, researchers have introduced methods to improve interpretability, allowing scientists to better understand the cause of predictions (e.g., Molnar 2022).

Terminology

Being a highly interdisciplinary field, different terms have been established in research and practice. Inspired by Kuhn and Johnson (2013, p.6), the following terminology is used throughout this dissertation:

- The term *target* variable refers to the criterion or response variable that is predicted.
- Predictive models can be categorized into two types: *unsupervised* machine learning methods (target values unknown) and *supervised* machine learning methods (target values available). This dissertation focuses on supervised learning.
- *Features* or *predictors* describe the input data used in the prediction models and can be continuous, categorical, or binary *variables*.
- The prediction *model* is defined by the specific machine learning *algorithm* and *hyperparameter* setting used. A hyperparameter is a parameter that is set before the learning process begins. These parameters, such as the number of branches in a decision tree, are tunable and can directly affect how well a model trains.
- Predicting categorical (or binary) data is called a *classification* task. The so-called classifier maps the input data to categorical targets. In contrast, *regression* is used for numerical targets, mapping input data to continuous targets using a so-called regressor.

Predictive Modeling Process

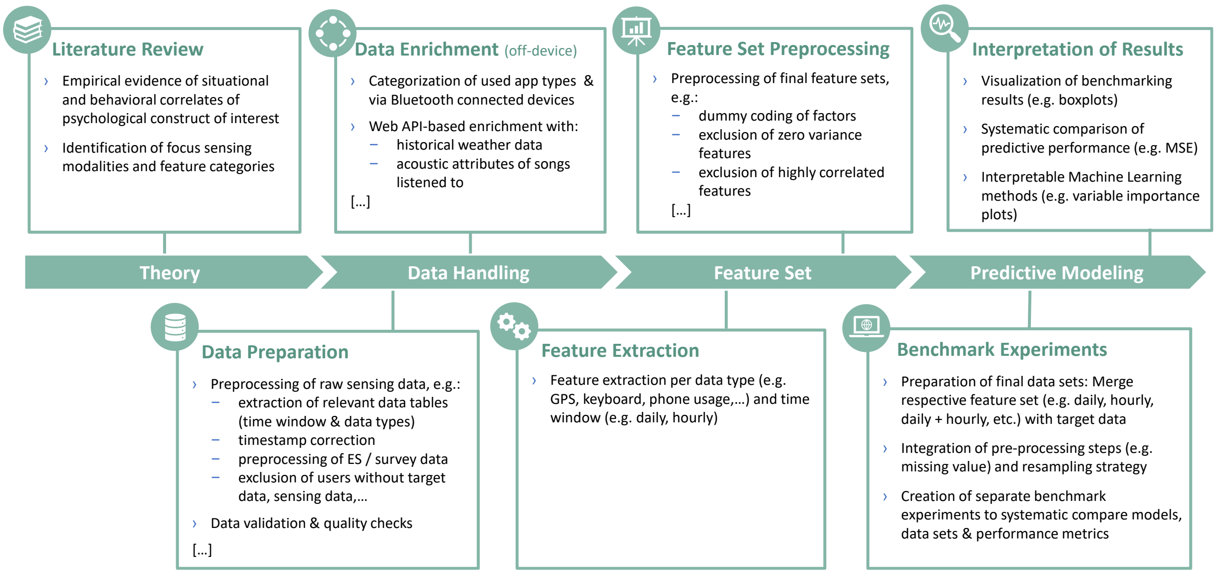
The predictive modeling process includes various phases, a few of which are outlined below. A more in-depth chronology of data preparation, feature engineering, and data analysis steps applied in this dissertation is illustrated in Figure 1.2.[1]

The first step is a thorough *literature review* to gain an overview of smartphone-sensed indicators that can be captured through smartphone sensors and log data. After identifying the relevant sensing data types, raw data is preprocessed in the *data preparation* phase by extracting relevant time windows and ensuring data quality through filter criteria, such as excluding participants or study days with insufficient data. To capture a broad range of behavioral and situational indicators, the smartphone data can be enriched with additional sources such as merging with location information from web-based APIs, during the *data enrichment* phase.

[1] For a further theoretical and methodological introduction to predictive modeling, please refer to James, et al. (2013) as well as Kuhn and Johnson (2013).

The feature engineering process comprises the *feature extraction* per time window and data type of interest.

Figure 1.2: Illustration of the Applied Predictive Modeling Process



The resulting feature sets are then prepared for analysis by ensuring proper coding of variables or excluding features with zero variance. Afterward, the actual predictive modeling process begins. Based on the research question, various data sets are utilized and compared using performance measures suited to the target variable being predicted. To evaluate the effectiveness of preprocessing pipelines, algorithms, and hyperparameter settings, *benchmark experiments* are used to analyze different machine learning models on one or more data sets. To avoid overfitting, resampling strategies split the data set into training and testing portions. Model performance can be assessed using various performance measures and visualizing the results offers insight into the interpretation of the results, such as the importance of specific features.

Machine learning approaches are increasingly being introduced in applied psychology, particularly in psychological assessment. So far, psychologists have especially focused on using machine learning in the field of personality prediction, for instance using digital traces on social media (e.g., Azucar et al., 2018; Settanni et al., 2018) or smartphone sensing data (e.g., Stachl et al., 2020; Y. Wang et al., 2018). However, when trying to understand the dynamics of complex psychological phenomena like affective experience in daily life, incorporating predictive approaches can be equally promising (Kuppens et al., 2022).

1.2. Studies of this Dissertation

1.2.1. Rationale

In summary, the field of psychoinformatics has made significant progress in recent years and has been established particularly in personality psychology. However, there is still huge potential for other areas of psychological research to benefit from this advancement. Leveraging new digital technologies to collect and analyze large amounts of unobtrusively sensed data, this dissertation pursues two goals.

First, this dissertation seeks to provide a comprehensive investigation of the capabilities of smartphone sensing for psychological research. By applying smartphone sensing methods to situation research (Study 1) and affect research (Study 2), this work aims to explore the potential of new digital technologies for research questions beyond the concept of personality. Previous research has already begun to investigate the relationship between specific types of smartphone-sensed data and psychological phenomena, such as GPS data and mental well-being (Müller et al., 2020) or smartphone-mediated communication and sociability (Harari, Müller, et al., 2020). While there is a growing body of large-scale smartphone sensing studies, particularly in personality computing (e.g., Phan & Rauthmann, 2021; Rüegger et al., 2020; Stachl et al. 2020), research often still focuses on examining a handful of sensing data types - probably also due to the technical limitations as well as the high effort and expertise required for data collection and analysis. Using data collected with the PhoneStudy app during the Smartphone Sensing Panel Study (SSPS; Schoedel & Oldemeier, 2020), the studies of the present dissertation comprise a large number of different smartphone-sensed data types. Moreover, by combining these smartphone logs with smartphone-based experience sampling as well as classical questionnaire data, this dissertation highlights the importance of sophisticated multi-method study designs to gain valuable insights into individual behavior, thinking, and feeling in everyday life.

Second, following Yarkoni and Westfall's (2017) call for psychological research to become a more predictive rather than purely explanatory science, principles, and techniques from the field of predictive modeling are applied. The use of machine learning algorithms facilitates efficient analysis of the large, high-dimensional data sets of this dissertation. As the collected sensor data consisted of time-stamped event data, it was preprocessed in a sophisticated feature engineering process to extract interpretable variables for the applied predictive modeling approach. Furthermore, interpretable machine learning algorithms were selected and compared in terms of their predictive performances to gain deeper insights into the nature of the smartphone sensing data.

Specifically, two empirical studies focusing on different psychological constructs were conducted as part of this dissertation project.

1.2.2. Study 1: Sensing Psychological Situation

Person-situation-behavior dynamics are influenced by the perceived characteristics of a situation (also called psychological situation; Rauthmann et al., 2014; Rauthmann et al., 2015). Therefore, Study 1 leverages smartphone-sensed data to predict an individual's in situ ratings of psychological characteristics of situations, such as the perceived level of sociability or intellectuality, in daily life. Previous literature using self-reports provided a selection of situational and behavioral indicators of psychological situations, which can be translated into meaningful smartphone-sensed variables. A supervised machine learning approach was applied to explore the predictability of the perceived psychological situation by introducing exploratory predictive methods to situation research. This study is a first of its kind in combining insights from smartphone sensing research and situation research.

1.2.3. Study 2: Sensing Affective Experience

Given its practical relevance for mental health research, Study 2 explores the application of smartphone sensing approaches in the field of affective computing. Concretely, the study exploratory investigates if affective experience in daily life can be predicted from unobtrusively collected smartphone data combining different types of sensing data. For this purpose, a broad range of smartphone-sensed indicators was derived from previous research on mental well-being and affect. Afterward, supervised machine learning algorithms were used to predict individuals' self-reported affect states and traits. Combining smartphone sensing with experience sampling, the study utilized a multi-method design to advance the interdisciplinary field of affective computing.

1.2.4. PhoneStudy Project

PhoneStudy Research App

The present dissertation uses data collected via a smartphone application (app) developed by the PhoneStudy project, an interdisciplinary project of psychologists, statisticians, and computer scientists at the Ludwig-Maximilian-University Munich.[2] The PhoneStudy research app enables the integration of research into everyday life and can be applied by social and behavioral scientists to study a large range of scientific questions. The Android-based smartphone app provides measurements of a variety of variables via tracking smartphone-use

[2] A detailed description of the research project can be found on the project's website https://phonestudy.org/.

patterns, such as call behavior, application use (e.g., social media), GPS, and many others. In addition, self-report questionnaires for conducting experience sampling of self-report data can be administered.

Smartphone Sensing Panel Study

The data for the described studies were collected during the *Smartphone Sensing Panel Study* (SSPS), which was carried out by the Ludwig-Maximilian-University (LMU) Munich in cooperation with the Leibniz Center for Psychological Information and Documentation (ZPID). The study was taking place from May 15th to November 11th, 2020. A quota sample of $N = 850$ persons was recruited and participants were asked to install the *PhoneStudy* research application (app) for three or six months. This app, which was available for Android OS version 5 or higher, has been developed by the *PhoneStudy* team of LMU Munich to continuously collect sensor and usage data in the background. The study procedure has been approved by the institutional ethics committees and review boards of LMU and ZPID and was conducted according to EU laws and ethical standards. The panel study combines three data collection modalities: (1) smartphone sensing, (2) experience sampling, and (3) monthly online surveys. Thus, the data set included high-dimensional and longitudinal behavioral and situational sensing data, in situ self-report data, as well as traditional questionnaire data measuring a variety of psychological traits and phenomena. A detailed study protocol can be also found in Schoedel and Oldemeier (2020).

1.2.5. Research Transparency and Openness

This dissertation embraces the values of openness and transparency in science. The studies included in this dissertation project follow the Journal Article Reporting Standards (JARS) for Quantitative Research in Psychology (Appelbaum et al., 2018). Moreover, the research questions and methodological procedures of the exploratory studies of this dissertation were preregistered before data analysis. All deviations from the preregistrations are described in detail in the respective online supplemental material. For data protection reasons, the raw data from smartphone sensors cannot be made publicly available. While access to the aggregated data sets is provided upon request, the code scripts for data exclusions, preprocessing, measures, and analysis, are published together with various supplemental materials on the respective Open Science Framework (OSF) project pages to allow the reproducibility of the studies. In addition, a detailed codebook was created for each study to provide interested researchers with as much information as possible on the data sets and procedures.

1.3. References

Aknin, L. B., Van de Vondervoort, J. W., & Hamlin, J. K. (2018). Positive feelings reward and promote prosocial behavior. *Current opinion in psychology, 20*, 55-59. https://doi.org/10.1016/j.copsyc.2017.08.017

American Psychological Association. (n.d.). Subjective well-being (SWB). In *APA dictionary of psychology*. Retrieved February 18, 2022, from https://dictionary.apa.org/subjective-well-being

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3. https://doi.org/10.1037/amp0000191

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and individual differences, 124*, 150-159. https://doi.org/10.1016/j.paid.2017.12.018

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. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Beck, E. D., & Jackson, J. J. (2022). Personalized Prediction of Behaviors and Experiences: An Idiographic Person-Situation Test. *Psychological Science, 33*(10), 1767-1782. https://doi.org/10.1177/09567976221093

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric rehabilitation journal, 38*(3), 218. https://doi.org/10.1037/prj0000130

Ben-Zeev, D., Young, M. A., & Madsen, J. W. (2009). Retrospective recall of affect in clinically depressed individuals and controls. *Cognition and Emotion, 23*(5), 1021-1040. https://doi.org/10.1080/02699930802607937

Canzian, L., & Musolesi, M. (2015, September). Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 1293-1304).

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2011). Who's who with big-five: Analyzing and classifying personality traits with smartphones. *5th Annual International Symposium on Wearable Computers*, 29-36. https://doi.org/10.1109/ISWC.2011.29

Conner, T. S., & Mehl, M. R. (2015). Ambulatory assessment: Methods for studying everyday life. In R. A. Scott, M. C. Buchmann, & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences* (pp. 1-15). https://doi.org/10.1002/9781118900772.etrds0010

De Vries, L. P., Baselmans, B. M., & Bartels, M. (2021). Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies. *Journal of Happiness Studies, 22*(5), 2361-2408. http://dx.doi.org/10.1007/s10902-020-00324-7

Ebner-Priemer, U. W., & Trull, T. J. (2009). Ambulatory assessment: An innovative and promising approach for clinical psychology. *European psychologist, 14*(2), 109-119. https://doi.org/10.1027/1016-9040.14.2.109

Ellison, W. D., Trahan, A. C., Pinzon, J. C., Gillespie, M. E., Simmons, L. M., & King, K. Y. (2020). For whom, and for what, is experience sampling more accurate than retrospective report?. *Personality and Individual Differences, 163*, 110071. https://doi.org/10.1016/j.paid.2020.110071

Epstein, S., & O'Brien, E. J. (1985). The person-situation debate in historical and current perspective. *Psychological Bulletin, 98*(3), 513-537. https://doi.org/10.1037/0033-2909.98.3.513

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of personality and social psychology, 80*(6), 1011 -1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W., & Law, M. K. (2015). Trait enactments as density distributions: The role of actors, situations, and observers in explaining stability and variability. *Journal of personality and social psychology, 109*(6), 1090-1104. https://doi.org/10.1037/a0039517

Fleeson, W., & Noftle, E. (2008). The end of the person-situation debate: An emerging synthesis in the answer to the consistency question. *Social and Personality Psychology Compass, 2*(4), 1667-1684. https://doi.org/10.1111/j.1751-9004.2008.00122.x

Funder, D. C. (2009). Persons, behaviors and situations: An agenda for personality psychology in the postwar era. *Journal of Research in Personality, 43*(2), 120-126. https://doi.org/10.1016/j.jrp.2008.12.041

Goldstein, S. P., Hoover, A., Evans, E. W., & Thomas, J. G. (2021). Combining ecological momentary assessment, wrist-based eating detection, and dietary assessment to characterize dietary lapse: A multi-method study protocol. *Digital Health, 7*, 2055207620988212. https://doi.org/10.1177/20552076209882

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. https://doi.org/ 10.1037//0022-3514.74.5.1337.

Harari, G. M., Gosling, S. D., Wang, R., & Campbell, A. T. (2015). Capturing Situational Information with Smartphones and Mobile Sensing Methods. *European Journal of Personality, 29*(5), 509-511. https://doi.org/10.1002/per.2032

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science, 11*(6), 838-854. https://doi.org/10.1177/174569161665028

Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences, 18*, 83-90. https://doi.org/10.1016/j.cobeha.2017.07.018

Harari, G. M., Müller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., ... & Gosling, S. D. (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of personality and social psychology, 119*(1), 204-228. https://doi.org/ 10.1037/pspp0000245

Harari, G. M., Vaid, S. S., Müller, S. R., Stachl, C., Marrero, Z., Schoedel, R., ... & Gosling, S. D. (2020). Personality sensing for theory development and assessment in the digital age. *European Journal of Personality, 34*(5), 649-669. https://doi.org/10.1002/per.2273

Hoemann, K., Khan, Z., Feldman, M. J., Nielson, C., Devlin, M., Dy, J., ... & Quigley, K. S. (2020). Context-aware experience sampling reveals the scale of variation in affective experience. *Scientific reports, 10*(1), 1-16.| https://doi.org/10.1038/s41598-020-69180-y

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., ... & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature, 595*(7866), 181-188. https://doi.org/ 10.1038/s41586-021-03659-0.

Huckins, J. F., DaSilva, A. W., Wang, R., Wang, W., Hedlund, E. L., Murphy, E. I., ... & Campbell, A. T. (2019). Fusing mobile phone sensing and brain imaging to assess depression in college students. *Frontiers in neuroscience, 13*, 248. https://doi.org/10.3389/fnins.2019.00248

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). New York City, NY: Springer.

Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American psychologist, 43*(1), 23-34. Https://doi.org/ 10.1037/0003-066X.43.1.23

Kuhn, M. & Johnson, K. (2013). *Applied predictive modeling* (Vol. 26). New York City, NY: Springer.

Kuppens, P., Dejonckheere, E., Kalokerinos, E. K., & Koval, P. (2022). Some recommendations on the use of daily life methods in affective science. *Affective Science*, *3*(2), 505-515. Https://doi.org/ 10.1007/s42761-022-00101-0

Lane, N., Lin, M., Rabbi, M., Yang, X., lu, H., Cardone, G., Ali, S., Doryab, A., Berke, E., Campbell, A., & Choudhury, T. (2014). BeWell: Sensing sleep, physical activities and social interactions to promote wellbeing. *Mobile Networks and Applications, 19*(3), 345-359. https://doi.org/10.1007/s11036-013-0484-5

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: Using smartphones to link happiness and physical activity. *PloS one, 12*(1), e0160589. https://doi.org/10.1371/journal.pone.0160589

Lazer, D. M., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... & Jebara, T. (2009). Computational Social Science. *Science, 323*, 721-723. https://doi.org/10.1126/science.1167742

Lewin, K. (1936). *Principles of topological psychology*. New York-London. New York: McGraw Hill.

Marzano, L., Bardill, A., Fields, B., Herd, K., Veale, D., Grey, N., & Moran, P. (2015). The application of mHealth to mental health: opportunities and challenges. *The Lancet Psychiatry, 2*(10), 942-948. http://dx.doi.org/10.1016/S2215-0366(15)00268-0

Matz, S. C., & Harari, G. M. (2021). Personality-place transactions: Mapping the relationships between Big Five personality traits, states, and daily places. *Journal of Personality and Social Psychology, 120*(5), 1367. https://doi.org/ 10.1037/pspp0000297

Mehrotra, A., Müller, S. R., Harari, G. M., Gosling, S. D., Mascolo, C., Musolesi, M., & Rentfrow, P. J. (2017). Understanding the role of places and activities on mobile phone interaction and usage patterns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*(3), 1-22. https://doi.org/10.1145/3131901

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science, 7*(3), 221-237. https://doi.org/10.1177/1745691612441215

Mohr, D. C., Tomasino, K. N., Lattie, E. G., Palac, H. L., Kwasny, M. J., Weingardt, K., ... & Schueller, S. M. (2017). IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. *Journal of medical Internet research, 19(*1), e6645. http://dx.doi.org/10.2196/jmir.6645

Molnar, C. (2022). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Retrieved from https://christophm.github.io/interpretable-ml-book/

Montag, C., Baumeister, H., Kannen, C., Sariyska, R., Meßner, E. M., & Brand, M. (2019). Concept, possibilities and pilot-testing of a new smartphone application for the social and life sciences to study human behavior including validation data from personality psychology. *Multidisciplinary Scientific Journal, 2*(2), 102-115. https:/doi.org/10.3390/j2020008

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. https://doi.org/10.1027/ 1614-0001/ a000139

Montag, C., Duke, É., & Markowetz, A. (2016). Toward psychoinformatics: Computer science meets psychology. *Computational and mathematical methods in medicine*, *2016*, 1-10. https://doi.org/10.1155/2016/2983685

Montag, C., & Elhai, J. D. (2019). A new agenda for personality psychology in the digital age?. *Personality and Individual Differences, 147*, 128-134.

Müller, S. R., Peters, H., Matz, S. C., Wang, W., & Harari, G. M. (2020). Investigating the relationships between mobility behaviours and indicators of subjective well-being using smartphone-based experience sampling and GPS tracking. *European Journal of Personality, 34*(5), 714-732. http://dx.doi.org/10.1002/per.2262

Naslund, J. A., Marsch, L. A., McHugo, G. J., & Bartels, S. J. (2015). Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. *Journal of mental health, 24*(5), 321-332. http://dx.doi.org/10.3109/09638237.2015.1019054

Neubauer, A. B., Scott, S. B., Sliwinski, M. J., & Smyth, J. M. (2020). How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *Journal of personality and social psychology, 119*(1), 185-203. https://doi.org/10.1037/pspp0000248

Oliveira, E. M. R., Viana, A. C., Sarraute, C., Brea, J., & Alvarez-Hamelin, I. (2016). On the regularity of human mobility. *Pervasive and Mobile Computing, 33*, 73-90. https://doi.org/10.1016/j.pmcj.2016.04.005

Onnela, J. P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology, 41*(7), 1691-1696. https://doi.org/10.1038/npp.2016.7

Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine learning in psychometrics and psychological research. *Frontiers in psychology, 10*, 2970. https://doi.org/10.3389/fpsyg.2019.02970

Ozer, D. J., & Benet-Martínez, V. (2006). Personality and the prediction of consequential outcomes. *Annual review of psychology, 57*, 401. https://doi.org/ 10.1146/annurev.psych.57.102904.190127

Pal, S., Mukhopadhyay, S., & Suryadevara, N. (2021). Development and progress in sensors and technologies for human emotion recognition. *Sensors, 21*(16), 5554. https://doi.org/10.3390/s21165554

Pargent, F., & Albert-von der Gönna, J. (2018). Predictive Modeling With Psychological Panel Data. *Zeitschrift Für Psychologie, 226* (4), 246-258. https://doi.org/10.1027/ 2151-2604/ a000343

Paulhus D. L., Vazire S. (2007). The self-report method. In Robins R. W., Fraley R. C., Kruger R. F. (Eds.), *Handbook of research methods in personality psychology* (pp. 224-239). New York, NY: Guilford.

Pervin, L. A. (1989). Persons, situations, interactions: The history of a controversy and a discussion of theoretical models. *Academy of Management Review, 14*(3), 350-360. https://doi.org/10.5465/amr.1989.4279066

Raballo, A. (2018). Digital phenotyping: an overarching framework to capture our extended mental states. *The lancet. Psychiatry, 5*(3), 194-195. https://doi.org/ 10.1016/S2215-0366(18)30054-3

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., ... & Funder, D. C. (2014). The Situational Eight DIAMONDS: a taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology, 107*(4), 677. https://doi.org/10.1037/a0037250

Rauthmann, J. F., & Sherman, R. A. (2020). The situation of situation research: Knowns and unknowns. *Current Directions in Psychological Science, 29(*5), 473-480. https://doi.org/10.1177/0963721420925546

Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality, 29*(3), 363-381. https://doi.org/10.1002/per.1994

Renner, K. H., Klee, S., & von Oertzen, T. (2020). Bringing back the person into behavioural personality science using big data. *European Journal of Personality, 34*(5), 670-686. http://dx.doi.org/10.1002/per.2303

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), 313-345. https://doi.org/ 10.1111/j.1745-6916.2007.00047.x.

Rojas IV, M. B., Sadeghvaziri, E., & Jin, X. (2016). Comprehensive review of travel behavior and mobility pattern studies that used mobile phone data. Trans*portation Research Record, 2563*(1), 71-79. https://doi.org/10.3141/2563-11

Rüegger, D., Stieger, M., Nißen, M., Allemand, M., Fleisch, E., & Kowatsch, T. (2020). How are personality states associated with smartphone data? *European Journal of Personality, 34*(5), 687-713. https://doi.org/10.1002/per.2309

Saeb S., Lattie E. G., Kording K. P., & Mohr D. C. (2017). Mobile phone detection of semantic location and its relationship to depression and anxiety. *JMIR mHealth and uHealth, 5*, e112. https://doi.org/10.2196/mhealth.7297.

Sandstrom, G. M., Lathia, N., Mascolo, C., & Rentfrow, P. J. (2017). Putting mood in context: Using smartphones to examine how people feel in different locations. *Journal of Research in Personality, 69*, 96-101. http://dx.doi.org/10.1016/j.jrp.2016.06.004

Saucier, G., Bel-Bahar, T., & Fernandez, C. (2007). What modifies the expression of personality tendencies? Defining basic domains of situation variables. *Journal of personality, 75*(3), 479-504. https://doi.org/10.1111/j.1467-6494.2007.00446.x

Schoedel, R. (2020). *At the Interface of Personality Psychology and Computational Science* [Doctoral Dissertation, Ludwig-Maximilians-University Munich].

Schoedel, R., & Oldemeier, M. (2020). Basic protocol: Smartphone sensing panel study. Leibniz Institut für Psychologische Information und Dokumentation (ZPID). https://doi.org//10.23668/PSYCHARCHIVES.2901

Schoedel, R., Pargent, F., Au, Q., Völkel, S. T., Schuwerk, T., Bühner, M., & Stachl, C. (2020). To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day-night behaviour patterns. *European Journal of Personality, 34*(5), 733-752. https://doi.org/10.1002/per.2258

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Strengths and weaknesses of the experience sampling technique. *Journal of Happiness Studies, 4*, 5-34. https://doi.org/10.1023/A:1023605205115

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking, 21*(4), 217-228. https://doi.org/10.1089/cyber.2017.0384

Shepard, C., Rahmati, A., Tossell, C., Zhong, L., & Kortum, P. (2011). *LiveLab: Measuring wireless networks and smartphone users in the field (Technical Report)*. Rice University, Texas. Retrieved from http://www.ruf.rice.edu/~mobile/publications/shepard10hotmetrics.pdf

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology, 109*(5), 872-888. https://doi.org/10.1037/pspp0000036

Shmueli, G. (2010). "To Explain or to Predict?" *Statist. Sci. 25*(3), 289 - 310. https://doi.org/10.1214/10-STS330

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., ... & Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality, 34*(5), 613-631. https://doi.org/10.1002/per.2257

Stachl, C., Boyd, R. L., Horstmann, K. T., Khambatta, P., Matz, S. C., & Harari, G. M. (2021). Computational personality assessment. *Personality Science, 2*, 1-22. https://doi.org/10.5964/ps.6115

Steptoe, A., Dockray, S., & Wardle, J. (2009). Positive affect and psychobiological processes relevant to health. *Journal of personality, 77*(6), 1747-1776. https://doi.org/10.1111/j.1467-6494.2009.00599.x

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology, 4*, 1-32. http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091415

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological bulletin, 146*(1), 30-90. http://dx.doi.org/10.1037/bul0000217

Tzafilkou, K., Economides, A. A., & Protogeros, N. (2022). Mobile sensing for emotion recognition in smartphones: a literature review on non-intrusive methodologies.

International Journal of Human-Computer Interaction, 38(11), 1037-1051. https://doi.org/10.1080/10447318.2021.1979290

Van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys (CSUR), 50*(6), 1-40. https://doi.org/10.1145/3123988

van der Krieke, L., Blaauw, F. J., Emerencia, A. C., Schenk, H. M., Slaets, J. P., Bos, E. H., et al. (2017). Temporal dynamics of health and well-being: A crowdsourcing approach to momentary assessments and automated generation of personalized feedback. *Psychosomatic Medicine, 79*(2), 213-223. https://doi.org/10.1097/PSY.0000000000000378

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A. T. (2014). StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3-14. https://doi.org/10.1145/2632048.2632054

Wang, W., Harari, G. M., Wang, R., Müller, S. R., Mirjafari, S., Masaba, K., & Campbell, A. T. (2018). Sensing behavioral change over time: Using within-person variability features from Mobile sensing to predict personality traits. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 2*, 1-21. https://doi.org/10.1145/3264951

Wang, Y., Lin, J., Annavaram, M., Jacobson, Q. A., Hong, J., Krishnamachari, B., & Sadeh, N. (2009). A framework of energy efficient mobile sensing for automatic user state recognition. In *Proceedings of MobiSys '09* (pp. 179-192). New York, NY: ACM.

Wichers, M., Groot, P. C., Psychosystems, ESM-Group, & EWS-Group. (2016). Critical slowing down as a personalized early warning signal for depression. *Psychotherapy and Psychosomatics, 85*, 114-116. https://doi.org/10.1159/000441458

Yarkoni, T. (2012). Psychoinformatics. *Current Directions in Psychological Science, 21*(6), 391-397. https://doi.org/10.1177/0963721412457362

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

2. Study 1: Sensing Psychological Situation

Predicting Psychological Situations in Daily Life

Using Smartphone Sensing

Note: Preregistration for this study is available at http://dx.doi.org/10.23668/psycharchives.4928 and the preprocessing and analysis code, as well as supplemental materials of this manuscript, are published on the corresponding Open Science Framework: https://osf.io/kwd82/. The password to access the aggregated data can be provided upon request.

At the time of the submission of this dissertation, a revised version of this manuscript was accepted for publication in the Journal of Personality and Social Psychology (JPSP). I contributed to the accepted paper as the second author, with the following roles according to the credit taxonomy (https://credit.nis.org): Data curation, Formal analysis, Methodology, Validation, Writing - original draft.

2.1. Abstract

The investigation of the perceived psychological situation (i.e., situational characteristics) and its objective counterparts (i.e., situational cues) is crucial for the comprehensive understanding of the dynamic interplay between an individual's characteristics and behavior. However, little is known about the relationship between perceived psychological situations and their specific situational cues in everyday life. Our study contributes to situation research by introducing smartphone sensing for the objective assessment of a wide range of situational cues in everyday situations. Specifically, we investigated whether smartphone-generated situational data can predict individuals' ratings of perceived situational characteristics. Combining smartphone sensing and experience sampling data collected over two weeks, our data set comprised 675 sensed indicators for situational cues and a total of 11,506 situational ratings from 510 participants. Results of an explorative machine learning approach indicate that certain dimensions of psychological situations (e.g., Duty, Intellectuality, Mating, Sociality) can be predicted with moderate accuracy from smartphone sensing data (e.g., logged events of connectivity, phone usage, mobility/activity, and timestamps). Our study provides initial evidence for the relationship between perceived characteristics and objectively assessed cues. In addition, our findings provide novel insights into the correlates of psychological situations and highlight the potential of smartphone sensing for situation research.

Keywords: psychological situation, situational characteristics, situational cues, smartphone sensing, predictive modeling, machine learning

2.2. Introduction

Individuals' behaviors are embedded in a situational context. Accordingly, psychological researchers assume that not only person-related variables (e.g., personality traits) but also situational aspects shape human behavior (Asendorpf & Rauthmann, 2020). While behavior has been regarded as a function of person and situation for almost a century (e.g., Funder, 2006; Lewin, 1936; Schmitt et al., 2013), the investigation of situations has traditionally received little attention (Hogan, 2009). Acknowledging this gap, personality science has recently witnessed an increase in the study of situations resulting in clarity on the conceptualization and assessment of situations. Yet, situations in daily life are still a big unknown (Rauthmann & Sherman, 2020).

In this context, smartphone sensing methods, which are increasingly utilized in psychological research, seem to hold great potential for investigating situations (Harari et al., 2016; Harari, Müller, & Gosling, 2020; Wrzus & Mehl, 2020). While smartphone sensing has been successfully applied to assess personality traits (e.g., Harari, Müller, Stachl, Au, et al., 2020; Montag et al., 2014, 2015; Mønsted, et al., 2018; Schoedel et al., 2018; Stachl, Au, et al., 2020; Stachl et al., 2017), its application in psychological situation research is still pending (Wrzus & Mehl, 2020).

2.2.1. The Psychological Situation

In general, situations can be conceptualized in terms of the information they offer to an individual (Rauthmann et al., 2014; Rauthmann, Sherman, & Funder, 2015). The situation consists of physical and objectively quantifiable stimuli, so-called *situational cues* (e.g., the presence of others, the ringing of a phone, the workplace, or the current time of the day). These cues may be perceived and interpreted by the individual experiencing the situation, resulting in psychological *situational characteristics* (e.g., whether work needs to be done or how intellectually stimulating a situation is; Rauthmann et al., 2014, Rauthmann, Sherman, & Funder, 2015). Accordingly, situational cues do not possess intrinsic psychological meaning but represent the immediate objective environment of an individual. In contrast, situational characteristics are based on perceived and processed cues and capture the cognitive representation of the situation (i.e., psychological situation) and its attributes (e.g., dutiful, positive, or social; Rauthmann et al., 2014, Rauthmann, Sherman, & Funder, 2015).

Depending on the psychological situation, specific behaviors are presented which, in turn, can impact situational cues (Brown et al., 2017). Accordingly, psychological situation characteristics have been shown to predict intra-individual differences in behavior (e.g., Horstmann et al., 2021; Sherman et al., 2015). For example, higher levels of perceived Sociality were related to more extroverted behavior, while higher levels of perceived Deception were associated with less honest and humble behavior (Sherman et al., 2015).

Situational Characteristics

The characteristics of a situation are crucial for understanding and modeling the dynamics between a person, situation, and behavior (Rautmann et al., 2015a). Previous situation research in psychology focused on investigating the content and structure of situational characteristics. One central conclusion was that individuals describe (and presumably perceive) characteristics of everyday situations on various dimensions. Accordingly, several taxonomies of situation characteristics have emerged, differing in both the number and labels of situational dimensions (e.g., Brown, et al., 2015; Gerpott et al., 2018; Griffo & Colvin, 2019; Oreg, et al., 2020; Parrigon et al., 2017; Rauthmann et al., 2014; Ziegler et al., 2019; see also Rauthmann and Sherman (2020) for an overview).

The prevailing taxonomy appears to be the situational eight DIAMONDS, which integrates the most commonly identified dimensions from literature and enables the reliable assessment of individuals' psychological situations (Rauthmann et al., 2014; Rauthmann, Sherman, & Funder, 2015). In multiple cross-cultural studies, eight major dimensions for perceiving daily situations were identified: (1) *Duty* (i.e., Does something need to be done?), (2) *Intellect* (i.e., Is deep thinking required or desired?), (3) *Adversity* (i.e., Are there external threats?), (4) *Mating* (i.e., Is the situation sexually or romantically charged?), (5) *pOsitivity* (i.e., Is the situation enjoyable?), (6) *Negativity* (i.e., Does the situation elicit unpleasant feelings?), (7) *Deception* (i.e., Is someone being untruthful or dishonest?), and (8) *Sociality* (i.e., Are social interaction and relationship formation possible, required or desired?; Rauthmann et al., 2014; Rauthmann, Sherman, & Funder, 2015).

Situational Cues

The DIAMONDS taxonomy has been a fertile ground for empirical research fostering a better understanding of how individual behavior unfolds within different situations. This research investigated whether and how objective situational cues are associated with perceived situational characteristics (e.g., Blake et al., 2020; Rauthmann et al., 2014; Serfass & Sherman, 2015).

As shown in Table 1.1, the most common and comprehensive way to categorize different cues of everyday situations is to distinguish (1) *persons/interactions,* (2) *objects*, (3) *events/activities*, (4) *places*, and (5) *time* (Harari, Müller, & Gosling, 2020; Rauthmann &

Sherman, 2020). For each of these cue types, associations with situational characteristics have been identified. For (1) *persons/interactions,* the sheer presence of people was related to the perceived level of a situation's Mating, pOsitivity, Negativity, and Sociality (Rauthmann et al., 2014). Moreover, several studies reported that socialness and talkativeness are associated with a situation's Duty, Mating, pOsitivity, Negativity, and Sociality (Breil et al., 2019; Horstmann et al., 2021; Rauthmann et al., 2014). Regarding (2) *objects*, studies found that the presence of music is related to a situation's Intellect and Sociality (Blake et al., 2020; Rauthmann et al., 2014). As an example of (3) *events/activities*, the activity of commuting was associated with how dutiful, intellectual, adverse, and positive situations were perceived (Rauthmann et al., 2014). Similarly, engagement in physical activities like moving around was reported to be linked to a situation's level of Duty, Intellect, Mating, pOsitivity, and Sociality (Rauthmann et al., 2014). In addition, the (4) *place* of a situation was found to be systematically related to individuals' psychological situation. For example, bars, cafes, and restaurants were associated with a situation's Duty, Mating, pOsitivity, Negativity, and Sociality (Breil et al., 2019; Rauthmann et al., 2014). Lastly, the weekday and daytime of a situation can be affiliated with the situation's Duty, pOsitivity, Negativity, and Sociality ((5) *time*; Serfass & Sherman, 2015).

2.2.2. Smartphone Sensing in Situation Research

Sensing Situational Cues

The above-described situational cues can be measured objectively (e.g., via cameras, microphones, life-logging systems, and sensors; Brown et al., 2017) or subjectively (e.g., inquiring about perceived or remembered cues from participants; Rauthmann & Sherman, 2020). However, previous studies have mostly relied on self-report measures, collecting situational variables via either day reconstruction (e.g., Rauthmann et al., 2014) or experience sampling (e.g., Breil et al., 2019; Horstmann et al., 2021; Sherman et al., 2015). Such self-reports are prone to a series of well-known biases and may differ significantly from the actual cues (e.g., Baumeister et al., 2007; Rauthmann et al., 2014; Schwarz, 2012).

A novel way to collect more objective information about individuals' situations is through smartphone sensing (Wrzus & Mehl, 2020). Custom research applications can access the native mobile sensors and system logs embedded in a user's device and retrieve data from a variety of modalities (e.g., GPS, app usage, accelerometer; Harari et al., 2016). This smartphone sensing approach allows for an unobtrusive and ecologically valid assessment of an individual's current environment (e.g., location, connected Bluetooth devices) and activities (e.g., phone usage, movement patterns) in a scalable manner (Harari, Müller, & Gosling, 2020; Miller, 2012).

Concerning situational cues for (1) *persons/interactions*, smartphone sensing research has successfully predicted social interactions from combinations of multiple data types such as ambient noise, Bluetooth environment, or WiFi environment (e.g., Lane et al., 2014; Harari et al., 2017; Rüegger et al., 2020). An example of sensing (2) *objects* is the detection of music. The presence of music can be inferred not only by logging a smartphone's internal audio player records (e.g., Harari, Müller, & Gosling, 2020) but also by using specially developed sound classifiers to analyze ambient noise (e.g., Lu et al., 2009). Regarding (3) *events/activities*, individuals' activity levels (e.g., jogging or cycling) have previously been inferred via data from smartphones' accelerometers and GPS logs (e.g., Y.-P. Chen et al., 2008; Do & Gatiza-Perez, 2014; Lane et al., 2014; R. Wang et al., 2014). Furthermore, smartphone usage itself has repeatedly been found to be highly context-specific, suggesting that smartphone usage behaviors may also be associated with individuals' psychological situations (e.g., Karikoski & Soikkeli, 2013; Oulasvirta et al., 2012; Xu et al., 2013). For example, social media postings (i.e., tweets), which are frequently made via smartphone, have been found to predict situational characteristics (Serfass & Sherman, 2015). Furthermore, smartphones accompany their users almost 90% of the time, making them a good proxy for their (4) *location* (Dey et al., 2011). Accordingly, patterns in a smartphone's GPS logs (e.g., Do & Gatica-Perez, 2014; Oliveira et al., 2016; Wolf & Jacobs, 2010), Bluetooth environment (e.g., Z. Chen et al., 2014, 2013), or ambient lighting (e.g., Azizyan et al., 2009) have been shown to predict whether individuals are visiting places such as bars, cafes, and restaurants. Finally, the (5) *time* point of a situation is arguably the most straightforward cue manifesting in smartphone sensing data, as smartphones are typically equipped with a timestamp tracker (Harari et al., 2015, Harari, Müller, & Gosling, 2020).

Sensing Situational Characteristics

The abundance of findings reflects the growing research interest in the relationship between smartphone-sensed data and situational information. However, past studies have often been conducted in the field of ubiquitous computing and have mainly focused on the automated inference of situational cues such as the current social (persons/interactions) or geographical context (locations; e.g., Azizyan, et al., 2009; Z. Chen et al., 2014, 2013; Do & Gatica-Perez, 2014; Lane et al., 2014; Lu et al., 2009; Min et al., 2013). Little is known about whether these smartphone-sensed situational cues also reflect the perceived characteristics of the corresponding psychological situation (Wrzus & Mehl, 2020).

Nevertheless, as Table 1.1 illustrates, it is reasonable to assume that smartphone sensing data can contribute to capturing an individual's psychological situation. Self-reports have exposed correlations between situational characteristics and situational cues (e.g., Horstmann et al., 2021), which, in turn, can be captured by smartphone sensing data (e.g., Harari et al., 2017). Thus, studying situations using a smartphone sensing approach may yield new insights into objectively sensed situational cues and their counterparts in perceived situational characteristics. The utilization of smartphone-sensed data has the potential to significantly increase the level of information density in psychological research investigating the interplay between an individual and the situational context. In contrast to burdensome experience samplings, the unobtrusive nature of smartphone sensing could enable the convenient collection of situational cues and characteristics for a wide range of everyday situations with high temporal resolution across large samples.

2.2.3. Rationale

In summary, our study aims to bridge the gap between the two fields of situation and smartphone sensing research. We present, to the best of our knowledge, the first empirical study that integrates psychological situation research with smartphone sensing by exploring whether objective situational data collected via smartphones can predict individuals' psychological situations. Overcoming self-report measures used in previous research, diverse and granular smartphone sensing data are used to investigate the link between situational cues and psychological situations. Thereby, we consider a wide range of situational cues on the one hand, and different types of situations on the other, taking into account the diversity and complexity of real-life situations. The great complexity of daily situations, in turn, requires flexible data modeling techniques to extract the maximum amount of information from the sensing data. Thus, we apply explorative machine learning approaches to investigate how well psychological situations can be predicted from objectively assessed situational cues. In doing so, we follow researchers who have highlighted the potential of big data and machine learning methods for psychological research (e.g., Orrù et al., 2020; Yarkoni & Westfall, 2017), especially in the context of situation research (Wrzus & Mehl, 2020).

Table 1.1: Overview of Situational Cues and Associated Smartphone-Sensed Indicators

Situational Cues		**Smartphone-Sensed Indicators**	
Persons/Interactions			
Presence of people	e.g., Breil et al. (2019); Horstmann et al. (2021); Rauthmann et al. (2014)	Ambient noise	e.g., Harari et al. (2017); Lane et al. (2014); Lu et al. (2009; 2012); Rabbi, et al. (2011); Rüegger et al. (2020); H. Wang et al. (2014); Wang & Marsella (2017); R. Wang et al. (2014)
Social interaction	e.g., Serfass & Sherman (2015)	Connectivity (e.g., Bluetooth, WiFi)	e.g., Z. Chen et al. (2014, 2013); Harari et al. (2015); Kalimeri et al. (2013); Min et al. (2013); Wang & Marsella (2017); R. Wang et al. (2014); Rüegger et al. (2020);
		Smartphone usage (e.g., screen logs, app logs)	e.g., Chatterjee et al., 2020; Harari et al. (2017); Harari, Müller, Stachl, Pargent, et al. (2020); Lane et al. (2014); Stachl, Au, et al. (2020); Stachl, et al. (2017)
		Communication (e.g., calls, text messages)	e.g., Harari et al. (2016; Harari, Müller, Stachl, et al. (2020b); Kalimeri et al. (2013); Lane et al. (2014); Min et al. (2013); Montag et al. (2014); Rüegger et al. (2020); Verkasalo (2009); Servia-Rodríguez et al. (2017); Stachl et al. (2017); R. Wang et al. (2014)
Objects			
Music	e.g., Blake et al. (2020); Rauthmann et al. (2014)	Ambient noise	e.g., Lane et al. (2014); Lu et al. (2009); H. Wang et al. (2014)
		Smartphone usage (e.g., app logs)	e.g., Harari, Müller, Stachl, et al. (2020b); Stachl et al. (2017); Stachl, Au, et al. (2020)
		Connectivity (e.g., Bluetooth, WiFi, headphone)	e.g., Blake et al. (2020); Lu et al. (2009)
Events/Activities			
Mobility (e.g., commuting, physical activity)	e.g., Blake et al. (2020); Rauthmann et al. (2014)	GPS/ accelerometer	e.g., Barnett et al. (2018); Canzian & Musolesi (2015); Y.-P. Chen et al. (2008); Do & Gatiza-Perez (2014); Harari et al. (2016, 2017); Lane et al. (2014); Lathia et al. (2017); Montoliu, et al. (2013); Müller et al. (2020); Rojas et al. (2016); Saeb et al. (2017, 2016, 2015); Servia-Rodríguez et al. (2017); R. Wang et al., 2014
		Ambient noise Connectivity (e.g., Bluetooth, WiFi)	e.g., Lu et al. (2009); Wang & Marsella (2017)
Studying/ working	e.g., Blake et al. (2020); Breil et al. (2019); Rauthmann et al. (2014; 2016)	Connectivity (e.g., Bluetooth, WiFi, flight mode)	e.g., Breil et al. (2019); Z. Chen et al. (2013)
Location			
Places	e.g., Breil et al. (2019); Rauthmann et al. (2014)	Connectivity (e.g., Bluetooth, WiFi)	e.g., Blake et al. (2020); Breil et al. (2019); Z. Chen et al. (2014, 2013); Montoliu, et al. (2013); Rauthmann et al. (2014); Wang & Marsella (2017)
		GPS/ accelerometer	e.g., Barnett et al. (2018); Do & Gatica-Perez, 2014; Fillekes et al. (2019); Harari et al. (2015, 2016); Mehl, et al. (2006); Müller et al. (2020); Mehrotra et al. (2017); Montoliu, et al. (2013); Oliveira et al. (2016); Rauthmann et al. (2014); Saeb et al. (2017, 2015); Sandstrom et al. (2017); Servia-Rodríguez et al. (2017); Verkasalo (2009); R. Wang et al. (2014)
Time			
Weekday & daytime	e.g., Serfass & Sherman (2015)	Timestamp	e.g., Harari et al. (2015), Harari, Müller, Stachl, et al. (2020b); Lu et al. (2009); Min et al. (2013); Saeb et al. (2015); Servia-Rodríguez et al. (2017); Verkasalo (2009); R. Wang et al. (2014)

Note. The left side of the table shows previous studies that have identified different situational cues of psychological situations. For every situational cue, studies on smartphone-sensed indicators of the respective cue are listed on the right side of the table.

2.3. Method

All data in this study were collected within the Smartphone Sensing Panel Study (SSPS; Schoedel & Oldemeier, 2020). The data collection procedures were carried out following the General Data Protection Regulation (GDPR) and received ethical approval. Additional material and analysis code can be found in the corresponding Open Science Framework project (OSF).[3] Please note that the tables and figures marked with A are in the Appendix section and those marked with S are available in the online supplemental material.

2.3.1. Procedures

Within the SSPS, a sample of 850 participants was recruited by a provider for non-probability online panels according to quotas that were representative of the German population in terms of gender, age, education, income, confession, and relationship status. Participants were required to be between 18 and 65 years old, fluent in German, and for technical reasons, use a smartphone running on Android 5 or higher (see Schoedel and Oldemeier (2020) for more details). After recruitment, participants were randomly assigned to two groups, one of which participated in the study for three months (n = 191) and the other for six months (n = 659). Afterward, they were asked to install an Android-based application (app) called *PhoneStudy*[4] on their personal smartphones, which continuously collected smartphone sensing data in the background for the respective study duration. The app sent participants a link to a 25-30-minute online survey each month (i.e., a total of three surveys for the three-month group, and six surveys for the six-month group). In addition, two 14-day experience sampling (ES) waves (the first one in July/August, the second one in September/October) were conducted as part of the SSPS.

During these experience sampling waves, participants were asked to complete short five-minute questionnaires two to four times a day. The schedule for the presentation of experience sampling questionnaires was pseudo-randomized as follows: each day was divided into four equally sized sections (on weekdays from 7 a.m. to 10 p.m. and weekends from 9 a.m. to 11 p.m.). Within each segment, the time was chosen randomly, but the interval between two consecutive questionnaires had to be at least 60 minutes. Participants were informed about the questionnaire via a notification as soon as the smartphone was actively used for the first time after the calculated time. This procedure was chosen to increase the respondents' commitment without artificially provoking smartphone use (van Berkel et al., 2019). In order to increase the

[3] https://osf.io/kwd82/
[4] https://phonestudy.org/

motivation of the participants in the context of such a data-intensive study, the different parts of the study were compensated separately (e.g., continuous granting of logging permissions; online surveys).

Drawing from this extensive data set, this study focuses on the smartphone sensing data collected within the first experience sampling wave (27th of July 2020 to 9th of August 2020). Additionally, the study uses self-report data collected within survey 1 (18th to 24th of May 2020; demographics), survey 2 (15th to 21st of June 2020; Big Five personality traits) as well as experience sampling phase 1 (27th of July 2020 to 9th of August 2020; Situational Eight DIAMONDS). For a comprehensive overview of all measures and related data collection procedures as well as details on compensation within the SSPS, please refer to the basic study protocol provided by Schoedel and Oldemeier (2020). The data analysis procedures applied in this study were preregistered as part of a master's thesis prior to data preprocessing and analyses. The corresponding preregistration form is provided in our PsychArchives project.[5]

2.3.2. Sample

As preregistered, participants were only included in our data analyses if they reported at least 14 experience samples of the respective criterion variable (i.e., Duty, Intellectuality, Adversity, Mating, pOsitivity, Negativity, Deception, Sociality). This resembles an average of one completed experience sample per day during the two-week experience sampling phase. In this way, we wanted to make sure that we were only analyzing data from participants who were seriously taking part in the study. Participants who met these inclusion criteria for one of the DIAMONDS did so for all others. Applying all exclusion criteria including the preprocessing steps as described below, resulted in a final sample of $N = 510$ participants for the data analysis presented in this article. Participants' demographic characteristics were collected within survey 1 of the SSPS. Demographic characteristics were provided by 443 participants. This subsample included 199 females (45%) and 243 males (55%) aged between 18 and 65 years with an average age of 41.39 years (SD = 12.45). As for their highest level of education, three participants (< 1%) reported having a doctoral degree, 96 (22%) a university degree, 133 (30%)

[5] The preregistration can be accessed at https://dx.doi.org/10.23668/psycharchives.4928. All deviations from the preregistration are described in detail in the online supplemental material published on the OSF project page. Please note that the preregistration was written as part of a master's thesis, in which I was involved as supervisor, and therefore had certain limitations in terms of scope. This manuscript extended the preregistration to include the prediction of all DIAMONDS dimensions (only four were selected in the master's thesis) and the inclusion/modification of additional sensing-based situation variables. We also added additional exploratory analysis for validation purposes. However, the preregistered preprocessing pipeline and analysis methods remained the same.

a high school degree, 175 (34%) a secondary school degree, 63 (14%) a secondary general school degree, and two (< 1%) no school degree.

2.3.3. Measures

Self-Report Measures

Situational Eight DIAMONDS

Characteristics of participants' psychological situations were assessed for 14 consecutive days in experience sampling measures via the German version of the S8-I ultra-brief measure for the situational eight DIAMONDS (S8-I; Rauthmann & Sherman, 2016a). The S8-I consists of eight items (one for each DIAMONDS dimension) on which participants indicate the extent to which they apply to their current situation (i.e., Duty: "*work has to be done*", Intellectuality: "*deep thinking is required*", Adversity: "*someone is threatened, accused, or criticized*", Mating: "*potential romantic partners are present*", pOsitivity: "*the situation is pleasant*", Negativity: "*the situation contains negative feelings (e.g., stress, anxiety, guilt, etc.)*", Deception: "*someone is being deceived*", Sociality: "*social interactions are possible or required*") in our data analyses. Diverging from Rauthmann and Sherman's (2016a) S8-I, all items were measured on a binary scale (0 = *does not apply*, 1 = *applies*). This was done to keep the participants' burden on an acceptable level, as items were presented repeatedly among other self-report measures.

Big Five Personality Traits

The Big Five personality traits were assessed using the Big Five Structure Inventory (BFSI; Arendasy et al., 2011), a hierarchical personality inventory that defines the five broad dimensions (factors) of personality (i.e., *Openness, Conscientiousness, Extraversion, Agreeableness, Emotional Stability*), each consisting of six facets. The questionnaire consists of 300 adjectives describing personality, which were rated on a 4-point Likert scale (ranging from 0 = *untypical for me* to 3 = *typical for me*). With Cronbach's α values ranging between .72 and .92, the BFSI shows favorable psychometric properties (Arendasy, 2009). Since the construction of the BFSI does not follow the classical test theory, but the item response theory framework, the person parameter estimates of the partial credit model were used as personality scores in our analyses.

Sensing Measures: Situational Cues

Throughout the study, a broad range of smartphone-sensed situational measures (e.g., connected Bluetooth devices, location, app usage) were assessed for up to 25 weeks using the Android logging app *PhoneStudy*. All collected data was timestamped as it was logged on an event-based basis by occurrence and/or in predefined time intervals. Depending on the class of the logged event (e.g., calls), it was further specified by additional information (e.g., outgoing). A comprehensive overview of all logging events and their specifications is provided by Schoedel and Oldemeier (2020). In the following sections, we briefly describe how we extracted the smartphone-sensed situational cues that were relevant to our research question.

Concretely, the raw sensing data logged over the two-week ES period were used to extract various sensing variables which were selected according to the following principles: First, situational correlates of individuals' psychological situations (i.e., DIAMONDS) were identified based on an extensive literature review (see Table 2.1). Second, we derived manifestations of these correlates in smartphone-sensed data based on previous literature and theoretical reasoning. For instance, we identified the activity of studying or working as a correlate of perceived situation characteristics (Rauthmann et al., 2014). This activity in turn has been successfully predicted via machine learning based on Bluetooth connectivity data (Z. Chen et al., 2013). Additionally, one may infer such activities based on the user's smartphone usage (e.g., less general app usage). Consequently, we included connected Bluetooth devices and general app usage as predictor variables in our study. Third, our theoretical selection of predictor variables was dependent on the data. For example, we did not include noise-related predictors in our study, although they are likely to be related to the DIAMONDS, because such data type was only recorded in the evening.

The resulting ideas for extracting situational cue variables in our study were based on events of five different clusters of data types (referred to as *sensing modalities* in the following): (1) *connectivity* (e.g., Bluetooth status or connected headphones), (2) *smartphone usage* (e.g., app usage or screen times), (3) *communication* (e.g., calls, text messages), (4) *GPS/accelerometer*, (e.g., total distance covered or types of places visited), and (5) *timestamp* (e.g., daytime or weekday). Table 2.1 presents an overview of these sensing modalities and derived variable classes. Please note that in machine learning jargon, variables are called *features*, so we use this term in the following.

Table 2.1: Overview of Different Feature Classes per Smartphone Sensing Modality

Sensing Modality	Feature Class
Connectivity (con)	Bluetooth, flight mode, headphones, power cable, WiFi
Smartphone usage (sma)	Apps, music, notifications, screen
Communication (com)	Calls & text messages, keyboard logs
GPS/accelerometer* (GPS)	Activity states, altitude, displacement, distance covered, GeoHashs, places, speed, trips
Timestamp (timestamp)	Daytime, weekday

Note. The abbreviations of the sensing modalities as reflected in the feature names are shown in brackets.
* GPS data were enriched via different APIs to identify the types of places visited.

Categorizations

The raw sensor data collected was directly interpretable for most data types, but some data types were not inherently meaningful. Thus, certain types of raw logging data had to be clustered into a finite number of categories before psychologically meaningful features could be extracted. Concretely, we derived categorizations for the sensing data on (a) connected Bluetooth devices, (b) apps used, and (c) places visited. In order to not go beyond the scope of this paper, detailed information on the newly developed categorizations and the respective definitions of categories are provided in Table A2.1 and the codebook available in the OSF project.

The raw sensing data on the *Bluetooth* connectivity status of the devices were grouped using a newly developed two-level category system. At level 1, the Bluetooth events were grouped as turned “on/connected”, turned “on/disconnected”, or turned “off”. For “connected” events, we further differentiated the types of connected Bluetooth devices (e.g., “headset”, “computer”, or “car”).

To gather data on participants' smartphone use, the names of the *apps* used in each session were recorded and categorized into meaningful app types, such as “social media” or “gaming”, based on a category system developed by Schoedel et al. (2022). To ensure data quality, only categories with at least moderate interrater agreement (Cohen’s Kappa > .60) were used, while “system apps” that automatically run in the background were excluded.

To track participants' locations, *GPS* and *accelerometer* data were logged (on-device) using the *Fused Location Provider API*.[6] For example, we employed the GeoHash algorithm (Niemeyer, 2008), which uses GPS latitude and longitude data to identify frequently visited geographic regions and extract indicators of individual location visiting patterns per participant.

[6] https://developers.google.com/location-context/fused-location-provider

Furthermore, using all available GPS data points, participants' home and workplace locations were identified by applying a sophisticated clustering algorithm as described in the online supplemental material. Additionally, the GPS data points logged at the time point of the experience sampling were annotated with the closest point of interest (POI) by using the *Google Places Nearby Search* [7], *HERE Geocoding & Search API* [8], and *Foursquare Places API* [9]. Moreover, a total of 1,457 unique latitude/longitude data pairs were categorized by iteratively developing a categorization based on the *HERE Places Category System* [10]. This resulted in different types of places, ranging from “food” to “outdoors” or “nightlife” (see Table A2.1). Finally, we enriched the data by incorporating for example the population density of the respective city in 2020, using the GENESIS-Online database of the German Federal Statistical Office (Destatis, 2020).

In addition, the processing of certain types of data was already integrated into the PhoneStudy app. First, the keyboard logs were preprocessed on-device using the LanguageLogger app (Bemmann & Buschek, 2020). Moreover, physical activity recognition was embedded into the PhoneStudy app using the Google Activity Recognition API.[11] Further details on the data enrichment processes can be found in Schoedel et al. (2022) and the online supplemental material in the OSF project.

Feature Extraction

The smartphone-sensed situational cues were linked to self-reported situational characteristics by aggregating the logged events in relation to their respective experience sampling (ES) reports. Given the exploratory nature of the research and the lack of previous studies on the appropriate time window to observe situational cues, two different temporal perspectives were applied. Specifically, we decided to extract features from the raw sensing data onto two types of aggregation modes: *status* and *timeframe* aggregation.

So-called *status* features reflect variables that were extracted at the exact time the experience sampling was recorded. The status features reflect binary features, i.e., status is present (1) or absent (0). For instance, the feature *Bluetooth status “on”* indicates whether a participant’s Bluetooth was turned on at the time the specific experience sampling questionnaire was answered. In contrast, so-called *timeframe* features are variables that quantify events within

[7] https://developers.google.com/maps/documentation/places/web-service/search-nearby
[8] https://developer.here.com
[9] https://location.foursquare.com/developer/reference/places-api-overview
[10] https://developer.here.com/documentation/places/dev_guide/topics/place_categories/places-category-system.html
[11] https://developers.google.com/location-context/activity-recognition

a certain timeframe around the respective experience sampling record. For instance, the timeframe feature *total duration of app usage* quantifies the total duration of app usage within one hour around the specific experience sampling questionnaires. As the start and end of daily situations can vary drastically (Rauthmann & Sherman, 2016b), a range of different timeframes may be informative about an individual's current psychological situation. Based on previous research on the average length of situations (Rauthmann & Sherman, 2016b) and the prevalence of situational manifestations (Andone et al., 2016; Wilcockson et al., 2018), a time window of 30 minutes before and 30 minutes after the start of the related experience sampling questionnaire was chosen for the extraction of timeframe features. That is, the raw logging events that occurred within these 60-minute time intervals were aggregated to quantify the distribution of the number, frequency, and duration of logging events for each timeframe feature (i.e., using the sum, median, median absolute deviation, minimum, and maximum). A detailed explanation of all quantification metrics can be found in Table A2.2.

Description of Features

The feature extraction procedure resulted in a total number of 675 features (58 *status features* and 617 *timeframe features*) for our predictive modeling approach. In order not to go beyond the scope of this paper, we do not provide a step-by-step description of our exact preprocessing procedures here, but refer the interested reader to the supplemental material on the project's OSF repository for more detailed information.[12] However, for a better understanding of the derived smartphone-sensed situational cues, some examples are given in the following:

The *connectivity* features comprise situational cues that reflect the connectivity of a participant's device by sensing the current connectivity status via Bluetooth, WiFi, or power cable. For example, the number and type of connected devices at the time of experience sampling are considered (status feature).

The *smartphone usage* features reflect the participant's phone-related behavior, including for example the number or type of apps that are used. Moreover, the number or duration of screen sessions as well as response times to notifications in and around the situation of interest are measured (timeframe feature).

The *communication* sensing modality considers for example features related to calls or text messages, as well as data collected from keyboard logging. For instance, the total number of outgoing calls or the average duration of a call was calculated (timeframe features). In

[12] https://osf.io/kwd82/

addition, the average sentiment of a text message or the total amount of words per text message are included in this sensing modality, following Pennebaker et al. (2015).

The features developed based on the *GPS/accelerometer* data cover information about the current location (or POI; point of interest), such as the type of place (e.g., home, work, shop, or restaurant), as well as other characteristics such as the population-density in the current district. For example, the type of place being visited during the experience sampling is extracted (status feature). In addition, speed- or accelerometer-based activity states (e.g., driving vs. walking) in a current situation are estimated (status feature). Other geographical metrics, such as the spatial coverage of the participant around the situation, are also included in this sensing modality (timeframe feature).

Finally, the *timestamp* features include all time-related information, from the time of day to the weekday of the current situation (status feature).

2.3.4. Data Preprocessing

Following the preregistration, a series of preprocessing steps were applied to prepare the data for predictive modeling. To ensure high data quality, data were excluded based on criteria at the level of complete study days, experience sampling (ES) records, and single observations. First, days with low levels of smartphone usage (less than ten unlocks of their screen or total usage time below 15 minutes) were excluded. Second, ES reports with a completion time above 15 minutes were excluded. This represents the maximum time participants were instructed to spend on the questionnaire. Third, due to technical logging errors, single observations of features could reach extreme values unrelated to the participant's situation. We therefore inspected the distributions of all extracted features prior to predictive modeling and replaced extreme outliers (4 standard deviations from the mean) in extracted features with missing values. Fourth, following the recommendations of Kuhn and Johnson (2013, p. 42), features with (a) more than 90% missing values, (b) zero or near-zero variance (10% cut-off), or (c) strong correlations with other features ($r > .90$) were removed, and missing values were imputed using median imputation. Please note that this step was performed within each inner resampling iteration to avoid overfitting.

2.3.5. Data Analyses

Machine Learning Algorithms

We employed machine learning methods to predict the reported presence (or absence) of individuals' psychological situations from smartphone-sensed status and timeframe features. Concretely, we framed the prediction of the eight perceived situational characteristics (i.e., the

DIAMONDS dimensions) as eight binary classification problems. The predictions were made on the level of experience sampling records.

To determine the best model, the predictive ability of a random forest and a logistic Lasso regression was compared to a featureless (naive guessing) model by conducting a statistical benchmark experiment. The random forest model is capable of handling non-linear relationships and complex interactions (Breiman, 2001). The model is based on multiple bootstrapped and decorrelated decision trees, reflecting an all-rounder that is widely used in machine learning research (Wright & Ziegler, 2017). The logistic Lasso regression, a linear model with L1 regularization, provides interpretability and sparsity of coefficients (Tibshirani, 1996).

Evaluation Metric

To assess the predictive performance of the machine learning models, we focused on the area under the receiver operating characteristic curve (AUC). This cutoff-independent metric balances the trade-off between sensitivity and specificity (Hosmer et al., 2013).[13] A naïve guessing model without features (featureless model) shows a linear relationship between the sensitivity and specificity, manifested in an AUC of .50. In prediction tasks, a model can perform better or worse than the featureless model. Accordingly, the AUC is defined between zero and one, while the higher the value above .50, the better the model fit (James et al., 2013, p. 150; Kuhn & Johnson, 2013, p. 264).

Resampling

To prevent overfitting, we applied repeated ten-fold cross-validation as a resampling scheme. Due to the repeated experience sampling measures per person, our analyzed data comprised two levels: experience sampling records and participants. Since the observations of a participant belong together and must not be separated during resampling, we applied a blocked resampling procedure using the participant's identifier as the grouping factor. Both the random forest and the Lasso models were trained with the standard settings of the mlr3 package (Lang et al., 2019), using the *ranger* (Wright & Ziegler, 2017) and *glmnet* implementations (Friedman et al., 2010).[14]

[13] The sensitivity (or true positive rate) measures the proportion of correctly classified positive observations, while the specificity (or true negative rate) reflects the correctly classified negative observations.

[14] In classification, the default setting of the random forest function (classif.ranger) conveys an estimation of 500 decision trees with the Gini coefficient as a loss function. In default, the Lasso (cv.glmnet) optimizes the value of the L1 penalty (i.e., λ) in a ten-fold cross-validation with the model's deviance as a loss function. In default settings, the optimal λ is the value that gives the most regularized model such that the cross-validated error (e.g., deviance) is within one standard error of the minimum.

Feature Importance

To gain a better understanding of the importance of specific features, we evaluated the feature importance of the prediction models that outperformed the featureless model. Concretely, the penalized logistic Lasso regression and the random forest model were fit to the full data sets of Duty (n = 11,506), Intellectuality (n = 11,492), Mating (n = 11,475), and Sociality (n = 11,446) scores. For the analysis of the single feature importance in the Lasso model, we extracted the z-standardized regression weights per feature and assessed their absolute values. For the random forest model, feature importance was assessed by permutation method, using the AUC loss across ten permutations as the evaluation metric (Breiman, 2001).

In addition, grouped feature importance scores were calculated for the Lasso and random forest models to compare the importance of the five different sensing modality groups. Concretely, the grouped feature importance scores represent the difference in the average dropout AUC loss of the respective feature group compared to the full model (including all features) across ten permutations (Biecek, 2018).

Linear Mixed Models for Validation

Moreover, we validated the predicted DIAMONDS scores using Big Five personality traits, that are associated with psychological situations (e.g., Jonason & Sherman, 2020; Rauthmann et al., 2014; Sherman et al., 2015). Linear multilevel models (LMMs) were calculated for dimensions that were predicted better than chance, using two models with the same predictors (Big Five traits) but different targets, namely (1) self-reported scores and (2) predicted scores. Due to the repeated measurements of participants' psychological situation scores, we used multilevel regression modeling with repeated measures of the DIAMONDS as level 1 variables nested within individuals (level 2). We specified a *random-intercept-fixed-slope model* to predict each experience-sampled DIAMONDS score (binary target) and included Big Five personality traits as level 2 predictors (Bates et al., 2015; Matuschek et al., 2017).

Statistical Software

All data preprocessing and analysis were performed using R 4.0.3 (R Core Team, 2021) and Python 3.10 (van Rossum & Drake, 2009). R packages used included *dplyr* (Wickham, François, et al., 2019), *httr* (Wickham, 2019), *jsonlite* (Ooms, 2014), *lubridate* (Grolemund & Wickham, 2011), and *tidyverse* (Wickham, Averick, et al., 2019). Missing imputation was conducted using *zoo* (Zeileis & Grothendieck, 2005). The packages *dbscan* (Hahsler, et al., 2019) and *geosphere* (Hijmans et al., 2017) were used for geographic clustering and spherical trigonometric analysis of the GPS data. The Python libraries *pandas* (McKinney et al., 2010) and *NumPy* (Harris et al., 2020) were applied for GPS-based trip identification with the *infostop* library (Aslak & Alessandretti, 2020). Predictive modeling was conducted with *mlr3* (Lang et al., 2019) and linear multilevel modeling was done with *lme4* (Bates et al., 2015). The packages *DALEX* (Biecek, 2018) and *DALEXtra* (Maksymiuk, et al., 2020) were used for grouped feature importance analysis. Visualization was performed using *ggplot2* (Wickham, 2016), *corrplot* (Wei et al., 2017), and *fmsb* (Nakazawa, 2021). All R packages used in the analysis are listed in the *renv.lock* file in the OSF project to ensure reproducibility (Ushey, 2020).

2.4. Results

2.4.1. Descriptive Statistics

Psychological Situation

After preprocessing, the final data set comprised 176 predictor variables (28 status features and 148 timeframe features) from 510 participants with 11,506 (Duty) to 11,175 experience sampling records (Sociality). On average, 19 reports were completed per participant, with 4% (Deception) to 73% (pOsitivity) of the DIAMONDS dimensions rated as present in the experience-sampled situations (see Table 2.2). Based on the random-effects multilevel models, we found that between 34% and 47% of the variance in the DIAMONDS ratings can be attributed to between-person differences. The person-level intercorrelations between the DIAMONDS dimensions ranged from -.46 (for pOsitivity and Negativity) to .50 (for Duty and Intellectuality) (see Table 2.2).

Situational Cues

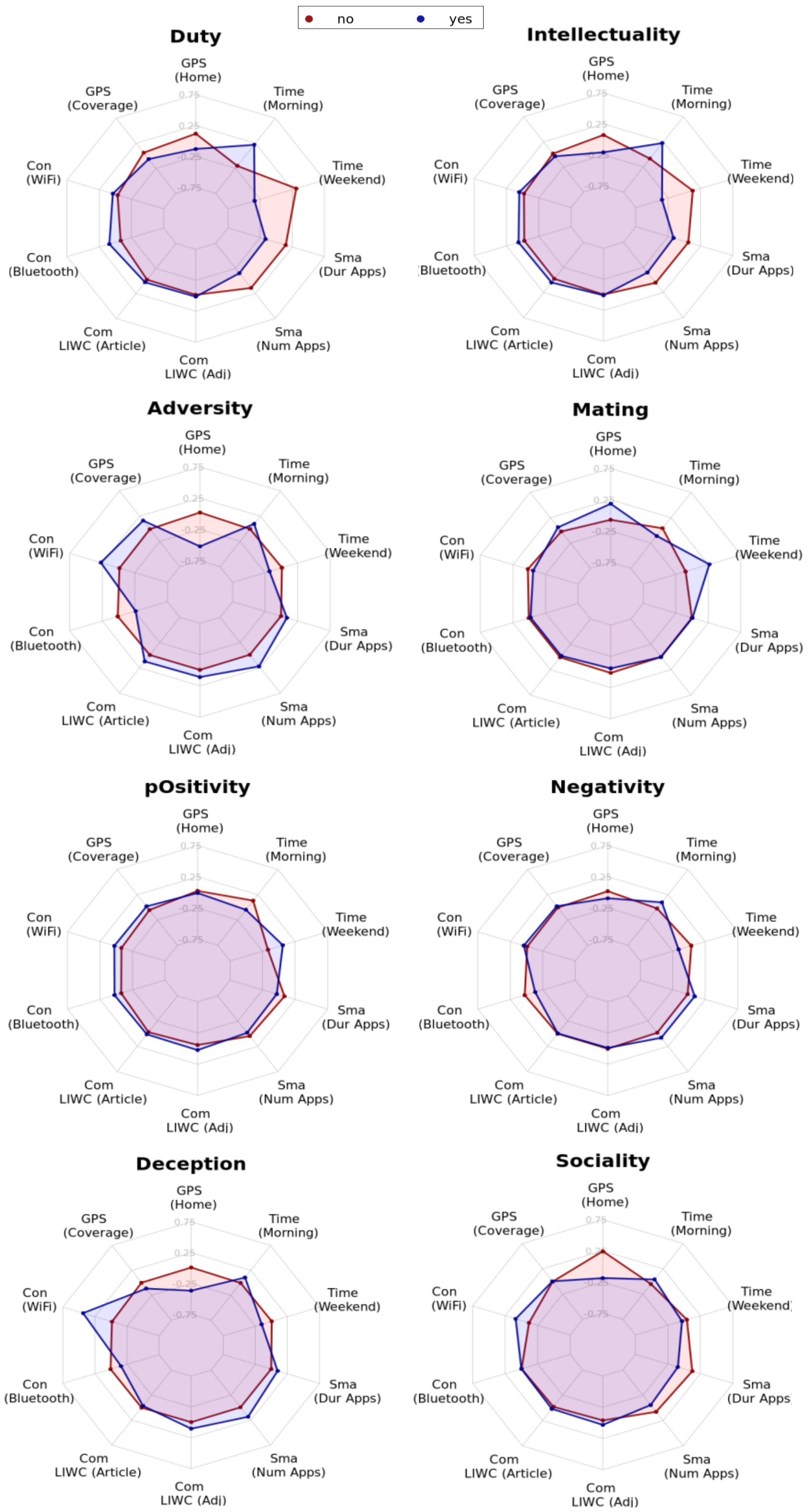
In this manuscript, we present only selected examples of situational cues. A comprehensive overview of descriptive statistics and correlations for all situational cues can be found in Table S5 in the online supplemental material. Figure 2.1 shows the average profiles of selected cues across participants and experience sampling reports for each DIAMONDS dimension. The blue polygons indicate the average situational cues for situations scoring high, and the red polygons indicate situations scoring low on the respective dimension. For example, situations rated high on Duty are more likely to be associated with situations outside the home, early on weekdays, and with lower app usage. Figure 2.1 illustrates that differential patterns of situational cues can be observed for some DIAMONDS dimensions (e.g., Duty, Intellectuality), while there is an overlap for others (e.g., pOsitivity, Negativity).

Table 2.2: Descriptive Statistics and Intercorrelations for DIAMONDS Ratings

	n	*M*	*SD*	ICC	Intercorrelations							
					D	I	A	M	O	N	De	S
Duty	11,506	0.48	0.50	.34	-							
Intellectuality	11,492	0.28	0.45	.40	.50	-						
Adversity	11,481	0.05	0.21	.46	.11	.21	-					
Mating	11,475	0.28	0.45	.47	-.06	-.03	.08	-				
pOsitivity	11,469	0.73	0.44	.42	-.11	-.12	-.16	.20	-			
Negativity	11,463	0.19	0.40	.43	.18	.26	.35	-.02	-.46	-		
Deception	11,453	0.04	0.19	.47	.09	.16	.45	.08	-.08	.27	-	
Sociality	11,446	0.56	0.50	.40	.22	.21	.08	.25	.07	.09	.08	-

Note. ICC = Intra-class correlations reflect the proportion of variance between persons divided by total variance (i.e., the higher the value, the more is the variance attributable to the person); D = Duty; I = Intellectuality; A = Adversity; M = Mating; O = pOsitivity; N = Negativity; De = Deception; S = Sociality; Intercorrelations resemble Pearson correlation on the person-level (N = 510); DIAMONDS were measured on a binary scale with 0 = *does not apply* and 1 = *applies*.

Figure 2.1: Visualization of Selected Smartphone-Sensed Situational Cues per DIAMONDS Dimension



Note. GPS (Coverage) = spatial coverage by convex hall (timeframe feature); GPS (Home) = situation at home (status feature); Time (Morning) = situation in the morning (status feature); Time (Weekend) = situation at the weekend (status feature); Sma (Dur Apps) = total duration of all app usages; Sma (Num Apps) = total number of all apps used; Com (LIWC (Adj)) = avg. score of LIWC dimension *Adjective* in keyboard logs (timeframe feature); Com (LIWC (Article)) = avg. score of LIWC dimension *Article* in keyboard logs (timeframe feature); Con (Bluetooth) = total duration of Bluetooth disconnected; Con (WiFi) = total duration of WIFI status off (timeframe feature). The average of the feature calculated across all experience samplings after z-standardizing the raw data ($N = 510$) is shown. The DIAMONDS were measured on a binary scale with 0 = *does not apply* (*no* in red) and 1 = *applies (yes* in blue*)*.

2.4.2. Prediction of Psychological Situations

Benchmark Experiments

Table 2.3 summarizes the results of the benchmark experiments for each DIAMONDS dimension, with the models ranked by highest predictive performance (mean AUC) from lowest to highest: featureless model, random forest, and Lasso. The random forest and Lasso models outperformed the featureless model, with higher mean AUC levels for Duty (.680 and .689), Intellectuality (.592 and .631), Mating (.559 and .586), and Sociality (.589 and .574), respectively. However, the random forest and Lasso models did not perform better than the featureless model for Adversity, pOsitivity, Negativity, and Deception at the descriptive level.

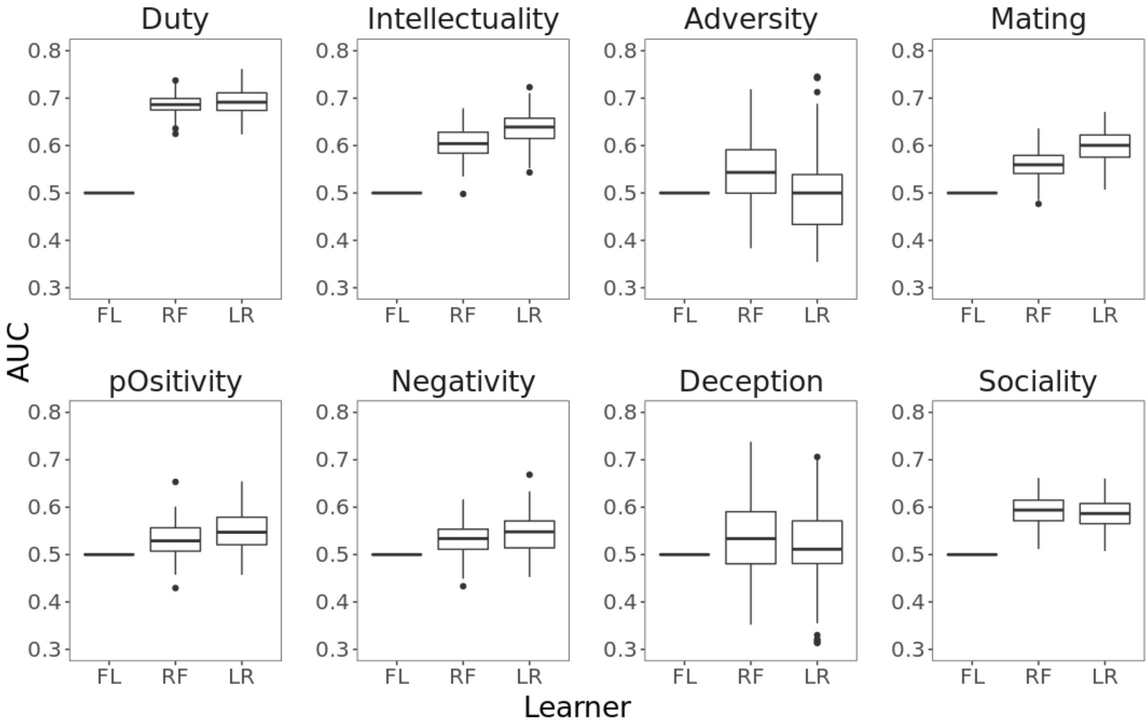
Table 2.3: Summary of Mean AUCs per Target Variable and Algorithm

Targets	Algorithm		
	Featureless Model	Random Forest	Logistic Lasso Regression
Duty	.500	.680	.689
Intellectuality	.500	.592	.631
Adversity	.500	.550	.534
Mating	.500	.559	.586
pOsitivity	.500	.517	.541
Negativity	.500	.515	.530
Deception	.500	.515	.531
Sociality	.500	.589	.574

Note. The coefficients represent the mean values of the AUC and are based on a ten-times repeated ten-fold cross-validation.

This impression is confirmed by the distribution of AUCs over the 100 iterations (i.e., repeated 10x10-fold CV) of the benchmark experiment shown in Figure 2.2. The AUCs for Adversity, pOsitivity, Negativity, and Deception predictions by both random forest and Lasso were close to the featureless model, while both models outperformed the featureless model for Duty, Intellectuality, Mating, and Sociality. The range of the box plots across all iterations shows that the predictions were robustly better than a naïve guessing approach. Both random forest and Lasso performed more or less similarly for all DIAMONDS dimensions.

Figure 2.2: Distribution of AUCs across Resampling Iterations per Algorithm

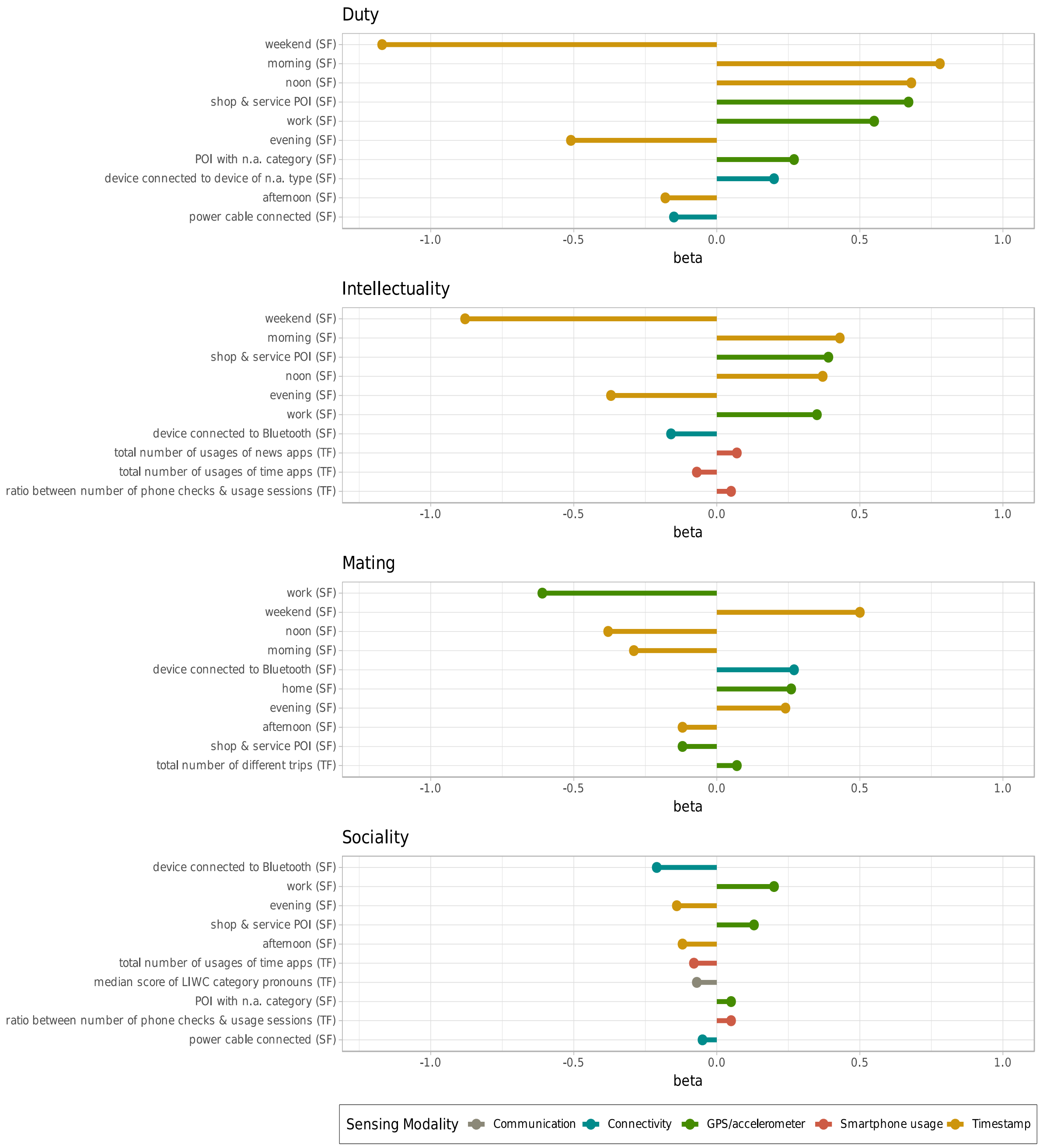


Note. AUC = Area under the operating characteristic curve; FL = featureless model; RF = Random Forest; LR = logistic Lasso regression; Distributions are based on a ten-times repeated ten-fold cross-validation.

Feature Importance

We analyzed the feature importance for the dimensions of Duty, Intellectuality, Mating, and Sociality based on the benchmark experiment results. We focused on the results of the linear prediction model (i.e., Lasso) because it had a similar predictive ability to the non-linear model (i.e., random forest). The absolute z-standardized logistic regression coefficients of the top ten most important features for each dimension are shown in Figure 2.3. The feature importance results for the random forest are shown in Table S6. Of the 175 features included in the final modeling after preprocessing, 99 had non-zero coefficients for Duty, 80 for Sociality, 79 for Mating, and 69 for Intellectuality, while no sensing modality particularly dropped out of the model. However, it is important to interpret the differences in feature importance with caution, as the differences in the regression coefficients are rather small.

Figure 2.3: Regression Coefficients of Lasso Model per Sensing Modality



Note. beta = z-standardized coefficients of the logistic Lasso regression; SF = status feature (reflecting time point of experience sampling), TF = timeframe feature (time window around experience sampling); POI = point of interest. The penalized logistic Lasso regression was fit to the full data set of available ratings for Duty (n = 11,506), Intellectuality (n = 11,492), Mating (n = 11,475), and Sociality (n = 11,446). For visualization purposes, only the features with the ten highest coefficients (absolute values) are shown.

As can be seen in Figure 2.3, the top ten features for *Duty* predictions were dominated by the timestamp, GPS/accelerometer, and connectivity features. The model was more likely to predict a dutiful situation if it occurred on a weekday, in the morning or at noon, in a shop/service or at work, and with a connected Bluetooth device. For *Intellectuality*, weekday mornings, shop/service or work location, disconnected Bluetooth, less use of time-related (e.g., timer or calendar) or news apps, and more short phone checks were predictive. *Mating* was more likely to be predicted if the location was at home on a weekend evening, with a connected Bluetooth device and a higher number of trips completed. *Sociality* was likely when the location was a shop/service or work, the phone's Bluetooth or power cable was not connected, the situation was earlier in the day, and less time-related app usage but more quick phone checks occurred.

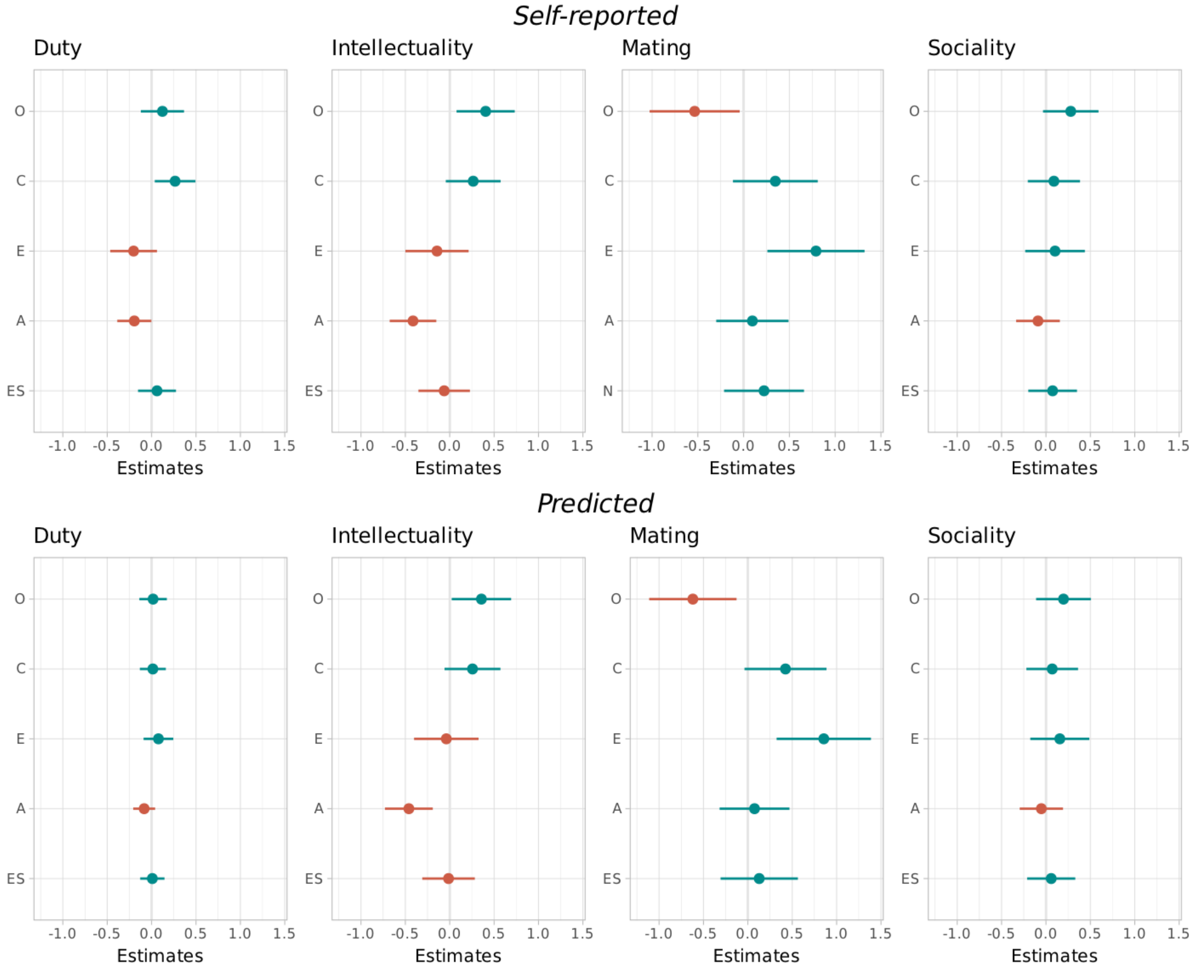
In summary, similarities in the importance of features across the four dimensions were observed. In particular, the time of the situation (especially weekday vs. weekend; morning vs. evening), the current location (at work or shop/service), and the phone's Bluetooth connectivity played an important role in the predictions. The grouped feature importance, calculated by the mean AUC loss across ten permutations, indicated that no sensing modality was more important than the others for prediction performance (see Table A2.8).

2.4.3. External Construct Validity of Predictions

The study also exploratively evaluated the external construct validity of the predicted DIAMONDS scores using a nomological net perspective (Bleidorn & Hopwood, 2019; Bornstein, 2009; Cronbach & Meehl, 1955). External discriminant validity refers to the associations between scores predicted by machine learning models and other constructs. Building on previous studies of associations between personality traits and characteristics of psychological situations, we inspected self-reported Big Five personality traits to examine external discriminant validity (Jonason & Sherman, 2020; Rauthmann et al., 2014; Sherman et al., 2015). As we did not perform formal hypothesis testing, confidence intervals are provided to indicate the robustness of the results.

In general, the Big Five personality traits showed similar associations with both self-reported and predicted DIAMONDS scores (Figure 2.4). For example, the personality trait Conscientiousness was positively associated with Duty or Intellectuality, while Extraversion was associated with higher Mating. Openness was associated with predicted and self-reported Intellectuality and Sociality scores. A more detailed overview of the results can be found in the Appendix (Tables A2.3-A2.6) and in the online supplemental material.

Figure 2.4: Estimated Fixed Model Effects of the Linear Mixed Models for Duty, Intellectuality, Mating, and Sociality



Note. Estimates = fixed effects estimates of linear mixed models for self-reported (left) vs. predicted (right) scores; O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, ES = Emotional Stability; 95% confidence intervals of estimated coefficients are shown. The red color indicates negative, the blue color positive coefficient estimates.

2.5. Discussion

We used a machine learning approach to investigate whether situational cues collected via smartphone sensing can predict perceived psychological characteristics (i.e., Duty, Intellectuality, Adversity, Mating, pOsitivity, Negativity, Deception, Sociality; DIAMONDS) of everyday life situations. For this purpose, we analyzed data collected in a large-scale panel study during a two-week experience sampling phase using a representative quota sample. By applying a multi-method approach, this study reflects a first step toward an ideal study design, as proposed by Rauthmann, Sherman, and Funder (2015): Situational characteristics were sampled from many people (i) in their everyday lives (Sherman et al., 2010), with (ii) different assessment methods (Rauthmann et al., 2014), from (iii) multiple situations using experience sampling to depict moment-to-moment changes in situation perception (Ram et al., 2012; Sherman et al., 2015), and (iv) across a variety of relevant groups (Guillaume et al., 2016). Our findings highlight the importance of multi-method study designs for generating new insights and show that certain, but not all, characteristics of the psychological situation can be predicted from data collected via smartphones. In the following sections, we critically discuss the findings of our exploratory study and provide post-hoc explanations for the prediction results. However, these interpretations should not be easily generalized and need to be confirmed by hypothesis testing in future research.

2.5.1. Varying Predictability of Different Situational Characteristics

This study extends previous situation research by using smartphone-sensed situational data and machine learning techniques to predict individuals' psychological situations. To the best of our knowledge, this study is the first to respond to recent calls from scholars to address the potential of smartphone sensing and machine learning for situation research (e.g., Harari, Müller, & Gosling, 2020; Wrzus & Mehl, 2020). By applying different prediction models, we found that the situational characteristics Duty, Intellectuality, Mating, and Sociality could be predicted above chance. However, the predictability of pOsitivity and Negativity was not very robust overall, only performing better than naïve guessing for some of the benchmark iterations. On the other hand, the scatter of the prediction performance across iterations for Adversity and Deception was very large, leading us to conclude that the prediction did not work for these dimensions.

Prediction of Duty, Intellectuality, Mating, and Sociality

For Duty, Intellectuality, Mating, and Sociality, our machine learning-based prediction model outperformed the featureless model on average, with improvements ranging from 7% (Sociality) to 19% (Duty). To illustrate, if we had randomly picked a dutiful and a non-dutiful situation from our sample, the Lasso model would have estimated a higher probability of the dutiful situation (69%) compared to the featureless model (50%). For comparison, previous studies using smartphone-sensed features to predict individuals' level of Big Five personality traits in a binary classification setting (e.g., high vs. low scores) reported similar mean prediction accuracies between 59% and 75% (e.g., Chittaranjan et al., 2011; Küster et al., 2018). However, it is important to note that direct comparisons between prediction results are complicated due to differences in stability and delineation between individuals and situations.

Prediction of pOsitivity and Negativity

Our findings for pOsitivity and Negativity suggest that our algorithms modeled weak associations, with prediction performances better than naïve guessing in most iterations. However, based on the magnitude of associations between pOsitivity or Negativity and situational cues reported in previous self-report-based studies, we would have expected higher prediction performances (e.g., Breil et al., 2019; Horstmann et al., 2021; Rauthmann et al., 2014; Rauthmann & Sherman, 2016a). One reason for the low predictive performance in our study may be that we used different methods to assess situational characteristics and cues, which freed our correlations from common method biases and probably resulted in lower correlations (Podsakoff et al., 2003). Furthermore, previous findings are based solely on questionnaire data and can be subject to cognitive biases that may inflate correlations (Baumeister et al., 2007; Schwarz, 2012).

Additionally, small associations may require a larger sample size to detect them, as previously noted in research on the relationship between smartphone sensing data and affect (Horstmann & Ziegler, 2019). For instance, Sandstrom et al. (2017) examined a sample of 3,646 participants with an average of 111 experience sampling records and found only weak associations between smartphone-sensed locations and affect. In contrast, our sample of 510 subjects with an average of 19 experience sampling records only accounted for 2% of the sample size studied by Sandstrom et al. (2017). As a result, an even larger sample may be required to accurately predict the pOsitivity and Negativity of a situation from smartphone-sensed situational cues. Another reason for the discrepancies with previous findings could be the different sample composition. While we investigated a quota-recruited sample representative of the German population, previous studies mostly focused on student samples (e.g., Blake et al., 2020; Breil et al., 2019; Horstmann et al., 2020; Rauthmann et al., 2014; Rauthmann & Sherman, 2016b; Shermann et al., 2015). The generalizability of students' situational perceptions to other populations is still an open question for future research.

Prediction of Adversity and Deception

The two dimensions Adversity and Deception could not be predicted by our smartphone-sensed situational cues. We suspect the under-sampling of these situational dimensions in our study to be one reason, making it difficult to train a good prediction model. In line with previous research, only a few situations were perceived as adverse or deceptive (Horstmann et al., 2021; Sherman et al., 2015). In addition, the Big Five traits of Agreeableness and Neuroticism, which are linked to Adversity and Deception (Jonason & Sherman, 2020; Rauthman et al., 2015b), also showed low predictability using smartphone data (Stachl, Au, et al., 2020), consistent with our findings.

2.5.2. Differential Fit of (Smartphone-Sensed) Situational Cues

Moreover, our prediction results indicate that the smartphone-sensed situational cues were differentially informative about different dimensions of psychological situations. We see two reasons for this. First, this study supports prior self-report-based findings on the differential suitability of situational cues for predicting situational characteristics (Blake et al., 2020; Breil et al., 2019; Horstmann et al., 2021; Rauthmann et al., 2014). Our interpretable machine learning analysis revealed that the dimension Duty was well-detected from objectively measurable situational cues, such as the time of the day or week and location. For example, individuals were more likely to rate a situation as dutiful if it was in the morning or at work. On the other hand, situational characteristics such as Adversity or Deception were not easily perceived from an external perspective (and were not accurately predicted in our study). This pattern of results is consistent with previous findings which observed the lowest agreement between participants' (internal) ratings and external raters' ratings for Adversity and Deception (Rauthmann et al., 2014). Similarly, in our study, the external raters (i.e., machine learning algorithms) were unable to extract information from the objective situational cues to accurately predict (internal) ratings of Adversity and Deception. Thus, our findings underscore the importance of considering the external detectability of situational characteristics when predicting psychological situations from situational cues. Based on our results, we conclude that situational cues are more effective for prediction the more readily apparent they are from the outside.

The second reason relates to the operationalization of situational cues using smartphones. Prior research has shown that situational cues provide differential information across different dimensions of psychological situations. Accordingly, the smartphone-sensed cues used in our study were not able to capture all dimensions equally well. While they were very well suited to quantifying aspects of a situation such as time, frequency, and duration of activities, they were limited in their ability to depict the quality of a situation. For example, except for the few keyboard features, the indicators we used contained little information about affective processes in each situation. However, previous studies have demonstrated a significant overlap between measures of affect and situation perception (Horstmann et al., 2021; Horstmann & Ziegler, 2019). Accordingly, the dimensions most associated with positive and negative affect (pOsitivity, Negativity, Deception, and Adversity) (Horstmann et al., 2021) were poorly predicted by our smartphone-sensed situational cues.

At this point, we already anticipate a limitation of our study that will be discussed later. Our study comprised situational cues from five different sensing modalities (i.e., connectivity, smartphone usage, communication, GPS/accelerometers, and timestamp). Yet, due to technical constraints, we lacked other types of data (e.g., ambient noise, light, photos of the surrounding) that have been proven to reveal information about affect (e.g., Lu et al., 2009; Rachuri et al., 2010; Wampfler et al., 2020). The use of smartphones as a research tool holds great promise for situational researchers to gather situational cues at a more sophisticated level in the future. By utilizing more advanced analysis techniques for audio data, such as speaker identification or sound and ambient noise analysis (Kalimeri et al., 2013; Lane et al., 2015; H. Wang et al., 2014; W. Wang et al., 2018), researchers will be able to obtain more in-depth situational cues. Instead of just detecting that a (digital) social interaction has taken place, future models will then be more capable of identifying the emotional tone (i.e., valence) of the interaction and the group of people involved (such as friends, family, colleagues, etc.).

2.5.3. The Diversity of Psychological Situations

Examining the ‘black box’ of machine learning algorithms, our study found that features from all five sensing modalities (connectivity, smartphone usage, communication, GPS/accelerometer, timestamp) were important in predicting at least one of the dimensions of Duty, Intellectuality, Mating, and Sociality. The analysis showed that no single feature (or group of features) was dominant in predicting psychological situations; instead, it was the combination of all features that contributed to the predictions. Similar results have been observed in previous prediction studies in the field of personality computing (Chittaranjan et al., 2011, 2013; Stachl, Au, et al., 2020). This highlights the complexity and diversity of psychological situations, requiring multiple types of sensed situational cues for reliable predictions.

For example, *connectivity* patterns, such as a higher use of the phone’s Bluetooth function, were associated with higher levels of Duty and Mating, but lower levels of Intellectuality and Sociality. Lower *app usage* and shorter phone checks were positively associated with a situation’s Intellectuality and Sociality. Similarly, prior studies have already reported longer periods of concentration and shorter screen times in highly intellectual situations (Breil et al., 2019; Rauthmann & Sherman, 2020). In real-life social contexts, people tend to check their devices briefly rather than use them for longer sessions, being more sociable and outgoing in such situations (Rauthmann et al., 2014). Moreover, our results underpin previous findings on the relevance of the *location* on the perceived Sociality of a situation. Being at work or studying in the classroom together with other colleagues was perceived as social (Blake et al., 2020; Breil et al., 2019; Rauthmann et al., 2014), while being at home was associated with rather unsocial situations except for Mating activities (Blake et al., 2020; Rauthmann et al., 2014). The weekday and time of the day also had an impact on all four dimensions (Duty, Intellectuality, Mating, and Sociality), with later days being lower in Intellectuality and higher in Mating (Blake et al., 2020). These findings support previous research reporting, for example, that these two cues are related to the level of Duty in situations described in social media posts (Serfass & Sherman, 2015).

In conclusion, our results demonstrate the complexity and diversity of real-life situations and highlight the need for a wide range of situational cues to accurately capture a situation. Our results also underscore the potential of smartphone sensing methods to extract as much situational information as possible by considering multiple data types (i.e., sensing modalities) within and around the situation. Nevertheless, the interpretation of the feature importance presented here should be treated with caution due to certain limitations. First, the standardized regression coefficients were relatively small and some features were highly correlated, leading to potential multicollinearity that may have distorted the feature importance scores (Farrar & Glauber, 1967). Secondly, binary-coded status features were more often included among the most important features in our linear model (Lasso) than non-binary-coded timeframe features. This may be a methodological artifact resulting from the model's preferential selection of binary input variables when predicting binary criterion variables. Third, the communication modality was the least represented across dimensions, possibly because people are more likely to use corresponding apps (which were included in the phone usage category) than traditional phone calls or text messages. This resulted in a lack of calls or texts in the selected time frame (60 minutes) and many communication modality features were dropped in the preprocessing pipeline. However, this should not be interpreted to mean that communication is unimportant for situational perception.

2.5.4. Methodological Lessons Learned for Predicting Psychological Situations

Benchmarking of Different Model Classes

Benchmarking different models, our results indicate that the flexible non-linear model (random forest) did not outperform the linear Lasso model. While this is a common finding in psychological research applying machine learning (e.g., Jacobucci et al., 2021; Pargent & Albert-von der Gönna, 2018; Rüegger et al., 2020), this finding does not support prior studies in situation research suggesting non-linear and interaction effects (Rauthmann et al., 2014;

Wrzus & Mehl, 2020). Future research is needed to determine whether this can be generalized to other situation studies.

One possible explanation for this pattern, as noted by Jacobucci and Grimm (2020), is that measurement errors in questionnaires may affect the ability of machine learning algorithms to accurately model true non-linear relationships, regardless of the sample size. This can explain why linear models (e.g., Lasso) are often found to perform just as well as flexible non-linear models (e.g., random forest) in psychological research. In our study, we assessed the criterion variable using the S8-I developed by Rauthmann and Sherman (2016a), which is highly suitable for experience sampling studies but shows limited reliability. This may have influenced our results in favor of the linear model. Additionally, Pargent et al. (2022) found that non-linear and linear models showed different predictive abilities based on the encoding applied to the categorical features. In our study, we used dummy coding for all categorical variables, which might have contributed to the pattern found in our benchmarking results and may be a promising area for further research and investigation.

External Construct Validity of Predictions

We conducted an explorative validation analysis to examine the associations between the Big Five personality traits and the predicted versus self-reported situational characteristics. The aim was to understand the confidence of the predictions and the inferences made based on the estimates of the constructs (Bleidorn & Hopwood, 2019; Rauthmann & Sherman, 2020; Stachl, Pargent, et al., 2020). We found comparable patterns of association between the Big Five personality traits and the self-reported and (above chance) predicted situational characteristics, with slightly higher effects for self-reports, potentially caused by common method biases (Podsakoff et al., 2003). This provides important evidence for the external construct validity of the predicted situational characteristics, with implications for future situation research.

While the additional analysis was exploratory and primarily for validation purposes, we briefly classify the prediction-based association patterns in relation to previous literature. The evidence was mixed, with some findings in line with previous research (e.g., positive associations between Openness and situational Intellectuality (Rauthmann et al., 2014; Sherman et al., 2015)), while others were not consistent (e.g., no positive associations between Sociality and Intellectuality with Extraversion and Agreeableness (Jonason & Sherman, 2020; Rauthmann et al., 2014; Serfass & Sherman, 2015; Sherman et al., 2015)). Further research is needed to confirm these findings and to investigate the differences in outcomes.

2.5.5. Implications for Situation Research

In recent psychological research debates, there has been increasing criticism of scientific psychology's almost exclusive focus on developing mechanistic and complex models to explain and understand psychological phenomena that have little (or unknown) ability to predict future behavior (Stachl, Pargent, et al., 2020; Yarkoni & Westfall, 2017). Our study takes up this debate. By combining smartphone sensing and machine learning techniques in situation research, our findings illustrate the ability of psychological situation theories, such as the situational eight DIAMONDS taxonomy of Rauthmann et al. (2014), not only to provide explanations but also to make (acceptable) predictions. In doing so, our findings make an essential contribution to demonstrating the empirical and practical applicability and relevance of this very well-established theory of individuals' psychological situations and their association with situational cues.

2.5.6. Limitations and Outlook

Finally, our study has some limitations that highlight areas for improvement in future research. Some limitations are related to the use of smartphones for data collection, including potential biases in the situation and person coverage. Participants may not always carry their smartphones close by, leading to a selection bias in the situations sampled. Additionally, our sample only includes Android smartphone users, which might have introduced a bias in terms of personality traits or demographics. However, based on the findings of previous studies, we consider this bias to be a secondary concern (Götz, et al., 2017; Keusch et al., 2020).

Another challenge was to accurately identify the logging data that belonged to a specific situation. In our study, we extracted features based on a one-hour timeframe symmetrically around the experience sampling record. The rationale behind this strategy was to maximize the probability of including all logging events that belong to the respective situation, i.e., to create a representative snapshot of a situation that is oriented toward the average length of situations (Rauthmann & Sherman, 2016b). However, the length of situations can vary, and our feature extraction method might have missed the beginning or end of a situation by artificially truncating it (Rauthmann & Sherman, 2016b). To overcome these limitations, future research could combine smartphone sensing methods with self-reports, such as asking participants to indicate the duration of their current situation (Rauthmann & Sherman, 2016b). Thus, further studies are needed to investigate the most reliable combination of sensing modalities and logging timeframes to accurately identify the situational cues related to the respective situation.

Third, the study focused on five sensing modalities to extract situational cues. Although a comprehensive set of cues was obtained, further advancements in smartphone sensing technology can provide more detailed data. For example, the content of text or information visible on the screen may be valuable predictors of psychological situations (Serfass & Sherman, 2015). The range of features could also be expanded by incorporating environmental information, such as the weather and traffic (von Stumm & d'Apice, 2021), or activity data from smartwatches or personal computers (Grover & Mark, 2017; Mehrotra, et al., 2014).

Furthermore, our study bears some limitations in measuring psychological situational characteristics. To minimize participant burden in our repeated measurement study design, we used a binary scale instead of the Likert scale of the S8-I by Rauthmann and Sherman (2016a). Moreover, we predicted each DIAMONDS dimension separately, but psychological situations are naturally characterized by the combination of different dimensions. For instance, a highly positive and intellectual situation differs from a highly positive and mating situation. While our models only focused on predicting psychological situations using the DIAMONDS dimensions, other taxonomies exist (e.g., Parrigon et al., 2017; Rauthmann et al., 2020), with some critics favoring theoretically derived taxonomies (e.g., Neel et al., 2020; Reis, 2018; Schönbrodt & Hagemeyer, 2015).

It is also important to note that our data were collected during the Covid-19 pandemic, which affected the daily lives of individuals in Germany (Kuper et al., 2021). For example, participants might have encountered fewer social situations, with the workplace reflecting the main social anchor during the pandemic. Additionally, mobility patterns might have varied depending on the individual personality (Chan et al., 2021; Elarde et al., 2021). Therefore, future research needs to examine the replicability of our findings in a post-pandemic context.

2.6. Conclusion

The present research combines smartphone data and self-reports to investigate the relationship between situational cues and psychological situations. The study illustrates that perceived situational characteristics can be partially predicted by objectively sensed situational cues, highlighting the potential of smartphone sensing and machine learning approaches in situation research. Herewith, our findings extend previous findings on the correlates of psychological situations and provide empirical evidence for the practicality of the DIAMONDS taxonomy to gain deeper insights into the human perception of situations in daily life. The study contributes to situation research in psychology by emphasizing the need for multi-method study designs that integrate multiple types of data to fully understand the complexity of everyday life situations.

2.7. Acknowledgments

A special thanks to Ramona Schoedel, Maximilian Bergmann, Florian Bemmann, and Markus Bühner for their input to this study, as well as Larissa Sust for her valuable feedback on some parts of this manuscript. Moreover, we would also like to thank the PhoneStudy team for their continuous and diligent work on the PhoneStudy app. For the data collection in the context of the Smartphone Sensing Panel Study, a special thank you to the cooperation partner, the Leibniz Institute of Psychology (ZPID), especially Michael Bosnjak and Holger Steinmetz, as well as Clemens Stachl. Finally, we would like to thank Michelle Oldemeier and Katja Frey, who provided study support during data collection.

2.8. References

Andone, I., Błaszkiewicz, K., Eibes, M., Trendafilov, B., Montag, C., & Markowetz, A. (2016). How age and gender affect smartphone usage. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 9-12. https://doi.org/10.1145/2968219.2971451

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

Arendasy, M., Sommer, M., & Feldhammer, M. (2011). *Manual Big-Five Structure Inventory BFSI*. Mödling, Austria: Schuhfried GmbH.

Asendorpf, J. B., & Rauthmann, J. F. (2020). States and situations, traits and environments. In G. Matthews & P. J. Corr (Eds.), *The Cambridge handbook of personality psychology* (pp. 56-68). Cambridge University Press. https://doi.org/10.1017/9781108264822.007

Aslak, U., & Alessandretti, L. (2020). *Infostop: Scalable stop-location detection in multi-user mobility data*. ArXiv preprint arXiv:2003.14370. https://doi.org/10.48550/arXiv.2003.14370

Azizyan, M., Constandache, I., & Choudhury, R. R. (2009). SurroundSense: Mobile phone localization via ambience fingerprinting. *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, 261-272. https://doi.org/10.1145/1614320.1614350

Barnett, I., Torous, J., Staples, P., Sandoval, L., Keshavan, M., & Onnela, J. P. (2018). Relapse prediction in schizophrenia through digital phenotyping- a pilot study. *Neuropsychopharmacology, 43*(8), 1660-1666. https://doi.org/10.1038/s41386-018-0030-z

Bates, D., Kliegl, R., Vasishth, S., Baayenc, H., & Baayen, R. H. (2015). *Parsimonious mixed models*. ArXiv preprint arXiv:1506.04967. https://doi.org/10.48550/arXiv.1506.04967

Bates D., Mächler M, Bolker B, Walker S (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/ 10.18637/jss.v067.i01.

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. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Bemmann, F., & Buschek, D. (2020). Languagelogger: A mobile keyboard application for studying language use in everyday text communication in the wild. *Proceedings of the ACM on Human-Computer Interaction, 4*, 1-24. http://dx.doi.org/10.1145/3397872

Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research, 19*(84), 1-5. https://jmlr.org/papers/v19/18-416.html.

Blake, A. B., Lee, D. I., De La Rosa, R., & Sherman, R. A. (2020). Wearable cameras, machine vision, and big data analytics: Insights into people and the places they go. In S. E. Woo, L. Tay, & R. W. Proctor (Eds.), *Big data in psychological research* (pp. 125-143). American Psychological Association. https://doi.org/10.1037/0000193-007

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review, 23*(2), 190-203. https://doi.org/10.1177/1088868318772990

Bornstein, R.F. (2009). Heisenberg, Kandisnsky, and the heteromethod convergence problem: Lessons from within and beyod psychology. *Journal of Personality Assessment, 91*, 1-10. https://doi.org/10.1177/1088868318772990

Breil, S. M., Geukes, K., Wilson, R. E., Nestler, S., Vazire, S., & Back, M. D. (2019). Zooming into real-life extraversion - how personality and situation shape sociability in social interactions. *Collabra: Psychology, 5*(1). https://doi.org/10.1525/collabra.170

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Brown, N. A., Blake, A. B., & Sherman, R. A. (2017). A snapshot of the life as lived: Wearable cameras in social and personality psychological science. *Social Psychological and Personality Science, 8*(5), 592-600. https://doi.org/10.1177/1948550617703170

Brown, N. A., Neel, R., & Sherman, R. A. (2015). Measuring the Evolutionarily Important Goals of Situations: Situational Affordances for Adaptive Problems. *Evolutionary Psychology, 13*, 1-15. https://doi.org/10.1177/1474704915593662

Chan, H. F., Moon, J. W., Savage, D. A., Skali, A., Torgler, B., & Whyte, S. (2021). Can psychological traits explain mobility behavior during the COVID-19 pandemic?. *Social Psychological and Personality Science, 12*(6), 1018-1029. https://doi.org/10.1177/1948550620952572

Chatterjee, S., Bhowmik, A., Singh, A., Ghosh, S., Mitra, B., & Chakraborty, S. (2020). Detecting mobility context over smartphones using typing and smartphone engagement patterns. *2020 IEEE International Conference on Pervasive Computing and Communications*, 1-8. https://doi.org/10.1109/PerCom45495.2020.9127359

Chen, Y.-P., Yang, J.-Y., Liou, S.-N., Lee, G.-Y., & Wang, J.-S. (2008). Online classifier construction algorithm for human activity detection using a tri-axial accelerometer. *Applied Mathematics and Computation, 205*(2), 849-860. https://doi.org/10.1016/j.amc.2008.05.099

Chen, Z., Chen, Y., Hu, L., Wang, S., Jiang, X., Ma, X., Lane, N. D., & Campbell, A. T. (2014). ContextSense: Unobtrusive discovery of incremental social context using dynamic bluetooth data. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 23-26. https://doi.org/10.1145/2638728.2638801

Chen, Z., Chen, Y., Wang, S., Liu, J., Gao, X., & Campbell, A. T. (2013). Inferring social contextual behavior from bluetooth traces. *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing*, 267-270. https://doi.org/10.1145/2494091.2494176

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2011). Who's who with big-five: Analyzing and classifying personality traits with smartphones. *5th Annual International Symposium on Wearable Computers*, 29-36. https://doi.org/10.1109/ISWC.2011.29

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

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin, 52*(4), 281. https://doi.org/10.1037/h0040957

Destatis (2020). *Erwerbsbeteiligung der Bevölkerung Ergebnisse des Mikrozensus zum Arbeitsmarkt*. Federal Statistical Office of Germany (Destatis). Retrieved from https://www.statistischebibliothek.de/mir/receive/DESerie_mods_00003420

Dey, A. K., Wac, K., Ferreira, D., Tassini, K., Hong, J.-H., & Ramos, J. (2011). Getting closer: An empirical investigation of the proximity of user to their smart phones. *Proceedings of the 13th International Conference on Ubiquitous Computing*, 163-172. https://doi.org/10.1145/2030112.2030135

Do, T. M. T., & Gatica-Perez, D. (2014). The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. *IEEE Transactions on Mobile Computing, 13*(3), 638-648. https://doi.org/10.1109/TMC.2013.19

Elarde, J., Kim, J. S., Kavak, H., Züfle, A., & Anderson, T. (2021). Change of human mobility during COVID-19: A United States case study. *PloS one, 16*(11), e0259031. https://doi.org/10.1371/journal.pone.0259031

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics, 49*(1), 92-107. https://doi.org/10.2307/1937887

Fillekes, Kim, Trumpf, Zijlstra, Giannouli, & Weibel (2019): Assessing Older Adults' Daily Mobility - A Comparison of GPS-Dervied and Self-Reported Mobility Indicators. *Sensors, 19*(20), 4551. https://doi.org/10.3390/s19204551

Friedman, J. H., Hastie, T. J., Tibshirani, R. J., Narasimhan, B., Tay, K., Simon, N., & Qian, J. (2010a). glm*net: Lasso and elastic-net regularized generalized linear models*, R package version 1.9-5. Retrieved from http://CRAN.R-project.org/package=glmnet

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality, 40*(1), 21-34. https://doi.org/10.1016/j.jrp.2005.08.003

Gerpott, F. H., Balliet, D., Columbus, S., Molho, C., & de Vries, R. E. (2018). How do people think about interdependence? A multidimensional model of subjective outcome interdependence. *Journal of Personality and Social Psychology, 115*, 716-742. http://dx.doi.org/10.1037/pspp0000166

Götz, F. M., Stieger, S., & Reips, U. D. (2017). Users of the main smartphone operating systems (iOS, Android) differ only little in personality. *PLoS ONE, 12*(5), e0176921. https://doi.org/10.1371/journal.pone.0176921

Griffo, R., & Colvin, C. R. (2019). An exploration of subjective situation dimensions associated with situation-specific behavior. *Social Psychological and Personality Science, 10*, 504-513. https://doi.org/10.1177/1948550618765073

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of statistical software, 40*, 1-25. Https://doi.org/ 10.18637/jss.v040.i03

Grover, T., & Mark, G. (2017). *Digital footprints: Predicting personality from temporal patterns of technology use.* Paper presented at the Proceedings of the 2017 International Joint Converence on Pervasive and Ubiquitous Computing and Processing 2017 ACM International Symposium on Wearable Computers, Maui, Hawaii. https://doi.org/10.1145/3123024.3123139

Guillaume, E., Baranski, E., Todd, E., Bastian, B., Bronin, I., Ivanova, C., ... & Funder, D. C. (2016). The world at 7: 00: Comparing the experience of situations across 20 countries. *Journal of Personality, 84*(4), 493-509. https://doi.org/10.1111/jopy.12176

Hahsler, M., Piekenbrock, M., & Doran, D. (2019). dbscan: Fast density-based clustering with R. *Journal of Statistical Software, 91*(1), 1-30.

Harari, G. M., Gosling, S. D., Wang, R., & Campbell, A. T. (2015). Capturing situational information with smartphones and mobile sensing methods. *European Journal of Personality, 29*(5), 509-511. https://doi.org/10.1002/per.2032

Harari, G. M., Gosling, S. D., Wang, R., Chen, F., Chen, Z., & Campbell, A. T. (2017). Patterns of behavior change in students over an academic term: A preliminary study of activity and sociability behaviors using smartphone sensing methods. *Computers in Human Behavior*, 67, 129-138. https://doi.org/10.1016/j.chb.2016.10.027

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science, 11*(6), 838-854. https://doi.org/10.1177/1745691616650285

Harari, G. M., Müller, S., & Gosling, S. (2020). Naturalistic assessment of situations using mobile sensing methods. In J. F. Rauthmann, R. A. Sherman, & D. C. Funder (Eds.), *The Oxford Handbook of Psychological Situations* (pp. 299-312). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.14

Harari, G. M., Müller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., ... & Gosling, S. D. (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of personality and social psychology, 119*(1), 204-228. https://doi.org/ 10.1037/pspp0000245

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature, 585*, 357-362. https://doi.org/10.1038/s41586-020-2649-2

Hijmans, R. J., Williams, E., Vennes, C., & Hijmans, M. R. J. (2017). Package 'geosphere'. *Spherical trigonometry, 1*(7).

Hogan, R. (2009). Much ado about nothing: The person-situation debate. *Journal of Research in Personality, 43*(2), 249. https://doi.org/10.1016/j.jrp.2009.01.022

Horstmann, K. T., Rauthmann, J. F., Sherman, R. A., & Ziegler, M. (2021). Unveiling an exclusive link: Predicting behavior with personality, situation perception, and affect in a preregistered experience sampling study. *Journal of Personality and Social Psychology*, 120(5), 1317-1343. https://doi.org/10.1037/pspp0000357

Horstmann, K. T., & Ziegler, M. (2019). Situational perception and affect: Barking up the wrong tree? *Personality and Individual Differences, 136*, 132-139. https://doi.org/10.1016/j.paid.2018.01.020

Horstmann, K. T., & Ziegler, M. (2020). Assessing Personality States: What to consider when constructing personality state measures. *European Journal of Personality, 34*(6), 1037-1059. https://doi.org/10.1002/per.2266

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression.* John Wiley and Sons.

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science, 15*(3), 809-816. https://doi.org/10.1177/1745691620902467

Jacobucci, R., Littlefield, A. K., Millner, A. J., Kleiman, E. M., & Steinley, D. (2021). Evidence of inflated prediction performance: A commentary on machine learning and suicide research. *Clinical Psychological Science, 9*(1), 129-134. https://doi.org/10.1177/2167702620954216

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112) (pp. 129-169). New York, NY: Springer.

Jonason, P. K., & Sherman, R. A. (2020). Personality and the perception of situations: The five and dark triad traits. *Personality and Individual Differences, 163*, 110081. https://doi.org/https://doi.org/10.1016/j.paid.2020.110081

Kalimeri, K., Lepri, B., & Pianesi, F. (2013). Going beyond traits: Multimodal classification of personality states in the wild. *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, 27-34. https://doi.org/10.1145/2522848.2522878

Karikoski, J., & Soikkeli, T. (2013). Contextual usage patterns in smartphone communication services. *Personal and Ubiquitous Computing, 17*(3), 491-502. https://doi.org/10.1007/s00779-011-0503-0

Keusch, F., & Bähr, S., Haas1 G.-C., Kreuter, F., & Trappmann, M. (2020). Coverage Error in Data Collection Combining Mobile Surveys With Passive Measurement Using Apps: Data From a German National Survey. *Sociological Methods & Research,* 1-38. https://doi.org/10.1177/0049124120914924

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (pp. 27-58, pp. 247 -266). Springer. https://doi.org/10.1007/978-1-4614-6849-3

Kuper, N., Modersitzki, N., Phan, L. V., & Rauthmann, J. (2021). The situation during the COVID-19 pandemic: A snapshot in Germany. *PLOS ONE, 16*(2), e0245719. https://doi.org/10.1371/journal.pone.0245719

Küster, L., Trahms, C., & Voigt-Antons, J. (2018). Predicting personality traits from touchscreen based interactions. *Tenth International Conference on Quality of Multimedia Experience*, 1-6. https://doi.org/10.1109/QoMEX.2018.8463375

Lane, N., Lin, M., Rabbi, M., Yang, X., lu, H., Cardone, G., Ali, S., Doryab, A., Berke, E., Campbell, A., & Choudhury, T. (2014). BeWell: Sensing sleep, physical activities and social interactions to promote wellbeing. *Mobile Networks and Applications, 19*, 345-359. https://doi.org/10.1007/s11036-013-0484-5

Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A. T. (2010). A survey of mobile phone sensing. *IEEE Communications Magazine, 48*(9), 140-150. https://doi.org/10.1109/MCOM.2010.5560598

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L., & Bischl, B. (2019). mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software, 4*(44), 1903. https://doi.org/10.21105/joss.01903

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: Using smartphones to link happiness and physical activity. *PloS one, 12*(1), e0160589. https://doi.org/10.1371/journal.pone.0160589

Lewin, K. (1936). *Principles of topological psychology*. McGraw-Hill. https://doi.org/10.1037/10019-000

Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., Gatica-Perez, D., & Choudhury, T. (2012). StressSense: Detecting stress in unconstrained acoustic environments using smartphones. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 351-360. https://doi.org/10.1145/2370216.2370270

Lu, H., Pan, W., Lane, N. D., Choudhury, T., & Campbell, A. T. (2009). SoundSense: Scalable sound sensing for people-centric applications on mobile phones. *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, 165-178. https://doi.org/10.1145/1555816.1555834

Maksymiuk, S, Gosiewska A, Biecek P (2020). *Landscape of R packages for eXplainable Artificial Intelligence*. ArXiv preprint arXiv:2003.14370.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of memory and language, 94*, 305-315. https://doi.org/10.1016/j.jml.2017.01.001

McKinney, W. et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference* (Vol. 445, pp. 51-56).

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology, 90*(5), 862. https://doi.org/10.1037/0022-3514.90.5.862

Mehrotra, A., Müller, S. R., Harari, G. M., Gosling, S. D., Mascolo, C., Musolesi, M., & Rentfrow, P. J. (2017). Understanding the role of places and activities on mobile phone interaction and usage patterns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*(3), 1-22. https://doi.org/10.1145/3131901

Mehrotra, A., Pejovic, V., & Musolesi, M. (2014). SenSocial: A middleware for integrating online social networks and mobile sensing data streams. *Proceedings of the 15th International Middleware Conference*, 205-216. https://doi.org/10.1145/2663165.2663331

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science, 7*(3), 221-237. https://doi.org/10.1177/1745691612441215

Min, J.-K., Wiese, J., Hong, J. I., & Zimmerman, J. (2013). Mining smartphone data to classify life-facets of social relationships. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 285-294. https://doi.org/10.1145/2441776.2441810

Montag, C., Błaszkiewicz, 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. https://doi.org/10.1027/ 1614-0001/ a000139

Montag, C., Błaszkiewicz, K., Lachmann, B., Sariyska, R., Andone, I., Trendafilov, B., & Markowetz, A. (2015). Recorded behavior as a valuable resource for diagnostics in mobile phone addiction: evidence from psychoinformatics. *Behavioral sciences, 5*(4), 434-442. https://doi.org/10.3390/bs5040434

Montoliu, R., Blom, J., & Gatica-Perez, D. (2013). Discovering places of interest in everyday life from smartphone data. *Multimedia Tools and Applications, 1*(62), 179-207. https://doi.org/10.1007/s11042-011-0982-z

Mønsted, B., Mollgaard, A., & Mathiesen, J. (2018). Phone-based metric as a predictor for basic personality traits. *Journal of Research in Personality, 74*, 16-22. https://doi.org/doi.org/10.1016/j.jrp.2017.12.004

Müller, S. R., Peters, H., Matz, S. C., Wang, W., & Harari, G. M. (2020). Investigating the Relationships between Mobility Behaviours and Indicators of Subjective Well-Being

Using Smartphone-Based Experience Sampling and GPS Tracking. *European Journal of Personality, 34*(5), 714-732. https://doi.org/10.1002/per.2262

Nakazawa, M. (2021). *Package 'fmsb'*. Retrieved from https://cran.r-project.org/web/packages/fmsb/index.html

Neel, R., Brown, N. A., & Sng, O. (2020). Evolutionary perspectives on situations. In J. F. Rauthmann, R. A. Sherman, & D. C. Funder (Eds.), *The Oxford handbook of psychological situations* (pp. 112-123). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.9

Niemeyer. G. (2008). *Geohash Algorithm*. Retrieved from http://geohash.org

Oliveira, E. M. R., Viana, A. C., Sarraute, C., Brea, J., & Alvarez-Hamelin, I. (2016). On the regularity of human mobility. *Pervasive and Mobile Computing, 33*, 73-90. https://doi.org/10.1016/j.pmcj.2016.04.005

Ooms, J. (2014). *The jsonlite package: A practical and consistent mapping between json data and r objects*. ArXiv preprint arXiv:1403.2805. https://doi.org/10.48550/arXiv.1403.2805

Oreg, S., Edwards, J. A., & Rauthmann, J. F. (2020). The situation six: Uncovering six basic dimensions of psychological situations from the Hebrew language. *Journal of personality and social psychology, 118*(4), 835. https://doi.org/10.1037/pspp0000280

Orrù, G., Monaro, M., Conversano, C., Gemignani, A., & Sartori, G. (2020). Machine learning in psychometrics and psychological research. *Frontiers in Psychology, 10*(2970). https://doi.org/10.3389/fpsyg.2019.02970

Oulasvirta, A., Rattenbury, T., Ma, L., & Raita, E. (2012). Habits make smartphone use more pervasive. *Personal and Ubiquitous Computing, 16*(1), 105-114. https://doi.org/10.1007/s00779-011-0412-2

Pargent, F., & Albert-von der Gönna, J. (2018). Predictive modeling with psychological panel data. *Zeitschrift für Psychologie, 226*(4), 246-258. https://doi.org/10.1027/2151-2604/a000343

Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2022). Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. *Computational Statistics, 37*, 2671-2692. https://doi.org/10.1007/s00180-022-01207-6

Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology, 112*(4), 642-681. https://doi.org/10.1037/pspp0000111

Pennebaker, J.W., Boyd, R.L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin. HTTPS://DOI.ORG/10.15781/T29G6Z

Phan, L. V., & Rauthmann, J. F. (2021). Personality computing: New frontiers in personality assessment. *Social and personality psychology compass, 15*(7), e12624. https://doi.org/10.1111/spc3.12624

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. *Journal of Appled Psychology, 88*(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rabbi, M., Ali, S., Choudhury, T., & Berke, E. (2011, September). Passive and in-situ assessment of mental and physical well-being using mobile sensors. In *Proceedings of the 13th international conference on Ubiquitous computing* (pp. 385-394). https://doi.org/ 10.1145/2030112.2030164

Rachuri, K. K., Musolesi, M., Mascolo, C., Rentfrow, P. J., Longworth, C., & Aucinas, A. (2010). EmotionSense: A mobile phones based adaptive platform for experimental social psychology research. *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, 281-290. https://doi.org/10.1145/1864349.1864393

Ram, N., Conroy, D. E., Pincus, A. L., Hyde, A. L., & Molloy, L. E. (2012). Tethering theory to method: Using measures of intraindividual variability to operationalize individuals' dynamic characteristics. In G. Hancock & J. Harring (Eds.), *Advances in longitudinal modeling in the social and behavioral sciences* (pp. 81-110). New York: Routledge.

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The situational eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology, 107*(4), 677-718. https://doi.org/10.1037/a0037250

Rauthmann, J. F., Horstmann, K. T., & Sherman, R. A. (2020). The psychological characteristics of situations: Towards an integrated taxonomy. In J. F. Rauthmann, R. A. Sherman, & D. C. Funder (Eds.), *The Oxford handbook of psychological situations* (pp. 389-406). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.19

Rauthmann, J. F., Jones, A. B., & Sherman, R. A. (2016). Directionality of person-situation transactions: Are there spillovers among and between situation experiences and personality states? *Personality and Social Psychology Bulletin, 42*(7), 893-909. https://doi.org/10.1177/0146167216647360

Rauthmann, J. F., & Sherman, R. A. (2020). The situation of situation research: Knowns and unknowns. *Current Directions in Psychological Science, 29*(5), 473-480. https://doi.org/10.1177/0963721420925546

Rauthmann, J. F., & Sherman, R. A. (2016a). Ultra-brief measures for the situational eight DIAMONDS domains. *European Journal of Psychological Assessment, 32*(2), 165-174. https://doi.org/10.1027/1015-5759/a000245

Rauthmann, J. F., & Sherman, R. A. (2016b). Situation change: Stability and change of situation variables between and within persons. *Frontiers in Psychology, 6*(1938). https://doi.org/10.3389/fpsyg.2015.01938

Rauthmann, J. F., & Sherman, R. A. (2018). S8-I - Situational Eight-I - deutsche Fassung [Fragebogen]. In Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID) (Ed.), *Elektronisches Testarchiv* (PSYNDEX Tests-Nr. 9007481). ZPID. https://doi.org/10.23668/psycharchives.823

Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality, 29*(3), 363-381. https://doi.org/10.1002/per.1994

Rauthmann, J. F., Sherman, R. A., Nave, C. S., & Funder, D. C. (2015). Personality-driven situation experience, contact, and construal: How people's personality traits predict characteristics of their situations in daily life. *Journal of Research in Personality, 55*, 98-111. https://doi.org/10.1016/j.jrp.2015.02.003

Reis, H. T. (2018). Why bottom-up taxonomies are unlikely to satisfy the quest for a definitive taxonomy of situations. *Journal of Personality and Social Psychology, 114*(3), 489-492. https://doi.org/10.1037/pspp0000158

Remus, R., Quasthoff, U., & Heyer, G. (2010, May). Sentiws-a publicly available german-language resource for sentiment analysis. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

Rojas IV, M. B., Sadeghvaziri, E., & Jin, X. (2016). Comprehensive review of travel behavior and mobility pattern studies that used mobile phone data. *Transportation Research Record, 2563*(1), 71-79. https://doi.org/10.3141/2563-11

Rüegger, D., Stieger, M., Nißen, M., Allemand, M., Fleisch, E., & Kowatsch, T. (2020). How are personality states associated with smartphone data? *European Journal of Personality, 34*(5), 687-713. https://doi.org/10.1002/per.2309

Saeb, S., Lattie, E. G., Kording, K. P., & Mohr, D. C. (2017). Mobile phone detection of semantic location and its relationship to depression and anxiety. *JMIR mHealth and uHealth, 5*(8), e7297. https://doi.org/10.2196/mhealth.7297

Saeb, S., Lattie, E. G., Schueller, S. M., Kording, K. P., & Mohr, D. C. (2016). The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ, 4*, e2537. https://doi.org/10.7717/peerj.2537

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), e175. https://doi.org/10.2196/jmir.4273

Sandstrom, G. M., Lathia, N., Mascolo, C., & Rentfrow, P. J. (2017). Putting mood in context: Using smartphones to examine how people feel in different locations. *Journal of Research in Personality, 69*, 96-101. https://doi.org/10.1016/j.jrp.2016.06.004

Schmitt, M., Gollwitzer, M., Baumert, A., Blum, G., Gschwendner, T., Hofmann, W., & Rothmund, T. (2013). Proposal of a nonlinear interaction of person and situation (NIPS) model. *Frontiers in Psychology, 4*(499). https://doi.org/10.3389/fpsyg.2013.00499

Schoedel, R., Au, Q., Völkel, S. T., Lehmann, F., Becker, D., Bühner, M., Bischl, B., Hussmann, H., & Stachl, C. (2018). Digital footprints of sensation seeking: A traditional concept in the big data era. *Zeitschrift für Psychologie, 226*(4), 232-245. https://doi.org/10.1027/2151-2604/a000342

Schoedel, R., Kunz, F., Bergmann, M., Bemmann, F., Bühner, M., & Sust, L. (in press). *Snapshots of Daily Life: Situations Investigated Through the Lens of Smartphone Sensing.* PsyArXiv, https://psyarxiv.com/f3htz/

Schoedel, R., & Oldemeier, M. (2020). Basic protocol: Smartphone sensing panel study. *Leibniz Institut für Psychologische Information und Dokumentation (ZPID)*. https://doi.org//10.23668/PSYCHARCHIVES.2901

Schoedel, R., Oldemeier, M., Bonauer, L., & Sust, L. (2022). Dataset for: Systematic categorisation of 3091 smartphone applications from a large-scale smartphone sensing dataset[Data set]. PsychArchives. https://doi.org/10.23668/psycharchives.5362

Schönbrodt, F., & Hagemeyer, B. (2015). The value of past and present cues: An evolutionary perspective. *European Journal of Personality*, *29*. https://doi.org/10.1002/per.2005

Schwarz, N. (2012). Why researchers should think "real-time": A cognitive rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life.* (pp. 22-42). The Guilford Press.

Serfass, D. G., & Sherman, R. A. (2015). Situations in 140 characters: Assessing real-world situations on Twitter. *PLOS ONE*, *10*(11), e0143051. https://doi.org/10.1371/journal.pone.0143051

Servia-Rodríguez, S., Rachuri, K. K., Mascolo, C., Rentfrow, P. J., Lathia, N., & Sandstrom, G. M. (2017). *Mobile sensing at the service of mental well-being: A large-scale longitudinal study*. Paper presented at the Proceedings of the 26th International Conference on World Wide Web, Perth, Australia. https://doi.org/10.1145/3038912.3052618

Sherman, R. A., Nave, C. N., & Funder, D. C. (2010). Situational similarity and personality predict behavioral consistency. *Journal of Personality and Social Psychology, 99*(2), 330-343. https://doi.org/10.1037/a0019796

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology, 109*(5), 872-888. https://doi.org/10.1037/pspp0000036

Spathis, D., Servia-Rodríguez, S., Farrahi, K., Mascolo, C., & Rentfrow, J. (2019). Passive mobile sensing and psychological traits for large scale mood prediction. *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 272-281. https://doi.org/10.1145/3329189.3329213

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., Völkel, S. T., Schuwerk, T., Oldemeier, M., Ullmann, T., Hussmann, H., Bischl, B., & Bühner, M. (2020). Predicting personality from patterns of behavior collected with smartphones. *Proceedings of the National Academy of Sciences, 117*(30), 17680-17687. https://doi.org/10.1073/pnas.1920484117

Stachl, C., Hilbert, S., Au, J. Q., Buschek, D., De Luca, A., Bischl, B., ... & Wrzus, C. (2017). Personality traits predict smartphone usage. *European Journal of Personality, 31*(6), 701-722. https://doi.org/10.1002/per.2113

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., Gosling, S. D., & Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality, 34*(5), 613-631. https://doi.org/10.1002/per.2257

Statista (2019). *Umfrage zur Nutzung von sozialen Netzwerken nach Endgeräten in Deutschland bis 2018*. Statista Research Department. Retrieved from https://de.statista.com/statistik/daten/studie/676050/umfrage/nutzung-von-sozialen-netzwerken-nach-endgeraeten-in-deutschland/

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, 58*(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Torous, J., Larsen, M. E., Depp, C., Cosco, T. D., Barnett, I., Nock, M. K., & Firth, J. (2018). Smartphones, sensors, and machine learning to advance real-time prediction and interventions for suicide prevention: A review of current progress and next steps. *Current Psychiatry Reports, 20*(7), 51. https://doi.org/10.1007/s11920-018-0914-y

Ushey K. (2020). Renv: Project Environments (Version 0.12.0). Retrieved from https://CRAN.R-project.org/package=renv.

Ushey, K., Allaire, J., & Tang, Y. (2020). reticulate: Interface to 'Python'. https://CRAN. R-project.org/package=reticulate

van Berkel, N., Goncalves, J., Lovén, L., Ferreira, D., Hosio, S., & Kostakos, V. (2019). Effect of experience sampling schedules on response rate and recall accuracy of objective self reports. *International Journal of Human-Computer Studies, 125*, 118-128. https://doi.org/10.1016/j.ijhcs.2018.12.002

Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Verkasalo, H. (2009). Contextual patterns in mobile service usage. *Personal and Ubiquitous Computing, 13*(5), 331-342. https://doi.org/10.1007/s00779-008-0197-0

von Stumm, S., & d'Apice, K. (2022). From genome-wide to environment-wide: Capturing the environome. *Perspectives on Psychological Science*, *17*(1), 30-40. https://doi.org/10.1177/1745691620979803

Wampfler, R., Klingler, S., Solenthaler, B., Schinazi, V. R., & Gross, M. (2020). Affective state prediction based on semi-supervised learning from smartphone touch data. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-13. https://doi.org/10.1145/3313831.3376504

Wang, H., Lymberopoulos, D., & Liu, J. (2014). Local business ambience characterization through mobile audio sensing. *Proceedings of the 23rd International Conference on World Wide Web*, 293-304. https://doi.org/10.1145/2566486.2568027

Wang, H., & Marsella, S. (2017, October). Assessing personality through objective behavioral sensing. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)* (pp. 131-137). IEEE.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A. T. (2014). StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3-14. https://doi.org/10.1145/2632048.2632054

Wang, W., Harari, G. M., Wang, R., Müller, S. R., Mirjafari, S., Masaba, K., & Campbell, A. T. (2018). Sensing behavioral change over time: Using within-person variability features from Mobile sensing to predict personality traits. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 2*, 1-21. https://doi.org/10.1145/3264951

Wei, T., Simko, V., Levy, M., Xie, Y., Jin, Y., & Zemla, J. (2017). Package 'corrplot'. *Statistician*, *56*(316), e24.

Wickham, H. (2016). ggplot2: *Elegant graphics for data analysis*. New York: Springer-Verlag. Retrieved from http://ggplot2.org

Wickham, H. (2019). *httr: Tools for Working with URLs and HTTP*. R package version 1.4.1, URL https://CRAN.R-project.org/package=httr.

Wickham, H., François, R., Henry, L., & Müller, K. (2019). *Dplyr: A grammar of data manipulation*. R package version 0.8.0.1, https://CRAN.R-project.org/package=dplyr

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of open source software, 4*(43), 1686. https://doi.org/10.21105/joss.01686

Wilcockson, T. D. W., Ellis, D. A., & Shaw, H. (2018). Determining typical smartphone usage: What data do we need? *Cyberpsychology, Behavior, and Social Networking, 21*(6), 395-398. https://doi.org/10.1089/cyber.2017.0652

Wolf, P. S., & Jacobs, W. J. (2010). GPS technology and human psychological research: a methodological proposal. *Journal of Methods and Measurement in the Social Sciences, 1*(1), 1-7. https://doi.org/10.2458/v1i1.74

Wright, N. M, & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, *77*(1), 1-17. https://doi.org/10.18637/jss.v077.i01

Wrzus, C., & Mehl, M. R. (2020). Ecological sampling methods for studying everyday situations. In J. F. Rauthmann, R. A. Sherman, & D. C. Funder (Eds.), *The Oxford* *handbook of psychological situations* (pp. 312-330). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.13

Xu, Y., Lin, M., Lu, H., Cardone, G., Lane, N., Chen, Z., Campbell, A., & Choudhury, T. (2013). Preference, context, and communities: A multi-faceted approach to predicting smartphone app usage patterns. *Proceedings of the 2013 International Symposium on Wearable Computers*, 69-76. https://doi.org/10.1145/2493988.2494333

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

Zeileis, A. & Grothendieck, G. (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software, 14*(6), 1-27. https://doi.org/ 10.18637/jss.v014.i06.

Ziegler, M., Horstmann, K. T., & Ziegler, J. (2019). Personality in situations: Going beyond the OCEAN and introducing the situation five. *Psychological Assessment, 31*(4), 567-580. https://doi.org/10.1037/pas0000654

2.9. Appendix

Table A2.1: Overview of Categorizations Applied for Feature Extraction

Category	Description	
Connectivity: Bluetooth devices		
Watch	The connected device is a wearable watch.	
Headset	The connected device is a headset or headphone.	
Phone	The connected device is another smartphone or cordless phone.	
Computer	The connected device is a laptop or desktop computer.	
Health	The connected device is a health-related wearable, such as weighing or pulse rate measure device.	
Car	The connected device is a in car entertainment system.	
HiFi	The connected device is a HiFi system or loudspeaker.	
Uncategorized	Other connected devices.	
n.a.	The connected device reveals no information about its type	
GPS/accelerometer: POIs		***Cohen's Kappa***
Arts & Entertainment	Places providing entertainment and arts such as theatre, museum, or tourist attractions.	1.0
College & University	Educational places such as schools, colleges, universities, or libraries.	1.0
Nightlife Spot	Places visited in the context of nightlife such as bars, pubs, or clubs.	1.0
Food	Places providing food and drinks such as restaurants, cafes, or food markets.	1.0
Outdoors & Recreation	Places visited for outdoor or recreation activities such as sport facilities, parks, or other natural areas.	1.0
Professional & Other Places	Places providing healthcare or governmental services such as administration departments, hospitals, or COVID-19 testing stations.	.82
Religion	Places visited for religious activities such as churches or mosques.	1.0
Residence	Places providing accommodations such as hotels or other lodgings.	1.0
Shop & Service	Places visited for shopping or other services, such as stores, supermarkets, repair, or financial services.	.93
Travel & Transport	Places related to travel and transport activities such as public transport stations, parking sites, or airports.	.98
GPS/accelerometer: Activity states		
Still	The device is still (not moving).	
In a vehicle	The device is in a vehicle, such as a car.	
In a road vehicle	The device is in a vehicle on the road.	
In a four-wheeler vehicle	The device is in a vehicle with four wheels (e.g., car)	
In a two-wheeler vehicle	The device is in a vehicle with two wheels (e.g., motorcycle)	
In a rail vehicle	The device is in a vehicle on rails.	
On a bicycle	The device is on a bicycle.	
On foot	The device is on a user who is walking or running.	
walking	The device is on a user who is walking.	
running	The device is on a user who is running.	
unknown	Unable to detect the current activity.	

Note. POIs = poinst of interest. Cohen's Kappa was calculated per place category to measure the level of agreement between the two independent raters.

Table A2.2: Name and Specification of Key Terms Used in Feature Description

Key term	Specification
Feature types	
Status feature	Prediction variable based on one logging event at the timepoint of the experience sampling
Timeframe feature	Prediction variable based on several logging events within a timeframe of one hour around the experience sampling
Features	
App	Mobile applications that are actively used by the user (e.g., no system applications running in the background)
POI	Points of interests visited by the user, reflected by the place's category merged using different places application programming interfaces (APIs)
Quantifiers	
Min	Minimum value
Max	Maximum value
Average	Measure of central tendency: Median
Variation	Measure of variation: Median absolute deviation around the median
Days	
Weekday	Monday, 07:00 am - Friday, 6:14 pm
Weekend	Friday, 6:15 pm - Monday, 06:59 am
Daytime	
Morning	7:00 am - 10:44 am (on Saturdays and Sundays: 9:00 am - 12:29 am)
Noon	10:45 am - 2:29 pm (on Saturdays and Sundays: 12:30 am - 3:59 pm)
Afternoon	2:30 pm - 6:14 pm (on Saturdays and Sundays: 4:00 pm - 7:29 pm)
Evening	6:15 pm - 10:00 pm (on Saturdays and Sundays: 7:30 pm - 11:00 pm)

Table A2.3: Linear Mixed Model Results for Duty

	Self-reported			Predicted		
Fixed Effects	Estimate	[95% CI]	SE	Estimate	[95% CI]	SE
Intercept	-0.14	[-0.28, 0.00]	0.07	-0.12	[-0.21, -0.03]	0.05
Openness	0.12	[-0.12, 0.37]	0.12	0.02	[-0.14, 0.17]	0.08
Conscient-iousness	0.26	[0.04, 0.50]	0.12	0.01	[-0.13, 0.16]	0.07
Extra-version	-0.20	[-0.47, 0.06]	0.13	0.08	[-0.01, 0.24]	0.09
Agreea-bleness	-0.19	[-0.39, 0.00]	0.10	-0.08	[-0.21, 0.04]	0.06
Emotional Stability	0.06	[-0.15, 0.28]	0.11	0.01	[-0.13, 0.15]	0.07

Table A2.4: Linear Mixed Model Results for Intellectuality

	Self-reported			Predicted		
Fixed Effects	Estimate	[95% CI]	SE	Estimate	[95% CI]	SE
Intercept	-1.32	[-1.53, -1.13]	0.10	-1.43	[-1.64, -1.23]	0.10
Openness	0.40	[0.08, 0.74]	0.17	0.36	[0.02, 0.70]	0.17
Conscient-iousness	0.26	[-0.05, 0.58]	0.16	0.26	[-0.06, 0.58]	0.16
Extra-version	-0.14	[-0.51, 0.22]	0.18	-0.04	[-0.41, 0.34]	0.19
Agreea-bleness	-0.41	[-0.69, -0.15]	0.13	-0.46	[-0.74, -0.19]	0.14
Emotional Stability	-0.06	[-0.36, 0.23]	0.15	-0.01	[-0.32, 0.29]	0.15

Table A2.5: Linear Mixed Model Results for Mating

	Self-reported			Predicted		
Fixed Effects	Estimate	[95% CI]	SE	Estimate	[95% CI]	SE
Intercept	-1.93	[-2.23, -1.63]	0.15	-2.3	[-2.44, -1.83]	0.16
Openness	-0.53	[-1.02, -0.04]	0.25	-0.62	[-1.11, -0.13]	0.25
Conscient-iousness	0.35	[-0.12, 0.81]	0.23	0.42	[-0.04, 0.89]	0.24
Extra-version	0.79	[0.26, 1.32]	0.27	0.86	[0.32, 1.39]	0.27
Agree-ableness	0.10	[-0.30, 0.49]	0.20	0.08	[-0.32, 0.47]	0.20
Emotional Stability	0.22	[-0.21, 0.66]	0.22	0.13	[-0.31, 0.56]	0.22

Table A2.6: Linear Mixed Model Results for Sociality

	Self-reported			Predicted		
Fixed Effects	Estimate	[95% CI]	SE	Estimate	[95% CI]	SE
Intercept	0.42	[0.23, 0.60]	0.09	0.41	[0.23, 0.59]	0.09
Openness	0.28	[-0.03, 0.59]	0.16	0.20	[-0.11, 0.51]	0.16
Conscient-iousness	0.09	[-0.21, 0.38]	0.15	0.07	[-0.22, 0.36]	0.15
Extra-version	0.10	[-0.24, 0.44]	0.17	0.16	[-0.18, 0.49]	0.17
Agree-ableness	-0.09	[-0.34, 0.16]	0.13	-0.05	[-0.30, 0.19]	0.12
Emotional Stability	0.07	[-0.20, 0.35]	0.14	0.06	[-0.21, 0.33]	0.14

Table A2.7: Regression Coefficients of Top Ten Features for the Logistic Lasso Regression

Target	Sensing Modality	Situational Cue	β
Duty			
	Timestamp	weekend (vs. weekday) (SF)	-1.17
	Timestamp	morning (SF)	0.78
	Timestamp	noon (SF)	0.68
	GPS/accelerometer	current place is a shop and service POI (SF)	0.67
	GPS/accelerometer	current location is at work (SF)	0.55
	Timestamp	evening (SF)	-0.51
	GPS/accelerometer	current place is a POI with unknown category (SF)	0.27
	Connectivity	device is connected to a device of unknown type (SF)	0.20
	Timestamp	afternoon (SF)	-0.18
	Connectivity	power cable status is connected (SF)	-0.15
Intellectuality			
	Timestamp	weekend (SF)	-0.88
	Timestamp	morning (SF)	0.43
	GPS/accelerometer	current place is a shop and service POI (SF)	0.39
	Timestamp	evening (SF)	-0.37
	Timestamp	noon (SF)	0.37
	GPS/accelerometer	current location is at work (SF)	0.35
	Connectivity	device is connected to Bluetooth (SF)	-0.16
	Smartphone usage	total number of usages of time apps (TF)	-0.07
	Smartphone usage	total number of usages of news apps (TF)	-0.07
	Smartphone usage	ratio number of phone checks and usage sessions (TF)	0.05
Mating			
	GPS/accelerometer	current location is at work (SF)	-0.61
	Timestamp	weekend (SF)	0.50
	Timestamp	noon (SF)	-0.38
	Timestamp	morning (SF)	-0.29
	Connectivity	device is connected to Bluetooth (SF)	0.27
	GPS/accelerometer	current location is at home (SF)	0.26
	Timestamp	timepoint of experience sampling is at evening (SF)	0.24
	GPS/accelerometer	current place is a shop and service POI (SF)	-0.12
	Timestamp	timepoint of experience sampling is at afternoon (SF)	-0.12
	GPS/accelerometer	total number of different trips (TF)	0.07
Sociality			
	Connectivity	device is connected to Bluetooth (SF)	-0.21
	GPS/accelerometer	current location is at work (SF)	0.20
	GPS/accelerometer	current place is a shop and service POI (SF)	0.13
	Timestamp	evening (SF)	-0.14
	Timestamp	afternoon (SF)	-0.12
	Smartphone usage	total number of usages of time apps (TF)	-0.08
	Communication	median score of LIWC category "pronouns" (TF)	-0.07
	Connectivity	power cable status is connected (SF)	-0.05
	Smartphone usage	ratio number of phone checks and usage sessions (TF)	0.05
	GPS/accelerometer	current place is a POI with n.a. category (SF)	0.05

Note. β = z-standardized coefficient of logistic regression; SF = status feature (moment of experience sampling), TF = timeframe feature (around experience sampling); POI = point of interest; The penalized logistic Lasso regression was fit to the full data set of available Duty (n = 11,506), Intellectuality (n = 11,492), Mating (n = 11,475), and Sociality (n = 11,446) ratings, respectively.

Table A2.8: Grouped Feature Importance of Sensing Modalities for In-sample Predictions of the Logistic Lasso Regression

Target	Sensing Modality	VI
Duty		
	Timestamp	0.05
	GPS/accelerometer	0.01
	Smartphone usage	0.01
	Connectivity	1.40 e^{-04}
	Communication	6.81 e^{-05}
Intellectuality		
	Timestamp	0.07
	GPS/accelerometer	0.02
	Smartphone usage	0.01
	Connectivity	2.42 e^{-04}
	Communication	4.91 e^{-05}
Mating		
	Timestamp	0.03
	GPS/accelerometer	0.02
	Smartphone usage	0.01
	Communication	0.34 e^{-03}
	Connectivity	0.11 e^{-03}
Sociality		
	GPS/accelerometer	-0.02
	Smartphone Usage	-0.01
	Connectivity	-1.62 e^{-03}
	Timestamp	-1.56 e^{-03}
	Communication	1.99 e^{-04}

Note. *VI* = grouped feature importance score representing the difference of the mean dropout AUC loss of the respective feature group compared to the full model (including all features) across ten permutations; The penalized logistic Lasso regression was fit to the full data set of available Duty (n = 11506), Intellectuality (n = 11492), Mating (n = 11475), and Sociality (n = 11446) ratings, respectively; The loss function used to assess feature importance is 1-AUC; Grouped feature importance was calculated using the DALEX (Biecek, 2018) and DALEXtra package (Maksymiuk, et al., 2020).

3. Study 2: Sensing Affective Experience

Smartphones as Mental Well-being Barometers:

Predicting Affect in Daily Life Using Different Sensing Modalities

Note: This study was preregistered: https://doi.org/10.23668/psycharchives.6895. For data privacy reasons, raw smartphone sensing data is not made publicly available. However, to ensure transparency of the analyses, the preprocessing and analysis code, as well as supplemental material of the study are published on the corresponding Open Science Framework: https://osf.io/a6wtc/. The password for the aggregated data can be provided upon request.

3.1. Abstract

Momentary experiences of positive and negative affect are core components of human well-being and performance. A growing number of smartphone-based mental health applications rely on affective experiences in daily life as an important proxy for mental well-being and as an early warning signal for personalized interventions. Therefore, this study investigates whether passively sensed data can be used to recognize individuals' self-reported affective states and traits based on their smartphone sensing data. This exploratory study uses data collected from $N = 453$ participants in a two-week experience sampling wave which was part of the Smartphone Sensing Panel Study (SSPS; Schoedel & Oldemeier, 2020). Different cross-validated machine learning algorithms were compared to predict participants' momentary affect states and traits from a variety of situational and behavioral indicators. In contrast to previous studies, a broad range of different smartphone-sensed data types were combined to capture the complexity and diversity of affective experience. Our findings show that none of the prediction models based solely on smartphone sensing data for affect states and traits performed notably better than our baseline models. Thus, this study reveals the limits and challenges of using smartphone sensing data alone to predict affective experiences in everyday life, underscoring the importance of multi-modal studies to capture the volatility and complexity of human affect.

Keywords: affect, affective experience, mental well-being, smartphone sensing, affective computing, predictive modeling, machine learning

3.2. Introduction

Affective well-being plays a crucial role in our daily lives, as it can influence what we do, how we do it, and how we experience it. In the long run, our affective well-being can impact our quality of life and even our life expectancy (Aichele et al., 2016; Friedman, Kern, et al., 2010; Koopmans et al., 2010). Due to the essential role of affect, many applications have emerged in recent decades to promote mental health by improving our mood (Marzano et al., 2015). In turn, in order to develop interventions to improve affective well-being, it is essential to properly track and detect affect as accurately and unobtrusively as possible.

With the rise of modern technologies, such as smart wearables or smartphones, a growing body of research has emerged. However, the experience of affect is highly complex and volatile, and can be influenced by the interplay of various factors, including situational factors such as the weather (e.g., Kööts et al., 2011; Park et al., 2013) and person-related factors such as biological underpinnings (e.g., cortisol levels, cardiovascular functions) (e.g., Steptoe et al., 2009) or personality (e.g., Cheng & Furnham, 2003). For example, just because the weather is bad today, we do not necessarily feel bad because, despite the bad weather, we are enjoying a nice evening with our loved ones. Our favorite place, like the park in our neighborhood, may be associated with positive feelings at the weekend, but negative ones after a tough day at work. In order to accurately and unobtrusively capture an individual's affective experience, a comprehensive multi-method approach is needed, gathering as much information as possible about the current situation, behavior, as well as person-related characteristics. Therefore, this study combines situational and behavioral information collected from smartphones and external sources with self-reported feelings to investigate how smartphones can be leveraged to predict affective experiences in real-life situations.

3.2.1. Importance of Affect for Mental Well-being

Mental well-being issues have become a global concern due to the fast-paced lifestyles of modern society. According to the World Health Organization (WHO), the burden of mental health disorders is expected to cost the global economy nearly $1 trillion annually by 2030 (World Health Organization, 2017). Affect-related mental health problems such as depression, bipolar disorder, and other mood disorders negatively affect people's quality of life (Fusar-Poli et al., 2015; Jansen et al., 2013). The COVID-19 pandemic has further highlighted the importance of mental health, as it has had a significant impact on people's lifestyles and mental well-being (Aknin et al., 2022; Giuntella et al., 2021). The Global Burden of Disease (GBD)

2020 study estimates that the COVID-19 pandemic has led to a 25% to 30% increase in cases of major depressive disorder worldwide in 2020 (Santomauro et al., 2020).

From a psychobiological perspective, positive affect serves as an important resource for mental well-being, as it helps individuals to withstand daily stress and cope with mental disorders such as depression or anxiety (e.g., Kashdan & Steger, 2006; Ong et al., 2006; Wichers et al., 2020).[15] Good mental health, in turn, has a positive impact on physical health outcomes, bolstering the immune system functioning and buffering the effects of stress (Aichele et al., 2016; Howell et al., 2007, Steptoe et al., 2009; Veenhoven, 2008).

3.2.2. Conceptualization of Affect

The construct of affect is complex and therefore conceptually difficult to grasp. Different disciplines have various theoretical conceptualizations and definitions of affect, with no clear terminology used consistently across studies.[16]

Core Affect

According to Russell (2003), at the heart of any emotionally charged experience is the so-called *core affect*, which can be perceived as good or bad, excited, or unnerved. This affective experience unconsciously influences our perception, cognition, and behavior and can be affected by many internal and external factors (Russell, 2003). Core affect can therefore either be experienced as free-floating (mood) or attributed to a cause (triggering an emotional episode). Russell (2003) describes the concept of core affect as the simplest neurophysiological state, similar to what Thayer (1986) called *activation*, what Watson and Tellegen (1985) called *affect*, and what Morris (1989) called *mood*. In this framework, mood is simply defined as "a prolonged core affect with no object" (Russell, 2003, p.149).

While some studies focus on the effects of mild, non-specific positive and negative affect on thinking and behavior, others concentrate on specific emotional states such as anger. In contrast to affective experience, these distinct emotions are defined as more object-oriented, intense, conscious, and short-lived experiences, such as fear, anger, or disgust (Forgas & Koch, 2013). This study follows the free-floating conceptualization of affective experience and

[15] In the following, we distinguish between the term *mental well-being* (as a broader state of happiness and contentment with low levels of distress and overall good mental health) and *affective well-being* (which refers to the frequency and intensity of positive and negative affective experiences) (American Psychological Association, n.d.).
[16] For a comprehensive discussion of the history and conceptualization of affective structure, see for example Russell (2003) or Watson et al. (1999).

focuses on the terminology of (core) affect. In research, however, the terms (core) affect and mood are often used interchangeably.

Structure of Affect

An empirically well-established conceptualization of affect is the circumplex ordering of stimuli around the dimensions of *valence* (unpleasant vs. pleasant) and *arousal* (low vs. high) (Feldman, 1995; Larsen & Diener, 1992, Russell, 1980; Russell et al., 1989; Watson & Tellegen, 1985). The valence dimension describes the hedonic quality of an affective experience, while arousal refers to the level of stimulation associated with it (Feldman, 1995; Russell, 1980). In response to stimuli, a person can experience an event as neutral (the center), moderate, or extreme (the periphery) (Russell, 2003).

Another structure of affect is based on the dual structure of positive affect (PA) and negative affect (NA) as the general dimensions of affective experience (Watson, et al., 1988). This two-factor approach is widely supported by research and has been linked to psychological constructs such as stress, anxiety, depression, and self-esteem (Merz & Roesch, 2011). Although PA and NA are distinct and separate dimensions, they are modestly and negatively related (Crawford & Henry, 2004; Lonigan et al., 1999; Tellegen et al., 1999, Terraciano et al., 2003).

Trait and State Affect

Affect can be both a (dispositional) trait and a (situational) state, although assessments of affect have typically relied on approaches that overestimate the role of stable traits and underestimate situational variation (Mischel, 2004; Shiffman, et al., 2008). Scholars have argued that affect can be measured either as a trait (e.g., "How do you *usually* feel") or as a state (e.g., "How do you feel *right now*") by modifying survey instructions accordingly (Hufford, 2007; Watson & Clark, 1994). Clearly, trait affect and state affect are interconnected and aggregations of states over time can approximate a trait. For example, a person with a high level of negative affect trait is likely to experience more states of negative affect in daily life. However, although trait and state scores appear to share some variability, traits do not fully explain all momentary affective experiences (Merz & Roesch, 2011; Vaidya, et al., 2002; Watson & Clark, 1997).

3.2.3. Mental Health in the Digital Age

Digital Mental Health Interventions

Even though there is often no lifelong cure for serious mental illness, effective and timely interventions can improve long-term mental well-being (Morriss, et al., 2013). The rise of mobile health technologies has made effective health interventions more accessible, as the number of mobile health (mHealth) apps focusing on mental health has grown rapidly in recent years. According to a 2015 World Health Organization (WHO) survey, 29% of the 15,000 mHealth apps focused on mental health diagnosis, treatment, or support (Anthes, 2016). Meta-reviews have found that these mental health apps promise clear clinical benefits as stand-alone or complementary treatments (e.g., Cornet & Holden, 2018; East & Havard, 2015; Mehrotra & Tripathi, 2018).

Well-designed apps can effectively help people manage serious mental health issues, such as depression, stress, and anxiety (Mohr et al., 2017; Stawarz et al., 2019), as well as psychosis (Marzano et al., 2015; Naslund, et al., 2015; O'Hanlon et al., 2016). A wide range of mHealth apps have been used in intervention studies to increase engagement in activities that alleviate their symptoms, improve treatment adherence, and support mental health self-management (Ben-Zeev et al., 2015; Caldeira et al., 2017). For instance, self-tracking of affective well-being can help users to gain more awareness of their affective experiences, enabling proactive self-regulation and maintenance of mental well-being (Gay et al., 2011; Murnane et al., 2016). This proactive approach to mental health management empowers users to be active and makes mental health resources more accessible (Caldeira et al., 2017). Positive affect can be a valuable resource, enhancing cardiovascular, hormonal, and immune functions (Howell et al., 2007; Steptoe et al., 2009), promoting healthy behaviors like healthy eating and physical activity (Schultchen et al., 2019), and fostering open-minded thinking and effective problem solving (Nelson et al., 2014).

While further research is needed to fully endorse the use of mental health and well-being apps for all populations, several technical, methodological, and privacy challenges still exist (Anthes, 2016; Chandrashekar, 2018; van Ameringen et al., 2017). For example, one methodological challenge is the lack of empirical evidence on mental health apps available on the market, while their efficiency and effectiveness are highly dependent on user adherence (Anthes 2016; Torous et al., 2018). An important improvement factor that has been shown to increase user engagement and adherence is the tailoring of the app content to individual needs (Jakob et al., 2022; Valentine et al., 2022). Accordingly, personalized early warning signs can enable more timely and effective interventions and preventive measures (Abdullah & Choudhury, 2018; Morriss et al., 2013; Paraschakis, 2017).

Digital Mental Health Assessment

Experience Sampling

Accurate and granular monitoring of symptoms and early-warning triggers is crucial for personalized interventions. However, unobtrusively assessing and analyzing individual affective experiences in daily life is challenging, which is why most mental health apps still rely on self-reported tracking (Marzano et al., 2015). The experience sampling (ES) method enables multiple assessments of momentary affect integrated into daily life. The increasing use of technologies such as smartphones has enabled the collection of ES assessments in real-time, leading to the development of mobile apps that prompt users to assess their affect several times a day. These apps use instruments such as two-dimensional affect grids (e.g., LiKamWa et al., 2013; Servia-Rodríguez et al., 2017) or items describing positive or negative emotional experiences (e.g., Zhang et al., 2018). However, self-report measures of affect require the users to correctly reflect on, recall, and report their experiences, even though people may not fully understand all the factors related to their affective experience (Watson, 2020). For example, they may not correctly perceive and interpret their internal physiological cues, such as heart rate or temperature regulation (Ventura-Bort et al., 2021). Finally, their affect may also be related to their willingness to participate in ES assessments, potentially leading to missed assessments of negative affect (Rintala et al., 2020).[17] Thus, alternative assessment approaches to self-reported experience sampling are becoming increasingly important in the development of digital mental health interventions.

Smartphone Sensing

Sensors in devices, such as mobile phones, wearables, and computers, leave a stream of digital traces. This has given rise to an interdisciplinary field of research called *affect recognition* or *affective computing*, aiming to detect a person's affective experience through wearable technology (see Schmidt et al. (2019) for a review). Physiological markers of affective experience have been identified, measured via EEG (Gable et al., 2021; Petrantonakis & Hadjileontiadis, 2010; Stikic et al., 2014), skin conductance, or temperature (Sano et al., 2015, 2018; Steptoe et al., 2009, 2005). However, sensing these markers typically requires using external sensors or devices, which limits their practical applicability in daily life.

The widespread adoption of smartphones and the increasing computational and communication capabilities of these devices, along with the growing number of embedded

[17] A detailed overview of measurement errors in the context of self-reported affect assessments is provided by Gray and Watson (2007, pp. 171-183), as well as Humrichouse et al. (2007, pp. 13-34).

sensors, have made them a viable alternative to wearables. Having become an integral part of people's daily lives, smartphones can be used to passively collect data from large and diverse populations in an unobtrusive, efficient, and ecologically valid way (Dey et al., 2011; Harari et al., 2020; Lane et al., 2010; Miller, 2012). This can be achieved by tapping into the many sensors embedded in the phones that can measure the location, activity, communication, light, sound, and digital device usage, among others. The digital traces people leave behind when using their smartphones can provide valuable information about where they are, what they are doing, and what they are seeing and hearing.

The unobtrusively collected data can then be translated into situational and behavioral indicators of affective experience (Cornet & Holden, 2018; de Vries et al., 2021; Mohr et al., 2017; Onnela & Rauch, 2016). "Putting mood into context", research has shown that timely and accurate collection of data from different sources can serve as reliable indicators for people's affective experience (Sandstrom et al., 2017, p.96). However, most previous studies have only considered the combination of a few selected indicators, often focusing on specific types of data such as GPS (e.g., Ben-Zeev et al., 2015; Canzian & Musolesi, 2015) or touchscreen data (e.g., Wampfler et al., 2020). Therefore, the goal of this study is to leverage a combination of various situational, behavioral, and dispositional indicators identified in previous research to investigate the manifestations of affective well-being in daily life and to give future research an idea of how data collection for personalized interventions in digital mental health could be designed.

3.2.4. Indicators of Affective Well-being

Researchers have identified various behavioral, situational, and dispositional correlates of affective well-being, ranging from the characteristics of the environment or activities to person-related attributes. However, existing studies have mainly focused on small pathological samples (e.g., Ren et al., 2022; R. Wang et al., 2016) or used adolescent samples (e.g., Ben-Zeev et al., 2015; LiKamWa et al., 2013; MacLeod et al., 2021; Messner et al., 2019; Sano et al., 2018), while the generalizability of the findings to larger populations is still to be investigated.

Social Interaction

For instance, social interaction is one of the factors known to promote pleasant feelings (Nauta et al., 2019; Ram et al., 2014). Prior research has observed that pedestrians might feel better in places with a lot of activity going on and many people being around (e.g., Ettema & Smajic, 2015). Thus, high levels of social interaction, including digital communication such as phone calls or text messages, indicate high levels of positive affect (e.g., LiKamWa et al., 2013; R.

Wang et al., 2016, 2014), while a decrease in communication activity indicates increased levels of depressed or stressed feelings (Ben-Zeev et al., 2015; Madan et al., 2010; Messner et al., 2019; Servia-Rodríguez et al., 2017). Accordingly, social media usage that promotes feelings of connectedness with others can boost affective well-being (Verduyn et al., 2017). However, extensive passive usage of social networks (e.g., scrolling) can tip over into negative feelings related to social comparison and envy (Meier & Johnson, 2022).

Places and Mobility

Moreover, how and where individuals move may also be associated with affective well-being. Studies have revealed that the types of places visited are associated with affective well-being (Cai et al., 2018; Chow et al., 2017; Müller et al., 2020; Sandstrom et al., 2017). For example, social places such as restaurants and bars (Cai et al., 2018; Müller et al., 2020) and time spent in nature have been shown to improve affective well-being (Russell et al., 2013). Spending more time at home, in turn, is linked to more negative and less positive affect (e.g., Cai et al., 2018; Chow et al., 2017; Ren et al., 2020) and may even manifest in mental health problems (R. Wang et al., 2016). Decreased or irregular geospatial activities can also serve as a trigger of affective well-being issues (Ben-Zeev et al., 2015; Cai et al., 2018; Canzian & Musolesi, 2015; de Vries, 2021; DeMasi et al., 2017; Lee et al., 2017; Servia-Rodríguez et al., 2017; Spathis et al., 2019).

Smartphone Usage

Smartphone usage data, such as screen time (Messner et al., 2019; Ren et al., 2022; Sano et al., 2018), touch data and typing dynamics (Cao et al., 2017; Wampfler et al., 2020), or app usage patterns (LiKamWa et al., 2013) can serve as important proxies of affective experience. Some studies have also indicated that extended smartphone usage may be associated with lower levels of affective well-being, at least for younger or predisposed groups (MacLeod et al., 2021; Messner et al., 2019; R. Wang et al., 2016). For example, people with higher anxiety scores often spent more time looking at their phone screens (MacLeod et al., 2021).

Sleep Patterns

In addition to our daily activities, our nightly activities may also be affiliated with our affective experiences. Studies have shown that sufficient, uninterrupted sleep is linked to positive affect (Ben-Zeev et al., 2015; DeMasi et al., 2017; Lane et al., 2011; R. Wang et al., 2014). Analyzing screen time during the night, for example, can provide important contextual information about affective experiences (MacLeod et al., 2021).

Diurnal Patterns

Prior research has also found that people tend to exhibit more positively associated behaviors, such as socializing or laughing, in the first eight to ten hours after waking (Hasler et al., 2008). Data from millions of public Twitter messages revealed that people around the world experience higher positive affect after waking time and on weekends (Golder & Macy, 2011), while a review of experience sampling studies found higher affective well-being in the evenings and on weekends (de Vries et al., 2021). In line with the finding of diurnal rhythms in affective experience, several studies have identified the weekday or time of the day as important indicators of affective experience (Cai et al., 2018; Servia-Rodríguez et al., 2017).

Weather

It is a common belief that sunny weather leads to happiness and rain brings sadness. Therefore, the role of weather as an important environmental indicator of people's affect has also been thoroughly investigated. Some previous studies have shown that positive and negative affect are related to factors such as temperature, sunlight, and humidity (Denissen et al., 2008; Kööts et al., 2011; Park et al., 2013). However, the relationship is very complex and can also be influenced by factors such as sociocultural background (Park et al., 2013) and time spent outdoors (Keller et al., 2005).

Personality

In addition to behavioral and situational indicators, personality traits, such as emotional stability, have a strong association with affect (Cheng & Furnham, 2003; Ching et al., 2014; Geukes et al., 2017; Pavot, et al., 1990). Individuals high in Extraversion tend to experience more positive affect, whereas Neuroticism is associated with higher levels of negative affect and arousal (Cheng & Furnham, 2003; Spathis et al., 2019). Thus, combining information on personality traits with smartphone sensing data can provide valuable insights into an individual's affective experience (Denissen et al., 2008; Kööts et al., 2011; Sandstrom et al., 2017; Spathis et al., 2019).

3.2.5. Rationale

Summing up, the diversity of the identified indicators observed for affective experience undermines that "to fully capture the richness and complexity of affective phenomena, we must also study [them] in the muddiness of daily life" (Kuppens et al., 2022, p. 3). Thus, measuring affect in different contexts and in response to different activities or events can help to understand its dynamic and multiple causes. Accordingly, the real-time data collection in the natural environment of participants can provide unique insights into patterns of affective experience. This deeper understanding can, in turn, foster the development of smartphone-based interventions to enhance well-being.

The subjectivity and complexity of affective experience in daily life make it challenging to predict. In line with recent advancements in sensor technology and apps, there is a growing call for applying machine learning methods to contribute to understanding the dynamics of affective experience in daily life (Kuppens et al., 2022). Our study therefore employs a multi-method design, combining smartphone logs with experience sampling assessments, and classical questionnaire data from a large representative sample in Germany. Using a sophisticated machine learning approach, this study aims to explore the accuracy of predicting affective experiences from unobtrusively collected smartphone data, in order to improve future mobile applications for mental well-being assessment and intervention.

Accordingly, the study makes two key contributions to affective computing research. First, it systematically compares different temporal perspectives of affective experience, including both affective experience as a relatively enduring feel-good experience and the prediction of momentary affective states. Second, we aim to fully exploit the potential of smartphone sensing to track affective experience by combining six categories of behavioral, situational, and dispositional indicators of affective well-being: (1) *social interaction*, (2) *places* and *mobility*, (3) *smartphone usage*, (4) *sleep patterns*, (5) *diurnal patterns*, (6) *weather*, and (7) *personality*. Building on previous research in both areas, this study seeks to bridge the gap between affective and smartphone sensing research.

3.3. Method

3.3.1. Procedures

Data collection was conducted as part of the six-month Smartphone Sensing Panel Study (SSPS) from May to December 2020. The study used three data collection modalities: (1) smartphone sensing, (2) experience sampling, and (3) monthly online surveys (for more details, see Study 1 of this dissertation as well as the study protocol of Schoedel and Oldemeier (2020)). The self-report measures analyzed in this study were collected during survey 1 (May 2020; demographics), survey 3 (July 2020; PANAS data), and survey 4 (August 2020; Big Five personality traits). A representative sample of N = 453 participants in Germany provided experience-sampled self-reports and passively collected smartphone data during the second ES wave (9th of September to 4th of October 2020) that was used in this study. The study was preregistered under https://doi.org/10.23668/psycharchives.6895. All deviations from the preregistration in data preprocessing and analysis and the rationale behind them are reported in the online supplemental material published on the OSF project page.

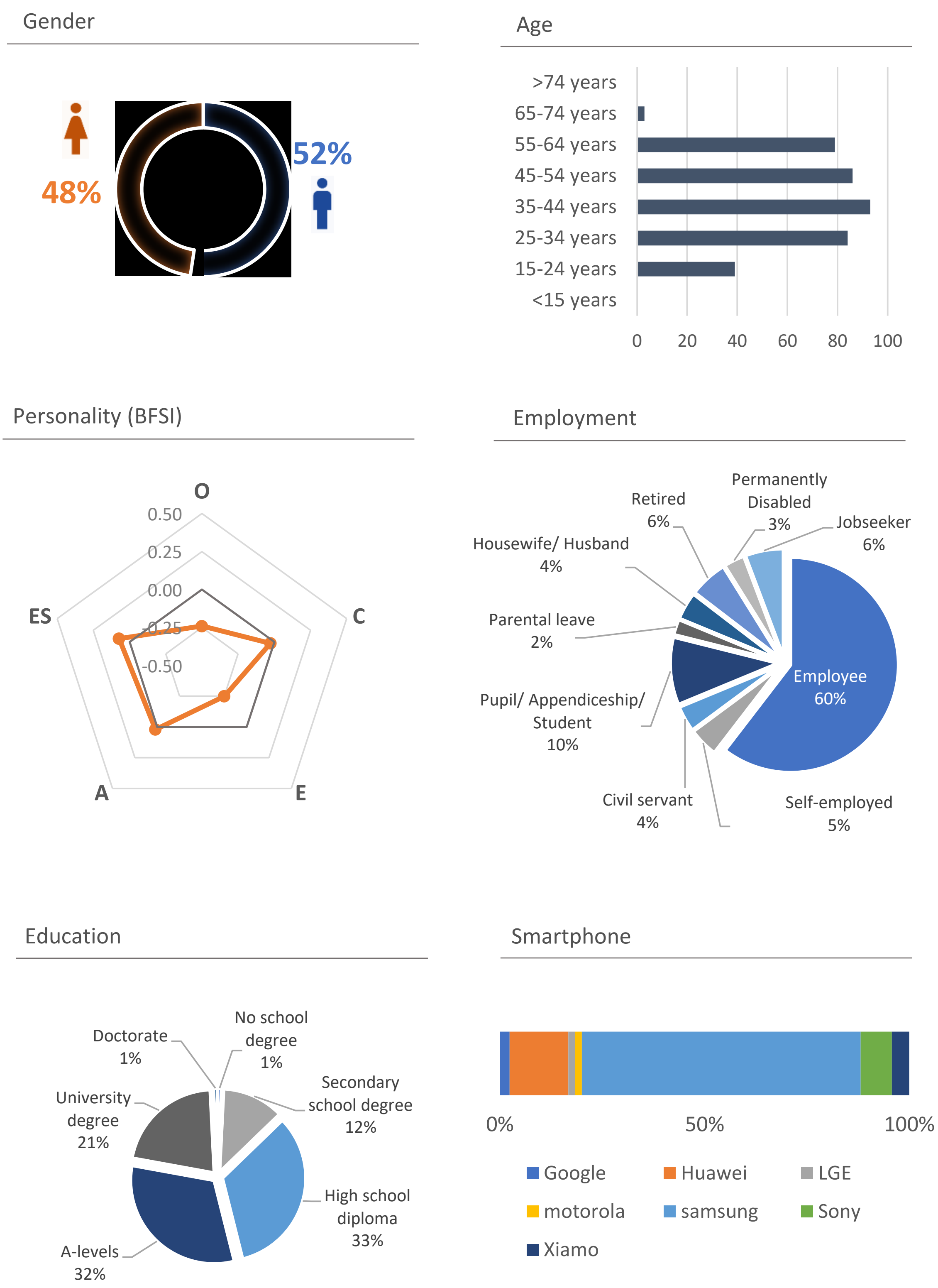
3.3.2. Sample

Sociodemographic characteristics were provided by a subsample of n = 384 participants and are visualized in Figure 3.1. The sample consisted of 48 % female (n = 184) and 52% male (n = 200) participants, aged between 19 and 65 years (M = 42.0, SD = 12.8). The highest level of education was reported by 22% of participants as a university or doctoral degree, 33% as a high school degree, 32% as a secondary school degree, 21% as a secondary general school degree, and less than 1% reported no school degree. Most of the smartphones used were manufactured by Samsung (64%), followed by Huawei (13%) and others (23%).

3.3.3. Data Preparation

Several data exclusion criteria were applied to ensure data quality, as described in the preregistration. First, participants who completed fewer than five experience sampling (ES) reports were excluded. Second, only study days with a minimum of phone usage, defined as a minimum of ten unlocks of the phone screen and a total usage time of 15 minutes, were considered in the analysis. Moreover, at least two experience sampling reports per day with response times of less than 15 minutes were required to enable meaningful predictions. After applying all exclusion criteria, including the preprocessing steps as described below, the final data set included N = 453 participants and 9,460 experience sampling reports, with an average of M = 20.9 (SD = 8.35 [1; 34]) ES reports per user.

Figure 3.1: Information on Study Sample Characteristics



Note. The graphic displays the composition of the sample in terms of gender (in % female vs. male); age (in % per category); Big Five personality traits (BFSI personality self-reports; Arendasy et al., 2009; z-standardized values; O = Openness, C = Conscientiousness, E = Extraversion, A = Agreeableness, ES = Emotional Stability); employment (in %); education (in %); smartphone manufacturers of participants' device (in %). Demographic and smartphone information was available for a subsample of $n = 384$, and the BFSI data was available for the full sample of $N = 453$ participants.

3.3.4. Experience Sampling Measures

During the two-week experience sampling phase, participants rated their momentary affect states two to four times per day. Participants were asked to indicate their *valence* ("How do you assess your current emotional state?") and *arousal* ("How do you assess your current activity level") level on a six-point Likert scale (0 = *very unpleasant/very inactive* to 5 = *very pleasant/very activated*). The self-designed single-item measures were based on Russell's (1980) circumplex model and were chosen to keep the participant burden low.

3.3.5. Survey Measures

Affect Traits

In survey 3 of the SSPS (July 2020), participants' affect traits were measured using the German version of the Positive and Negative Affect Schedule (PANAS; Breyer & Bluemke, 2016). The 20-item self-report measure separately evaluates a person's levels of positive affect (PA) and negative affect (NA) (Watson et al., 1988). Participants rated ten adjectives each for PA (e.g., *active, enthusiastic, proud*) and NA (e.g., *afraid, scared, ashamed*) on a five-point- Likert scale (ranging from 1 = *not at all* to 5 = *extremely*). The PANAS is a widely used measure that has demonstrated reliable and valid results (e.g., Serafini et al., 2016; Watson et al., 1988, Watson & Clark, 1994; Wedderhoff et al., 2021).

Personality Traits

We used the well-established Big Five personality trait theory for personality assessment. In survey 4 of the SSPS (August 2020), participants' Big Five personality traits were measured using the German version of the Big Five Structure Inventory (BFSI; Arendasy et al., 2009). This hierarchical personality inventory defines five broad dimensions of personality (*Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, and *Emotional Stability*). The questionnaire consisted of 300 adjectives describing personality, which were rated on a four-point Likert scale (ranging from *untypical for me* to *typical for me*). In our analyses, we used the person parameters of the partial credit model as the personality trait scores, as the construction of the BFSI follows item response theory.

3.3.6. Sensing Measures

Moreover, participants downloaded the PhoneStudy research app for Android OS version 5 or higher, which was developed to passively collect data from their phone sensors.[18] The app collected several data types, including (1) GPS logs, (2) phone logs (e.g., calls, texts, app

[18] See https://phonestudy.org for more information on the PhoneStudy project.

usage), (3) connectivity logs (e.g., Bluetooth status, connected devices), and (4) keyboard logs (e.g., number of words typed). Detailed information on the collected smartphone logging events is provided by Schoedel et al. (in press). The raw sensing data logged over the two-week ES period was used to extract various sensing variables (also known as *features*).

Feature Engineering Process

Our feature engineering process was guided by a theory-driven approach and began with a comprehensive review of smartphone-sensed indicators in affective computing research. This resulted in a focus on six sensing modalities (reflecting different data types): (1) c*ommunication*, (2) *location,* (3) *music,* (4) s*martphone usage,* (5) *time*, and (6) *weather*. These modalities can measure a range of behavioral and situational indicators of affective experience. An overview of previous literature is provided in Table 3.1, while the sensing modalities used in the present study are briefly introduced below.

Communication

Smartphone sensing research has linked (in-phone) communication patterns to affective experiences, considering the number and length of text messages or emails, as well as phone call records (e.g., Madan et al., 2010; Messner et al., 2019; Servia-Rodríguez et al., 2017). For example, a decrease in communication activities such as calling or texting has been associated with feeling sad or stressed among students (Madan et al., 2010). Other studies have also used systematic logging of keyboard data to gather communication-related information (Cao et al., 2017). Building on such previous research, this study also integrates in-phone communication data in the prediction models.

Location

GPS and accelerometer data are widely used to study affective experiences by tracking users' geolocation traces. This provides insights into mobility patterns and physical activity, such as location variance or distance traveled (de Vries, 2021; DeMasi et al., 2017; Lee et al., 2017; Servia-Rodríguez et al., 2017; Spathis et al., 2019). Moreover, the types of places visited have been shown to reveal information about affective experiences (Cai et al., 2018; Chow et al., 2017; Müller et al., 2020; Sandstrom et al., 2017).

Smartphone Usage

Various affective computing models also included smartphone usage data such as screen time (Messner et al., 2019; Ren et al., 2022; Sano et al., 2018) and app usage patterns (e.g., LiKamWa et al., 2013). In addition, we also extracted connectivity-related smartphone data, such as the device's Bluetooth, power charging, or headphone plugin status, to collect more behavioral and situational information.

Time

We also extracted time-related features such as the weekday and time of the day, as done in previous sensing studies (e.g., Cai et al., 2018; Servia-Rodríguez). By combining these features with others, such as screen time during the night, this study aims to provide important information related to affective computing, such as sleep patterns (MacLeod et al., 2021).

Weather

Weather information, such as temperature, wind power, and sunlight, was also identified in the literature review as an important environmental indicator (Denissen et al., 2008; Kööts et al., 2011). Although not directly recorded by the PhoneStudy app, external weather information was added using the time and GPS logs of the app to form this additional sensing modality (see *Data Enrichment* section).

Data Enrichment

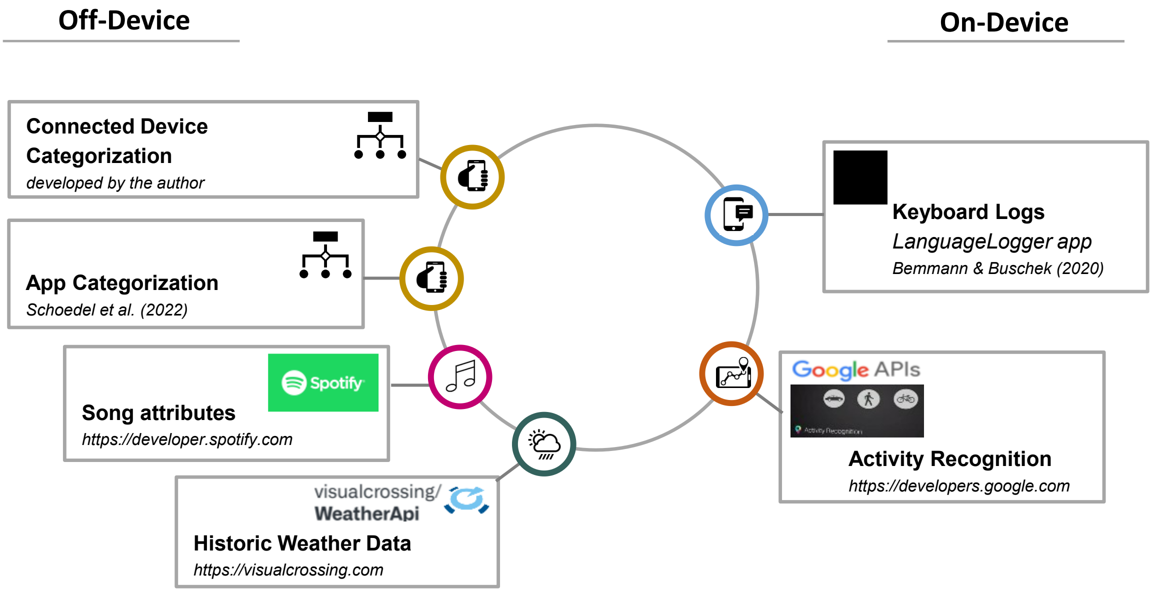
In order to reflect the participants' daily behavior and situations as comprehensively as possible and to cover all the feature types deemed important, the raw smartphone data was enriched with additional data from external sources. This includes both *on-device* data enrichment, which is already embedded in the PhoneStudy app, and *off-device* data enrichment, which is performed after data collection. An overview of the data enrichment steps applied per sensing modality is provided in Figure 3.2. While the categorizations of connected devices and apps were conducted after data collection (off-device), the keyboard logs were preprocessed on-device by integrating the LanguageLogger app (Bemmann & Buschek, 2020) in the PhoneStudy app. Moreover, physical activity recognition was embedded into the PhoneStudy App using the Google Activity Recognition API. Details of the data enrichment processes are provided in Schoedel et al. (in press), which is why we focus exclusively on describing two additional off-device data enrichment steps conducted in the present study: (1) enrichment of *music* features with song attributes and (2) integration of *weather* data. An overview of all raw data categorizations can be found in the Appendix Table A3.1.

Table 3.1: Previous Research on Smartphone-Sensed Indicators of Affective Well-being

Sensing Modality	Feature Class	Exemplary Features [a)]	References
Communication			
	Calls/ text messages	e.g., number of outgoing/ incoming calls, number of outgoing/ incoming text messages, emails	Cai et al., 2018; Lane et al., 2011; LiKamWa et al., 2013; Ma et al., 2012; MacLeod et al., 2021[b)]; Madan et al., 2010; Messner et al., 2019; Servia-Rodríguez et al., 2019; R. Wang et al., 2016[b)]
	Keyboard logs	e.g., semantic text characteristics, keyboard typing dynamics	Cao et al., 2017 [b)]; Carlier et al., 2022; Neviarouskaya, et al., 2011; Nguyen et al., 2014; Z. Wang et al., 2020
Location			
	Places (e.g., GPS, WiFi)	e.g., places visited, homestay, location changes	Cai et al., 2018; Chow et al., 2017; Müller et al., 2020; Ren et al., 2022; Sandstrom et al., 2017; Servia-Rodríguez et al., 2019; R. Wang et al., 2016 [b)]
	Movement (e.g., accelerometer, gyroscope)	e.g., transition time, speed/ acceleration, activity type, location entropy/variance	Ben-Zeev et al., 2015 [b)]; Cai et al., 2018; Canzian & Musolesi, 2015 [b)]; DeMasi et al., 2017; Lane et al., 2011; Lee et al., 2017; LiKamWa et al., 2013; Ma et al., 2012; MacLeod et al., 2021 [b)]; Müller et al., 2020; Ren et al., 2022; Sano et al., 2018; Servia-Rodríguez et al., 2019; Spathis et al., 2019; R. Wang et al., 2016 [b)] , 2014
Music			
	Music listening behavior	e.g., duration of music listening, acoustic characteristics	Miranda et al., 2009 [b)]; Till et al., 2016 [b)]; Zhang et al., 2018
Smartphone Usage			
	Phone screen (on/off)	e.g., screen usage, screen checks	Ben-Zeev et al., 2015 [b)]; DeMasi et al., 2017; Lane et al., 2011; MacLeod et al., 2021[b)]; Messner et al., 2019; Ren et al., 2022; Sano et al., 2018; Wampfler et al., 2020; R. Wang et al., 2016 [b)], 2014
	Phone apps	e.g., app categories used, duration of app usage	LiKamWa et al., 2013; Messner et al., 2019; R. Wang et al., 2016 [b)]
Time			
	Timestamp	e.g., morning, evening, night, weekend vs. weekday	Cai et al., 2018; de Vries et al., 2021; Servia-Rodríguez et al., 2019
Weather			
		e.g., temperature, wind power, sunlight, humidity, barometric pressure	Denissen et al., 2008; Keller et al., 2005; Kööts et al., 2011; Park et al., 2013

Note. The studies included in the literature review considered different conceptualizations and operationalizations of affective experience (e.g., momentary positive/ negative state affect, distinct emotions, mental well-being). [a)] The exemplary features shown were also extracted in the feature engineering process and included in the prediction models of this study. [b)] Studies focused on mental health and disorders (e.g., depression, schizophrenia, suicidality, ...) as target variables.

Figure 3.2: Overview of Off-Device and On-Device Data Enrichment Steps per Sensing Modality



Note. An exemplary overview of off-device data enrichment (i.e., post-hoc data collection) and on-device data enrichment steps (i.e., embedded in the PhoneStudy app) conducted for this study.

Song Attributes

The PhoneStudy app logged the titles of the songs played, which were enriched with audio attributes using the Spotify Web API.[19] The audio features include characteristics such as acoustic, danceability, instrumentality, and liveliness. This allowed us to extract music features, such as the average danceability of songs listened to by a user. All related features are described in detail in Appendix Table A3.2.

Historic Weather Data

By combining GPS and timestamp logs, the hourly and daily weather was retrieved using the Visual Crossing Weather API, which provides historical weather data.[20] The weather data includes information on the cloud coverage, temperature, humidity, moon, precipitation, sun, or wind. Features such as minimum and maximum temperature, average solar radiation, and UV index were calculated from this data. A summary of all weather-related features is provided in the Appendix Table A3.2.

[19] https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-audio-features
[20] https://www.visualcrossing.com/resources/documentation/weather-api/weather-api-documentation/#history

Feature Extraction

Time Windows

The features were extracted by aggregating raw smartphone sensing data for different time frames in order to compare the predictive power of features based on different time windows of smartphone data. In line with theoretical discussions on the duration of affective experience sequences in daily life, previous studies have used various time perspectives, ranging from daily or (bi-)weekly aggregation (Müller et al., 2020; Spathis et al., 2019) to time windows of one to twelve hours (Chow et al., 2017). The time point of the selected time window may also impact predictive accuracy, as historically contextual features may be more predictive than instantaneous features (Cai et al., 2018). Therefore, we decided to systematically compare different time windows of smartphone data before and around the experience-sampled self-reports.

Figure 3.3: Illustration of Feature Types Extracted for Different Time Windows

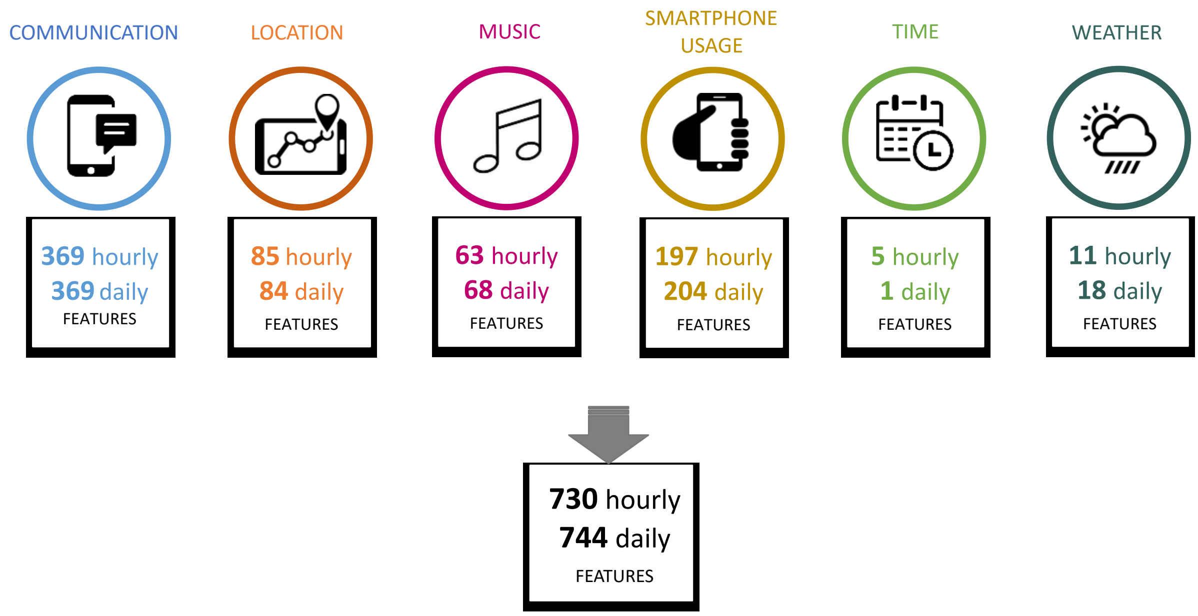
Note. This figure exemplarily outlines the raw sensing data (e.g., GPS) logged during the study. Three different feature types were extracted for different time windows: (1) hourly, (2) daily, and (3) two weeks.

Figure 3.3 exemplarily visualizes smartphone logs, such as GPS data points, that we collected during the two-week experience sampling period. The logs were used to extract two different types of sensing features. The first type of feature is based on data aggregated from the hour before each experience sampling (ES) report and is called *hourly feature*. The second type of feature is based on the sensing data available for the entire 24-hour period of the respective study day and is called *daily feature*. Third, we created a so-called *two-week* feature set by calculating the aggregated value of the daily features per participant over the 14 study days. Different quantification measures (mean, median, mean absolute deviation, minimum, maximum) were used to aggregate the daily features into these two-week feature sets. Including one aggregated observation per smartphone sensing feature per participant, this feature set was only used to predict participants' affect traits. To illustrate the different feature sets and their temporal perspectives, a sketch of three example data sets is provided in Figure A3.1 in the Appendix.

Sensing Modalities

In total, several hundred features were extracted, which can be grouped into the following six sensing modalities: (1) *communication*, (2) *smartphone usage*, (3) *location*, (4) *music*, (5) *time*, and (6) *weather*. Except for some daily weather features that were not available on an hourly basis, similar features were extracted for the hourly and daily time windows. This resulted in 730 hourly sensing features and 744 daily sensing features. In addition, daily features were aggregated for the full two-week time window of the experience sampling study, resulting in 744 two-week features. While Table 3.2 provides an overview of the features and selected examples, a full description of the features extracted from the raw smartphone sensing logs is provided in the Appendix Table A3.2.

Figure 3.4: Total Number of Features Extracted per Sensing Modality and Time Window



Note. This figure illustrates the total number of extracted features per sensing modality and time window of raw sensing data used for feature extraction. The total number of features extracted for the two-week time window equals the total number of daily features ($n_{two\text{-}week}$ = 744 features).

Table 3.2: Overview of Smartphone-Sensed Features per Sensing Modality

Sensing Modality	Feature Class	Feature Examples
Communication	calls	mean duration of outgoing calls
	text messages	variation of text message length
	keyboard usage logs in communication apps	minimum amount of words per message
Location	places [b)]	total time spent *at home*
	altitude	mean altitude of locations visited
	Geohash	total number of different GeoHashs visited
	displacement	radius of gyration of GPS logs
	speed	maximum speed
	mobility/ activity [b)]	mean probability of activity *walking*
Music	listening behavior	variation of duration of sessions listened to songs
	songs	mean level of *danceability* of songs listened to
Smartphone usage	connectivity [b)]	
	power plug status	total duration of power cable status *connected*
	flight mode status	total duration of flight mode status *on*
	Bluetooth connectivity	total duration of smartphone Bluetooth *connected*
	WiFi connectivity	total duration of WiFi status *on* and *disconnected*
	headphones plug status	total duration of headphone status *plugged*
	screen	minimum duration of screen usage session
	apps [b)]	total number of different app categories used
	notification	mean latency of notification caused app usage
Time	weekday	current timestamp is *at the weekend*
	daytime [b)]	current timestamp is *at night*
Weather	clouds	cloud coverage of sky
	temperature	maximum temperature
	humidity	relative humidity
	moon [a)]	daily moonphase
	precipitation	total amount of precipitation
	sun	solar radiation power
	wind	mean windspeed

Note. The table provides an overview of the six sensing modalities of situational and behavioral features. The feature examples reflect only a small selection of all extracted features. All features were created for an hourly (i.e., 60 minutes before experience sampling was started by the participant), daily (i.e., 24 hours of the respective study day), and two-week ES wave time window (i.e., all available study days). [a)] Features were only extractable on a daily and two-week basis due to the underlying database. All quantification metrics (i.e., total, ratio, min, max, mean, variation) of features were based on logging events within these time windows. [b)] Features were extracted for each category of the corresponding categorizations as shown in Appendix Table A2.1 (i.e.; device categories; app categories; LIWC dimensions; Spotify audio feature category; activity recognition categories). All categorical features are dummy coded with 1 = *yes*; 0 = *no*. Keyboard logging features were extracted only for keyboard logs entered in communication applications (i.e., average sentiment of a text message (SentiWS; Remus et al., 2010) and Linguistic Inquiry and Word Count (LIWC, Wolf et al., 2008)).

3.3.7. Data Analyses

All data preprocessing and analyses were performed using R Statistical Software (R Core Team, 2021). As R does not make use of parallelization by default, parallel computing capabilities supported by mlr3 were enabled by using the R-packages *future* (Bengtsson, 2022a) and *progressr* (Bengtsson, 2022b). For reproducibility, we used the package management tool *renv* (Ushey, 2020) and provide a complete list of all R packages used in this paper in the renv.lock file in the corresponding OSF repository via https://osf.io/a6wtc/.

Preprocessing

Due to potential technical recording errors caused by the PhoneStudy app, single observations might have reached extreme values that do not reflect real (extreme) situational or behavioral patterns. Therefore, we used robust estimators (e.g., median, mean absolute deviation) for feature engineering and applied a dedicated data preprocessing procedure.

First, categorical variables (factors) were re-coded as dummy variables and numeric variables were centered and scaled. Second, we decided to replace extreme observations exceeding four standard deviations from the mean with missing values. Additionally, irrelevant and redundant features were excluded from the analyses. Specifically, features with more than 90% missing values, zero or near-zero variance [21], and extremely highly correlated features ($r > .90$) were dropped, following the recommendations of Kuhn and Johnson (2013, p.42). A median imputation algorithm was used as the missing value imputation method to replace missing values.

Benchmark Experiments

Given the exploratory nature of the present study, a predictive modeling approach was applied. Specifically, several benchmark experiments were conducted to investigate our research questions. In predictive modeling, benchmark experiments allow the systematic comparison of different prediction models and data sets in terms of their predictive performance (Hothorn et al., 2005). Table 3.3 provides an overview of the data and machine learning algorithms used to predict the target variables in the benchmarks. As the predicted target variables were operationalized as affect states (repeated measures of valence and arousal at the level of

[21] Concretely, near-zero variance features fulfilled the following two criteria: (1) less than 10% of the observations had unique values, and (2) the frequency ratio of the most common value to the frequency of the second most common value was greater than 19 (95/5).

experience sampling) and affect traits (one measurement of positive and negative affect), the different benchmark experiments included different data sets.

Additional benchmark experiments were conducted for the prediction of both affect states and traits to further investigate whether predictive performance improved when only data collected on the weekend were included. As the results of this additional analysis did not differ from the main analyses reported here, we do not describe the results further in this manuscript.

Table 3.3: Overview of Benchmark Experiments per Target Variable

No.	Feature Set	Algorithm	Target Variable
Benchmark Experiments for Predicting Affect States			
1)	Hourly Daily Hourly + Daily BFSI Hourly + Daily + BFSI Previous Affect State	Featureless, Random Forest, Lasso	Valence
2)	Hourly Daily Hourly + Daily BFSI Hourly + Daily + BFSI Previous Affect State	Featureless, Random Forest, Lasso	Arousal
Benchmark Experiments for Predicting Affect Traits			
3)	Two-week Two-week + BFSI BFSI	Featureless, Random Forest, Lasso	Positive Affect
4)	Two-week Two-week + BFSI BFSI	Featureless, Random Forest, Lasso	Negative Affect

Note. Overview of the four statistical benchmark experiments conducted for the prediction of affect states (i.e., valence; arousal) and affect traits (i.e., positive affect; negative affect) and the feature sets and algorithms compared in the different benchmarks.

Data Sets

Prediction of Affect States. First, we modeled the self-reported momentary affect states assessed via experience sampling. Thus, two separate benchmark experiments were conducted to predict the self-reported scores of valence and arousal at the observational level. We aimed to systematically compare the predictability of smartphone sensing features extracted for different time windows (i.e., hourly and daily features) with self-reported BFSI personality scores. Therefore, the different sensing feature sets were included both individually and in various combinations with and without the personality data. Finally, we included an additional baseline model that included only the previous valence and arousal scores measured in the preceding experience sampling report as features.

Prediction of Affect Traits. Table 3.3 shows that in the two benchmark experiments for predicting self-reported positive and negative affect traits, the daily sensing features were averaged over the two-week ES wave. Additionally, the smartphone sensing feature sets were combined with the BFSI personality trait data, while the third model used only the personality trait scores for prediction.

Machine Learning Algorithms

Besides comparing different data sets, we also benchmarked different algorithms (hereafter referred to as *models*). Serving as the baseline, a featureless model was included in each benchmark experiment. This baseline algorithm does not consider any features, but constantly predicts the observed mean value of the target variable of the training data set for all cases in the test set. In addition, we fit a non-linear random forest (RF; Breiman, 2001) using the implementation called *ranger* (Wright & Ziegler, 2017), which is particularly suitable for high-dimensional data. The default number of $n_{trees} = 500$ regression trees were grown using the estimated variances of the target variable. Third, a regularized logistic linear regression (Lasso; Tibshirani, 1996; Zou & Hastie, 2005) was trained using the mlr3-integrated *glmnet* algorithm (Friedman, Hastie, et al., 2010).

Performance Metrics

To estimate the predictive performance of the models, the data sets were split into a training and a test set. As a resampling strategy, a ten-fold cross-validation scheme with ten repetitions (repeated 10x10 CV) was conducted to evaluate model performance. To avoid over-optimistic performance evaluation, preprocessing was performed within the resampling scheme whenever possible. Due to the nested data structure (i.e., repeated ES observations per person), the resampling procedures included blocking to avoid oversampling of participants with a larger number of ES observations. This method ensures that observations of the same participant are considered to belong together and are not split into the training and test sets.

All models were trained to reduce the mean squared error (MSE) for regression tasks. The MSE is one of the most commonly used metrics for evaluating the performance of models with numeric target variables. The closer a predicted target value r_i is to the true value t_i, the smaller the MSE:

$$\frac{1}{n} w_i \sum_{i=1}^{n} (t_i - r_i)^2 .$$

Moreover, we analyzed the coefficient of determination (R^2) as a measure of performance, which is defined as:

$$1 - \frac{\sum_{i=1}^{n} (t_i - r_i)^2}{\sum_{i=1}^{n} \left(t_i - \bar{t}\right)^2} .$$

The coefficient subtracts the root mean squared error (RSE) from 1, so it compares the squared error of the predictions relative to a naive model predicting the mean. It can be interpreted as the proportion of the variation in the target variable (i.e., affect states/ affect traits) in the data that can be explained by the respective prediction model.

3.4. Results

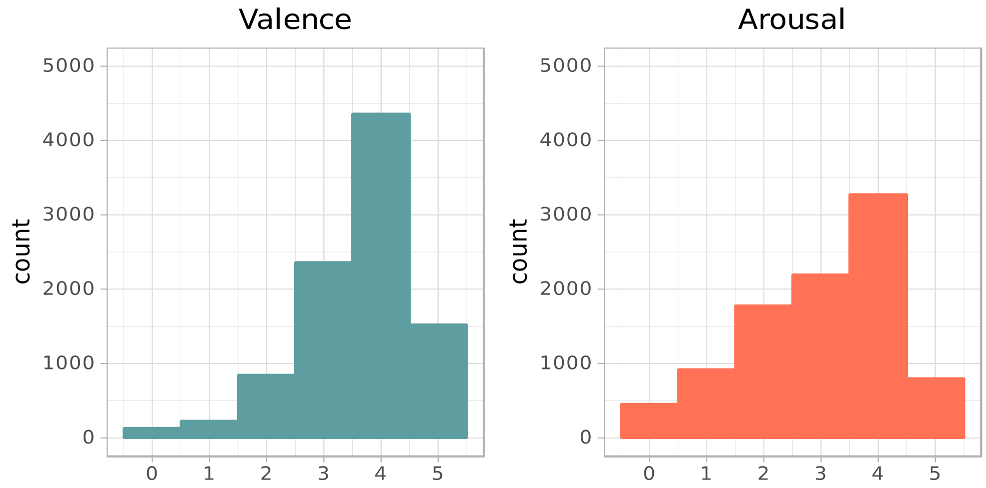
3.4.1. Descriptive Statistics

Depending on the respective target variable, the preprocessing resulted in different sizes of final feature sets and samples (see Table 3.4).

Affect States

A total of 9,460 self-reports of valence and arousal states were collected from 453 users over the 14-day ES wave, with an average score of *M* = 3.62 (*SD* = 1.03) for valence and *M* = 3.01 (*SD* = 1.30) for arousal. Figure 3.5 shows that the skewness of valence was -.96 indicating an extremely left-skewed distribution and -.54 for arousal reflecting a moderately left-skewed distribution across individuals. Intraclass correlations indicated that between-person differences accounted for 49% (for valence) to 39% (for arousal) of the variance in self-reported affect states across experience samplings.

Figure 3.5: Distribution of Affect States on Experience Sampling Level

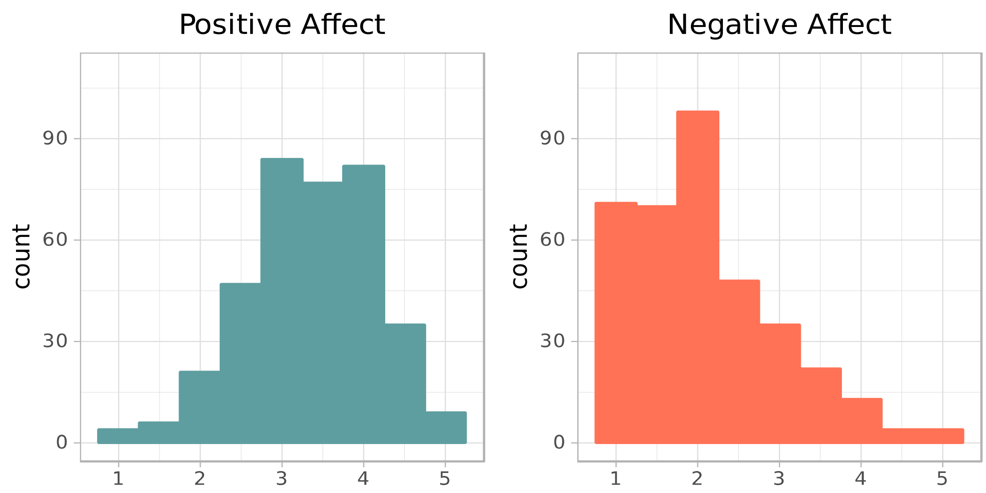


Note. X-axis reflects the scores on a six-point Likert scale ranging from 0 = *very pleasant* to 5 = *very unpleasant* for valence and 0 = *very inactive* to 5 = *very activated* for arousal. Count = number of experience sampling reports with the respective scores (N = 9,460).

Affect Traits

At the trait level, 363 users completed the positive and negative affect questionnaires. On average, participants reported a slightly more positive affect trait, with an average score of M = 3.36 (SD = 0.80) for positive affect, while their average score for negative affect was M = 2.11 (SD = 0.92). Accordingly, as can be seen in Figure 3.6, the skewness of the positive affect was found to be -.33, indicating an almost symmetrical distribution, and that of negative affect was .86, indicating a moderately right-skewed distribution. The kurtosis of the positive affect scores (kappa) was found to be κ = .10, reflecting an almost normal distribution. The kurtosis of the negative affect scores was found to be κ = .39, indicating a slightly heavier-tailed distribution compared to the normal distribution.

Figure 3.6: Distribution of Affect Traits on Experience Sampling Level



Note. X-axis reflects the total positive and negative affect scores which were measured by items participants rated on a five-point Likert scale ranging from 0 = *not at all* to 5 = *extremely*. Count = number of users with respective scores (N = 363).

Table 3.4: Descriptive Statistics of Benchmark Experiments

Benchmark Experiment 1: Prediction of Affect States

Feature set	Sample size	No. of features		No. of observations	
		Before preprocessing	After preprocessing	Valence	Arousal
Hourly	409	730	148	7,879	7,879
Daily	406	744	133	7,994	7,971
Hourly + Daily	406	1,474	281	7,852	7,852
Hourly + Daily + BFSI	349	1,509	317	6,830	6,830
BFSI	390	35	35	8,228	8,228
Previous Affect State	453	1	1	9,456	9,428

Benchmark Experiment 2: Prediction of Affect Traits

Feature set	Sample size	No. of features	No. of observations	
			PA	NA
Two-week	363	1,141	363	363
Two-week + BFSI	363	1,176	363	363
BFSI	363	35	363	363

Note. PA = positive affect, NA = negative affect. Sample sizes reflect the number of users with available feature and target data for each benchmark experiment after preprocessing. The number of observations reflects the number of ES reports after data preparation and preprocessing. The two-week feature sets were extracted by aggregating the preprocessed daily feature sets per participant, hence the number of observations equals the number of participants. The number of features in the two-week feature sets reflects the final number of features included in the prediction models.

Correlations

We found a moderate positive correlation between the arousal and valence states ($r(456) = .42$, $p < .01$). Moreover, there was a moderate positive correlation between positive affect and valence ($r(399) = .40$, $p < .01$) and arousal states ($r(399) = .34$, $p < .01$). Negative affect was moderately correlated with valence ($r(399) = -.30$, $p < .01$) and positive affect ($r(399) = -.37$, $p < .01$), while low correlations were found with arousal ($r(399) = -.11$, $p < .05$).

The detailed descriptive statistics for the self-reported affect states and traits, their intercorrelations, and their correlations with participants' age, gender, and Big Five personality traits are presented in Table A3.3 in the Appendix. In general, very low correlations were found between the affect states and traits with the demographic variables, while some moderate correlations were found with the Big Five personality traits. Furthermore, Pearson correlation coefficients of situational and behavioral features with self-reported affect states and traits were computed and are provided in the online supplemental material (Table S1).[22]

3.4.2. Prediction of Affect States

As the results for different performance metrics (MSE and R^2) were similar, we focus on reporting the MSE as the performance metric of the models in the following. The distributions of the R^2 across resampling iterations by affect state target are shown in Table A3.4 (Figure A3.2). When interpreting the results descriptively, it should be noted that a lower predictive performance of a model is not only expressed by a higher MSE, but also by a larger dispersion of the MSE (i.e., a wider spread of the points). Comparing the different models in our benchmark experiments, our results do not suggest a superior predictive performance of the non-linear random forest model compared to the linear prediction model (Lasso). Following the premise of the highest possible interpretability in the case of equal prediction performance (Kotsiantis, et al., 2007; Molnar, 2022), we focus on the results of the regression-based Lasso model in the following.

For valence, the distribution of the mean squared errors (MSE) across the resampling iterations of the applied 10x10 CV resampling scheme for the Lasso did not differ from the featureless (naïve guessing) models on a descriptive level. The y-axis in Figure 3.7 displays the different feature sets that we compared in the benchmark experiment. The boxplots show that the Lasso models were close to the featureless models for the feature sets containing exclusively sensing features (i.e., hourly, daily, daily + hourly), exclusively personality trait features (BFSI), as well as the combination of sensing and personality features (hourly + daily + BFSI).

[22] https://osf.io/a6wtc/

In contrast, the Lasso model including only the individuals' previous valence states as predictors showed a superior predictive performance (see Table 3.5).

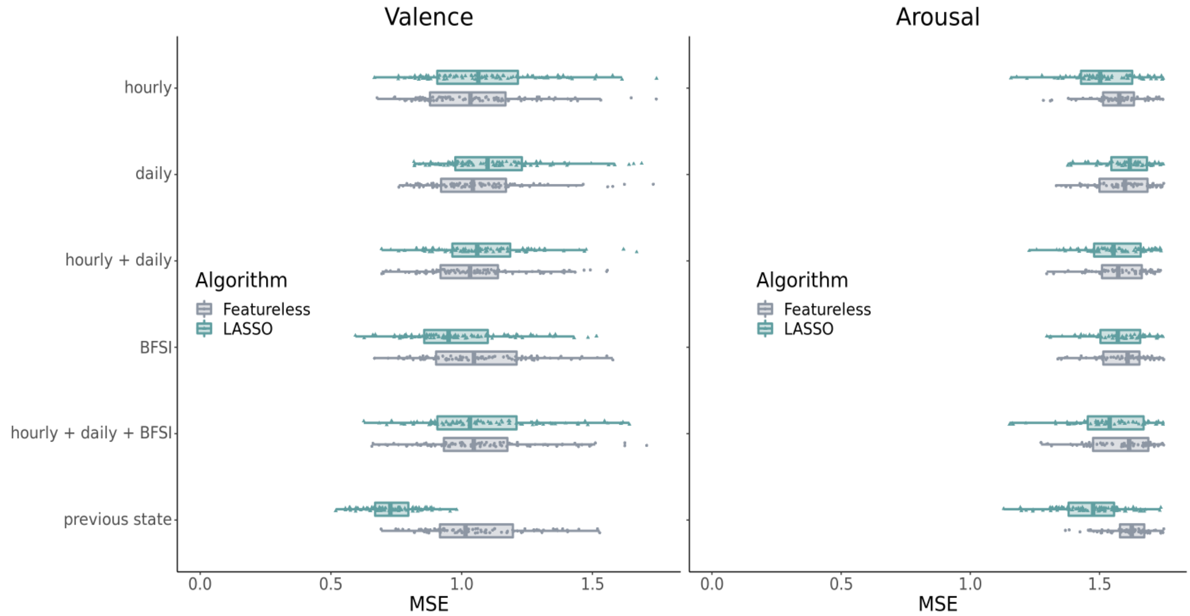
Similarly, for the prediction of arousal states, only the models that included the previous valence states as predictors showed slightly better predictive performance (i.e., lower MSE and lower dispersion of MSE) than the featureless baseline model (see Table 3.5).

Table 3.5: MSEs across Resampling Iterations for the Prediction of Affect States by Feature Set and Algorithm

Prediction of Valence				
Feature Sets	Featureless		Lasso	
	M_{MSE}	SD_{MSE}	M_{MSE}	SD_{MSE}
Hourly	1.06	0.22	1.08	0.22
Daily	1.06	0.19	1.14	0.27
Hourly + Daily	1.07	0.21	1.15	0.31
Hourly + Daily + BFSI	1.07	0.22	1.13	0.33
BFSI	1.07	0.21	0.98	0.19
Previous Affect State	1.06	0.21	0.73	0.10
Prediction of Arousal				
Feature Sets	Featureless		Lasso	
	M_{MSE}	SD_{MSE}	M_{MSE}	SD_{MSE}
Hourly	1.06	0.19	1.09	0.18
Daily	1.70	0.20	1.74	0.19
Hourly + Daily	1.70	0.19	1.62	0.18
Hourly + Daily + BFSI	1.71	0.21	1.68	0.25
BFSI	1.72	0.21	1.70	0.21
Previous Affect State	1.70	0.15	1.47	0.13

Note. This table presents the means and standard deviations of the mean squared error (MSE) measures across the 100 resampling iterations of the 10x10 CV scheme. The performance measures are reported according to the feature sets, which were compared in the benchmark experiment for predicting affect states (i.e., valence; arousal) with the Lasso versus the featureless (naïve guessing) model.

Figure 3.7: Distribution of MSEs across Resampling Iterations for Affect States per Feature Set for the Lasso Model



Note. Distribution of the mean squared errors (MSE) across the resampling iterations of the applied 10x10 CV scheme for Lasso and baseline featureless (naïve guessing) models. The y-axis reflects the different feature sets systematically compared in the benchmark experiment. MSEs of the single iterations are represented by single dots. The boundaries of the boxes in the boxplots indicate the 25th and the 75th percentile, while their middle lines indicate the median.

3.4.3. Prediction of Affect Traits

Similar to the prediction of affect states, we focus on reporting the regression-based Lasso models with the MSE as performance metric for the results of the benchmark experiments conducted to predict the affect traits (i.e., positive affect; negative affect). The distribution of R^2 values across resampling iterations for positive and negative affect is shown in Table A3.5 and Figure A3.3 in the Appendix.

The different feature sets included in the prediction models for positive affect traits revealed that the Lasso model including only the smartphone-sensed features extracted over the 14-day ES wave did not show superior predictive performance compared to the baseline featureless model (Table 3.6; Figure 3.8). The addition of the self-reported Big Five personality traits (BFSI) slightly improved the predictive performances compared to the featureless baseline model. The best predictive performance was observed when the Big Five personality traits were added to the smartphone sensing features. The boxplots in Figure 3.8 also show that the dispersion of the MSEs across iterations was smaller for the model combining smartphone sensing and personality traits than for the model based on personality traits alone.

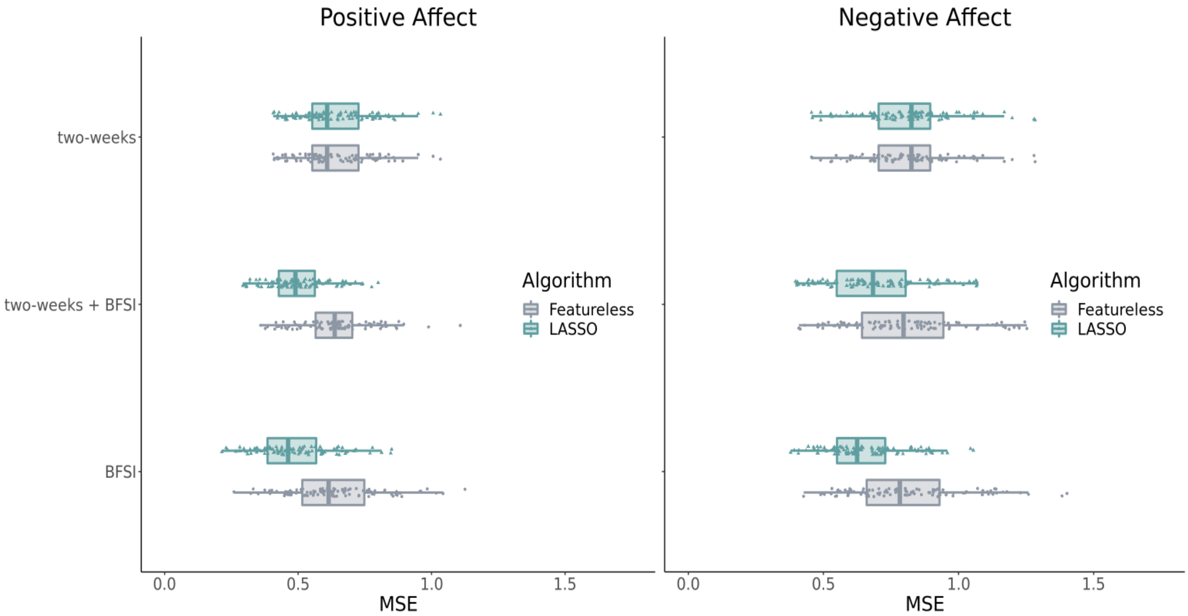
Table 3.6: MSEs across Resampling Iterations for the Prediction of Affect Traits by Feature Set and Algorithm

Prediction of Positive Affect				
Feature Sets	Featureless		Lasso	
	M_{MSE}	SD_{MSE}	M_{MSE}	SD_{MSE}
Two-week	0.64	0.14	0.64	0.13
Two-week + BFSI	0.64	0.14	0.50	0.11
BFSI	0.63	0.17	0.48	0.14
Prediction of Negative Affect				
Feature Sets	Featureless		Lasso	
	M_{MSE}	SD_{MSE}	M_{MSE}	SD_{MSE}
Two-week	0.81	0.17	0.81	0.17
Two-week + BFSI	0.81	0.22	0.70	0.18
BFSI	0.81	0.21	0.66	0.15

Note. This table provides means and standard deviations of the mean squared error (MSE) measures across the 1000 resampling iterations of the repeated 10x10 CV scheme. The performance measures are reported by feature set systematically compared in the benchmark experiment for the prediction of affect traits (i.e., positive affect; negative affect) with the Lasso versus the featureless (naïve guessing) model.

For the prediction of the negative affect traits, the Lasso model combining the 14-day sensing feature set with the Big Five personality traits (BFSI) showed slightly better prediction performance than the featureless baseline model at the descriptive level (Table 3.6; Figure 3.8). Similar results were observed when only the personality trait features (BFSI) were considered, while the comparison of MSEs in this benchmark experiment suggests that the two-week smartphone sensing features did not contribute to an increase in predictive performance.

Figure 3.8: Distribution of MSEs across Resampling Iterations for Affect Traits per Feature Set for the Lasso Model



Note. Distribution of the mean squared errors (MSE) across the resampling iterations of the applied repeated 10x10 CV scheme for Lasso and baseline featureless (naïve guessing) models. The y-axis reflects the different feature sets systematically compared in the benchmark experiment. MSEs of the single iterations are represented by single dots. The boundaries of the boxes in the boxplots indicate the 25th and the 75th percentile, while their middle lines indicate the median.

3.5. Discussion

Applying different machine learning algorithms, this study systematically examined whether passively sensed smartphone sensing data can be used to predict participants' momentary affect states (i.e., valence and arousal) and stable affect traits (i.e., positive and negative affect). Building on previous affect research, we extracted a wide range of situational and behavioral correlates of affective experience in everyday life. In a series of statistical benchmark experiments, the predictive performance of various smartphone sensing feature sets with different temporal perspectives was compared with the predictive performance of personality traits, which served as a baseline to systematically compare self-reported and unobtrusively sensed features. For the prediction of affect states, our prediction models incorporating smartphone sensing data did not perform better than chance. Our results suggest that most of the predictive information about the currently experienced valence and arousal state is provided by the individual's previous affect state. In contrast, the benchmarks for affect traits revealed that combining smartphone sensing data with self-reported personality traits led to an increase in the predictive performance of the prediction models, especially for positive affect, compared to the featureless baseline model.

3.5.1. Predictability of Affect States versus Traits

Previous research has suggested that people experience a more positive mood when being in social situations and that fluctuations in participants' negative state affect can be associated with different contextual aspects, such as the current physical state, geographic location, and time (e.g., Sandstrom et al., 2017; Servia-Rodríguez et al., 2017). However, the prediction results of the present study did not replicate previous findings on the associations between smartphone-sensed indicators and positive and negative affective experiences in daily life (e.g., Cai et al., 2018; Ren et al., 2022; Sandstrom et al., 2017; Servia-Rodríguez et al., 2017).

One possible explanation for the limited comparability of our results with prior studies is the different conceptualization and operationalization of affective experience. As previous research has emphasized, employing multiple assessment methods (e.g., different affect scales) can impact the observed associations and effect sizes (e.g., Lucas & Fujita, 2000). Our study focused on affective experiences as "relatively low-intensity […] and persistent affective states" (Forgas, 2006, pp. 6-7). In contrast, previous studies of affect recognition, particularly those involving smartphone-based communication and text analysis, have tended to concentrate on more distinct and intense emotions as extreme, short-lived experiences (e.g., Arevian et al., 2020; Carlier et al., 2022; Neviarouskaya, et al., 2011; Ren et al., 2022; Z. Wang et al., 2020).

Moreover, research has found substantial differences in predictive performance across different affect variables (e.g., highest for anger and lowest for sadness), suggesting that certain negative emotions may be more strongly linked to passive smartphone features than others (Ren et al., 2022). This also underscores the importance of future studies examining distinct negative emotional states (e.g., sadness, nervousness, or anger) separately, rather than aggregating them into a single negative affect variable, as it is commonly done (Ren et al., 2022).

In addition, our descriptive analyses revealed limited within- and between-person variation in the measured affect states. Similarly, a rather low inter-individual variance was observed for both positive and negative affect traits, suggesting that our data set primarily captured less extreme affective experiences, rather than more extreme cases. However, it is possible that typical day-to-day fluctuations in negative affect are not extreme enough to have a noticeable impact on an individual's cognitive resources (von Stumm, 2016). In other words, only considerable changes in negative affect or clinically low levels of affective experiences may manifest in behavioral and situational characteristics measured via smartphone sensing data, but such cases were rarely observed in the present study. While our research examined affective experiences in a sample representative of the German population in terms of age and gender, previous studies using passively collected data have mostly involved rather small or adolescent samples (e.g., Ben-Zeev et al., 2015; Cai et al., 2018; Cao et al., 2017; Chow et al., 2017; DeMasi et al., 2017; LiKamWa et al., 2013; Messner et al., 2019; Ren et al., 2022) - with some exceptions (e.g., Sandstrom et al., 2017; Servia-Rodríguez et al., 2017). Furthermore, studies using passively sensed data in the context of affective well-being have often focused on predicting more pathological mental well-being constructs such as stress, depression, or anxiety (e.g., Ben-Zeev et al., 2015; Cao et al., 2017; MacLeod et al., 2021; R. Wang et al., 2016). This may be another possible reason for the limited replicability of previous findings in the present study and highlights the importance of further exploring the generalizability of the findings in future research.

Moreover, our study found that models incorporating both smartphone-sensed indicators and personality traits were able to predict positive trait affect slightly better than chance. This suggests that personality traits can, at least to some extent, explain interindividual differences in trait affect (Cheng & Furnham, 2003; Sandstrom et al., 2017; Spathis et al., 2019; Wilt & Revelle, 2019). Previous research in personality psychology has already identified positive relationships between positive affect and Extraversion (e.g., Cheng & Furnham, 2003; Kuppens, et al., 2007; Lucas & Fujita, 2000; Wilt, et al., 2012), as well as Agreeableness and Conscientiousness (e.g., Besser & Shackelford, 2007; Komulainen et al., 2014; Steel, et al., 2008). In line with our findings, Spathis et al. (2019) demonstrated that combining passive sensing with traditional personality survey data can enable affect predictions with higher precision. However, our results did not show similar effects of personality traits on the prediction of affect states. Thus, our findings suggest that enduring personality traits may play a more substantial role in predicting more stable affect traits than in predicting fluctuating momentary affect states, which warrants further investigation in future research. This is inconsistent with previous studies reporting that the Big Five traits predict affect states across five cultures (Ching et al., 2014) and can explain approximately one third of the variance in a person's current momentary affect (with the valence dimension being more predictable than the arousal dimension) (Yik & Russell, 2001). For example, trait neuroticism has been associated with variability in both high- and low-arousal negative affect states, consistent with the trait definition of neuroticism as greater fluctuation in negative affect (Sandstrom et al., 2014). One explanation may be that individuals tend to select contexts (e.g., the company of others or physical activity) that are consistent with their personality traits; for example, extraverted individuals are more likely to be in the company of others, which in turn may influence their momentary affective experience (Wilt & Revelle, 2017). However, due to the ongoing Covid-19 pandemic, participants in our study might have had less flexibility to freely choose their contexts according to their personality traits, as they were constrained by external restrictions (Kuper et al., 2021). It is therefore important to investigate the replicability of the findings in a post-pandemic setting.

3.5.2. Idiographic Approach to Affect Prediction

Further, our findings suggest that the best predictor of an individual's current affective experience is the preceding observation of the affect state. The observed intra-class correlation coefficients also indicate a non-negligible intra-individual variability in affective experiences, especially for valence scores. Similarly, other studies have already observed that a participant is likely to be in a bad mood today if he or she was in a bad mood the day before (e.g., Beltz et al., 2016). In this respect, our findings confirm the importance of studying intra-individual variability in affect, which has already been highlighted in previous studies (e.g., Beltz et al., 2016; Bosley et al., 2020; Eid & Diener, 1999; Fisher et al., 2017; Rodriguez et al., 2022; von Stumm, 2016). For example, the association between social media use and depressive symptoms may vary considerably across individuals, emphasizing the importance of examining intra-individual relationships to predict mental well-being and personalize corresponding treatments (e.g., Rodriguez et al., 2022).

Despite increasing criticism of the underlying assumption of homogeneity across individuals and time, the *nomothetic* approach, i.e., the study of *inter*-individual variation, has long dominated human and social sciences (Molenaar et al., 2004), including affect research (Beltz et al., 2016; De Vries et al., 2021; Fisher et al., 2018). Thus, an often-criticized key limitation of existing studies is that they still primarily focus on predicting average affect scores across individuals, while intra-individual differences and patterns of affective experiences over time are often neglected (e.g., Fisher et al. 2018; Howe et al., 2020; Rodriguez et al., 2022). While there may be some overlap (so-called consensual variance) in affective experiences between individuals due to social expectations (e.g., feeling good when the weather is good), this normative component is likely to be small for assessing how we feel in a particular context. Accordingly, it might have been difficult for our prediction models to recognize and learn patterns across individuals, whereas more individual-specific, personalized prediction models may lead to better predictive performance. In the present study, we collected multiple types of data from a large sample of several hundred participants. However, our predictive modeling approach trained a 'one-size-fits-all' model, without explicitly considering individual-specific predictor combinations or the multi-level structure of an individual's repeated measures. Rather, predictions were made at the level of individual observations (i.e., of single experience samplings of affect).

In contrast, so-called *idiographic* models examine correlations within a single person over many points in time (Haynes et al., 2009; Molenaar, 2004). This approach has recently attracted attention for its ability to model relationships between symptoms and behaviors as they unfold over time within a single individual (e.g., Bosley et al., 2020; Epskamp, et al., 2018; Fisher, et al., 2017, 2018). Accordingly, affect researchers have increasingly turned their attention to the intra-individual variability of affective experiences and new analytical strategies have been developed to accommodate the idiographic-nomothetic paradigm (e.g., Beltz et al., 2016; Jahng et al., 2008; Wigman et al., 2013). Concretely, researchers have several options for considering idiographic prediction in future studies of affect prediction. These approaches include investigating the antecedents of affect through idiographic network-based analyses (e.g., Fisher et al., 2017; Howe et al., 2020; Rodriguez et al., 2022) or idiographic multilevel approaches (e.g., Bosley et al., 2020; Goetz et al., 2010). Moreover, nomothetic and idiographic machine learning models can either be combined (e.g., Jacobson & Chung, 2020), or their performance can be systematically compared with each other (e.g., Cheung et al., 2017). In addition, idiographic studies can apply classical machine learning techniques that are also used in nomothetic prediction, such as random forests (Cheung et al., 2017) or elastic nets (Fisher &

Soyster, 2019). To this end, prediction models can use a standard set of theoretical features or consider individual-specific predictors. For example, Beck and Jackson (2022) found that both personal and situational factors predict future behavior and experiences, although the key predictors can vary widely across individuals. To date, however, idiographic sensing studies have mainly originated from the clinical domain, such as the prediction of depressed mood (Jacobson & Chung, 2020), smoking behavior (Fisher & Soyster, 2019), or alcohol consumption (Soyster et al., 2022). While the discussion of idiographic approaches is accelerating, especially in personality psychology (e.g., Kuper et al., 2021; Renner et al., 2020), scientists should also drive it for other research areas in psychology. Accordingly, we argue that idiographic studies are important to complement nomothetic prediction models and provide a sophisticated understanding of the dynamic processes underlying individual differences in all areas of scientific psychology (Molenaar, 2004).

3.5.3. Challenges of Experience Sampling Affect States

Another possible reason for the limited predictability of affect states could be the experience sampling design underlying our study. While sampling from an individual's everyday life undoubtedly has various advantages, data quality in experience sampling studies is an intensely discussed phenomenon in affect research (see Scollon et al. (2003) for a review).

In contrast to sampling from a set of continuous and objective stimuli, such as smartphone data, in an experience sampling study, the individual participant decides whether or not to respond to a signal. However, one's attention to and willingness to accept new smartphone notifications may be strongly linked to the current situation (Mehrotra et al., 2017). For example, individuals are less likely to be receptive to notifications when they are exercising, away from home, or going to bed (Mehrotra et al., 2017; Rintala et al., 2020). The current activity and location can also be correlated with the momentary affective experience, with positive affect being more likely in social situations, such as at a party with friends (Breil et al., 2019), and negative affect being more likely at work or at home (Chow et al., 2017; Müller et al., 2020). However, people tend to use their smartphones less frequently in social situations outside the home (e.g., Mehrotra et al., 2017; Rintala et al., 2020), potentially leading to missed experience sampling signals and unreported positive affective experiences. Moreover, the time of day may have a salient effect on the probability of non-response, with participants most likely to ignore notifications early in the day (e.g., Courvoisier et al., 2012; Messiah et al., 2011; Silvia et al., 2014).

Moreover, the affective experience itself may also be related to the motivation to participate in a self-report measure, potentially influencing how compliant the participant is in responding to signals (Rintala et al., 2020). However, findings in this area are controversial and require further investigation. For instance, Rintala et al. (2020) observed that individuals with higher positive affect tended to be more compliant in studies using experience sampling, which might have also caused the left-skewed distribution of experience-sampled affect states that we observed in the present study. On the other hand, Courvoisier et al. (2012) found that when people reported high levels of enthusiasm at one signal, they were significantly less likely to respond to the subsequent signal, perhaps because they were engaged in enjoyable and engaging activities that interfered with responding to the subsequent signal.

In addition to the described situational predictors of non-response in experience sampling research, there may be systematic dispositional factors that influence participation compliance. In other words, are people with certain personality traits more likely to respond to self-reports? Thus, to understand the validity of inferences from experience sampling studies, it is important to know whether there are situation- or person-related predictors of missing experience sampling responses. But research on 'knowing the unknown' is challenging because the very data points needed to answer the question are the ones that are missing. While researchers have already gained interesting insights from previous experience sampling data points (e.g., Courvoisier et al., 2012; Rintala et al., 2020), more technically sophisticated methods to collect information about what participants are doing when they miss reports are on the rise. For example, using data from unobtrusive audio recordings, recent research has found very little evidence that missing data in experience sampling correlates with personality or emotion (e.g., Sun et al., 2021). As these findings provide only preliminary evidence for the validity of experience sampling in affect research, future systematic studies are essential to substantiate them.

3.5.4. Limitations and Outlook

In addition, the present study also bears some limitations, which are discussed below and can serve as inspiration for future research.

Smartphone Sensing Modalities

Although our study comprised a considerable number of different indicators, insights from mental health research suggest potential additional indicators that could be leveraged for affective computing research by combining different sensing technologies (Abdullah & Choudhury, 2018).

First, previous studies have highlighted the importance of physiological indicators of affective experience. Therefore, we propose to combine smartphone sensors with physiological sensor measurements such as skin conductance and temperature (e.g., Sano et al., 2018). As an example, wearables such as smartwatches can be leveraged to measure heart rate variability (e.g., Hennekens et al., 2005; Weiner et al., 2011) or electrodermal activity (e.g., Lanata et al., 2014; Schell et al., 2005). Other ideas include integrating data from in-phone cameras to measure the user's facial expressions or eye movements. For instance, research has shown improved emotion classification when the front-facing smartphone camera is used as a tool for emotion detection based on facial expressions (e.g., Kosch et al., 2020; Niforatos & Karapanos, 2015). While research has also already identified eye movement features as important indicators of affect recognition (e.g., Alghowinem et al., 2014; Lu et al., 2015; Partala & Surakka, 2003), the technical feasibility of deploying smartphone in-phone cameras to collect reliable eye-movement data is still under investigation (e.g., Brousseau et al., 2020; Lim et al., 2020).

Second, building on previous research (e.g., Madan et al., 2010; R. Wang et al., 2016, 2014), we incorporated various features of in-phone communication, for example, by applying sentiment analysis to the collected keyboard logging data. For privacy reasons, we did not record speech or log the raw data (i.e., written content) of a user's communication traces, but extracted meta-statistics such as the average sentiment score of the typed text (see Bemmann & Buschek, 2020). On the other hand, when analyzing voice recordings or microphone data, as well as written content such as social media texts, previous studies have found significant correlations between the communicated content and the affective experience, rather than between classical communication features such as the length or number of text messages and calls (e.g., Cai et al., 2018; Carlier et al., 2022; Nguyen et al., 2014; Servia-Rodriguze et al., 2017). In addition, more sophisticated speech analyses (e.g., on acoustic and other paralinguistic parameters of the voice) can reveal important information for tracking affective well-being (e.g., Arevian et al., 2020; R. Wang et al., 2014).

Third, ambient sounds measured by microphone data, such as background noise levels, were found to be significantly associated with affective experiences and to reveal important information about the current environment, for example in combination with GPS data (Servia-Rodríguez et al., 2017; Spathis et al., 2019). In addition, ambient light sensors can be used to determine the position of the phone or detect the brightness of the environment as a promising feature to gain more insights into the current environmental characteristics of an individual (Ben-Zeev et al., 2015; Ma et al., 2012). For instance, prior research has suggested that the noise level of the user's environment may be more informative about the user's mood than their sociability level in terms of communication behavior (e.g., Servia-Rodríguez et al., 2017).

Duration of Affective Experiences

While there is extensive literature on the causes of affective experiences, there is still a lack of empirical research on their interplay with cognition and behavior (Russell, 2003). In general, researchers agree that positive or negative affect influences the way we access and use information stored in our memory (affect congruence) (Forgas, 2017) and interact with others (Forgas, 2002). Nevertheless, it remains an essential question in psychology whether a negative or positive event (e.g., a situational trigger or behavioral pattern) causes momentary affective experiences or vice versa. In the present study, we extracted features using sensing data collected one hour before the moment of assessment (i.e., the experience sampling event), as well as data collected within 24 hours of the respective study day. Thus, the present study zoomed in on participants' patterns one hour before each experience sampling event. In doing so, we also drew on the findings from previous studies that have already demonstrated the predictive relevance of historical contextual features measured prior to the experience sampling (e.g., Cai et al., 2018). However, assuming that affective experiences also trigger the situations and events that we encounter in our daily lives, the time window after experience sampling may also contain valuable information related to the affective experience.

Moreover, surprisingly little research has yet systematically investigated the duration of affective states (Johnson et al., 2008). Research on emotional duration has highlighted that the duration of emotional experiences can be highly variable, with durations ranging from a few seconds to several hours, or even longer, depending on the events we encounter (Verduyn et al., 2009; 2015; Verduyn & Layrisen, 2015). Similarly, negative events were positively associated with negative affect six to nine hours later and were negatively associated with positive affect ratings for approximately three to six hours (Johnson et al., 2008). Thus, to account for the potentially different duration of affective experiences and their manifestations in daily life, future studies should additionally consider longer (e.g., six hours) windows of data collection before and around the time of experience sampling (Cai et al., 2018; Chow et al., 2017). In addition, empirical research can further contribute to the ongoing examination of the duration and dynamics of affective experiences and their behavioral manifestations in everyday life by systematically comparing the informativeness of different time windows of sensing data. For example, there may be systematic inter-individual differences in the duration and persistence of affective responses depending on gender (Johnson et al., 2008) or personality traits (Schimmack, 2000) that should be further explored in future research.

Self-Report Measures of Affect

Another important aspect to discuss is the operationalization of our affect target variables. Our self-report-based measures of momentary and general affective experience require the participants to (1) recognize, (2) reflect on and experience, and (3) quantify and report them accurately (Watson, 2000). This requires a high level of self-awareness of one's own affective experiences, as well as the ability and commitment to report it accurately. Moreover, self-reports of affective experience can be strongly influenced by social expectations, such as the assumption that Monday is the worst day of the week in terms of mood (also known as the 'Monday blues' stereotype) (e.g., Croft & Walker, 2001). Thus, our self-report data on affective traits and states may be prone to measurement errors. Some studies of well-being or emotional experience have attempted to overcome this limitation by integrating external rater scorings (e.g., Ponocny et al., 2016; Sikka et al., 2017; Weismayer, 2021). Nonetheless, given the conceptual understanding of affective experience as a deeply subjective psychological phenomenon, the reliability of such external raters merits additional investigation.

Furthermore, our study used different measures to assess affect traits (i.e., positive and negative affect) and states (i.e., valence and arousal). However, it is more common to ask participants to recall their affect in general (trait) as well as at a particular moment (state), with the same PANAS measure being administered twice (e.g., Hufford, 2007; Kashdan & Roberts, 2004). On the other hand, other studies, have assessed affect several times a day and used the average of the experience sampled affect states over a period of time (e.g., two weeks) as trait estimates (e.g., Merz & Roesch, 2011; Müller et al., 2020). Therefore, the impact of different conceptualizations and operationalizations on the prediction of self-reported affect states and traits based on smartphone sensing data requires further exploration in future studies.

3.6. Conclusion

The successful development of personalized, just-in-time interventions in mental health applications relies on reasonably accurate, real-time predictions of people's affective experiences in everyday life. By employing a comprehensive multi-method design, this study contributes to the growing interdisciplinary body of research on affect recognition. Our findings highlight that affect is a highly complex, volatile, and personal experience that is difficult to predict using passively collected data. While the predictability based solely on smartphone-sensed data needs to be further explored, our research highlights that reliable predictions may require a large set of highly sensitive data points that delve deeply into a person's daily life. Smartphones as data collectors can capture a piece of the complex system of a person's emotion and cognition, behavior, and environment. But only when combined with other sources of information, such as past events, can the full puzzle of our affective experiences in daily life be revealed.

3.7. Acknowledgments

A special thank you to Dr. Ramona Schoedel for her input to this study. My appreciation also goes to the PhoneStudy team for their continuous and diligent work on the PhoneStudy app, that made this research possible. For the data collection in the context of the Smartphone Sensing Panel Study, I would like to thank our cooperation partner, the Leibniz Institute of Psychology (ZPID) and Clemens Stachl, as well as Michelle Oldemeier and Katja Frey, who provided study support during the data collection.

3.8. References

Abdullah, S., & Choudhury, T. (2018). Sensing technologies for monitoring serious mental illnesses. *IEEE MultiMedia*, *25*(1), 61-75. http://dx.doi.org/10.1109/MMUL.2018.011921236

Aichele, S., Rabbitt, P., & Ghisletta, P. (2016). Think fast, feel fine, live long: A 29-year study of cognition, health, and survival in middle-aged and older adults. *Psychological science*, *27*(4), 518-529. http://dx.doi.org/10.1177/0956797615626906

Aknin, L. B., De Neve, J. E., Dunn, E. W., Fancourt, D. E., Goldberg, E., Helliwell, J. F., ... & Ben Amor, Y. (2022). Mental health during the first year of the COVID-19 pandemic: A review and recommendations for moving forward. *Perspectives on psychological science*, *17*(4), 915-936. http://dx.doi.org/10.1177/17456916211029964

Alghowinem, S., Alshehri, M., Goecke, R., & Wagner, M. (2014). Exploring eye activity as an indication of emotional states using an eye-tracking sensor. In L. Chen, S. Kapoor, & R. Bhatia (Eds.), *Intelligent Systems for Science and Information* (pp. 261-276). Springer. https://doi.org/ 10.1007/978-3-319-04702-7_15

Anthes, E. (2016). Mental health: There's an app for that. *Nature, 532*(7597), 20-23. http://dx.doi.org/10.1038/532020a

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

Arevian, A. C., Bone, D., Malandrakis, N., Martinez, V. R., Wells, K. B., Miklowitz, D. J., & Narayanan, S. (2020). Clinical state tracking in serious mental illness through computational analysis of speech. *PLoS one*, *15*(1), e0225695. http://dx.doi.org/10.1371/journal.pone.0225695

Beltz, A. M., Wright, A. G., Sprague, B. N., & Molenaar, P. C. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment*, *23*(4), 447-458. http://dx.doi.org/10.1177/1073191116648209

Bemmann, F., & Buschek, D. (2020). Languagelogger: A mobile keyboard application for studying language use in everyday text communication in the wild. *Proceedings of the ACM on Human-Computer Interaction*, *4*, 1-24. http://dx.doi.org/10.1145/3397872

Ben-Zeev, D., Scherer, E. A., Wang, R., Xie, H., & Campbell, A. T. (2015). Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health. *Psychiatric rehabilitation journal*, *38*(3), 218-226. https://doi.org/10.1037/prj0000130

Bengtsson, H. (2022a). future: Unified Parallel and Distributed Processing in R for Everyone. R package version 1.28.0. Retrieved from https://CRAN.R-project.org/package=future

Bengtsson, H. (2022b). progressr: An Inclusive, Unifying API for Progress Updates. R package version 0.11.0. Retrieved from https://CRAN.R-project.org/package=progressr

Besser, A., & Shackelford, T. K. (2007). Mediation of the effects of the big five personality dimensions on negative mood and confirmed affective expectations by perceived situational stress: A quasi-field study of vacationers. *Personality and Individual Differences*, *42*(7), 1333-1346. http://dx.doi.org/10.1016/j.paid.2006.10.011

Bosley, H. G., Sandel, D. B., & Fisher, A. J. (2020). Idiographic dynamics of positive affect in GAD: Modeling emotion regulation at the person level. *European Journal of Psychological Assessment, 36*(3), 500-509. https://doi.org/10.1027/1015-5759/a000580

Breil, S. M., Geukes, K., Wilson, R. E., Nestler, S., Vazire, S., & Back, M. D. (2019). Zooming into real-life extraversion - How personality and situation shape sociability in social interactions. *Collabra Psychology, 5*(1), Article 7. https://doi.org/10.1525/collabra.170

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5-32. https://doi.org/10.1023/A:1010933404324

Breyer, B., & Bluemke, M. (2016). *Deutsche Version der Positive and Negative Affect Schedule PANAS (GESIS Panel)*. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). https://doi.org/10.6102/zis242

Brousseau, B., Rose, J., & Eizenman, M. (2020). Hybrid eye-tracking on a smartphone with CNN feature extraction and an infrared 3D model. *Sensors, 20*(2), 543. https://doi.org/10.3390/s20020543

Cai, L., Boukhechba, M., Wu, C., Chow, P. I., Teachman, B. A., Barnes, L. E., & Gerber, M. S. (2018, September). State affect recognition using smartphone sensing data. In *Proceedings of the 2018 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies* (pp. 120-125). http://dx.doi.org/10.1145/3278576.3284386

Caldeira, C., Chen, Y., Chan, L., Pham, V., Chen, Y., & Zheng, K. (2017). Mobile apps for mood tracking: an analysis of features and user reviews. In *AMIA Annual Symposium Proceedings* (Vol. 2017, p. 495). American Medical Informatics Association.

Canzian L., & Musolesi M. (2015, September). Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In

Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (pp. 1293-1304). http://dx.doi.org/10.1145/2750858.2805845

Cao, B., Zheng, L., Zhang, C., Yu, P. S., Piscitello, A., Zulueta, J., ... & Leow, A. D. (2017, August). Deepmood: modeling mobile phone typing dynamics for mood detection. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 747-755). https://doi.org/10.1145/3097983.3098086

Carlier, C., Niemeijer, K., Mestdagh, M., Bauwens, M., Vanbrabant, P., Geurts, L., ... & Kuppens, P. (2022). In Search of State and Trait Emotion Markers in Mobile-Sensed Language: Field Study. *JMIR Mental Health*, *9*(2), e31724. http://dx.doi.org/10.2196/31724

Chandrashekar, P. (2018). Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. *Mhealth*, *4*(6). https://doi.org/10.21037/mhealth.2018.03.02

Cheng, H., & Furnham, A. (2003). Personality, self-esteem, and demographic predictions of happiness and depression. *Personality & Individual Differences, 34*, 921-942. http://dx.doi.org/10.1016/S0191-8869(02)00078-8

Cheung, Y. K., Hsueh, P. Y. S., Qian, M., Yoon, S., Meli, L., Diaz, K. M., ... & Davidson, K. W. (2017). Are nomothetic or ideographic approaches superior in predicting daily exercise behaviors?. Methods of information in medicine, 56(06), 452-460. Https://doi.org/ 10.3414/ME16-02-0051

Ching, C. M., Church, A. T., Katigbak, M. S., Reyes, J. A. S., Tanaka-Matsumi, J., Takaoka, S., ... & Ortiz, F. A. (2014). The manifestation of traits in everyday behavior and affect: A five-culture study. *Journal of Research in Personality*, *48*, 1-16. http://dx.doi.org/10.1016/j.jrp.2013.10.002

Chow, P. I., Fua, K., Huang, Y., Bonelli, W., Xiong, H., Barnes, L. E., & Teachman, B. A. (2017). Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. *Journal of medical Internet research*, *19*(3), e6820. http://dx.doi.org/10.2196/jmir.6820

Cornet, V. P., & Holden, R. J. (2018). Systematic review of smartphone-based passive sensing for health and wellbeing. *Journal of biomedical informatics*, *77*, 120-132. http://dx.doi.org/10.1016/j.jbi.2017.12.008

Courvoisier, D. S., Eid, M., & Lischetzke, T. (2012). Compliance to a cell phone-based ecological momentary assessment study: The effect of time and personality characteristics. *Psychological Assessment, 24*(3), 713–720. https://doi.org/10.1037/a0026733

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British journal of clinical psychology*, *43*(3), 245-265. http://dx.doi.org/10.1348/0144665031752934

Croft, G. P., & Walker, A. E. (2001). Are the monday blues ail in the mind? the role of expectancy in the subjective experience of mood. *Journal of Applied Social Psychology*, *31*(6), 1133-1145. https://doi.org/10.1111/j.1559-1816.2001.tb02666.x

DeMasi, O., Feygin, S., Dembo, A., Aguilera, A., & Recht, B. (2017). Well-being tracking via smartphone-measured activity and sleep: cohort study. *JMIR mHealth and uHealth, 5*(10), e7820. http://dx.doi.org/10.2196/mhealth.7820

De Vries, L. P., Baselmans, B. M., & Bartels, M. (2021). Smartphone-based ecological momentary assessment of well-being: A systematic review and recommendations for future studies. *Journal of Happiness Studies*, *22*(5), 2361-2408. http://dx.doi.org/10.1007/s10902-020-00324-7

Denissen, J. J. A., Butalid, L., Penke, L., & van Aken, M. A. G. (2008). The effects of weather on daily mood: A multilevel approach. *Emotion, 8*(5), 662-667. https://doi.org/10.1037/a0013497

Dey, A. K., Wac, K., Ferreira, D., Tassini, K., Hong, J.-H., & Ramos, J. (2011). Getting closer: An empirical investigation of the proximity of user to their smart phones. *Proceedings of the 13th International Conference on Ubiquitous Computing*, 163-172. https://doi.org/10.1145/2030112.2030135

East, M. L., & Havard, B. C. (2015). Mental health mobile apps: from infusion to diffusion in the mental health social system. *JMIR mental health*, *2*(1), e3954. http://dx.doi.org/10.2196/mental.3954

Eid, M., & Diener, E. (1999). Intraindividual variability in affect: Reliability, validity, and personality correlates. *Journal of personality and social psychology*, *76*(4), 662. http://dx.doi.org/10.1037/0022-3514.76.4.662

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior research methods*, *50*(1), 195-212. http://dx.doi.org/10.3758/s13428-017-0862-1

Ettema, D., & Smajic, I. (2015). Walking, places and wellbeing. *The geographical journal*, *181*(2), 102-109. http://dx.doi.org/10.1111/geoj.12065

Feldman, L. A. (1995). Variations in the circumplex structure of mood. *Personality and Social Psychology Bulletin*, *21*(8), 806-817. http://dx.doi.org/10.1177/0146167295218003

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of abnormal psychology, 126*(8), 1044-1056. https://doi.org/10.1037/abn0000311

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106-E6115. http://dx.doi.org/10.1073/pnas.1711978115

Fisher, A. J., & Soyster, P. (2019). *Generating accurate personalized predictions of future behavior: A smoking exemplar*. PsyArXiv, https://europepmc.org/article/ppr/ppr332908/. Https://doi.org/ 10.31234/osf.io/e24v6

Forgas, J. P., & Koch, A. (2013). Mood effects on cognition. In M. D. Robinson, E. R. Watkins, & E. Harmon-Jones (Eds.), *Handbook of emotion and cognition* (pp. 231-251). New York, London: The Guilford Press.

Forgas, J. P. (2017). Mood effects on cognition: Affective Influences on the Content and Process of Information Processing and Behavior. In J. Myounghoon (Ed.), *Emotions and Affect in Human Factors and Human-Computer Interaction* (pp. 89-122). Academic Press. http://doi.org/10.1016/B978-0-12-801851-4.00003-3

Forgas, J. P. (2002). Feeling and doing: Affective influences on interpersonal behavior. *Psychological inquiry, 13*(1), 1-28. https://doi.org/10.1207/S15327965PLI1301_01

Forgas, J. P. (Ed.) (2006). *Affect in social thinking and behavior*. New York: Psychology Press. https://doi.org/10.4324/9780203720752

Friedman, H. S., Kern, M. L., & Reynolds, C. A. (2010). Personality and health, subjective well-being, and longevity. *Journal of personality*, *78*(1), 179-216. http://dx.doi.org/10.1111/j.1467-6494.2009.00613.x

Friedman, J. H., Hastie, T. J., Tibshirani, R. J., Narasimhan, B., Tay, K., Simon, N., & Qian, J. (2010). *glmnet: Lasso and elastic-net regularized generalized linear models, R package version 1.9-5*. Retrieved from http://CRAN.R-project.org/package=glmnet

Fusar-Poli, P., Rocchetti, M., Sardella, A., Avila, A., Brandizzi, M., Caverzasi, E., ... & McGuire, P. (2015). Disorder, not just state of risk: meta-analysis of functioning and quality of life in people at high risk of psychosis. *The British Journal of Psychiatry*, *207*(3), 198-206. http://dx.doi.org/10.1192/bjp.bp.114.157115

Gable, P. A., Paul, K., Pourtois, G., & Burgdorf, J. (2021). Utilizing electroencephalography (EEG) to investigate positive affect. *Current Opinion in Behavioral Sciences*, *39*, 190-195. http://dx.doi.org/10.1016/j.cobeha.2021.03.018

Gay, V., Leijdekkers, P., & Barin, E. (2011). A mobile health monitoring-and-treatment system based on integration of the SSN sensor ontology and the HL7 FHIR standard. *BMC medical informatics and decision making, 11*(1), 1-15. https://doi.org/10.1186/1472-6947-11-58

Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C., & Back, M. D. (2017). Trait personality and state variability: Predicting individual differences in within-and cross-context fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research in Personality*, *69*, 124-138. http://dx.doi.org/10.1016/j.jrp.2016.06.003

Giuntella, O., Hyde, K., Saccardo, S., & Sadoff, S. (2021). Lifestyle and mental health disruptions during COVID-19. *Proceedings of the National Academy of Sciences*, *118*(9), e2016632118. http://dx.doi.org/10.1073/pnas.2016632118

Goetz, T., Frenzel, A. C., Stoeger, H., & Hall, N. C. (2010). Antecedents of everyday positive emotions: An experience sampling analysis. *Motivation and Emotion, 34*, 49-62. https://doi.org/10.1007/s11031-009-9152-2

Golder, S. A., & Macy, M. W. (2011). Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science*, *333*(6051), 1878-1881. http://dx.doi.org/10.1126/science.1202775

Gray, E. K., & Watson, D. (2007). Assessing positive and negative affect via self-report. In J. A. Coan & J. J. B. Allen (Eds.), *Handbook of emotion elicitation and assessment* (pp. 171-183). Oxford University Press.

Harari, G. M., Müller, S., & Gosling, S. (2020). Naturalistic assessment of situations using mobile sensing methods. In J. F. Rauthmann, R. A. Sherman, & D. C. Funder (Eds.), *The Oxford Handbook of Psychological Situations* (pp. 299-312). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.14

Hasler, B. P., Mehl, M. R., Bootzin, R. R., & Vazire, S. (2008). Preliminary evidence of diurnal rhythms in everyday behaviors associated with positive affect. *Journal of Research in Personality*, *42*(6), 1537-1546. http://dx.doi.org/10.1016/j.jrp.2008.07.012

Haynes SN, Mumma GH, & Pinson C (2009). Idiographic assessment: Conceptual and psychometric foundations of individualized behavioral assessment. *Clinical Psychology Review, 29*, 179-191. https://doi.org/10.1016/j.cpr.2008.12.003

Hennekens, C. H., Hennekens, A. R., Hollar, D., & Casey, D. E. (2005). Schizophrenia and increased risks of cardiovascular disease. *American heart journal*, *150*(6), 1115-1121. http://dx.doi.org/10.1016/j.ahj.2005.02.007

Hothorn, T., Leisch, F., Zeileis, A., & Hornik, K. (2005). The Design and Analysis of Benchmark Experiments. *Journal of Computational and Graphical Statistics, 14*(3), 675-699. http://dx.doi.org/10.1198/106186005X59630

Howe, E., Bosley, H. G., & Fisher, A. J. (2020). Idiographic network analysis of discrete mood states prior to treatment. *Counselling and Psychotherapy Research, 20*(3), 470-478. https://doi.org/10.1002/capr.12295

Howell, R. T., Kern, M. L., & Lyubomirsky, S. (2007). Health benefits: Meta-analytically determining the impact of well-being on objective health outcomes. *Health Psychology Review*, *1*(1), 83-136. http://dx.doi.org/10.1080/17437190701492486

Hufford, M. R. (2007). Special methodological challenges and opportunities in ecological momentary assessment. In A. Stone, S. Shiffman, A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (54-75). Oxford University Press.

Humrichouse, J., Chmielewski, M., McDade-Montez, E. A., & Watson, D. (2007). Affect Assessment Through Self-Report Methods. In J. Rottenberg & S. L. Johnson (Eds.), *Emotion and psychopathology: Bridging affective and clinical science* (pp. 13-34). American Psychological Association. https://doi.org/10.1037/11562-001

Insel, T. R. (2017). Digital phenotyping: Technology for a new science of behavior. *JAMA, 318(*13), 1215-1216. https://doi.org/10.1001/jama.2017.11295

Jacobson, N. C., & Chung, Y. J. (2020). Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors, 20*(12), 3572. https://doi.org/10.3390/s20123572

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modelling. *Psychological Methods, 13*(4), 354-375. http://dx.doi.org/10.1037/a0014173

Jakob, R., Harperink, S., Rudolf, A. M., Fleisch, E., Haug, S., Mair, J. L., ... & Kowatsch, T. (2022). Factors Influencing Adherence to mHealth Apps for Prevention or Management of Noncommunicable Diseases: Systematic Review. *Journal of Medical Internet Research*, *24*(5), e35371. https://dx.doi.org/10.2196/35371

Jansen, K., Mondin, T. C., de Azevedo Cardoso, T., da Costa Ores, L., de Mattos Souza, L. D., Pinheiro, R. T., ... & da Silva, R. A. (2013). Quality of life and mood disorder episodes: community sample. *Journal of affective disorders*, *147*(1-3), 123-127. http://dx.doi.org/10.1016/j.jad.2012.10.021

Johnson, E. I., Husky, M., Grondin, O., Mazure, C. M., Doron, J., & Swendsen, J. (2008). Mood trajectories following daily life events. *Motivation and Emotion*, *32*(4), 251-259. http://dx.doi.org/10.1007/s11031-008-9106-0

Kashdan, T. B., & Roberts, J. E. (2004). Trait and state curiosity in the genesis of intimacy: Differentiation from related constructs. *Journal of Social and Clinical Psychology*, *23*(6), 792-816. http://dx.doi.org/10.1521/jscp.23.6.792.54800

Kashdan, T. B., & Steger, M. F. (2006). Expanding the topography of social anxiety: An experience-sampling assessment of positive emotions, positive events, and emotion suppression. *Psychological Science*, *17*(2), 120-128. http://dx.doi.org/10.1111/j.1467-9280.2006.01674.x

Keller, M. C., Fredrickson, B. L., Ybarra, O., Côté, S., Johnson, K., Mikels, J., ... & Wager, T. (2005). A warm heart and a clear head: The contingent effects of weather on mood and cognition. *Psychological science, 16*(9), 724-731. http://dx.doi.org/10.1111/j.1467-9280.2005.01602.x

Komulainen, E., Meskanen, K., Lipsanen, J., Lahti, J. M., Jylhä, P., Melartin, T., ... & Ekelund, J. (2014). The effect of personality on daily life emotional processes. *PLoS One, 9*(10), e110907. https://doi.org/10.1371/journal.pone.0110907

Koopmans, T. A., Geleijnse, J. M., Zitman, F. G., & Giltay, E. J. (2010). Effects of happiness on all-cause mortality during 15 years of follow-up: The Arnhem Elderly Study. *Journal of Happiness Studies, 11*(1), 113-124. http://dx.doi.org/10.1007/s10902-008-9127-0

Kööts, L., Realo, A., & Allik, J. (2011). The influence of the weather on affective experience: An experience sampling study. *Journal of individual differences*, *32*(2), 74. http://dx.doi.org/10.1027/1614-0001/a000037

Kosch, T., Hassib, M., Reutter, R., & Alt, F. (2020, September). Emotions on the go: Mobile emotion assessment in real-time using facial expressions. In *Proceedings of the International Conference on Advanced Visual Interfaces* (pp. 1-9). https://doi.org/10.1145/3399715.3399928

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer* *engineering, 160*, 3-24. Retrieved from http://ahdehao.com/uploads/tmp/071c35cf43.pdf

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (pp. 27-58). Springer. https://doi.org/10.1007/978-1-4614-6849-3

Kuper, N., Modersitzki, N., Phan, L. V., & Rauthmann, J. (2021). The situation during the COVID-19 pandemic: A snapshot in Germany. *PLOS ONE, 16*(2), e0245719. https://doi.org/10.1371/journal.pone.0245719

Kuppens, P., Dejonckheere, E., Kalokerinos, E. K., & Koval, P. (2022). Some recommendations on the use of daily life methods in affective science. *Affective Science, 3*, 505-515. https://doi.org/10.1007/s42761-022-00101-0

Lanata, A., Greco, A., Valenza, G., & Scilingo, E. P. (2014, May). A pattern recognition approach based on electrodermal response for pathological mood identification in bipolar disorders. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3601-3605). IEEE.

Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., Choudhury, T., & Campbell, A. T. (2010). A survey of mobile phone sensing. *IEEE Communications Magazine, 48*(9), 140-150. https://doi.org/10.1109/MCOM.2010.5560598

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Emotion* (pp. 25-59). Sage Publications, Inc.

Lee, H., Cho, A., Jo, Y., & Whang, M. (2017). The Relationships Between Behavioral Patterns and Emotions in Daily Life. In *Advances in Computer Science and Ubiquitous Computing* (pp. 1332-1339). Springer, Singapore.

LiKamWa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013, June). Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services* (pp. 389-402). http://dx.doi.org/10.1145/2462456.2483967

Lim, J. Z., Mountstephens, J., & Teo, J. (2020). Emotion recognition using eye-tracking: taxonomy, review, and current challenges. *Sensors, 20*(8), 2384. https://doi.org/10.3390/s20082384

Lonigan, C. J., Hooe, E. S., David, C. F., & Kistner, J. A. (1999). Positive and negative affectivity in children: confirmatory factor analysis of a two-factor model and its relation to symptoms of anxiety and depression. *Journal of consulting and clinical psychology*, *67*(3), 374. http://dx.doi.org/10.1037/0022-006X.67.3.374

Lu, Y., Zheng, W. L., Li, B., & Lu, B. L. (2015). Combining eye movements and EEG to enhance emotion recognition. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence* (Vol. 15, pp. 1170-1176)

Lucas, R. E., & Fujita, F. (2000). Factors influencing the relation between extraversion and pleasant affect. *Journal of personality and social psychology, 79*(6), 1039-1056. https://doi.org/10.1037/0022-3514.79.6.1039

Ma, Y., Xu, B., Bai, Y., Sun, G., & Zhu, R. (2012, May). Daily mood assessment based on mobile phone sensing. In *2012 ninth international conference on wearable and implantable body sensor networks* (pp. 142-147). IEEE. http://dx.doi.org/10.1109/BSN.2012.3

MacLeod, L., Suruliraj, B., Gall, D., Bessenyei, K., Hamm, S., Romkey, I., ... & Meier, S. (2021). A mobile sensing app to monitor youth mental health: observational pilot study. *JMIR mHealth and uHealth, 9*(10), e20638. http://dx.doi.org/10.2196/20638

Madan, A., Cebrian, M., Lazer, D., & Pentland, A. (2010, September). Social sensing for epidemiological behavior change. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (pp. 291-300). http://dx.doi.org/10.1145/1864349.1864394

Marzano, L., Bardill, A., Fields, B., Herd, K., Veale, D., Grey, N., & Moran, P. (2015). The application of mHealth to mental health: opportunities and challenges. *The Lancet Psychiatry, 2*(10), 942-948. http://dx.doi.org/10.1016/S2215-0366(15)00268-0

Mehrotra, A., Müller, S. R., Harari, G. M., Gosling, S. D., Mascolo, C., Musolesi, M., & Rentfrow, P. J. (2017). Understanding the role of places and activities on mobile phone interaction and usage patterns. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(3), 1-22. http://dx.doi.org/10.1145/3131901

Mehrotra, S., & Tripathi, R. (2018). Recent developments in the use of smartphone interventions for mental health. *Current opinion in psychiatry*, *31*(5), 379-388. http://dx.doi.org/10.1097/YCO.0000000000000439

Meier, A., & Johnson, B. K. (2022). Social comparison and envy on social media: A critical review. *Current Opinion in Psychology, 45*, 101302-101302. https://doi.org/10.1016/j.copsyc.2022.101302

Merz, E. L., & Roesch, S. C. (2011). Modeling trait and state variation using multilevel factor analysis with PANAS daily diary data. *Journal of research in personality*, *45*(1), 2-9. http://dx.doi.org/10.1016/j.jrp.2010.11.003

Messiah, A., Grondin, O., & Encrenaz, G. (2011). Factors associated with missing data in an experience sampling investigation of substance use determinants. *Drug and Alcohol Dependence, 114*(2), 153-158. https://doi.org/10.1016/j.drugalcdep.2010.09.016

Messner, E. M., Sariyska, R., Mayer, B., Montag, C., Kannen, C., Schwerdtfeger, A., & Baumeister, H. (2019). Insights-future implications of passive smartphone sensing in the therapeutic context. *Verhaltenstherapie*, 1-10. http://dx.doi.org/10.1159/000501951

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science, 7*(3), 221-237. https://doi.org/10.1177/1745691612441215

Miranda, D., Dave, R., & Claes, M. (2009). Music listening, coping, peer affiliation and depression in adolescence. *Psychology of Music, 37*(2), 215-233. https://doi.org/10.1177/0305735607086041

Mischel, W. (2004). Toward an integrative science. *Annual review of psychology*, *55*, 1-22. https://doi.org/10.1146/annurev.psych.55.042902.130709

Mohr, D. C., Tomasino, K. N., Lattie, E. G., Palac, H. L., Kwasny, M. J., Weingardt, K., ... & Schueller, S. M. (2017). IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. *Journal of medical Internet research*, *19*(1), e6645. http://dx.doi.org/10.2196/jmir.6645

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. Measurement, 2(4), 201-218. https://doi.org/10.1207/s15366359mea0204_1

Molnar, C. (2022). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Retrieved from https://christophm.github.io/interpretable-ml-book/

Morris, W.N. (1989). The Concept of Mood. In: *Mood. Springer Series in Social Psychology*. Springer, New York, NY. https://doi.org/10.1007/978-1-4612-3648-1_1

Morriss, R., Vinjamuri, I., Faizal, M. A., Bolton, C. A., & McCarthy, J. P. (2013). Training to recognise the early signs of recurrence in schizophrenia. *Schizophrenia bulletin, 39*(2), 255-256. https://doi.org/10.1093/schbul/sbs192

Müller, S. R., Peters, H., Matz, S. C., Wang, W., & Harari, G. M. (2020). Investigating the relationships between mobility behaviours and indicators of subjective well-being using smartphone-based experience sampling and GPS tracking. *European Journal of Personality*, *34*(5), 714-732. http://dx.doi.org/10.1002/per.2262

Murnane, E. L., Cosley, D., Chang, P., Guha, S., Frank, E., Gay, G., & Matthews, M. (2016). Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. *Journal of the*

American Medical Informatics Association, 23(3), 477-484. https://doi.org/10.1093/jamia/ocv165

Naslund, J. A., Marsch, L. A., McHugo, G. J., & Bartels, S. J. (2015). Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. *Journal of mental health*, *24*(5), 321-332. http://dx.doi.org/10.3109/09638237.2015.1019054

Nauta, M. H., aan het Rot, M., Schut, H., & Stroebe, M. (2019). Homesickness in social context: An ecological momentary assessment study among 1st-year university students. International *Journal of Psychology, 55*(3), 392. https://doi.org/10.1002/ijop.12586

Nelson, D. W., & Sim, E. K. (2014). Positive affect facilitates social problem solving. *Journal of Applied Social Psychology*, *44*(10), 635-642. http://dx.doi.org/10.1111/jasp.12254

Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2011). Affect Analysis Model: Novel rule-based approach to affect sensing from text. *Natural Language Engineering, 17*(1), 95-135. https://doi.org/10.1017/S1351324910000239

Nguyen, T., Phung, D., Adams, B., & Venkatesh, S. (2014). Mood sensing from social media texts and its applications. *Knowledge & Information Systems, 39*(3), 667-702 (2014). https://doi.org/10.1007/s10115-013-0628-8

Niforatos, E., & Karapanos, E. (2015). EmoSnaps: a mobile application for emotion recall from facial expressions. *Personal and Ubiquitous Computing, 19*(2), 425-444. https://doi.org/ 10.1007/s00779-014-0777-0

O'Hanlon, P., Aref-Adib, G., Fonseca, A., Lloyd-Evans, B., Osborn, D., & Johnson, S. (2016). Tomorrow's world: current developments in the therapeutic use of technology for psychosis. *BJPsych advances*, *22*(5), 301-310. http://dx.doi.org/10.1192/apt.bp.115.014654

Ong, A. D., Bergeman, C. S., Bisconti, T. L., & Wallace, K. A. (2006). Psychological resilience, positive emotions, and successful adaptation to stress in later life. *Journal of personality and social psychology*, *91*(4), 730. http://dx.doi.org/10.1037/0022-3514.91.4.730

Onnela, J. P., & Rauch, S. L. (2016). Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology, 41*(7), 1691-1696. http://dx.doi.org/10.1038/npp.2016.7

Paraschakis, D. (2017, May). Towards an ethical recommendation framework. In *2017 11th international conference on research challenges in information science (RCIS)* (pp. 211-220). IEEE. http://dx.doi.org/10.1109/RCIS.2017.7956539

Park, K., Lee, S., Kim, E., Park, M., Park, J., & Cha, M. (2013). Mood and weather: Feeling the heat?. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 7, No. 1, pp. 709-712). http://dx.doi.org/10.1609/icwsm.v7i1.14451

Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International journal of human-computer studies, 59*(1-2), 185-198. https://doi.org/10.1016/S1071-5819(03)00017-X

Pavot, W., Diener, E. D., & Fujita, F. (1990). Extraversion and happiness. *Personality and individual differences*, *11*(12), 1299-1306. http://dx.doi.org/10.1016/0191-8869(90)90157-M

Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on affective computing*, *1*(2), 81-97. http://dx.doi.org/10.1109/T-AFFC.2010.7

Ponocny, I., Weismayer, C., Stross, B., & Dressler, S. G. (2016). Are most people happy? Exploring the meaning of subjective well-being ratings. *Journal of Happiness Studies, 17*(6), 2635-2653. http://dx.doi.org/10.1007/s10902-015-9710-0

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Ram, N., Conroy, D. E., Pincus, A. L., Lorek, A., Rebar, A., Roche, M. J., … & Gerstorf, D. (2014). Examining the interplay of processes across multiple time-scales: Illustration with the Intraindividual Study of Afect, Health, and Interpersonal Behavior (iSAHIB). *Research in Human Development, 11*(2), 142- 160. https://doi.org/10.1080/15427609.2014.906739

Ren, B., Balkind, E. G., Pastro, B., Israel, E., Pizzagalli, D. A., Rahimi-Eichi, H., ... & Webb, C. A. (2022). Predicting states of elevated negative affect in adolescents from smartphone sensors: A novel personalized machine learning approach. PsyArXiv. https://doi.org/10.31234/osf.io/zjm4d

Renner, K. H., Klee, S., & von Oertzen, T. (2020). Bringing back the person into behavioural personality science using big data. *European Journal of Personality*, *34*(5), 670-686. http://dx.doi.org/10.1002/per.2303

Rintala, A., Wampers, M., Myin-Germeys, I., & Viechtbauer, W. (2020). Momentary predictors of compliance in studies using the experience sampling method. *Psychiatry research, 286*, 112896. http://dx.doi.org/10.1016/j.psychres.2020.112896

Rodriguez, M., Aalbers, G., & McNally, R. J. (2022). Idiographic network models of social media use and depression symptoms. *Cognitive Therapy and Research, 46*, 124-132. https://doi.org/10.31234/osf.io/t4hc3, https://doi.org/10.1007/s10608-021-10236-2

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161. http://dx.doi.org/10.1037/h0077714

Russell, J. A. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, *110*(1), 145-172. http://dx.doi.org/10.1080/02699930902809375

Russell, R., Guerry, A. D., Balvanera, P., Gould, R. K., Basurto, X., Chan, K. M., ... & Tam, J. (2013). Humans and nature: how knowing and experiencing nature affect well-being. *Annual review of environment and resources*, *38*, 473-502. http://dx.doi.org/10.1146/annurev-environ-012312-110838

Russell, J. A., Lewicka, M., & Niit, T. (1989). A cross-cultural study of a circumplex model of affect. *Journal of personality and social psychology*, *57*(5), 848. http://dx.doi.org/10.1037/0022-3514.57.5.848

Sandstrom, G. M., Lathia, N., Mascolo, C., & Rentfrow, P. J. (2017). Putting mood in context: Using smartphones to examine how people feel in different locations. *Journal of Research in Personality*, *69*, 96-101. http://dx.doi.org/10.1016/j.jrp.2016.06.004

Sano, A., Phillips, A. J., Amy, Z. Y., McHill, A. W., Taylor, S., Jaques, N., Czeisler, C. A., Klerman, E. B., & Picard, R. W. (2015). Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. In *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 1-6. http://dx.doi.org/10.1109/BSN.2015.7299420

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., & Barger, L. K. (2018). Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study. *J Med Internet Res, 20*(6), e210. http://dx.doi.org/10.2196/jmir.9410

Santomauro, D. F., Herrera, A. M. M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D. M., ... & Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet, 398*(10312), 1700-1712. http://dx.doi.org/10.1016/S0140-6736(21)02143-7

Schell, A. M., Dawson, M. E., Rissling, A., Ventura, J., Subotnik, K. L., Gitlin, M. J., & Nuechterlein, K. H. (2005). Electrodermal predictors of functional outcome and negative symptoms in schizophrenia. *Psychophysiology*, *42*(4), 483-492. http://dx.doi.org/10.1111/j.1469-8986.2005.00300.x

Schimmack, U., Oishi, S., Diener, E., & Suh, E. (2000). Facets of affective experiences: A framework for investigations of trait affect. *Personality and Social Psychology Bulletin*, *26*(6), 655-668. https://doi.org/10.1177/0146167200268

Schmidt, P., Reiss, A., Dürichen, R., & Van Laerhoven, K. (2019). Wearable-based affect recognition—A review. *Sensors*, *19*(19), 4079. http://dx.doi.org/10.3390/s19194079

Schoedel, R., Kunz, F., Bergmann, M., Bemmann, F., Bühner, M., & Sust, L. (in press). Snapshots of Daily Life: Situations Investigated Through the Lens of Smartphone Sensing. *Journal of Personality and Social Psychology.*

Schoedel, R., & Oldemeier, M. (2020). Basic protocol: Smartphone sensing panel study. Leibniz Institut für Psychologische Information und Dokumentation (ZPID). https://doi.org//10.23668/PSYCHARCHIVES.2901

Schoedel, R., Oldemeier, M., Bonauer, L., & Sust, L. (2022). Dataset for: Systematic categorisation of 3091 smartphone applications from a large-scale smartphone sensing dataset[Data set]. PsychArchives. https://doi.org/10.23668/psycharchives.5362

Schultchen, D., Reichenberger, J., Mittl, T., Weh, T. R., Smyth, J. M., Blechert, J., & Pollatos, O. (2019). Bidirectional relationship of stress and affect with physical activity and healthy eating. *British Journal of Health Psychology*, *24*(2), 315. http://dx.doi.org/10.1111/bjhp.12355

Scollon, C. N., Kim-Prieto, C., & Diener, E. (2003). Experience Sampling: Promises and Pitfalls, Strengths and Weaknesses. *Journal of Happiness Studies*, *4*(1), 5-34. http://dx.doi.org/10.1023/A:1023605205115

Serafini, K., Malin-Mayor, B., Nich, C., Hunkele, K., & Carroll, K. M. (2016). Psychometric properties of the Positive and Negative Affect Schedule (PANAS) in a heterogeneous sample of substance users. *The American journal of drug and alcohol abuse*, *42*(2), 203-212. http://dx.doi.org/10.3109/00952990.2015.1133632

Servia-Rodríguez, S., Rachuri, K. K., Mascolo, C., Rentfrow, P. J., Lathia, N., & Sandstrom, G. M. (2017, April). Mobile sensing at the service of mental well-being: a large-scale longitudinal study. In *Proceedings of the 26th international conference on world wide web*, 103-112. http://dx.doi.org/10.1145/3038912.3052618

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4,* 1-32. http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091415

Sikka, P., Feilhauer, D., Valli, K., & Revonsuo, A. (2017). How you measure is what you get: Differences in self-and external ratings of emotional experiences in home dreams. *The American journal of psychology, 130*(3), 367-384. http://dx.doi.org/10.5406/amerjpsyc.130.3.0367

Silvia, P. J., Kwapil, T. R., Walsh, M. A., & Myin-Germeys, I. (2014). Planned Missing Data Designs in Experience Sampling Research: Monte Carlo Simulations of Efficient Designs for Assessing Within-Person Constructs. *Behavior research methods, 46(*1), 41-54. https://doi.org/10.3758/s13428-013-0353-y

Soyster, P. D., Ashlock, L., & Fisher, A. J. (2022). Pooled and person-specific machine learning models for predicting future alcohol consumption, craving, and wanting to drink: A demonstration of parallel utility. *Psychology of Addictive Behaviors, 36*(3), 296-306. https://doi.org/10.1037/adb0000666

Spathis, D., Servia-Rodríguez, S., Farrahi, K., Mascolo, C., & Rentfrow, J. (2019). Passive mobile sensing and psychological traits for large scale mood prediction. *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*, 272-281. https://doi.org/10.1145/3329189.3329213

Stawarz, K., Preist, C., & Coyle, D. (2019). Use of smartphone apps, social media, and web-based resources to support mental health and well-being: online survey. *JMIR mental health*, *6*(7), e12546. http://dx.doi.org/10.2196/12546

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological bulletin, 134*(1), 138-161. https://doi.org/10.1037/0033-2909.134.1.138

Steptoe, A., Dockray, S., & Wardle, J. (2009). Positive affect and psychobiological processes relevant to health. *Journal of personality*, *77*(6), 1747-1776. http://dx.doi.org/10.1111/j.1467-6494.2009.00599.x

Steptoe, A., & Wardle, J. (2005). Positive affect and biological function in everyday life. *Neurobiology of Aging*, *26*(1), 108-112. http://dx.doi.org/10.1016/j.neurobiolaging.2005.08.016

Stikic, M., Johnson, R. R., Tan, V., & Berka, C. (2014). EEG-based classification of positive and negative affective states. *Brain-Computer Interfaces*, *1*(2), 99-112. http://dx.doi.org/10.1080/2326263X.2014.912883

Sun, J., Rhemtulla, M., & Vazire, S. (2021). Eavesdropping on missing data: What are university students doing when they miss experience sampling reports?. *Personality and*

Social Psychology Bulletin, 47(11), 1535-1549. http://dx.doi.org/10.1177/0146167220964639

Tellegen, A., Watson, D., & Clark, L. A. (1999). On the dimensional and hierarchical structure of affect. *Psychological science*, *10*(4), 297-303. http://dx.doi.org/10.1111/1467-9280.00157

Terraciano, A., McCrae, R. R., & Costa Jr, P. T. (2003). Factorial and construct validity of the Italian Positive and Negative Affect Schedule (PANAS). *European journal of psychological assessment*, *19*(2), 131. http://dx.doi.org/10.1027//1015-5759.19.2.131

Thayer, R. E. (1986). Activation-deactivation adjective check list: Current overview and structural analysis. *Psychological reports*, *58*(2), 607-614. http://dx.doi.org/10.2466/pr0.1986.58.2.607

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, 58*(1), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Till, J., Tan, L., & Louis, W. R. (2016). Music and well-being. In R. J. Scott & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource* (pp. 1-14). Wiley. https://doi.org/10.1002/9781118900772.etrds0237

Torous, J., Nicholas, J., Larsen, M. E., Firth, J., & Christensen, H. (2018). Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evidence-Based Mental Health, 21*(3), 116-119. https://doi.org/10.1136/eb-2018-102891

Ushey K. (2020). Renv: Project Environments (Version 0.12.0). Retrieved from https://CRAN.R-project.org/package=renv.

Vaidya, J. G., Gray, E. K., Haig, J., & Watson, D. (2002). On the temporal stability of personality: evidence for differential stability and the role of life experiences. *Journal of personality and Social Psychology*, *83*(6), 1469. http://dx.doi.org/10.1037/0022-3514.83.6.1469

Valentine, L., D'Alfonso, S., & Lederman, R. (2022). Recommender systems for mental health apps: advantages and ethical challenges. *AI & society*, 1-12. http://dx.doi.org/10.1007/s00146-021-01322-w

Van Ameringen, M., Turna, J., Khalesi, Z., Pullia, K., & Patterson, B. (2017). There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. *Depression and anxiety*, *34*(6), 526-539. http://dx.doi.org/10.1002/da.22657

Veenhoven, R. (2008). Healthy happiness: Effects of happiness on physical health and the consequences for preventive health care. *Journal of happiness studies*, *9*(3), 449-469. http://dx.doi.org/10.1007/s10902-006-9042-1

Ventura-Bort, C., Wendt, J., & Weymar, M. (2021). The role of interoceptive sensibility and emotional conceptualization for the experience of emotions. *Frontiers in Psychology, 12*, 712418. https://doi.org/10.3389/fpsyg.2021.712418

Verduyn, P., Delaveau, P., Rotgé, J. Y., Fossati, P., & Van Mechelen, I. (2015). Determinants of emotion duration and underlying psychological and neural mechanisms. *Emotion Review*, *7*(4), 330-335. http://dx.doi.org/10.1177/1754073915590618

Verduyn, P., Delvaux, E., Van Coillie, H., Tuerlinckx, F., & Van Mechelen, I. (2009). Predicting the duration of emotional experience: two experience sampling studies. *Emotion*, *9*(1), 83. http://dx.doi.org/10.1037/a0014610

Verduyn, P., & Lavrijsen, S. (2015). Which emotions last longest and why: The role of event importance and rumination. *Motivation and Emotion*, *39*(1), 119-127. http://dx.doi.org/10.1007/s11031-014-9445-y

Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review*, *11*(1), 274-302. http://dx.doi.org/10.1111/sipr.12033

von Stumm, S. (2016). Is day-to-day variability in cognitive function coupled with day-to-day variability in affect?. *Intelligence, 55,* 1-6. http://dx.doi.org/10.1016/j.intell.2015.12.006

Wampfler, P., Rohner, P., Grabherr, L., & Mast, F. W. (2020). The effect of touchscreen latency on gaze-supported pointing and affective responses. *Journal of Experimental Psychology: Applied, 26*(1), 107-117. https://doi.org/10.1037/xap0000231

Wang R., Aung, M.S., Abdullah, S., Brian, R., Campbell, A.T., Choudhury, T., … , Tseng, V.W. (2016). CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 886-897. http://dx.doi.org/10.1145/2971648.2971740

Wang, R., Chen, F., Chen, Z., Li, T., Harari G., Tignor, S., Zhou, X., Ben-Zeev, D., & Campbell, A.T. (2014). Studentlife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 3-14. http://dx.doi.org/10.1145/2632048.2632054

Wang, Z., Ho, S. B., & Cambria, E. (2020). A review of emotion sensing: categorization models and algorithms. *Multimedia Tools and Applications, 79*(47), 35553-35582. https://doi.org/10.1007/s11042-019-08328-z

Watson, D. (2020). *Mood and temperament*. New York: Guilford Press.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule - Expanded Form*. https://doi.org/10.17077/48vt-m4t2

Watson, D., & Clark, L. A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of personality assessment, 68*(2), 267-296. https://doi.org/10.1207/s15327752jpa6802_4

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063-70. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological bulletin*, *98*(2), 219. http://dx.doi.org/10.1037/0033-2909.98.2.219

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of personality and social psychology*, *76*(5), 820. http://dx.doi.org/10.1037/0022-3514.76.5.820

Wedderhoff, N., Gnambs, T., Wedderhoff, O., Burgard, T., & Bošnjak, M. (2021). On the structure of affect: A meta-analytic investigation of the dimensionality and the cross-national applicability of the Positive and Negative Affect Schedule (PANAS). *Zeitschrift für Psychologie*, *229*(1), 24. http://dx.doi.org/10.1027/2151-2604/a000434

Weiner, M., Warren, L., & Fiedorowicz, J. G. (2011). Cardiovascular morbidity and mortality in bipolar disorder. *Annals of clinical psychiatry: official journal of the American Academy of Clinical Psychiatrists*, *23*(1), 40.

Weismayer, C. (2021). Investigating the affective part of subjective well-being (SWB) by means of sentiment analysis. *International Journal of Social Research Methodology, 24*(6), 697-712. http://dx.doi.org/10.1080/13645579.2020.1816251

Wichers, M., Smit, A. C., & Snippe, E. (2020). Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: a confirmatory single-subject time-series study. *Journal for Person-Oriented Research*, *6*(1), 1. http://dx.doi.org/10.17505/jpor.2020.22042

Wigman, J. T., van Os, J., Thiery, E., Derom, C., Collip, D., Jacobs, N., & Wichers, M. (2013). Psychiatric diagnosis revisited: towards a system of staging and profiling combining nomothetic and idiographic parameters of momentary mental states. *PLoS One*, *8*(3), e59559. http://dx.doi.org/10.1371/journal.pone.0059559

Wilt, J., Noftle, E. E., Fleeson, W., & Spain, J. S. (2012). The dynamic role of personality states in mediating the relationship between extraversion and positive affect. *Journal of Personality, 80*(5), 1205-1236. https://doi.org/10.1111/j.1467-6494.2011.00756.x

Wilt, J., & Revelle, W. (2019). The Big Five, everyday contexts and activities, and affective experience. *Personality and individual differences, 136*, 140-147. http://dx.doi.org/10.1016/j.paid.2017.12.032

Wolf, M., Joye, Y., Gatti, E., & Pollet, T. (2008). Automatic emotion detection from text using linguistic inquiry and word count. *Proceedings of the 2008 ACM symposium on Applied computing* (pp. 1556-1560). https://doi.org/10.1145/1363686.1364102

World Health Organization (2017). *Depression and other common mental disorders: global health estimates* [Technical report].

Wright, N. M, & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software*, *77*(1), 1-17. https://doi.org/10.18637/jss.v077.i0

Yik, M. S., & Russell, J. A. (2001). Predicting the Big Two of affect from the Big Five of personality. *Journal of research in personality*, *35*(3), 247-277. https://doi.org/10.1006/jrpe.2001.2322

Zhang, X., Li, W., Chen, X., & Lu, S. (2018). Moodexplorer: Towards compound emotion detection via smartphone sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*(4), 1-30. https://doi.org/10.1145/3161414

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, *67*(2), 301-320. http://dx.doi.org/10.1111/j.1467-9868.2005.00527.x

3.9. Appendix

Table A3.1: Overview of Categorizations Applied for Feature Extraction

Category	Sub-category	Description
Bluetooth devices		
	Watch	The connected device is a wearable watch.
	Headset	The connected device is a headset or headphone.
	Phone	The connected device is another smartphone or cordless phone.
	Computer	The connected device is a laptop or desktop computer.
	Health	The connected device is a health-related wearable, such as weighing or pulse rate measure device.
	Car	The connected device is a in-car entertainment system.
	HiFi	The connected device is a HiFi system or loudspeaker.
	Uncategorized	Other connected devices.
	n.a.	The connected device does not reveal information about its type.
Apps		
	Social network, gaming, etc.	Adapted from Schoedel et al. (2022)
Spotify song attributes		
	Danceability	Danceability describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.
	Energy	Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy. For example, death metal has high energy, while a Bach prelude scores low on the scale. Perceptual features contributing to this attribute include dynamic range, perceived loudness, timbre, onset rate, and general entropy.
	Loudness	Relative loudness of a track compared to other Spotify tracks. The overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks.
	Mode	Mode indicates the modality (major (1) or minor (0)) of a track, the type of scale from which its melodic content is derived.
	Speechiness	Speechiness detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g., talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably variatione entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.

	Liveliness	Liveliness detects the presence of an audience in the recording. Higher liveliness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
	Valence	A measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g., happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g., sad, depressed, angry).
	Tempo	The overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.
	Instrumentality	Measurement of the likelihood the track is instrumental vs. Predicts whether a track contains no vocals. "Ooh" and "aah" sounds are treated as instrumental in this context. Rap or spoken word tracks are clearly "vocal". The closer the instrumentality value is to 1.0, the greater likelihood the track contains no vocal content. Values above 0.5 are intended to represent instrumental tracks, but confidence is higher as the value approaches 1.0.
Mobility activity		
	Still	The device is still (not moving).
	In a vehicle	The device is in a vehicle, such as a car.
	In a road vehicle	The device is in a vehicle on the road.
	In a four-wheeler vehicle	The device is in a vehicle with four wheels (e.g., car).
	In a two-wheeler vehicle	The device is in a vehicle with two wheels (e.g., motorcycle).
	In a rail vehicle	The device is in a vehicle on rails.
	On a bicycle	The device is on a bicycle.
	On foot	The device is on a user who is walking or running.
	Walking	The device is on a user who is walking.
	Running	The device is on a user who is running.
	Unknown	Unable to detect the current activity.

Note. The description of the Spotify song attributes were taken from https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-audio-features.

Table A3.2: Overview of Features Extracted per Sensing Modality and Feature Class

Sensing Modality	Feature Class	Feature(s)
Communication	Calls***	• total number of calls; total duration of calls; min duration of calls; max duration of calls; mean duration of calls; variation of duration of calls • total number of outgoing calls; total duration of outgoing calls; min duration of outgoing calls; max duration of outgoing calls; mean duration of outgoing calls; variation of duration of outgoing calls • total number of incoming calls; total duration of incoming calls; min duration of ringing of incoming calls; max duration of ringing of incoming calls; mean duration of ringing of incoming calls; variation of ringing of incoming calls; min duration of incoming calls; max duration of incoming calls; mean duration of incoming calls; variation of duration of incoming calls • total number of missed calls; total duration of ringing of missed calls; min duration of ringing of missed calls; max duration of ringing of missed calls; mean duration of ringing of missed calls; variation of duration of ringing of missed calls • total number of rejected calls; total duration of ringing of missed calls; min duration of ringing of rejected calls; max duration of ringing of rejected calls; mean duration of ringing of rejected calls; variation of duration of ringing of rejected calls
	Text messages	• total number of texts; min length of texts; max length of texts; mean length of texts; variation of length of texts • total number of outgoing texts; min length of outgoing texts; max length of outgoing texts; mean length of outgoing texts; variation of length of outgoing texts • total number of incoming texts; min length of incoming texts; max length of incoming texts; mean length of incoming texts; variation of length of incoming texts
	Keyboard logging[a)]	• min amount of words per message; max amount of words per message; mean amount of words per message; variation of amount of words per message • min average sentiment of words; max average sentiment of words; mean of average sentiment of words; variation of average sentiment of words • min score of "LIWC dimension"*; max score of "LIWC dimension"*; mean score of "LIWC dimension"*; variation of score of "LIWC dimension"*

Location	Places	• total time spent at home; min distance from home; max distance from home; mean distance from home; variation of distance from home • total time spent at work; min distance from work; mean distance from work; variation of distance from work
	Altitude	• min altitude; max altitude; mean altitude; variation of altitude; altitude change; altitude positive change; altitude negative change
	Geohash	• total number of different Geohashs visited, min time spent per Geohash visited, max time spent per Geohash visited
	Displacement	• radius of gyration; location variance; total distance covered; spatial coverage by convex hull
	Speed	• min speed; max speed; mean of speed; variation of speed; speed change; total time spent in transit
	Activity state	• min probability of “activity category”*; max probability of “activity category”*; mean probability of “activity category”*; variation of probability of “activity category”*
Music consumption	Listening behavior	• total duration of sessions listened to songs; min duration of sessions listened to songs; max duration of sessions listened to songs; mean duration of sessions listened to songs; variation of duration of sessions listened to songs • total number of songs skipped
	Songs	• total number of unique songs listened to • total number of unique artists listened to • total duration of songs listened to; min duration of songs listened to; max duration of songs listened to; mean duration of songs listened to; variation of duration of songs listened to • total level of “Spotify audio feature category” of listened songs*; min level of “Spotify audio feature category” of listened songs*; max level of “Spotify audio feature category” of listened songs*; mean level of “Spotify audio feature category” of listened songs*; variation of level of “Spotify audio feature category” of listened songs*
Smartphone usage	Connectivity	• total duration of power plug status *connected* • total duration of flight mode status *on* • total number of Bluetooth status changes • total duration of Bluetooth status *on* and *connecting/-ed* • total duration of Bluetooth status *on* and *disconnected* • total duration of Bluetooth status *off* • total duration connected with "device category"* • total number of WiFi status changes • total duration of WiFi status *on* and *connecting/-ed* • total duration of WiFi status *on* and *disconnected*

		• total duration of WiFi status *off* • total duration of headphone status *plugged*
	Screen	• total number of sessions; total duration of sessions; min duration of sessions; max duration of sessions; mean duration of sessions; variation of duration of sessions • total number of checks; total duration of checks; min duration of checks; max duration of checks; mean duration of checks; variation of duration of checks • ratio between total number of checks and total number of sessions; ratio between total duration of checks and total duration of sessions • first screen time of day; last screen time of day; total duration of screen inactivity at night before; total number of checks during night before; total duration of sessions during night before
	Apps	• total number of all app usages; total duration of all app usages; total number of different apps used; min total number of usages per app; max total number of usages per app; mean total number of usages per app; variation of total number of usages per app; min total usage duration per app; max total usage duration per app; mean total usage duration per app; variation of total usage duration per app • total number of different app categories used; min number of total usages per app category*; max number of total usages per app category*; mean number of total usages per app category*; variation of number of total usages per app category*; min total usage duration per app category*; max total usage duration per app category*; mean total usage duration per app category*; variation of total usage duration per app category*
	Notification	• total latency of notification caused app usage; min latency of notification caused app usage; max latency of notification caused app usage; mean latency of notification caused app usage; variation of latency of notification caused app usage
Time	Weekday	• current timestamp is at the weekend (vs. weekday)
	Daytime	• timepoint at morning; timepoint at noon; timepoint at afternoon; timepoint at evening
Weather	Clouds	• cloud coverage of sky; visibility (distance at which objects are visible)
	Temperature	• dew point temperature • temperature at the location; min temperature; max temperature • feelslike temperature; min feelslike temperature; max feelslike temperature
	Humidity	• relative humidity

Moon	• daily moonphase[a)] (fractional portion through current moon lunation cycle)
Precipitation	• total amount of precipitation • snowdepth
Sun	• solar radiation power • UV index • daily sunrise epoch[a)]; daily sunset epoch[a)]; daily time difference between sunrise & sunset[a)]
Wind	• mean windspeed • wind direction (in metereological degrees)

Note. All features are created for an hourly time window (i.e., 60 minutes before the experience sampling was started by the participant), daily time window (i.e., 24 hours of the respective study day), and two-week ES wave time window (i.e., all study days available for the respective participant). The features labeled with [a)] are only extractable on daily and two-week basis due to their nature. Accordingly, all quantification metrics (i.e.; total; ratio; min; max; mean; variation) of features are based on logging events within these time windows; All categorical features are dummy coded with 1 = *yes*; 0 = *no*; [b)] Keyboard logging features were only extracted for keyboard logs entered in communication apps (inter alia the average sentiment of a text message (SentiWS; Remus et al., 2010) and the Linguistic Inquiry and Word Count (LIWC, Wolf et al., 2008)); * indicates that this feature will be extracted for each category of the related categorization (i.e.; device categories; app categories; LIWC dimensions; Spotify audio feature category; activity categories) as shown in *Table 2*; ** Level of the Spotify audio features are weighted by listening duration of the respective song; *** Some of the features of the categories *calls, text messages, keyboard logs, apps, screen*, and *notification* were built on previous work by colleagues of the Chair for Psychological Methods and Diagnostics at the Department of Psychology at Ludwig-Maximilians-University of Munich. All other categorizations and feature engineering code was developed by the author.

Figure A3.1: Different Data Sets Compared in the Benchmark Experiments of the Study

1 Hourly Dataset

user_id	ES_id	day_id	mean_dur_calls	...	tot_time_home	...	tot_num_song	...	min_dur_screen	...	weekend	...	cloud coverage	...	Valence ES score	Arousal ES score
1	1	1	5	...	...	...	...	...	...	...	...	...	...	...	3	2
1	2	1	0	...	...	...	...	...	...	...	...	...	...	...	1	2
1	3	1	3	...	...	...	...	...	...	...	...	...	...	...	3	3
1	4	2	7	...	...	...	...	...	...	...	...	...	...	...	4	3
1	5	2	0	...	...	...	...	...	...	...	...	...	...	...	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2	30	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...

identifier variable
communication feature
location feature
music feature
smarthpone usage feature
time feature
weather feature
target variable (affect states/ traits)

avg. of day 1 of user 1

2 Daily Dataset

user_id	ES_id	day_id	mean_dur_calls	...	tot_time_home	...	tot_num_song	...	min_dur_screen	...	weekend	...	cloud coverage	...	Valence ES score	Arousal ES score
1	1	1	2,7	...	...	...	...	...	...	...	...	...	...	...	3	2
1	2	1	2,7	...	...	...	...	...	...	...	...	...	...	...	1	2
1	3	1	2,7	...	...	...	...	...	...	...	...	...	...	...	3	3
1	4	2	3,5	...	...	...	...	...	...	...	...	...	...	...	4	3
1	5	2	3,5	...	...	...	...	...	...	...	...	...	...	...	2	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2	30	1	...	...	...	...	...	...	...	...	...	...	...	...	...	...

3 Two-Weeks Dataset

avg. of user 1 across all days 1 to 13

user_id	mean_dur_calls	...	tot_time_home	...	tot_num_song	...	min_dur_screen	...	weekend	...	cloud coverage	...	PANAS Pos. Affect score	PANAS Neg. Affect score
1	3,6	...	...	...	...	...	...	...	...	...	...	...	3	2
2		...	...	...	...	...	...	...	...	...	...	...	1	4
3	...	...	...	...	...	...	...	...	...	...	...	...	2	2
4	...	...	...	...	...	...	...	...	...	...	...	...	3	3
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
450	...	...	...	...	...	...	...	...	...	...	...	...	3	4

Note. This figure exemplarily outlines the different data sets included in the benchmark experiments conducted to predict affect states ((1) hourly data set and (2) daily data set) and affect traits (3) two-week data set).

Table A3.3: Descriptive Statistics and (Inter)Correlations for Self-Reported Affect States, Affect Traits and Related Constructs

Variable	***M***	***SD***	***ICC***	**Valence**	**Arousal**	**Positive Affect**	**Negative Affect**
Valence	3.60	1.02	0.49				
Arousal	2.99	1.30	0.39	0.41 [0.34, 0.49]			
Positive Affect	3.36	0.80		0.40 [0.31, 0.48]	0.34 [0.25, 0.42]		
Negative Affect	2.11	0.92		-0.30 [-0.39, -0.21]	-0.11 [-0.21, -0.01]	-0.37 [-0.45, -0.28]	
Age	42.03	12.84		0.05 [-0.05, 0.14]	0.02 [-0.08, 0.11]	0.09 [-0.01, 0.19]	-0.07 [-0.17, 0.03]
Gender	0.48	0.50		-0.09 [-0.02, 0.17]	-0.12 [-0.17, 0.01]	-0.08 [-0.18, 0.02]	0.12 [0.01, 0.22]
Openness	-0.21	0.77		0.15 [0.04, 0.25]	0.21 [0.11, 0.31]	0.44 [0.35 0.52]	-0.05 [-0.15; 0.06]
Consciousness	-0.01	0.79		0.22 [0.12, 0.32]	0.20 [0.09, 0.30]	0.50 [0.41, 0.57]	-0.16 [-0.26, -0.05]
Extraversion	-0.22	0.76		0.26 [0.16, 0.36]	0.20 [0.10, 0.30]	0.51 [0.43, 0.59]	-0.22 [-0.32, -0.12]
Agreeableness	0.05	0.84		0.16 [0.06, 0.27]	0.12 [0.01, 0.22]	0.30 [0.20, 0.39]	-0.51 [-0.58, 0-.43]
Emotional Stability	-0.22	0.84		0.34 [0.25, 0.43]	0.21 [0.11, 0.31]	0.49 [0.41, 0.57]	-0.02 [-0.13, 0.08]

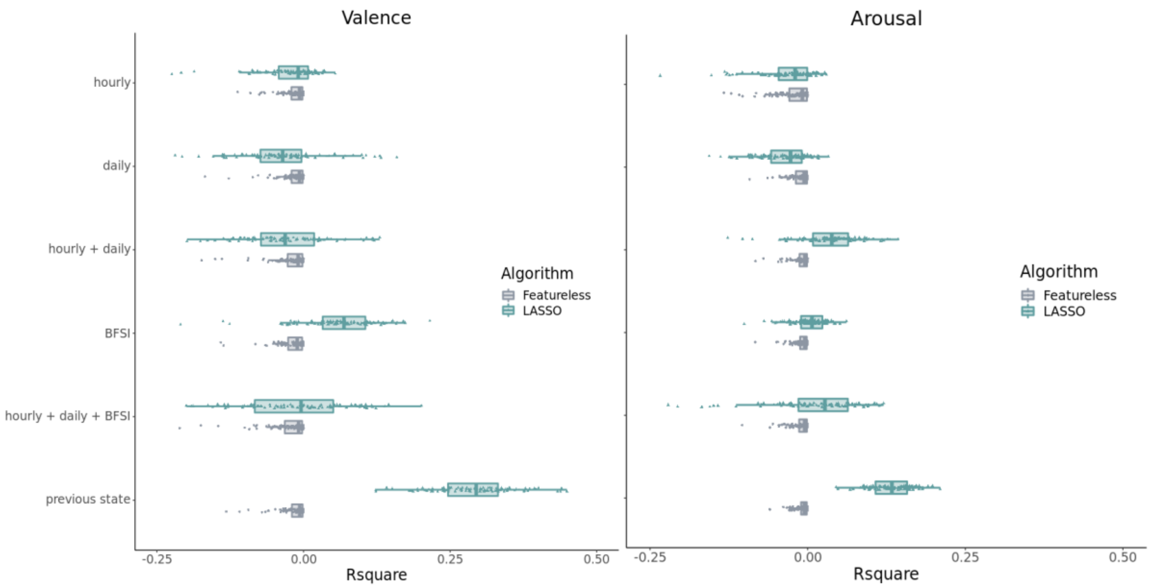
Note. Intra-class correlations (*ICC*) reflect the proportion of variance in state measures attributable to between-person effects. Affect state scores for valence and arousal were averaged per person to calculate Pearson correlations and 95% bootstrapped confidence intervals (in parentheses) at the person level. Due to the availability of questionnaire data, the sample size ($n = 453$) was reduced to $n = 384$ for demographics (age; gender) and to $n = 453$ for BFSI personality traits and PANAS (positive and negative affect) data. Gender was coded as 0 = *male* and 1 = *female*. Valence was coded from 0 = *very pleasant* to 5 = *very unpleasant.* Arousal was coded from 0 = *very activated* to 5 = *very inactive.* Openness, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability reflect BFI-2-XS facet scores coded from 0 = *do not agree at all* to 5 = *agree completely*. Positive and negative affect was measured with the PANAS questionnaire on a scale ranging from 0 = *not at all* to 4 = *extremely*.

Table A3.4: R^2 across Resampling Iterations for Affect States per Feature Set for the Lasso Model

	Prediction of Valence		Prediction of Arousal	
Feature Sets	M_{R2}	SD_{R2}	M_{R2}	SD_{R2}
Hourly	-0.039	0.088	-0.045	0.075
Daily	-0.103	0.241	-0.035	0.036
Hourly + Daily	-0.116	0.279	0.035	0.047
Hourly + Daily + BFSI	-0.091	0.251	0.007	0.089
BFSI	0.062	0.066	0.003	0.027
Previous Affect State	0.291	0.068	0.132	0.034

Note. This table shows the means and standard deviations of the R^2 measures across the 1000 resampling iterations of the repeated 10x10 CV scheme. The performance measures are reported per feature set that were systematically compared with the Lasso model in the benchmark experiment used to predict affect states (i.e., valence; arousal).

Figure A3.2: Distribution of R^2 across Resampling Iterations for Affect States per Feature Set for the Lasso Model



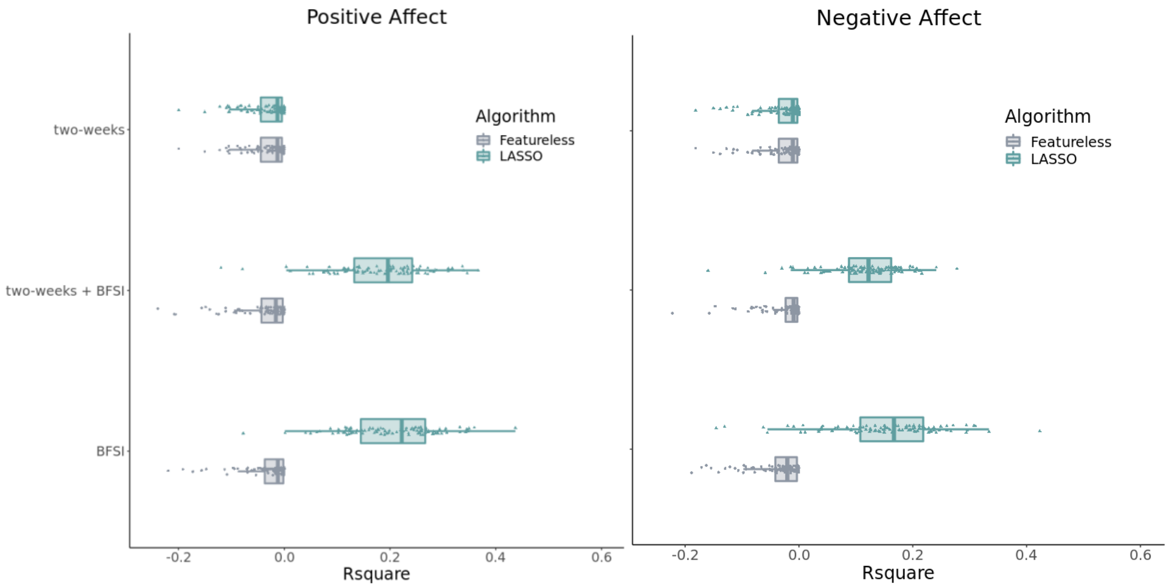
Note. Distribution of the R^2 measures over the resampling iterations of the applied repeated 10x10 CV scheme for Lasso and baseline featureless models. The y-axis reflects the different feature sets systematically compared in the benchmark experiment. The R^2s of the single iterations are represented by single dots. The boxes of the boxplots contain all values between the 25% and 75% quantiles, while their middle line indicates the median.

Table A3.5: R^2 across Resampling Iterations for Affect Traits per Feature Sets for the Lasso Model

Feature Sets	Prediction of Pos. Affect		Prediction of Neg. Affect	
	M_{R2}	SD_{R2}	M_{R2}	SD_{R2}
Two-week	-0.026	0.037	-0.033	0.052
Two-week + BFSI	0.115	0.067	0.182	0.087
BFSI	0.157	0.096	0.208	0.086

Note. This table shows the means and standard deviations of the R^2 measures across the 1000 resampling iterations of the repeated 10x10 CV scheme. The performance measures are reported per feature set systematically compared with the Lasso model in the benchmark experiment used to predict affect traits (i.e., positive affect; negative affect).

Figure A3.3: Distribution of R^2 across Resampling Iterations for Affect Traits per Feature Set for the Lasso Model



Note. Distribution of the R^2 measures across the 1000 resampling iterations of the applied ten times repeated 10x10 CV scheme for Lasso and baseline featureless models. The y-axis reflects the different feature sets systematically compared in the benchmark experiment. R^2s of the single iterations are represented by single dots. The boxes of the boxplots contain all values between the 25% and 75% quantiles, while their middle line indicates the median.

4. General Discussion

Tying up, the present dissertation explored the potential of smartphones as a research tool in psychology. While digital technologies have already been widely used in the field of personality assessment, this research sheds light on two additional potential areas of application in psychology, namely situation and affect research. The multi-method design of the two empirical studies combined experience sampling reports with passively collected smartphone sensor data. In doing so, the dissertation zoomed in on real-life experiences of psychological constructs in everyday situations. Study 1 demonstrated that certain characteristics of psychological situations can be predicted better than chance from smartphone-sensed data, such as phone usage, mobility patterns, or timestamp information. Study 2 investigated the use of different types of smartphone sensing data to recognize self-reported affective experiences in daily life. Despite the use of a wide range of sensing modalities, reflecting a comprehensive literature-driven selection of situational and behavioral indicators of affective well-being, predictions based solely on smartphone sensing data were not successful. These findings underscore the volatility and complexity of affective experiences and highlight the limitations and challenges of relying exclusively on objective data to predict highly personal and subjective experiences such as affect.

The subsequent discussion addresses the overall contribution of the dissertation, examines the limitations and boundaries of the studies, and derives potential directions for future research in the field of psychoinformatics.

4.1. Contribution of the Dissertation

4.1.1. Leveraging Smartphones in Psychological Research

This dissertation explored how smartphones can be leveraged in empirical research in two ways: collecting data through smartphone sensing and experience sampling self-reports. Thus, this work contributes to the burgeoning field of sensing research across various areas of applied psychology.

The Potential of Smartphone Sensing

A notable aspect of the two empirical studies is the incorporation of diverse smartphone sensors and log files, as well as external data sources. Prior research has mainly focused on individual data sources such as GPS and accelerometer (e.g., Müller et al., 2020; Sandstrom et al., 2017) or phone usage patterns (e.g., Montag et al., 2014). Only a few studies have combined a limited set of data types, including in-phone communication (LiKamWa et al., 2013; Messner et al., 2019) or physical activity and ambient sound (Wang et al., 2018). Moreover, only a handful of studies to date, mainly in the field of personality prediction, have employed a comprehensive range of different smartphone-sensed data types (e.g., Rüegger et al., 2020; Stachl et al., 2020). This dissertation advances this line of research by utilizing a broad spectrum of different smartphone-collected data types to facilitate more objective and unobtrusive measures of psychological constructs in real-life settings.

In Study 1, the situational DIAMONDS theory (Rauthmann et al., 2014; Rauthmann, Sherman, & Funder, 2015) was applied to examine real-life behaviors and situations. The study illustrated that specific situational cues, such as social interactions, activities, places, and time, can be assessed using passively collected data. This investigation corresponds with recent calls for more sensor-based situation research (Harari et al., 2020; Rauthmann et al., 2020) and builds on prior research on the associations between situational cues and psychological situations (e.g., Breil et al., 2019; Horstmann et al., 2021; Rauthmann et al., 2014). In support of previous self-report-based findings (e.g., Blake et al., 2020; Breil et al., 2019; Horstmann et al., 2021; Rauthmann et al., 2014), the smartphone-sensed variables demonstrated varying efficacy in capturing different situational characteristics. In particular, dimensions strongly linked with positive and negative affect (e.g., pOsitivity, Negativity, Deception, and Adversity; Horstmann et al., 2021) proved difficult to predict using smartphone-sensed situational cues. The results suggest that the smartphone-sensed data have a limited capability to represent the emotional quality or valence of a situation. These findings were also further corroborated in Study 2, which revealed limited predictability of affect traits and states, that are closely related to the valence-annotated psychological characteristics of situations, such as the degree of pOsitivity or Adversity of a situation (Horstmann et al., 2021; Horstmann & Ziegler, 2019; Kritzler et al., 2020). Previous research has already identified a significant overlap between measures of affect and situation perception (Horstmann et al., 2021; Horstmann & Ziegler, 2019). On the one hand, such person-situation interactions may arise from individuals engaging with situations in ways that correspond to their personality and mood (e.g., Horstmann & Ziegler, 2019; Rauthmann, 2021; Rauthmann & Sherman, 2016; Rauthmann, Sherman, Nave, et al., 2015). On the other hand, daily situations may also elicit specific affective experiences (e.g., Horstmann & Ziegler, 2019; Kuppens, 2009).

Combining the results of the two empirical studies of this dissertation, it can be concluded that the employed smartphone-sensed behavioral and situational indicators used did not adequately capture information about the affective components of daily situations. Identifying the emotional tone (i.e., valence) of an interaction or emotional experience evoked by a particular activity using solely objective data appears to be more challenging than simply detecting the occurrence of a social interaction or activity. Furthermore, the unsatisfactory predictive performance of certain models suggests that smartphone sensing research may require a substantial amount of data in terms of sensor types, granularity, and volume to achieve acceptable accuracy.

Integrating Multiple Data Sources

Although the technological developments of digital technologies in research are fascinating, the dissertation reveals that they have limitations and will not be the solution to all research questions posed in psychology. While smartphone sensing has shown potential in predicting certain dimensions of psychological situations, self-report measures remain crucial for certain areas of research. Accordingly, smartphone sensing data should complement, but not replace, self-reports (Montag & Elhai, 2019; Montag et al., 2021; Rauthmann, 2020). For instance, unobtrusive sensor-based measures may be suitable for studying traits such as personality or behavioral patterns such as sleep and sociability (Harari et al., 2020; Schoedel et al., 2020; Stachl et al., 2020). However, other psychological constructs such as mood or affective well-being are highly subjective experiences that probably hardly manifest in smartphone sensing data. Therefore, questionnaires are likely to remain an important source of information, as personal feelings and experiences play a crucial role. Smartphones and related data can only capture a fraction of a person's characteristics, behavior, and daily situations. They are only one piece of the puzzle. Relying on smartphones alone may therefore limit our understanding of the psychological constructs we are trying to capture, as much relevant information remains unknown. For example, neurochemical hormone levels (e.g., Alexander et al., 2021) or previous events, such as stressful incidents at work (e.g., Sonnentag, 2001), can also significantly influence how people think, feel, and behave in everyday life. Accordingly, in line with prior research (e.g., Cao et al., 2017), this dissertation also found that combining self-reports with passively sensed measures, such as smartphone sensing, can improve the predictive performance of psychometric sores. Therefore, future research should incorporate multiple data sources, including indirect measures, observations, and neurophysiological and biochemical indicators, to gain a more comprehensive understanding of psychological traits and states (Rauthmann, 2020).

Furthermore, the compatibility of algorithm-generated scores with established psychological constructs, typically assessed via self-report measures, requires further investigation (Montag & Elhai, 2019; Phan & Rauthmann, 2021; Rauthmann, 2020; Tempelaar et al., 2020). While smartphone-collected data and related metrics can provide useful information on certain aspects of traits and states, they should not be relied upon alone. Self-report measures will remain critical in relating smartphone data to established psychological constructs, while differences between an individual's self-perception and actual behavior can provide valuable insights.

4.1.2. Predictive Modeling in Psychological Research

Another important contribution of this thesis to psychological research is the exemplary application of predictive modeling approaches to two research areas in applied psychology. The emergence of new types and volumes of sensing data in psychological research requires appropriate analysis and interpretation of the results. Machine learning methods offer enormous flexibility for modeling not only large numbers of predictor variables but also non-linear associations between predictors and targets. However, focusing solely on maximizing predictive accuracy is not sufficient when applying predictive modeling in psychological assessment, as several challenges and pitfalls need to be cautiously considered (Fokkema et al., 2022; Rauthmann, 2020; Stachl et al., 2020). Thus, knowing the strengths and weaknesses of machine learning methods and how to harness their opportunities in psychological research is essential.

Challenges and Pitfalls

Therefore, the following section discusses some challenges of predictive modeling in psychological research and how they have been handled in the present research. In doing so, this dissertation fits into the current discourse on how psychological research should be conducted in the era of digitalization and machine learning (e.g., Fokkema et al., 2022; Rauthmann, 2020). Furthermore, this dissertation contributes to the discussion on the applicability of machine learning methods to traditional psychological science.

First, the choice of data preprocessing procedures, such as the way categorical variables are coded, can have a significant impact on the predictive performance of a model (Pargent et al., 2022). Thus, following the example of previous studies (e.g., Schoedel et al., 2020), this dissertation comprised a detailed description of the applied preprocessing pipeline (e.g., data transformation, missing value imputation, variable selection). In addition, the use of large amounts of data bears the risk of overemphasizing false relationships, which means that researchers may discover false effects by chance. Therefore, the predictive modeling approaches in both studies incorporated appropriate cross-validated resampling strategies to avoid overfitting and correctly estimate model performance (Pargent & Albert-von der Gönna, 2018; Renner et al., 2020; Yarkoni & Westfall, 2017). On the other hand, measurement error and heterogeneity among predictors can negatively impact the accuracy of predictive models, particularly in traditional research fields such as psychology (Luijken et al., 2019). Specifically, poor measurement quality can lead to severe underfitting of true relationships (Jacobucci & Grimm, 2020). The findings of this dissertation also suggest that large sample sizes may be required to assign significance to individual predictors, particularly when data sets contain an increasing number of predictors with modest effects (Efron, 2020).

Whereas predictive modeling methods have improved our ability to make predictions, the appropriate attribution and interpretation remains a challenging task. With a large number of predictors and non-linear associations between predictors and targets, fitted models are inherently difficult to interpret. Therefore, it is crucial to avoid a one-sided focus on maximizing predictive performance, as this neither adds value to understanding and explaining behavior nor enables hypothesis generation (Fokkema et al., 2022). While various tools have been developed to interpret black box models, their precision is not always quantifiable (Carvalho et al., 2019; Molnar, 2022) and their functioning is not always easy to understand, potentially leading to misapplication and misinterpretation (Kaur et al., 2020; Rudin, 2019). The studies in this dissertation compared algorithms with different levels of complexity, namely linear and non-linear models, to shed light on the nature of associations in the data. In line with prior research, no (substantial) benefit of more sophisticated non-linear machine learning models over classical linear models was observed (e.g., Fokkema et al., 2022; Hand et al., 2006). Accordingly, the linear model was chosen to apply interpretable machine learning techniques to contribute to the theoretical matching of situational cues and psychological situations, promoting the use of more interpretable and sparse models (e.g., Kotsiantis et al., 2007; Molnar, 2022; Rudin, 2019).

Moreover, this work follows the increasing call from researchers highlighting that the scores predicted from smartphone sensing data should be subject to rigorous psychometric testing for reliability and validity (e.g., Alexander et al., 2020; Bleidorn & Hopwood, 2019; Rauthmann, 2020). The use of newer and larger data sources requires even greater validity of assessment procedures to avoid limited content and validity of results. In particular, construct validity must be demonstrated to ensure that the measures accurately represent the intended concept (Ziegler, 2014). Improving the understanding of the causes and consequences of smartphone-sensed and algorithm-based assessments requires a strong focus on nomological validity (Rauthmann, 2020). Therefore, the application of nomological networks (Cronbach & Meehl, 1955) in Study 1 was an important first step in contextualizing these scores and identifying what they do and do not capture. Future research should continue to refine, adapt, and develop theory, evidence, and methods to open the 'black box' of predictive models (Alexander et al., 2020; Bleidorn & Hopwood, 2019).

Interdisciplinarity and Collaboration

Finally, collecting, analyzing, and interpreting the high-dimensional and complex data sets that emerge from sensing studies in real-world settings requires additional skills in both statistics and computer science. In addition, both studies of this dissertation have confirmed the essential role of incorporating domain knowledge and prior research to enable a theory-driven feature engineering process (as illustrated in Figure 1.2) and to develop meaningful and interpretable models. Bringing together different perspectives and competencies ensures that the models are accurate, responsible, ethical, and easy to understand (Phan & Rauthmann, 2021; Renner et al., 2020). Thus, intensified interdisciplinary collaboration is essential to not only collect, but also to correctly analyze and interpret smartphone sensing data. While we agree with Rauthmann et al. (2020) that the primary role of psychologists should be the psychological meaning of the digitally sensed variables, statistical competencies are also essential for researchers to understand and correctly analyze the vast amounts of data generated by digital technologies. Thus, big data-specific skills should be integrated into the common methodological and statistical training of psychologists (Jacobucci & Grimm, 2020; Yarkoni & Westfall, 2017).

4.2. Limitations and Future Directions for Research

Despite the progress this dissertation has made in using smartphone sensing for situation and affect research, there are still some limitations that need to be addressed. Therefore, this chapter discusses some key challenges and suggests future research directions to advance psychological research in the digital age.

4.2.1. Integrative Research in Psychology

The rise of new technologies and data sources has increased the focus on exploring patterns in psychological data. To fully harness the potential of these data, it is crucial to strike a balance between identifying *what* works and understanding *why* it works (Fried, 2020; Rauthmann, 2020; Robinaugh, et al., 2021). This requires a theoretical and formalized approach, especially when collecting and analyzing large amounts of data in an exploratory way, bearing the challenge of controlling for multiple testing and avoiding false positive findings. Big data can reveal new findings that need to be supported and explained by theory, which is why "theory must be also the guiding light in a digital age" (Montag & Elhai, 2019; p.132). Thus, smartphone sensing-based research, whether explanatory or predictive, or some combination, should specify its scientific goals and theoretical framework (Hofman et al., 2021). The present dissertation responds to the appeal for more predictive approaches in applied psychology (Yarkoni & Westfall, 2017). While explanation and prediction have often been cast as opposing scientific goals, scientists have increasingly outlined that the approaches should be used in a complementary manner (Hofman et al., 2021; Mahmoodi, et al., 2017; Yarkoni & Westfall, 2017). Accordingly, any exploratory predictive research should be cyclically followed by more explanatory research, deriving testable hypotheses from the theory and operationalizing them in quantifiable terms (Kuppens et al., 2022).

As described in the respective discussions of Study 1 and Study 2, this dissertation raises a variety of research questions that should be addressed by future studies in a hypothesis-driven manner. To give just one example, an interesting research question to be pursued in future studies could focus on investigating the role of personality traits in the interplay of situational indicators such as locations and specific dimensions of psychological situations. The additional nomological analyses in Study 1 have already indicated that the Big Five personality traits are associated with psychological situations, adding to previous studies (e.g., Jonason & Sherman, 2020; Rauthmann et al., 2014; Sherman et al., 2015). Other studies have linked individuals' momentary affect to their location, providing preliminary evidence that the relationship between affective experience and location may be moderated by personality (e.g., Sandstrom et al., 2017). Therefore, an interesting hypothesis to test may be whether the associations between situational indicators and psychological characteristics of situations identified in this dissertation (such as the location and perceived sociability of a situation) are moderated by personality traits. In addition, the following two sections provide additional suggestions for future research questions that can be answered in an explanatory manner.

4.2.2. Big Data as Fuel for Idiographic Approaches in Psychology

Traditionally, predictive models in psychology have focused primarily on identifying features that differ between individuals in order to predict aggregated behaviors (e.g., Puterman et al., 2020; van Mens et al., 2020). Also, neither of the two empirical studies in this dissertation systematically considered within-person variability in their analyses. While so-called nomothetic approaches focus on differentiating people from one another and using such differences for prediction, idiographic approaches focus on understanding the dynamics within individuals (Molenaar, 2004). However, as already discussed in Study 2 in the context of affect research, idiographic approaches are essential for understanding the highly individual and dynamic nature of human behavior and cognition.

Nevertheless, personalized idiographic models come at the cost of increased data requirements, as users must provide their own training data to personalize their models (LiKamWa et al., 2013). Recent technological advances in both data collection and computationally intensive analyses can fuel idiographic approaches in psychological research (Renner et al., 2020). For instance, smartphone sensing methods can enable unobtrusive data collection across diverse contexts (e.g., Ren et al., 2022), while (smartphone-based) experience sampling methods enable dynamic, real-world assessments of psychological constructs (e.g., Conner et al., 2009; Scollon et al., 2003). Leveraging big data for idiographic models, a study by Cheung et al. (2017) found that idiographic prediction models using machine learning algorithms were more accurate than nomothetic models for predicting physical activity and exercise behavior. Similar approaches have already been applied in preventive medicine to predict drug use (e.g., Boyer et al., 2012) and in clinical psychology to predict smoking behavior (e.g., Fisher & Soyster, 2019). While scholars have already highlighted the potential of recent technological advances for more idiographic personality research (Renner et al., 2020), it is still underutilized in other psychological research areas and should be further leveraged to "bring the person back into scientific psychology" (Molenaar, 2004, p.202).

On the other hand, the increasing popularity of person-specific idiographic models in psychological science has also been met with criticism for their potential lack of generalizability to other individuals or situations (e.g., Beltz et al., 2016; Connor et al., 2009; Spencer & Schöner, 2003). While idiographic models offer highly personalized approaches, they may not be able to distinguish between individual-specific patterns and general processes that all individuals engage in, at least to some extent (Wright & Zimmermann, 2019). For example, almost everyone experiences daily stress and negative emotions as a result from time to time.

Similarly, there is an overlap (consensual variance) in situational perceptions that we all share (Rauthmann, Sherman, Nave, et al., 2015; Sherman et al., 2015; Wagerman & Funder, 2009). For instance, it is common sense that being at a party is typically considered a social situation, whereas reading a book at home is not. Furthermore, research in the field of genetic analysis suggests that positive affect may be more situational, whereas negative affect may be more dispositional - underlining the relevance of intra- and inter-individual variability (e.g., Zheng et al., 2016).

This dissertation therefore highlights the importance of balancing the ability to develop personalized models with ensuring comparability across individuals in order to fully realize the value of big data (Wright & Zimmermann, 2019). Rather than treating the choice between nomothetic and idiographic approaches as dichotomous, they should be seen as complementary, as both can provide valuable insights (Beltz et al., 2016; Wright & Zimmermann, 2019). Consequently, statistical approaches have been developed that integrate person-specific models (i.e., developing a unique model per individual) with a data-driven search for the optimal weighting across all subject-specific models (i.e., 'borrowing' information from prediction models of other individuals) (e.g., Ren, et al., 2022, 2021). Such ideographically weighted machine learning approaches have demonstrated superior performance to fully idiographic models, suggesting that meaningful shared signals exist across individuals in the predictive relationships between smartphone sensor data and affective experiences (e.g., Jacobson & Chung, 2020; Ren et al., 2022). Nonetheless, substantial variation across individuals in the directionality and strength of the association between predictors and outcomes has been observed, underscoring the importance of personalized models. Similarly, Beck and Jackson (2022) found that both personal and situational factors predicted future behavior and experiences, with key predictors varying significantly across individuals. Thus, the value of multi-method studies using smartphones to collect big data will be most fully realized by balancing the ability to develop personalized models with ensuring comparability across individuals. However, systematically combining nomothetic and idiographic approaches is challenging and still requires future theoretical and empirical investigation (Beltz et al., 2016; Wright & Zimmermann, 2019).

4.2.3. Challenges of Smartphone-Based Studies

Finally, this dissertation has also identified some difficulties in using smartphones as data collectors that should be further addressed and investigated in future studies.

Generalizability of Findings

Although psychological assessment is increasingly focused on studies outside the laboratory, generalization of research findings remains difficult and external validity is crucial. Advances in predictive accuracy which have been achieved in controlled research settings can be negated by practical aspects of data issues such as coverage errors in samples, measurement errors, interpretability, or ethical issues (Efron, 2020; Fokkema et al., 2022; Luijken et al., 2019; Rauthmann, 2020).

Coverage Errors in Samples

For example, smartphone sensing studies that target the general population can face concerns about a sample coverage bias. In general, smartphone ownership can be correlated with substantive socioeconomic and demographic variables, as smartphone owners tend to be younger, more educated, and more likely to live in larger communities than non-smartphone owners (Keusch et al., 2020). Furthermore, data quality is highly dependent on the willingness of participants to download the research app and share their data, which can also be influenced by socio-demographic characteristics (Keusch et al., 2022). Although the sample used in the studies of this dissertation was carefully recruited and representative of the German population in terms of age and gender, a coverage bias might have occurred due to technical limitations of the research app, which was only available for certain Android systems (Schoedel & Oldemeier, 2020). Previous studies have found little coverage bias for key personality and socio-demographic characteristics exist due to smartphone ownership - even when the sample was restricted to Android smartphone owners only (Kreuter et al., 2020). However, future research is needed to further investigate how participants' characteristics, such as age, privacy concerns, or motivation, may affect data quality and how data analyses need to be adjusted to account for potential coverage and non-participation bias.

Determination of Causality

In addition, it is very difficult to determine causality in observational data from smartphones because it is not possible to randomly assign participants to specific conditions. In a laboratory research setting, participants can be assigned to specific treatment and control groups in a controlled manner. For example, randomly selected participants can be instructed to engage in a sporting activity prior to the interview, while others can be instructed to lie down. Because smartphone sensing studies take place in real-world settings outside the laboratory, participants cannot be randomly assigned to specific conditions, nor can their physical environment or activity be controlled. Therefore, smartphone-based studies have primarily focused on identifying relationships between smartphone data, rather than determining cause and effect (Tsapeli & Musolesi, 2015). However, an alternative quasi-experimental approach for smartphone sensing studies can be to compare participants with similar values on the confounding variables but different treatment levels (Hofman et al., 2021). For example, participants can be divided into a treatment group (e.g., high physical activity) and a control group (e.g., no physical activity) based on their accelerometer and location data (Tsapeli & Musolesi, 2015).

Biases in Smartphone Sensing Data

Another challenge of conducting research in real life is that there may be uncontrollable external factors that influence the behaviors and situations that are observed during data collection. The studies of this dissertation were conducted during the Covid-19 pandemic, which undoubtedly had a strong impact on everyday life in Germany (Kuper et al., 2021). Although the data in the present studies were collected during a period of comparatively relaxed restrictions in and across Germany (Hale et al., 2021; Steinmetz et al., 2022), the restrictions negatively affected the social and cultural activities (e.g. Kohls et al., 2021), mobility patterns (e.g., Anke et al., 2021; Destatis, 2023), as well as mental health and affective well-being (e.g., Ammar et al., 2020; Hajek & König, 2022; Lades et al., 2020). Thus, the replicability of our findings in post-pandemic settings needs to be further investigated in future studies.

Moreover, while the empirical Study 1 of this dissertation has shown that smartphones can be a valuable source of information for certain psychological constructs, the devices are primarily designed for social interaction. As Study 2 in particular has highlighted, this raises the question of how much of the information obtained from smartphones actually reflects an individual's internal processes, and how much it is influenced by their interactions with others via the device or with the device itself (Montag et al., 2016). The reliability and validity of sensor-based data in psychological research is still not well understood, which can result in inaccurate conclusions, as scientists warn (e.g., Fokkema et al., 2022; Phan & Rauthmann, 2021; Rauthmann, 2020; Ziegler, 2014). For example, the mere awareness of being monitored in a study can lead to distortions and trigger emotions, memories, or behavioral patterns (Kern et al., 2016; Kosinski et al., 2015; Whelan & DuVernet, 2015). In addition, the use of digital technology itself could influence the observed behavioral, emotional, as well as cognitive processes (e.g., Hadar et al., 2017; Hoehe & Thibaut, 2020). Therefore, another question for future research is to systematically investigate whether there are biases caused by smartphone-based study designs. Furthermore, establishing common research standards and guidelines is another important step to avoid false conclusions and to control for potential biases in smartphone sensing data.

Study Design Guidelines

Due to the growing number of smartphone sensing studies, researchers can already build on existing smartphone sensing measures of in-phone communication (e.g., number of outgoing calls or text messages) or mobility behavior (e.g., total distance traveled per day). However, common study guidelines and research standards are crucial to ensure comparable and reproducible results, as well as high data quality (Montag et al., 2016). Otherwise, poor measurement quality can lead to severe underfitting of true associations (Jacobucci & Grimm, 2020). Therefore, research-based recommendations are needed to guide the design of smartphone sensing data collection, including details on sampling frequency or collection period. In particular, the impact of the study duration and sampling frequency on the data quality for different data types needs to be systematically investigated. For example, a participant's phone usage behavior may only need to be observed for a few days or weeks, whereas less frequently used features may require longer observation periods (Montag et al., 2014). Furthermore, as discussed in Study 2, the selection of specific time windows of raw data used for feature extraction is an important but challenging decision. Studies on personality traits and affective well-being may need to focus on daily or weekly patterns (e.g., DeMasi et al., 2017; Sano et al., 2018; Stachl et al., 2020), whereas studies on state variables should rather use an hourly time scale (e.g., Ren et al., 2022). Although the studies presented in this dissertation compared different time scales in several benchmark experiments, only a fraction of all possible levels of aggregation could be investigated.

In addition, smartphone sensing studies require strict attention to data security and ethical guidelines, as they may reveal personal information about participants (Harari et al. 2016; Montag et al., 2016; Phan & Rauthmann, 2021; Renner et al., 2020). Therefore, very complex technical solutions are required to conduct sensing-based studies and handle the data. The PhoneStudy research app used in this dissertation used encryption and only shared aggregated data, excluding any personally identifiable information (Schoedel & Oldemeier, 2020). However, privacy concerns and risks are increasing with the growth of mental health care technologies (e.g., Iwaya et al., 2023). Adherence to ethical standards is not only morally crucial when collecting sensitive data but can also benefit sample quality by increasing willingness to participate (e.g., Bemmann et al., 2022; Harari et al., 2016; Kreuter et al., 2020).

Thus, researchers must continue to advance the ethical discourse and treat data privacy as a top priority in smartphone sensing research.

4.3. Conclusion

The digital age has opened up new opportunities for empirical research, also in applied psychology. This dissertation contributes to the emerging application of smartphone sensing methods by investigating the potential of smartphone-gathered data for situation and affect research. The two empirical studies combined active logging of experience-sampled self-reports with passively collected smartphone sensing data. In contrast to previous research, a broad range of different data types were combined in order to observe situational and behavioral indicators of the psychological concepts as comprehensively as possible. By providing two empirical examples of applying state-of-the-art machine learning methods to psychological research questions, this dissertation contributes to a better understanding of the potential of introducing more observational data. The research findings indicate that smartphones can serve as an important source of information for psychologists to better understand and predict person- and situation-specific characteristics - but the full potential is only realized in combination with other data sources and methods. Accordingly, self-report- and sensor-based research should not be seen and treated as competing (or even substituting) approaches, but rather as complementary methods.

In summary, smartphone sensing data can offer new opportunities for objective observation, but also introduce new challenges compared to traditional self-report measures. Mitigating these challenges requires sophisticated study design standards, interdisciplinary collaboration with experts, and multi-method approaches. Conducting high-quality research outside the laboratory needs a deep understanding and expertise in various fields and may necessitate increased communication and collaboration with other researchers and specialists. Accordingly, psychological research will require a great deal of openness, including “open data, open collaboration, and above all, open minds” (Ong, 2016, p. 6).

4.4. References

Alexander, L. III, Mulfinger, E., & Oswald, F. L. (2020). Using big data and machine learning in personality measurement: Opportunities and challenges. *European Journal of Personality, 34*(5), 632-648. https://doi.org/10.1002/per.2305

Alexander, R., Aragón, O. R., Bookwala, J., Cherbuin, N., Gatt, J. M., Kahrilas, I. J., ... & Styliadis, C. (2021). The neuroscience of positive emotions and affect: Implications for cultivating happiness and wellbeing. *Neuroscience & Biobehavioral Reviews, 121*, 220-249. https://doi.org/10.1016/j.neubiorev.2020.12.002

Ammar, A., Mueller, P., Trabelsi, K., Chtourou, H., Boukhris, O., Masmoudi, L., ... & ECLB-COVID19 Consortium. (2020). Psychological consequences of COVID-19 home confinement: The ECLB-COVID19 multicenter study. *PloS one, 15*(11), e0240204. https://doi.org/10.1371/journal.pone.0240204

Anke, J., Francke, A., Schaefer, L. M., & Petzoldt, T. (2021). Impact of SARS-CoV-2 on the mobility behaviour in Germany. *European Transport Research Review, 13*(1), 1-13. https://doi.org/10.1186/s12544-021-00469-3

Beck, E. D., & Jackson, J. J. (2022). Personalized Prediction of Behaviors and Experiences: An Idiographic Person-Situation Test. *Psychological Science, 33*(10), 1767-1782. https://doi.org/10.1177/09567976221093

Beltz, A. M., Wright, A. G., Sprague, B. N., & Molenaar, P. C. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment, 23*(4), 447-458. http://dx.doi.org/10.1177/1073191116648209

Bemmann, F., Windl, M., Erbe, J., Mayer, S., & Hussmann, H. (2022). The Influence of Transparency and Control on the Willingness of Data Sharing in Adaptive Mobile Apps. *Proceedings of the ACM on Human-Computer Interaction, 6*(MHCI), 1-26. https://doi.org/10.1145/3546724

Blake, A. B., Lee, D. I., De La Rosa, R., & Sherman, R. A. (2020). Wearable cameras, machine vision, and big data analytics: Insights into people and the places they go. In S. E. Woo, L. Tay, & R. W. Proctor (Eds.), *Big data in psychological research* (pp. 125-143). American Psychological Association. https://doi.org/10.1037/0000193-007

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review, 23*(2), 190-203. https://doi.org/10.1177/1088868318772990

Breil, S. M., Geukes, K., Wilson, R. E., Nestler, S., Vazire, S., & Back, M. D. (2019). Zooming into real-life extraversion - how personality and situation shape sociability in social interactions. *Collabra: Psychology, 5*(1). https://doi.org/10.1525/collabra.170

Cao, B., Zheng, L., Zhang, C., Yu, P. S., Piscitello, A., Zulueta, J., ... & Leow, A. D. (2017, August). Deepmood: modeling mobile phone typing dynamics for mood detection. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 747-755). https://doi.org/10.1145/3097983.3098086

Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics, 8*(8), 832. https://doi.org/10.3390/electronics8080832

Cheung, Y. K., Hsueh, P. Y. S., Qian, M., Yoon, S., Meli, L., Diaz, K. M., ... & Davidson, K. W. (2017). Are nomothetic or ideographic approaches superior in predicting daily exercise behaviors?.*Methods of information in medicine, 56*(06), 452-460. https://doi.org/ 10.3414/ME16-02-0051

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience sampling methods: A modern idiographic approach to personality research. *Social and personality psychology compass, 3*(3), 292-313. https://doi.org/10.1111/j.1751-9004.2009.00170.x

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin, 52*(4), 281. https://doi.org/10.1037/h0040957

DeMasi, O., Feygin, S., Dembo, A., Aguilera, A., & Recht, B. (2017). Well-being tracking via smartphone-measured activity and sleep: cohort study. *JMIR mHealth and uHealth, 5*(10), e7820. http://dx.doi.org/10.2196/mhealth.7820

Destatis (2023). Mobilität während der Corona-Pandemie. Federal Statistical Office of Germany (Destatis). Retrieved from https://www.destatis.de/DE/Methoden/WISTA-Wirtschaft-und-Statistik/2021/03/mobilitaet-corona-pandemie-032021.html

Efron, B. (2020). Prediction, estimation, and attribution. *International Statistical Review, 88*, 28-59. https://doi.org/10.1111/insr.12409

Fisher, A. J., & Soyster, P. (2019). *Generating accurate personalized predictions of future behavior: A smoking exemplar*. PsyArXiv. https://psyarxiv.com/e24v6

Fokkema, M., Iliescu, D., Greiff, S., & Ziegler, M. (2022). Machine learning and prediction in psychological assessment: Some promises and pitfalls. *European Journal of Psychological Assessment, 38*(3), 165. https://doi.org/10.1027/1015-5759/a000714

Fried, E. I. (2020). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry, 31*(4), 271-288. https://doi.org/10.1080/1047840X.2020.1853461

Hadar, A., Hadas, I., Lazarovits, A., Alyagon, U., Eliraz, D., & Zangen, A. (2017). Answering the missed call: Initial exploration of cognitive and electrophysiological changes associated with smartphone use and abuse. *PloS one, 12*(7), e0180094. https://doi.org/10.1371/journal.pone.0180094

Hajek, A., & König, H. H. (2022). Prevalence and correlates of loneliness, perceived and objective social isolation during the COVID-19 pandemic. Evidence from a representative survey in Germany. *Social psychiatry and psychiatric epidemiology, 57*(10), 1969-1978. https://doi.org/10.1007/s00127-022-02295-x

Hale, T., Angrist, N., Hale, A. J., Kira, B., Majumdar, S., Petherick, A., ... & Zhang, Y. (2021). Government responses and COVID-19 deaths: Global evidence across multiple pandemic waves. *PLoS One, 16*(7), e0253116. https://doi.org/10.1371/journal.pone.0253116

Hand, D. J. (2006). Classifier Technology and the Illusion of Progress. *Statistical Science, 21*(1), 1-15. HTTPS://DOI.ORG/ 10.1214/088342306000000060

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science, 11*(6), 838-854. https://doi.org/10.1177/1745691616650285

Harari, G. M., Müller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., ... & Gosling, S. D. (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *Journal of personality and social psychology, 119*(1), 204-228. https://doi.org/ 10.1037/pspp0000245

Hoehe, M. R., & Thibaut, F. (2020). Going digital: how technology use may influence human brains and behavior. *Dialogues in Clinical Neuroscience, 22*(2), 93-97. https://doi.org/ 10.31887/DCNS.2020.22.2/mhoehe

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., ... & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature, 595*(7866), 181-188. https://doi.org/10.1038/s41586-021-03659-0

Horstmann, K. T., Rauthmann, J. F., Sherman, R. A., & Ziegler, M. (2021). Unveiling an exclusive link: Predicting behavior with personality, situation perception, and affect in a preregistered experience sampling study. Journal of Personality and Social Psychology, 120(5), 1317-1343. https://doi.org/10.1037/pspp0000357

Horstmann, K. T., & Ziegler, M. (2019). Situational perception and affect: Barking up the wrong tree? *Personality and Individual Differences, 136*, 132-139. https://doi.org/10.1016/j.paid.2018.01.020

Iwaya, L. H., Babar, M. A., Rashid, A., & Wijayarathna, C. (2023). On the privacy of mental health apps: An empirical investigation and its implications for app development. *Empirical Software Engineering, 28*(1), 2. https://doi.org/10.1007/s10664-022-10236-0

Jacobson, N. C., & Chung, Y. J. (2020). Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. *Sensors, 20*(12), 3572. https://doi.org/ 10.3390/s20123572

Jacobucci, R., & Grimm, K. J. (2020). Machine learning and psychological research: The unexplored effect of measurement. *Perspectives on Psychological Science, 15*(3), 809-816. https://doi.org/10.1177/17456916209024

Jonason, P. K., & Sherman, R. A. (2020). Personality and the perception of situations: The five and dark triad traits. *Personality and Individual Differences, 163*, 110081. https://doi.org/10.1016/j.paid.2020.110081

Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Wortman Vaughan, J. (2020). Interpreting interpretability: Understanding data scientists' use of interpretability tools for machine learning. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-14. https://doi.org/10.1145/3313831.3376219

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods, 21*, 507-525. https://doi.org/10.1037/ met0000091

Keusch, F., Bähr, S., Haas, G. C., Kreuter, F., & Trappmann, M. (2020). Coverage error in data collection combining mobile surveys with passive measurement using apps: Data from a German national survey. *Sociological Methods & Research*, 0049124120914924. https://doi.org/10.1177/0049124120914924

Keusch, F., Bähr, S., Haas, G. C., Kreuter, F., Trappmann, M., & Eckman, S. (2022). Non-participation in smartphone data collection using research apps. *Journal of the Royal Statistical Society. Series A, Statistics in Society*, *2022*, 1-21. https://doi.org/ 10.1111/rssa.12827

Kohls, E., Baldofski, S., Moeller, R., Klemm, S. L., & Rummel-Kluge, C. (2021). Mental health, social and emotional well-being, and perceived burdens of university students during COVID-19 pandemic lockdown in Germany. *Frontiers in psychiatry, 12*, Article 643957. https://doi.org/10.3389/fpsyt.2021.643957

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American psychologist, 70*(6), 543-556. https://doi.org/10.1037/a0039210

Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering, 160*, 3-24. Retrieved from http://ahdehao.com/uploads/tmp/071c35cf43.pdf

Kreuter, F., Haas, G. C., Keusch, F., Bähr, S., & Trappmann, M. (2020). Collecting survey and smartphone sensor data with an app: Opportunities and challenges around privacy and informed consent. *Social Science Computer Review, 38*(5), 533-549. https://doi.org/10.1177/0894439318816389

Kritzler, S., Krasko, J., & Luhmann, M. (2020). Inside the happy personality: Personality states, situation experience, and state affect mediate the relation between personality and affect. *Journal of Research in Personality, 85*, 103929. https://doi.org/10.1016/j.jrp.2020.103929

Kuper, N., Modersitzki, N., Phan, L. V., & Rauthmann, J. (2021). The situation during the COVID-19 pandemic: A snapshot in Germany. *PLOS ONE, 16*(2), e0245719. https://doi.org/10.1371/journal.pone.0245719

Kuppens, P. (2009). The legacy of the person? situation debate for understanding variability in emotional experience. *Journal of Research in Personality, 43*(2), 255-256. https://doi.org/10.1016/j.jrp.2008.12.027

Kuppens, P., Dejonckheere, E., Kalokerinos, E. K., & Koval, P. (2022). Some recommendations on the use of daily life methods in affective science. A*ffective Science, 3*, 505-515. https://doi.org/10.1007/s42761-022-00101-0

Lades, L. K., Laffan, K., Daly, M., & Delaney, L. (2020). Daily emotional well-being during the COVID-19 pandemic. *British journal of health psychology, 25*(4), 902-911. https://doi.org/10.1111/bjhp.12450

LiKamWa, R., Liu, Y., Lane, N. D., & Zhong, L. (2013, June). Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services* (pp. 389-402). http://dx.doi.org/10.1145/2462456.2483967

Luijken, K., Groenwold, R. H., Van Calster, B., Steyerberg, E. W., & van Smeden, M. (2019). Impact of predictor measurement heterogeneity across settings on the performance of prediction models: A measurement error perspective. *Statistics in Medicine, 38*(18), 3444-3459. https://doi.org/10.1002/sim.8183

Mahmoodi, J., Leckelt, M., van Zalk, M. W., Geukes, K., & Back, M. D. (2017). Big Data approaches in social and behavioral science: four key trade-offs and a call for integration. *Current Opinion in Behavioral Sciences, 18*, 57-62. https://doi.org/10.1016/j.cobeha.2017.07.001

Messner, E. M., Sariyska, R., Mayer, B., Montag, C., Kannen, C., Schwerdtfeger, A., & Baumeister, H. (2019). Insights-future implications of passive smartphone sensing in the therapeutic context. *Verhaltenstherapie*, 1-10. http://dx.doi.org/10.1159/000501951

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement, 2*(4), 201-218. https://doi.org/10.1207/s15366359mea0204_1

Molnar, C. (2022). *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Retrieved from https://christophm.github.io/interpretable-ml-book/

Montag, C., Błaszkiewicz, 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. https://doi.org/10.1027/1614-0001/a000139

Montag, C., Duke, É., & Markowetz, A. (2016). Toward psychoinformatics: Computer science meets psychology. *Computational and mathematical methods in medicine, 2016*, 1-10. https://doi.org/10.1155/2016/2983685

Montag, C., & Elhai, J. D. (2019). A new agenda for personality psychology in the digital age?. *Personality and Individual Differences, 147*, 128-134. https://doi.org/10.1016/j.paid.2019.03.045

Montag, C., Dagum, P., Hall, B. J., & Elhai, J. D. (2021). Do we still need psychological self-report questionnaires in the age of the Internet of Things?.*Discover Psychology, 2*(1), 1-5. https://doi.org/10.1007/s44202-021-00012-4

Müller, S. R., Peters, H., Matz, S. C., Wang, W., & Harari, G. M. (2020). Investigating the relationships between mobility behaviours and indicators of subjective well-being using smartphone-based experience sampling and GPS tracking. *European Journal of Personality, 34*(5), 714-732. http://dx.doi.org/10.1002/per.2262

Ong, M. B. H. (2016). Biden designates nci's genomic data commons as foundation of moonshot. *The Cancer Letter, 42*, 1- 8

Pargent, F., & Albert-Von Der Gönna, J. (2018). Predictive Modeling with Psychological Panel Data. *Zeitschrift Fur Psychologie / Journal of Psychology, 226*(4), 246-258. https://doi.org/10.1027/2151-2604/a000343

Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2022). Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. *Computational Statistics, 37*, 2671-2692. https://doi.org/10.1007/s00180-022-01207-6

Phan, L. V., & Rauthmann, J. F. (2021). Personality computing: New frontiers in personality assessment. *Social and personality psychology compass, 15*(7), e12624. https://doi.org/10.1111/spc3.12624

Puterman, E., Weiss, J., Hives, B. A., Gemmill, A., Karasek, D., Mendes, W. B., & Rehkopf, D. H. (2020). Predicting mortality from 57 economic, behavioral, social, and psychological factors. *Proceedings of the National Academy of Sciences, 117*(28), 16273-16282. https://doi.org/10.1073/pnas.1918455117

Rauthmann, J. F. (2020). A (more) behavioural science of personality in the age of multi-modal sensing, big data, machine learning, and artificial intelligence. *European Journal of Personality, 34*(5), 593-598. https://doi.org/10.1002/per.2310

Rauthmann, J. F. (2021). Capturing interactions, correlations, fits, and transactions: A person-environment relations model. In *The handbook of personality dynamics and processes* (pp. 427-522). Academic Press. https://doi.org/10.1016/B978-0-12-813995-0.00018-2

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., ... & Funder, D. C. (2014). The Situational Eight DIAMONDS: a taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology, 107(*4), 677. https://doi.org/10.1037/a0037250

Rauthmann, J. F., & Sherman, R. A. (2016). Situation change: Stability and change of situation variables between and within persons. *Frontiers in Psychology, 6*(1938). https://doi.org/10.3389/fpsyg.2015.01938

Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality, 29*(3), 363-381. https://doi.org/10.1002/per.1994

Rauthmann, J. F., Sherman, R. A., Nave, C. S., & Funder, D. C. (2015). Personality-driven situation experience, contact, and construal: How people's personality traits predict characteristics of their situations in daily life. *Journal of Research in Personality*, *55*, 98-111. https://doi.org/10.1016/j.jrp.2015.02.003

Rauthmann, J. F., Horstmann, K. T., & Sherman, R. A. (2020). The psychological characteristics of situations: Towards an integrated taxonomy. In J. F. Rauthmann, R. A. Sherman, & D. C. Funder (Eds.), *The Oxford handbook of psychological situations* (pp. 389-406). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190263348.013.19

Ren, B., Balkind, E. G., Pastro, B., Israel, E., Pizzagalli, D. A., Rahimi-Eichi, H., ... & Webb, C. A. (2022). Predicting states of elevated negative affect in adolescents from smartphone sensors: A novel personalized machine learning approach. PsyArXiv. https://doi.org/10.31234/osf.io/zjm4d

Ren, B., Patil, P., Dominici, F., Parmigiani, G., & Trippa, L. (2021). Cross-study learning for generalist and specialist predictions. ArXiv:2007.12807 [Math, Stat]. Retrieved from http://arxiv.org/abs/2007.12807

Renner, K. H., Klee, S., & von Oertzen, T. (2020). Bringing back the person into behavioural personality science using big data. *European Journal of Personality, 34*(5), 670-686. http://dx.doi.org/10.1002/per.2303

Robinaugh, D. J., Haslbeck, J. M., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible Hands and Fine Calipers: A Call to Use Formal Theory as a Toolkit for Theory Construction. *Perspectives on psychological science: a journal of the Association for Psychological Science, 16*(4), 725. https://doi.org/10.1177/1745691620974

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence, 1*(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x

Rüegger, D., Stieger, M., Nißen, M., Allemand, M., Fleisch, E., & Kowatsch, T. (2020). How are personality states associated with smartphone data? *European Journal of Personality, 34*(5), 687-713. https://doi.org/10.1002/per.2309

Sandstrom, G. M., Lathia, N., Mascolo, C., & Rentfrow, P. J. (2017). Putting mood in context: Using smartphones to examine how people feel in different locations. *Journal of Research in Personality, 69*, 96-101. http://dx.doi.org/10.1016/j.jrp.2016.06.004

Sano, A., Taylor, S., McHill, A. W., Phillips, A. J., & Barger, L. K. (2018). Identifying Objective Physiological Markers and Modifiable Behaviors for Self-Reported Stress and Mental Health Status Using Wearable Sensors and Mobile Phones: Observational Study. J Med Internet Res, 20(6), e210. http://dx.doi.org/10.2196/jmir.9410

Schoedel, R., & Oldemeier, M. (2020). Basic protocol: Smartphone sensing panel study. Leibniz Institut für Psychologische Information und Dokumentation (ZPID). https://doi.org//10.23668/PSYCHARCHIVES.2901

Schoedel, R., Pargent, F., Au, Q., Völkel, S. T., Schuwerk, T., Bühner, M., & Stachl, C. (2020). To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day-night behaviour patterns. *European Journal of Personality, 34*(5), 733-752. https://doi.org/10.1002/per.22

Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology, 109*(5), 872-888. https://doi.org/10.1037/pspp0000036

Sonnentag, S. (2001). Work, recovery activities, and individual well-being: A diary study. *Journal of Occupational Health Psychology, 6*(3), 196-210. https://doi.org/10.1037/1076-8998.6.3.196

Spencer, J. P., & Schöner, G. (2003). Bridging the representational gap in the dynamic systems approach to development. *Developmental Science, 6*(4), 392-412. https://doi.org/10.1111/1467-7687.00295

Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S., ... & Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality, 34*(5), 613-631. https://doi.org/10.31234/osf.io/efnj8

Steinmetz, H., Batzdorfer, V., Scherhag, J., & Bosnjak, M. (2022). *The ZPID Lockdown Measures Dataset for Germany* [Data set]. PsychArchives. https://doi.org/10.23668/PSYCHARCHIVES.6676

Tempelaar, D., Rienties, B., & Nguyen, Q. (2020). Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application. *PLoS ONE, 15(*6), e0233977. https://doi.org/10.1371/journal.pone.0233977

Tsapeli, F., & Musolesi, M. (2015). Investigating causality in human behavior from smartphone sensor data: a quasi-experimental approach. *EPJ Data Science, 4*(1), 24. https://doi.org/10.1140/epjds/s13688-015-0061-1

van Mens, K., de Schepper, C. W. M., Wijnen, B., Koldijk, S. J., Schnack, H., de Looff, P., ... & De Beurs, D. (2020). Predicting future suicidal behaviour in young adults, with different machine learning techniques: A population-based longitudinal study. *Journal of affective disorders, 271*, 169-177. https://doi.org/10.1016/j.jad.2020.03.081

Wagerman, S. A., & Funder, D. C. (2009). Personality psychology of situations. In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (pp. 27-42). Cambridge University Press. https://doi.org/10.1017/CBO9780511596544.005

Wang, W., Harari, G. M., Wang, R., Müller, S. R., Mirjafari, S., Masaba, K., & Campbell, A. T. (2018). Sensing behavioral change over time: Using within-person variability features from Mobile sensing to predict personality traits. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 2*, 1-21. https://doi.org/10.1145/3264951

Whelan, T. J., & DuVernet, A. M. (2015). The big duplicity of big data. *Industrial and Organizational Psychology, 8*(4), 509-515. https://doi.org/10.1017/iop.2015.75

Wright, A. G., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological assessment, 31*(12), 1467-1480. https://doi.org/10.1037/pas0000685

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

Zheng, Y., Plomin, R., & Von Stumm, S. (2016). Heritability of intraindividual mean and variability of positive and negative affect: Genetic analysis of daily affect ratings over a month. *Psychological Science, 27*(12), 1611-1619. https://doi.org/10.1177/0956797616669994

Ziegler, M. (2014). Stop and state your intentions! *European Journal of Psychological Assessment, 30*(4), 239-242. https://doi.org/10.1027/1015-5759/a000228