

# Journal of Personality and Social Psychology

## Snapshots of Daily Life: Situations Investigated Through the Lens of Smartphone Sensing

Ramona Schoedel, Fiona Kunz, Maximilian Bergmann, Florian Bemann, Markus Bühner, and Larissa Sust  
Online First Publication, July 6, 2023. <https://dx.doi.org/10.1037/pspp0000469>

### CITATION

Schoedel, R., Kunz, F., Bergmann, M., Bemann, F., Bühner, M., & Sust, L. (2023, July 6). Snapshots of Daily Life: Situations Investigated Through the Lens of Smartphone Sensing. *Journal of Personality and Social Psychology*. Advance online publication. <https://dx.doi.org/10.1037/pspp0000469>

# Snapshots of Daily Life: Situations Investigated Through the Lens of Smartphone Sensing

Ramona Schoedel<sup>1</sup>, Fiona Kunz<sup>1</sup>, Maximilian Bergmann<sup>1</sup>, Florian Bemann<sup>2</sup>,  
Markus Bühner<sup>1</sup>, and Larissa Sust<sup>1</sup>

<sup>1</sup> Department of Psychology, Ludwig-Maximilians-Universität München

<sup>2</sup> Department of Computer Science, Ludwig-Maximilians-Universität München

Daily life unfolds in a sequence of situational contexts, which are pivotal for explaining people's thoughts, feelings, and behaviors. While situational data were previously difficult to collect, the ubiquity of smartphones now opens up new opportunities for assessing situations in situ, that is, while they occur. Seizing this opportunity, the present study demonstrates how smartphones can help establish associations between the psychological perception and physical reality of situations. We employed an intensive longitudinal sampling design and investigated 9,790 situational snapshots experienced by 455 participants for 14 consecutive days. These snapshots combined self-reported situation characteristics from experience samplings with their corresponding objective cues obtained via smartphone sensing. More precisely, we extracted a total of 1,356 granular cues from different sensing modalities to account for the complexity of real-world situations. We applied linear and nonlinear machine learning algorithms to examine how well these cues predicted the perceived characteristics in terms of the Situational Eight Duty, Intellect, Adversity, Mating, pOstivity, Negativity, Deception, Sociality (DIAMONDS), finding significant out-of-sample predictions for the five dimensions reflecting the situations' Duty, Intellect, Mating, pOstivity, and Sociality. In a series of follow-up analyses, we further explored the data patterns captured by our models, revealing, for example, that those cues related to time and location were particularly informative of the respective situation characteristics. We conclude by interpreting the mapping between cues and characteristics in real-world situations and discussing how smartphone-based situational snapshots may push the boundaries of psychological research on situations.

**Keywords:** psychological situation, situation characteristics, Situational Eight DIAMONDS, smartphone sensing, mobile sensing

Situational contexts play a key role in explaining intraindividual variations in our daily thoughts, feelings, and behaviors that are unaccounted for by personality traits and states (e.g., Fleeson, 2004; Funder, 2001, 2006; Lewin, 1936; Rauthmann, 2021; Sherman et al.,

2015). Yet, situations have traditionally received little attention in personality psychology, in part because their ecologically valid assessment was impeded by a lack of measurement tools, which can now be filled by smartphones capturing situations in situ, that is, while

Ramona Schoedel  <https://orcid.org/0000-0001-7275-0626>

Fiona Kunz  <https://orcid.org/0000-0002-9936-0529>

Maximilian Bergmann  <https://orcid.org/0000-0002-1272-0916>

Florian Bemann  <https://orcid.org/0000-0002-5759-4976>

Markus Bühner  <https://orcid.org/0000-0002-0597-8708>

Larissa Sust  <https://orcid.org/0000-0002-3389-1626>

Preliminary results from this manuscript were presented at the 2021 Biennial Convention of the German Psychology Society for Personality Psychology and Psychological Diagnostics (DPPD).

The authors 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 (SPSS), the authors specially thank their partners for their cooperation, the Leibniz Institute of Psychology (ZPID) and Clemens Stachl. The authors thank the Schuhfried GmbH for providing the Big Five Structure Inventory as digital version. The authors thank Heinrich Hussmann for his many years of support in the PhoneStudy project. The authors thank Florian Pargent for his helpful modeling advices and feedback to the article. The authors thank Sarah Kritzler for inspiring their collection of situation cues. Finally, the authors want to thank Michelle Oldemeier and Katja Frey, who provided study support during data collection.

The SPSS, which is presented in this article is a joint project of Ludwig-

Maximilians-Universität München and the ZPID. ZPID provided most of the funding for the implementation of the described panel study.

All materials, code, and data for this project are available on the Open Science Framework at <https://osf.io/b7krz/>. Raw smartphone sensing data cannot be made public due to their privacy-sensitive nature. But aggregated data is available via <https://doi.org/10.23668/psycharchives.12706>. Pre-registration for parts of this study is available via <https://doi.org/10.23668/psycharchives.4928>.

Ramona Schoedel played a lead role in conceptualization, data curation, investigation, project administration, visualization, writing—original draft, and writing—review and editing and an equal role in formal analysis, methodology, and software. Fiona Kunz played a supporting role in data curation and writing—original draft and an equal role in formal analysis, methodology, software, and validation. Maximilian Bergmann played a supporting role in methodology and validation. Florian Bemann played a lead role in resources and software. Markus Bühner played a lead role in funding acquisition and a supporting role in supervision. Larissa Sust played a lead role in writing—original draft, writing—review and editing; and an equal role in conceptualization, investigation and supervision.

Correspondence concerning this article should be addressed to Ramona Schoedel, Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstraße 13, 80802 Munich, Germany. Email: [schoedela@gmail.com](mailto:schoedela@gmail.com)

people experience them in a natural context. These devices accompany their users in nearly every instance of life and bear witness to a large spectrum of situations they encounter. For example, smartphones are by their users' side when they wake up in the morning, when they are on the train to visit an old friend, or when they enjoy a free afternoon in their favorite park. Given that about 66% of the global population owns a smartphone (Kemp, 2022) and that smartphones are in the same room as their owners 90% of the time (Dey et al., 2011), these small supercomputers can potentially become the perfect tool for investigating daily situations (Harari et al., 2015; Harari, Müller, & Gosling, 2020; Wrzus & Mehl, 2015, 2020).

Thereby, smartphones allow for collecting various types of situational information spanning the psychological experience and objective reality of situations by combining active logging with passive sensing (Wrzus & Mehl, 2015). While the advent of the experience sampling (ES) method has already added momentum to the investigation of the subjective experience of daily situations (e.g., Horstmann et al., 2021; Sherman et al., 2015), the implementation of passive sensing for the empirical assessment of objective situational parameters is still pending. Because of this imbalance in situation research, we still know little about the objective parameters and their role in the psychological experience of situations (Rauthmann et al., 2014). However, investigating the association between situations' objective reality and psychological experience is a necessary step to fully understand why situations are perceived in a certain way, which, in turn, lays the groundwork for uncovering intra- and interindividual differences in situation perception that give rise to other psychological outcomes such as feelings or behavior (Funder, 2016). Our study addresses this gap in research and demonstrates how smartphones can be used to relate objective parameters to the psychological experience of situations in daily life.

### Conceptualizing the Psychological Situation

The psychological definition of the term "situation" has been a point of contention for the longest time as researchers from social and personality psychology disagreed on whether to view situations as objective or subjective phenomena, that is, as physical realities or idiosyncratic perceptions (Hogan, 2009; Magnusson, 1981; Rauthmann, Sherman, & Funder, 2015; Reis, 2008). Only in the past decade, the field has witnessed a resurgence of theoretical interest in situations that resolved this controversy and provided clarity on the conceptualization of situations (Funder, 2016; Parrigon et al., 2017; Rauthmann, 2012; Rauthmann et al., 2014; Rauthmann, Sherman, & Funder, 2015; Ziegler et al., 2019).

Summarizing the new status quo of situation research, Rauthmann and Sherman (2021) concluded that situations can be defined on two levels, each representing a different type of information: Situations comprise objectively quantifiable stimuli, the so-called *situation cues* (e.g., the ringing of an alarm, people present, the flora in a park), which may be perceived by individuals, resulting in psychological representations of the so-called *situation characteristics* (e.g., adverse, sociable, pleasant, Block & Block, 1981; Fleeson, 2007; Funder, 2016; Magnusson, 1981; Mischel & Shoda, 1995; Murray, 1938; Nystedt, 1981; Rauthmann et al., 2014; Rauthmann, Sherman, & Funder, 2015; Rauthmann & Sherman, 2021). More precisely, individuals automatically filter, evaluate, and interpret situation cues through an evolutionary optimized perceptual process (Buss, 2009;

Edwards & Templeton, 2005; Miller, 2007), which is guided by stable and fluctuating person parameters (i.e., traits and states, Mischel & Shoda, 1995; Nystedt, 1981; Rauthmann et al., 2014). Thereby, individuals efficiently generate condensed representations of a situation's characteristics that ultimately guide their mental processes and behaviors and, in turn, potentially influence their physical reality and situation cues (Buss, 1987; Endler, 1981; Lewin, 1936; Magnusson, 1981; Mischel & Shoda, 1995; Murray, 1938; Rauthmann, 2021).

### Situation Cues

Situation cues represent the raw sensory information in the individual's physical environment (Block & Block, 1981; Murray, 1938; Rauthmann et al., 2014). Because there is a myriad of cues available in any given situation, researchers commonly categorize them via five W-questions (Harari, Müller, & Gosling, 2020; Mehl & Robbins, 2012; Pervin, 1978; Saucier et al., 2007), distinguishing (a) *interactions and persons* (i.e., Who is with you? Are you interacting with them?), (b) *objects* (i.e., Which objects are around you?), (c) *activities and events* (i.e., What is happening? What are you doing?), (d) *locations* (i.e., Where is it happening?), and (e) *time* (i.e., When is it happening?).

Traditionally, (experimental) changes in the objective nature of situations have been a central paradigm in social psychology when investigating factors that uniformly—regardless of person variables— influence psychological outcomes (see Funder, 2016; Reis, 2008; Richard et al., 2003). More recently, situation cues have also received attention in personality psychology when assessing person–environment transactions. For example, researchers successfully related people's locations and surrounding persons to their personality traits (Mehl et al., 2006) and states (Matz & Harari, 2021; Wilt & Revelle, 2019; Wrzus et al., 2016). While these studies provide interesting insights, their explanatory power is limited because, as pointed out above, situation cues only lay the environmental foundation for psychological situation representations (e.g., Block & Block, 1981; Mischel & Shoda, 1995; Rauthmann et al., 2014). Thus, perceived situation characteristics add complementary value as another interesting unit of analysis for personality research (Rauthmann, Sherman, & Funder, 2015; Rauthmann & Sherman, 2021).

### Situation Characteristics

Situation characteristics subsume the psychological meaning attributed to objective situation cues (Block & Block, 1981; Magnusson, 1981; Mischel & Shoda, 1995; Murray, 1938; Rauthmann, Sherman, & Funder, 2015). These perceptions can be described on continuous dimensions indicating the extent to which a psychological characteristic applies to a situation (de Raad, 2004; Edwards & Templeton, 2005). Accordingly, several taxonomies for organizing situation characteristics have emerged, differing in the number and labels of dimensions (e.g., Brown et al., 2015; Gerpott et al., 2018; Parrigon et al., 2017; Rauthmann et al., 2014; Ziegler et al., 2019). While the proposed dimensions exhibit conceptual overlap, the systematic integration of situation taxonomies is still pending (Horstmann et al., 2018; Rauthmann & Sherman, 2018a).

One of the prevailing taxonomies for describing broad, everyday situations was proposed by Rauthmann et al. (2014; Horstmann

et al., 2018), who reduced the comprehensive collection of situational descriptions provided by the Riverside Situational Q-Sort (Funder, 2016; Wagerman & Funder, 2009) to eight underlying dimensions. The Situational Eight Duty, Intellect, Adversity, Mating, pOsitivity, Negativity, Deception, Sociality (DIAMONDS) capture situations' perceived (a) *Duty* (i.e., Does something need to be done?), (b) *Intellect* (i.e., Is deep thinking required or desired?), (c) *Adversity* (i.e., Are there external threats?), (d) *Mating* (i.e., Is the situation sexually or romantically charged?), (e) *pOsitivity* (i.e., Is the situation enjoyable?), (f) *Negativity* (i.e., Does the situation elicit unpleasant feelings?), (g) *Deception* (i.e., Is someone being untruthful or dishonest?), and (h) *Sociality* (i.e., Are social interaction and relationship formation possible, required, or desired?; Rauthmann et al., 2014).

The DIAMONDS have corresponding dimensions in existing taxonomies for personality traits (Rauthmann et al., 2014) and mood states (Horstmann & Ziegler, 2019) and, thus, provide fertile ground for empirical research on the interplay between persons and situations (e.g., Fleeson, 2007). For example, researchers found that those high in the trait Openness experience more intellectual situations (e.g., Abrahams et al., 2021; Horstmann et al., 2021; Rauthmann et al., 2014; Sherman et al., 2015) and that those currently in a more positive affective state experience more situations high in pOsitivity and Sociality (Horstmann et al., 2021; Horstmann & Ziegler, 2019; Kritzler et al., 2020). All these person-situation associations may arise because people select, evoke, modify, create, or construe situations in personality- and mood-congruent ways (Buss, 1987; Caspi & Roberts, 2001; Horstmann & Ziegler, 2019; Ickes et al., 1997; Rauthmann, 2021; Rauthmann, Sherman, & Nave, 2015; Rauthmann & Sherman, 2016b). Alternatively, their daily situations may also elicit mood states (Griner & Smith, 2000; Horstmann & Ziegler, 2019; Kuppens, 2009) and, in the long run, shape personality traits by providing opportunities for traits to manifest and be reinforced (Fleeson, 2007; Ickes et al., 1997; Rauthmann, 2021).

### ***Situation Cues as Basis of Situation Characteristics***

As noted above, situation cues provide the objective situational information from which the psychological *experience* in terms of situation characteristics arises (e.g., Block & Block, 1981; Mischel & Shoda, 1995; Rauthmann et al., 2014). Therefore, an individual's situation representation contains idiosyncratic variance attributed to their perception process (i.e., the perceiver effect) and consensual variance attributed to the situation cues (i.e., the situation effect, e.g., Block & Block, 1981; Rauthmann, 2012; Rauthmann, Sherman, & Funder, 2015; Rauthmann & Sherman, 2021; Serfass & Sherman, 2013).

The idiosyncratic variance represents the subjective meaning of a situation, that is, how persons uniquely construe its situation cues. These nonshared situation interpretations may arise because individuals perceive cues that others miss or interpret these cues differently (Rauthmann, Sherman, & Funder, 2015). The resulting individual differences in situation perceptions may be related to persistent and momentary person factors such as personality traits or states (Mischel & Shoda, 1995; Nystedt, 1981; Rauthmann, 2012).

The consensual variance represents the normative meaning of a situation, that is, how situation cues are generally rated on situation characteristics across individuals (Rauthmann, 2012; Rauthmann,

Sherman, & Funder, 2015). These shared interpretations of the same situation emerge because (healthy) individuals perceive the same physically present external reality and interpret it via common mental schemata (e.g., cultural norms) for situations (Block & Block, 1981; Kenny, 1988; Rauthmann, Sherman, & Funder, 2015). Consequently, perceptions should exhibit considerable overlap between persons (Rauthmann, Sherman, & Funder, 2015; Sherman et al., 2015; Wagerman & Funder, 2009).

Thus, despite lacking intrinsic psychological meaning, situation cues should (at least to some degree) be indicative of psychological situation representations due to their consensual variance. Accordingly, some studies have started investigating the foundation of situation perceptions by mapping the situation cues and perceived characteristics of a given situation. Exemplary findings for some of the cue groups introduced above are that situations involving communication (i.e., interactions and persons) were reported to contain higher levels of the DIAMONDS dimensions of Intellect, pOsitivity, Deception, and Sociality (Breil et al., 2019; Rauthmann et al., 2014; Rauthmann & Sherman, 2016a). Furthermore, situations involving eating or drinking (i.e., activities and events) were rated as lower in Duty and Negativity and higher in Mating, pOsitivity, and Sociality (Breil et al., 2019; Rauthmann & Sherman, 2016a). Finally, situations occurring at home (i.e., locations) were found to contain lower levels of Mating, Deception, and Sociality, while situations occurring at the office or at university were associated with higher levels of Duty, Intellect, and Negativity (Blake et al., 2020; Rauthmann et al., 2014; Rauthmann & Sherman, 2016a).

These studies provided valuable insights, but only for a limited and potentially biased selection of situation cues. That is because a typical procedure for assessing situation cues was coding them from situation descriptions provided by participants in an open-ended response format (e.g., Breil et al., 2019; Rauthmann et al., 2014). In such unguided descriptions, however, participants could only report on aspects of the situation that they were consciously aware of or remembered and that they judged relevant enough to report. Thereby, perceptual limitations (e.g., attention, memory) and motivational factors may have biased participants' descriptions, and, thus, the selection of cues studied in the past. While Blake et al. (2020) recently assessed situation cues in a more direct manner via automatic object recognition from wearable camera recordings, they later had to limit their cue selection due to statistical modeling choices. As a consequence, situation research today still lacks the simultaneous and systematic assessment of various situation cues and their constellations and, ultimately, their comprehensive mapping to perceived characteristics for situations encountered in everyday life.

### **Situation Assessment via Smartphones**

Because situation cues and characteristics change dynamically throughout the day, their ecologically valid assessment presents a methodological challenge. While past studies investigated situations with retrospective surveys (e.g., Horstmann & Ziegler, 2019; Rauthmann, Sherman, & Nave, 2015) or laboratory experiments (e.g., Morse et al., 2015), situational information can now also be collected in situ by combining two types of ambulatory assessment via smartphones.

## Sampling Situations

Smartphones are currently the most used device for administering the ES methodology, which has gained considerable importance in psychological assessment over the last decades (Conner & Mehl, 2015; Larson & Csikszentmihalyi, 2014; van Berkel et al., 2017). When applying ES, situation researchers can use smartphones to present short questionnaires to repeatedly ask participants about their current situation's characteristics, for example, in terms of the DIAMONDS dimensions (e.g., Abrahams et al., 2021; Breil et al., 2019; Horstmann et al., 2021; Kritzler et al., 2020). Such self-reports are a useful tool to assess the psychological experience across different situations at different times throughout the day (Rauthmann, Sherman, & Funder, 2015).

Researchers have also applied ES to collect in situ self-reports of situation cues, either indirectly via open situation descriptions (Breil et al., 2019) or directly via closed questions on specific cues (Matz & Harari, 2021; Wilt & Revelle, 2019; Wrzus et al., 2016). While participants' free reports are biased in their selection of situation cues, as discussed above, closed questions can also not collect comprehensive information across all five groups of situation cues. In particular, due to the short scope of ESs, researchers typically had to limit their study to one or two cue groups and present a preselected number of specific cues as response options. For example, both Wrzus et al. (2016) and Wilt and Revelle (2019) assessed only participants' company (i.e., interactions and persons) and activity (i.e., activities and events) in their ES studies on objective situations. Thus, ES is not well suited for the systematic and wholesome collection of situation cues and their constellations. However, today's technologies are creating new opportunities to collect a broader range of situation cues, in a more objective manner and without burdening participants, through automatic measurements, for example, with recordings from wearable cameras (see Blake et al., 2020; Brown et al., 2017) or microphones (see Mehl et al., 2006) or, more conveniently, via smartphone sensing.

## Sensing Situations

Smartphone sensing refers to the collection of smartphone usage data via specifically developed research applications (short: apps). These sensing apps can access the system logs (e.g., call and screen logs) and native sensors (e.g., Global Positioning System [GPS], accelerometers) embedded in a user's smartphone and unobtrusively collect situational data over longer periods of time. The sensed data cover a wide range of modalities and all five groups of situation cues (Harari et al., 2015; Harari, Müller, & Gosling, 2020). For example, interactions can be inferred from call records (e.g., Harari, Müller, & Stachl, 2020; Servia-Rodríguez et al., 2017; Wang et al., 2016), objects from music listening records (e.g., Sust et al., 2022; Yang & Teng, 2015), (physical) activities from accelerometer sensors (e.g., Servia-Rodríguez et al., 2017; Wang et al., 2016; see, Ramanujam et al., 2021, for a review of human activity recognition), locations from GPS sensors (e.g., Canzian & Musolesi, 2015; Do & Gatica-Perez, 2014; Müller et al., 2020), and time frames from the timestamps of a given situation (Böhmer et al., 2011; Stachl et al., 2020).

The previous literature conceptualized situation cues as short-term aspects of the situation (unfolding within minutes) that are highly concrete, therefore salient, and thus likely processed by the human perception system (Rauthmann, 2021; Rauthmann, Sherman, & Funder,

2015; Rauthmann & Sherman, 2021). That means situation cues represent information about the situation that individuals in situ are probably aware of, such as *being at the workplace* or *driving a car*. Smartphones can sense these salient situation cues in an automated fashion, but they also provide the opportunity to go one step further and assess cues that are less salient and, therefore, less likely perceived by individuals in a given situation. These more subtle cues, however, may still play some role in forming psychological situation representations and may, thus, also contribute objective situational information related to perceived situation characteristics.

More specifically, smartphones provide access to two additional types of cues: smartphone and environment cues. Smartphone cues represent short-to-medium term aspects of the situation (spanning minutes to hours) that are concrete but less salient than classic situation cues because they require some complex reflection or aggregation that individuals could only approximate if explicitly asked to. They comprise rather computational features such as the *distance to the workplace* exhibited in a given situation. Finally, environment cues are stable aspects of the situation (spanning weeks to months) that are very abstract (Rauthmann & Sherman, 2021). These cues cover information that is hardly salient and that individuals in situ are most likely unaware of, such as the *number of inhabitants of the city* in which a situation takes place. However, because the consequences of this information (e.g., being in a crowded area) are perceptible, environment cues may still depict relevant information for the given situation perception.

To summarize, smartphones allow for capturing various cues in their momentary constellations and can therefore paint a comprehensive and objective picture of a situation's objective reality. Even though "the value of [smartphone] sensing methods for research on situations is derived from the ability to assess cues unobtrusively and continuously" (Harari, Müller, & Stachl, 2020, p. 15) and despite researchers repeatedly pointing out the method's advantages in terms of ecological validity (Breil et al., 2019; Harari et al., 2015; Harari, Müller, & Gosling, 2020; Wrzus & Mehl, 2015, 2020), the application of smartphone sensing in situation research is still pending.

## The Present Study

In an exploratory study, we investigated whether smartphone sensing can lift psychological situation research to a new level of information density—as repeatedly predicted by scholars. We applied an intensive longitudinal sampling design and collected 9,790 situational snapshots of 455 participants over 14 consecutive study days. In a multimethod approach, we assessed experience-sampled situation characteristics and smartphone-sensed situation cues. In doing so, our study aimed to draw a comprehensive picture of the mapping of raw objective cues to perceived characteristics of situations experienced in everyday life.

To account for the extensive scope and complex nature of smartphone sensing data, our preregistered exploratory analysis followed a data-driven approach and explored *how much* information the objective situational data from smartphones generally contain about psychological situation representations. We extracted a set of 1,356 objective and granular cues from various sensing modalities to accommodate the diversity of real-life situations. We applied linear and nonlinear machine learning algorithms and evaluated out-of-sample prediction performances to investigate how well our range of

situation cues predicts situation characteristics in terms of the Situational Eight DIAMONDS.

In a follow-up second analysis, we explored *what* kinds of objective situational information in smartphone sensing data are relevant for predicting perceived situation characteristics. We went beyond previous literature and investigated the relevance of the different subgroups of situation cues (interaction, objects, activities, location, time) but also of smartphone cues and environment cues for each specific dimension of situation experience. We applied interpretable machine learning techniques and analyzed the importance of individual cues and their constellations as groups. For a comprehensive mapping, we additionally explored whether different situation characteristics were identified by unique or shared patterns of cues.

## Method

We analyzed data collected in the SSPS, an interdisciplinary research project conducted by Ludwig-Maximilians-Universität München (LMU Munich) in cooperation with the Leibniz Institute for Psychology (Schoedel & Oldemeier, 2020). The SSPS obtained data via three modalities, namely online surveys, ES, and smartphone sensing, from a quota sample representative of the German population.<sup>1</sup> All procedures received approval from the ethics committee of the psychology department at LMU Munich under the study title “A longitudinal panel study combining smartphone sensing and survey methods.” Furthermore, all procedures adhered to the General Data Protection Regulation.

## Transparency and Openness

In this article, we focus our report on the procedures and measures of the SSPS that are relevant to our current research questions. A detailed description of the overall project is available in our preregistered study protocol at <https://doi.org/10.23668/psycharchives.2901>. We report how we determined our sample size, all data exclusions, and all measures of this study and we follow the Journal Article Reporting Standards (Appelbaum et al., 2018).

Parts of our exploratory study’s analysis plan were preregistered before conducting any data preprocessing and analyses under <https://doi.org/10.23668/psycharchives.4928>. Note that we made some changes to accommodate practical hurdles encountered during data preprocessing (e.g., to account for unforeseen data structures) and extended the preregistered analyses in several ways (e.g., by adding validation checks and comparisons of cue importance). We communicate all deviations from the preregistration in detail in our comprehensive codebook, which is available in our project’s Open Science Framework (OSF) repository under <https://osf.io/b7krz/>. The OSF project additionally contains our online supplemental materials and the code for data preprocessing and data analyses, which we conducted in the statistical software R (Version 4.1.2 for preprocessing and Version 4.2.1 for data analysis; R Core Team, 2022). For reproducibility, we used the package management tool *renv* (Ushey, 2022), and we provide a complete list of all R packages used in this article in the *renv.lock* file in the OSF project. While the privacy-sensitive nature of the smartphone data prevents us from sharing the raw logging data, we provide a data set of aggregated variables under <https://doi.org/10.23668/psycharchives.12706>.

## Procedure

With the help of a nonprobability online panel provider, we recruited an initial sample of 850 participants located across Germany. We chose sampling quotas to represent the German population with respect to gender, age, education, income, religion, and relationship status. Participants were required to be between 18 and 65 years old, to be fluent in German, and, for technical reasons, to be the sole user of a smartphone running on Android Version 5 or higher. After recruitment, we randomly assigned participants to one of two groups with a study duration of three ( $n = 191$ ) or 6 months ( $n = 659$ ) for practical reasons. Participation was compensated monetarily with up to EUR 131.50 depending on the number of completed study parts.

Data collection started in May 2020 and took place simultaneously for all participants. Figure 1 illustrates the full timeline of the study. We asked participants to install our self-developed Android-based mobile sensing app PhoneStudy<sup>2</sup> on their private smartphones, which continuously collected various data in the background for the respective study duration. Each month, the app sent participants a link to a 30-min online survey. Furthermore, the SSPS included two 14-day ES waves. During these ES waves, we asked participants to complete short 5-min questionnaires on two to four occasions per day. We pseudorandomized the schedule of the ES questionnaires via the following rationale: Each day (from 7 am to 10 pm on weekdays and from 9 am to 11 pm on weekends) was divided into four equally sized sections, and within these sections, the timing of the ES was chosen randomly while maintaining a minimum interval of 60 min between two consecutive questionnaires. As soon as participants actively used their smartphone for the first time after the selected point in time, the app informed them about the ES questionnaire via a notification. We chose this procedure to increase participants’ commitment and not provoke artificial smartphone usage events distorting the naturally occurring sensing data (van Berkel et al., 2019).

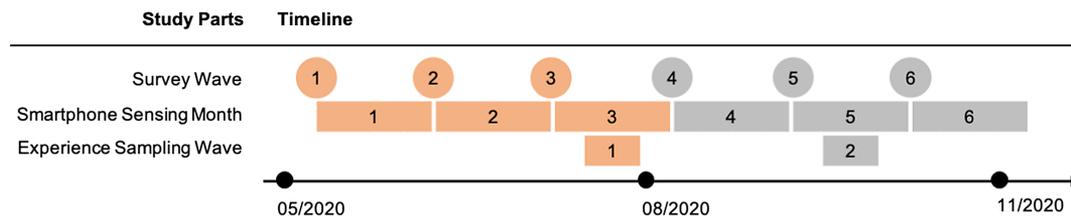
## Sample

In the present study, we used the questionnaire data from Survey 1 (May 2020; demographics), Survey 2 (June 2020, Big Five personality traits), Survey 3 (July 2020, mood traits), as well as the ES data (Situational Eight DIAMONDS, mood states) and the sensing data collected during the first ES wave (07/27/2020–08/09/2020). These parts of the SSPS exhibited a sample size of  $N = 455$  after applying several exclusion criteria (see Table S1, in our online supplemental materials), which we applied to reduce the risk of including fake participants (i.e., participants who installed the app, but then (un)intentionally did not participate in the remaining study parts). From our final sample, 431 participants provided their demographic information: Age ranged between 18 and 65, with an average of 41 years ( $SD = 12.2$ ), 45.0% ( $n = 194$ ) of participants indicated to be female and 55.0% to be male ( $n = 237$ ). Figure 2 provides an overview of further self-reported socioeconomic demographics (Panel A), home cities of participants (Panel B; detected via GPS data), and details on

<sup>1</sup> A small part of this data set has already been published in Schoedel et al. (2022). This methodological article introduces a categorization of smartphone apps that we also use in the present study (see the section *Smartphone Sensing Measures* in the following Method section) and contains some summary statistics on app usage.

<sup>2</sup> <https://phonestudy.org/en/>.

**Figure 1**  
*Timeline of the Smartphone Sensing Panel Study*



*Note.* The study started simultaneously for all participants in May 2020. After recruitment, participants were randomly assigned to one of two groups with a study duration of three or six months. The 3-month group completed only the orange-shaded parts of the study, and the 6-month group completed the orange- and gray-shaded parts. See the online article for the color version of this figure.

participants' smartphones, including manufacturers and Android versions (Panel C).

### COVID-19 Restrictions

Because the SSPS was conducted during the COVID-19 pandemic, we checked the containment measures implemented by the German government during the data collection period investigated in this study. We aimed to get an impression of the possible implications of restrictions in everyday life on the collected situational data. In a descriptive analysis, we inspected two composite measures subsuming different restriction indicators, such as school closures or travel bans: the country-level Stringency Index (Hale et al., 2021) and a self-created state-level index calculated from an open-access data set by Steinmetz et al. (2022). Both indices show that restrictions in Germany were relatively loose during our study period in the summer of 2020, compared to the onset and later (winter) stages of the pandemic (see Figure A1 and also the dashboard presentation by Mathieu et al., 2020). However, the Stringency Index—with a range of 0 (*no restriction*) to 100 (*full restriction*)—still exhibited values from 55 to 57 in July 2020, which are far higher than the value of 15 obtained in December 2022 when life was back to the “new normal.” Thus, we cannot rule out the possibility that the COVID-19 restrictions in place still affected daily life—and the situations encountered by participants—during our study. For example, social-distancing rules such as *keeping a minimum distance of 1.5 m from others* or *meeting only a limited number of persons at the same time in public space* (see Steinmetz et al., 2022) might have affected socializing in daily life. Moreover, in Germany, the number of employees working remotely has risen sharply due to the pandemic (Destatis, 2023), which, in turn, might be linked to changes in everyday routines and mobility.

Because our participants were recruited across the country, we also explored whether governmental restrictions differed within Germany (i.e., between its 16 federal states). We created a daily average score of different restriction measures provided by Steinmetz et al. (2022) and found minimal variation across all federal states and study days (see Figure A1). Small state-level differences resulted, for example, from varying levels of strictness in rules for wearing masks and visiting gastronomy (i.e., restaurants/cafés were open but had different requirements for negative test results or the number of people admitted; see Steinmetz et al., 2022). In sum, while containment measures potentially influenced daily life and the situations encountered, they did so similarly for all participants in our sample. Therefore, we did not

consider COVID-19-related restrictions further in our data analysis, but we interpreted our findings against the pandemic background in the Discussion section.

### Self-Report Measures

#### Survey Measure

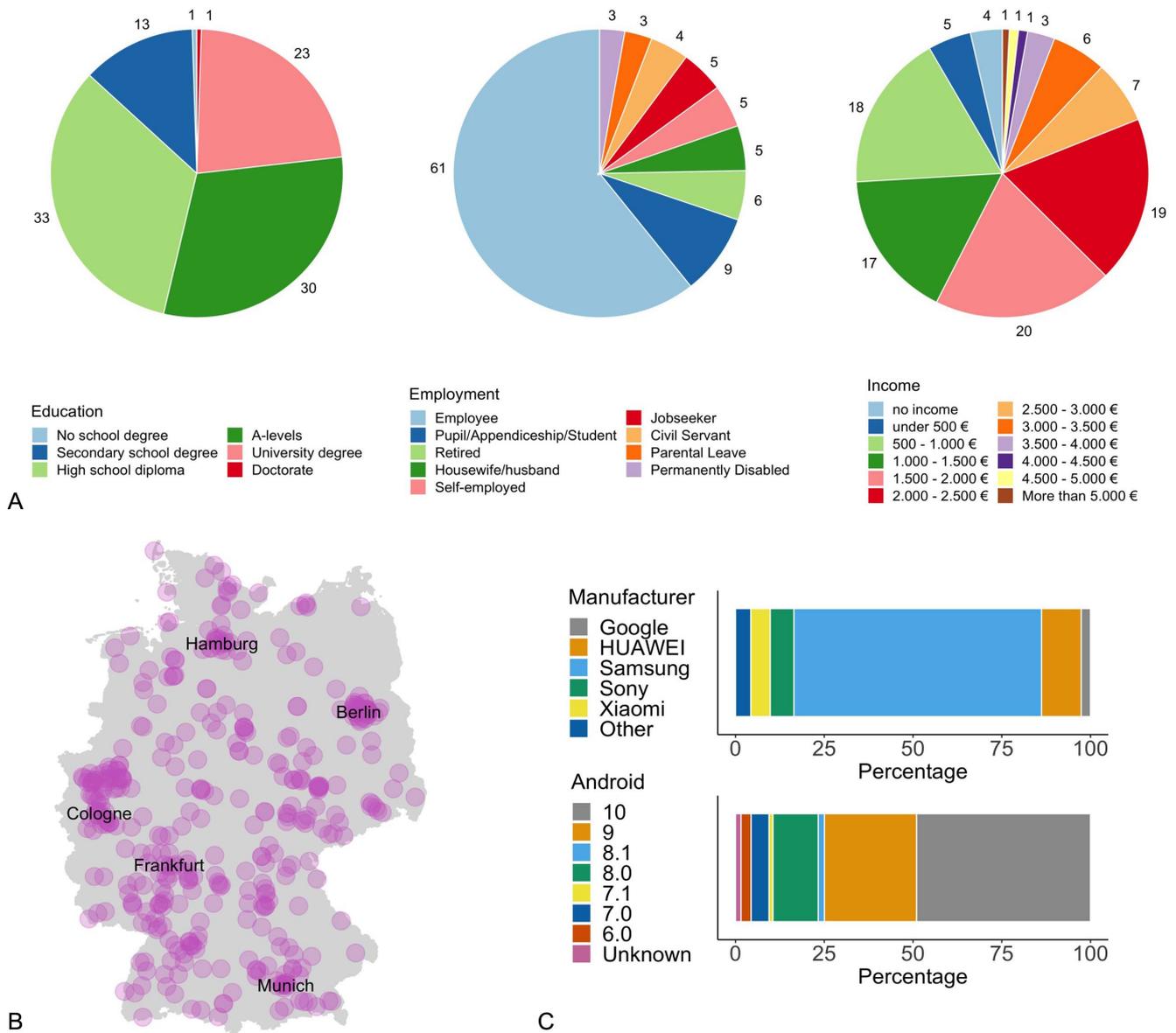
**Personality Traits.** We applied the Big Five Structure Inventory (BFSI; Arendasy, 2009) to assess participants' Big Five personality traits. The questionnaire comprises 300 items (personality describing adjectives and short phrases), which are rated on a 4-point Likert scale (0 = *untypical for me* to 3 = *typical for me*). Because the construction of the BFSI follows item response theory, we used participants' person parameters as estimates of their latent trait values (instead of sum scores) in our analyses.

#### ES Measures

**Situation Characteristics.** We used the German version of the S8-I scale to assess participants' situation experience in terms of the Situational Eight DIAMONDS (Rauthmann & Sherman, 2016c, 2018a). The S8-I comprises one item for each dimension and exhibits sufficient convergent and discriminant validity (Rauthmann & Sherman, 2016c). Thus, the ultra-brief measure ensured the economic in situ assessment of situations in our ES design. To further reduce the participants' burden, we deviated from the measure's original 7-point Likert scale and asked participants to indicate on a binary scale (0 = *does not apply*, 1 = *applies*) whether the respective dimension applied to their current situation.

**Mood States.** In line with the previous ES studies, we applied a single-item measure to assess the valence of participants' mood states (Kushlev & Heintzelman, 2018). According to the dimensional approach to affect, valence can be modeled on a bipolar scale ranging from displeasure to pleasure (Russell, 1980). Thus, we asked participants to indicate their current mood on a 6-point Likert scale (1 = *very unpleasant* to 6 = *very pleasant*). As a validity check for this single-item measure, we inspected Pearson correlations between the person-mean score of the item aggregated across ES questionnaires and trait affect scores assessed via the German version of the Positive and Negative Schedule (Janke & Glöckner-Rist, 2012; Watson et al., 1988). We found a high positive association ( $r = 0.53$ ) with the Positive

**Figure 2**  
*Characteristics of the Study Sample*



*Note.* Panel A displays the composition of the sample in terms of education, employment, and net income. Numbers in the margin represent percentages (%). These survey variables were only available for a subsample of  $n = 397$ . Panel B displays the detected home places throughout Germany based on participants' GPS data. Home places of four participants, which were located outside of Germany, were removed for this figure. The five largest German cities are marked as landmarks. Panel C displays the distribution of smartphone manufacturers and Android versions across participants. GPS = Global Positioning System. See the online article for the color version of this figure.

Affect Scale and a medium-to-high negative association ( $r = -0.44$ ) with the Negative Affect Scale.

### Smartphone Sensing Measures

Our app recorded many different types of smartphone sensing data providing information about participants' daily situations. Based on the data type availability, we defined a variety of smartphone sensing variables, which we refer to as *cues* or—to follow machine learning

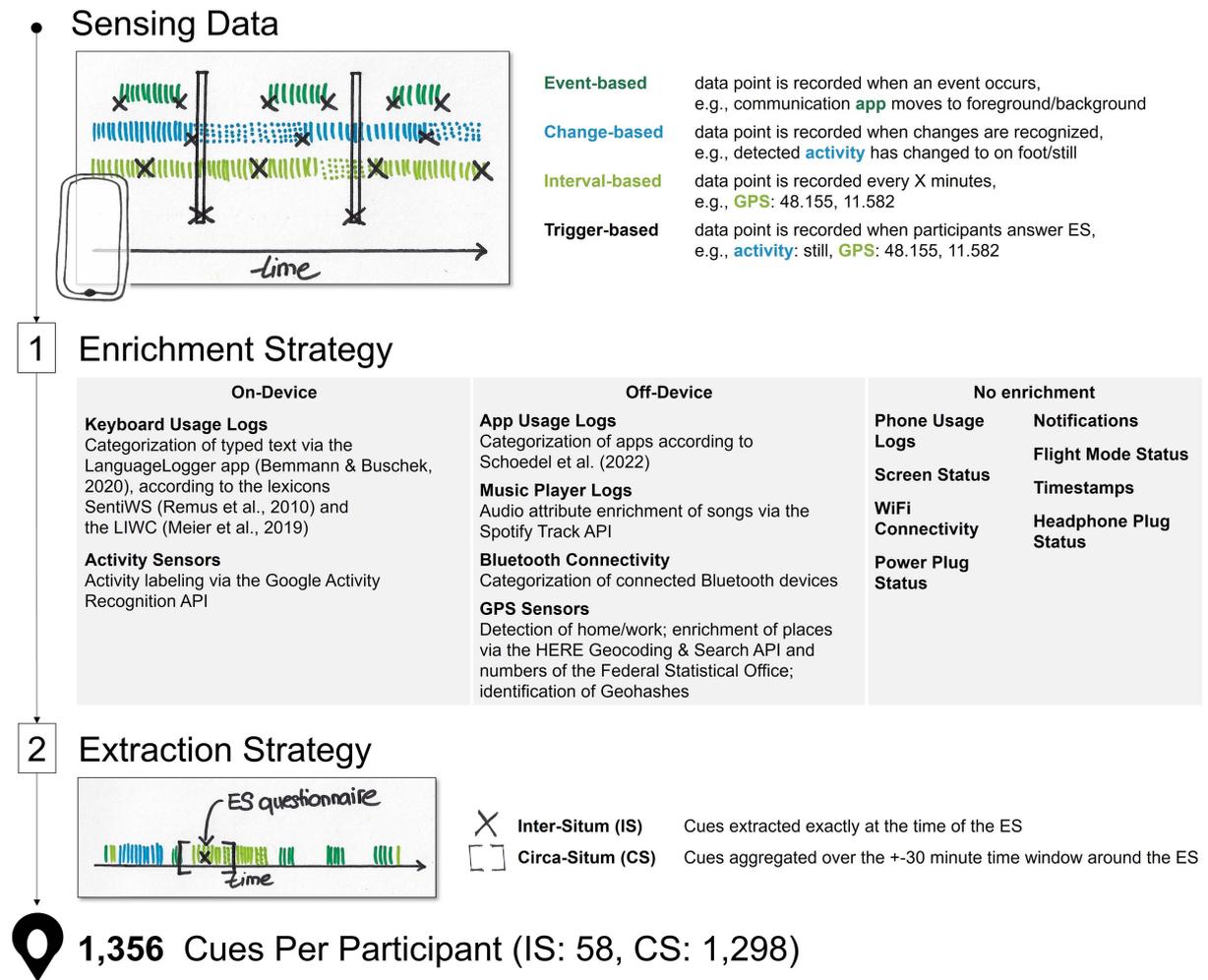
terminology (see the Data Analyses)—*features*. In total, we extracted 1,356 cues by applying a complex preprocessing workflow, described in Figure 3 and the following sections.

### Sensing Data as Starting Point

The PhoneStudy app logged raw sensing data as time-stamped data points and, depending on the respective data type, stored these data points with different specifications (e.g., app usage logs with

**Figure 3**

Preprocessing Workflow for Extracting Cues From Smartphone Sensing Data



*Note.* Different types of sensing data were collected via different logging modes and then enriched for further preprocessing as defined by our (1) Enrichment Strategy. Next, we applied our (2) Extraction Strategy to extract features inter-situm (IS) and circa-situm (CS) in relation to the experience sampling questionnaires (ES). API = application programming interface; LIWC = Linguistic Inquiry and Word Count. HERE = name of the online geodata service we used to enrich our data; GPS = Global Positioning System. See the online article for the color version of this figure.

the app name; GPS sensors with longitude and latitude). Thereby, the PhoneStudy app applied different logging modes (see the upper part of Figure 3): Data resulting from user-smartphone interactions (i.e., phone usage, app usage, keyboard usage, music player usage, notifications, screen status, flight mode status, Bluetooth connectivity, Wi-Fi connectivity, and power plug status) were logged in an *event-based* manner. This means the app recorded data points whenever they occurred (e.g., creating a screen status data point when the user turned on the screen). For collecting *context-based* data, our app combined three different logging modes to get the most accurate picture of the user's context while conserving battery power: (a) GPS data were logged *interval-based*, meaning that GPS data points were recorded at fixed time intervals (every 10–60 min, depending on the smartphone model); (b) physical activities, GPS data, and headphones plug status were logged *change-based* (i.e., whenever these parameters changed) via the Google Fence application programming interface (API)<sup>3</sup>; and (c) *trigger-based*,

that is, at the exact point in time when the participant opened an ES questionnaire via the Google Snapshot API.<sup>4</sup>

### Enrichment Strategy

For most types of sensing data, the raw data points were directly interpretable and made for meaningful cues. For example, phone usage logs contained information about whether calls were (a) outgoing, (b) incoming, (c) missed, or (d) rejected, and screen logs informed us whether the screen was (a) on versus off and (b) locked versus unlocked. However, several data types were not inherently meaningful. Thus, we enriched their raw sensing data with further information from external sources. Thereby, we distinguish between

<sup>3</sup> <https://developers.google.com/awareness/android-api/fence-api-overview>.

<sup>4</sup> <https://developers.google.com/awareness/android-api/snapshot-api-overview>.

(a) on-device enrichment (i.e., data were enriched directly while being logged on the device) and (b) off-device enrichment (i.e., data were enriched after data collection during the preprocessing).

#### **On-Device.**

**Keyboard Usage Logs.** We integrated the app LanguageLogger into the PhoneStudy app (Bemmann & Buschek, 2020). LanguageLogger provides several options for the real-time processing of typed text. Thereby, written language is abstracted on the fly in a privacy-friendly manner so there is no need to store raw text data. We extracted meta statistics of keyboard sessions (e.g., the number of submitted characters) and used two dictionaries of the LanguageLogger app's word categorization module: (a) The German-language affect dictionary Sentiment Wortschatz (SentiWS) served to extract sentiment scores ranging from very negative to very positive (e.g., the typed word joy was logged with its sentiment score of 0.65; Remus et al., 2010) and (b) the latest German version of the Linguistic Inquiry and Word Count (LIWC) served to extract psychologically meaningful word categories (e.g., the typed word joy was logged with its assigned LIWC categories affective processes and positive emotion; Meier et al., 2019; Tausczik & Pennebaker, 2010). The app extracted categorization results on a word-by-word basis per keyboard session (e.g., per text message in communication apps; per search query in internet apps) and stored them along with the app name in which the keyboard use occurred (e.g., in WhatsApp). For more technical details regarding the LanguageLogger app, please refer to the Bemmann and Buschek (2020).

**Activity Sensors.** We recorded participants' physical activities with the Google Activity Recognition API.<sup>5</sup> This API processes data from various smartphone sensors, such as the accelerometer or gyroscope, in real-time during logging and provides a list of detected activities sorted by probability. These activities include being still (i.e., not moving), in a vehicle (such as a car or a train), on a bicycle, on foot (running and walking), and an unknown category.

#### **Off-Device.**

**App Usage Logs.** We recorded participants' raw app usage logs containing the names of the app used in each event which, however, were not always informative of the app's functionality. Thus, we subsequently annotated the logged apps with psychologically meaningful labels based on a category system by Schoedel et al. (2022). Thereby, we excluded the category system apps because it contained only background apps while our focus was on active, user-initiated app use. Furthermore, we excluded the category spirituality apps because of its low interrater agreement of Cohen's  $\kappa < .60$  (Schoedel et al., 2022). Ultimately, we represented app usage sessions in terms of 24 categories (e.g., communication, games, social media).

**Music Player Logs.** The PhoneStudy app also logged participants' music player records containing the titles of their played songs. To describe these songs in terms of their intrinsic musical attributes, we enriched them with song-level variables provided by Spotify's Track API.<sup>6</sup> The resulting ten variables reflect the audio characteristics (e.g., danceability, energy, loudness) of songs listened to and are described in more detail in Sust et al.'s (2022) and our codebook in the OSF project.

**Bluetooth Connectivity.** The PhoneStudy app recorded data on participants' Bluetooth connectivity status, which we grouped according to a custom two-level category system (see our OSF repository, for the detailed coding scheme). At Level-1, Bluetooth events were categorized as turned on/connected, turned on/disconnected, or

turned off. If a Bluetooth event was categorized as on/connected, we further differentiated and grouped the event by the type of connected device informing us about present objects (i.e., car, computer, headset, health-related device, speaker, phone, watch, or other).

**GPS Sensors.** We logged GPS data (including latitude, longitude, altitude, and speed) via the Fused Location Provider API<sup>7</sup> and subjected them to different preprocessing pipelines. For example, latitude and longitude were converted to places using a two-step procedure. First, we clustered all available GPS data points across the whole study period of the SSPS per participant to identify their home and workplace (i.e., the center of the cluster in which a participant was present most frequently between 1 am–5 am and 8 am–4 pm on weekdays). The resulting clusters are displayed in Panel B of Figure 2 and highlight that our sample was well-distributed across Germany.

In a second step, for 2,551 GPS data points not labeled as home and/or work, we used the HERE Geocoding and Search API (discover and search service)<sup>8</sup> to annotate these places with the closest point of interest (e.g., restaurant, shop, building) within a radius of maximum 100 meters. If the distance was the same for more than one place in the HERE database, we assigned GPS points multiple place labels. Finally, we grouped the types of places based on the HERE Places Category System level 2<sup>9</sup> (see the OSF project, for the detailed coding scheme), resulting in 11 different place categories (e.g., education, shopping, arts/culture/entertainment).

In addition, the HERE Geocoding and Search API provided us with the name of the city where the extracted places were located. Hence, we enriched the city by the density of inhabitants who lived there in 2020 available via the German Federal Statistical Office (Destatis, 2021). For privacy reasons, we categorized the density of inhabitants into groups of 500 inhabitants per km<sup>2</sup> (<500, 500–999, 1,000–1,499, ..., >4,000).

Furthermore, we used the GPS longitude and latitude to engineer so-called Geohashes according to the Niemeyer algorithm (Chirico, 2020; Niemeyer, 2008). This GPS encoding technique divides geographic regions into a hierarchical grid structure. We used this algorithm to assign unique numerical codes to all the different locations each participant visited during the study period. These Geohashes, in turn, served us to extract indicators for individual location visiting patterns per participant (Roy & Pebesma, 2017).

Finally, we also extracted GPS-based features used in the previous literature, which were more straightforward and did not require data enrichment, such as the total distance covered, the radius of gyration, or location variance; see Müller et al. (2020) for an overview. Please refer to the codebook in our OSF project for more information.

### **Extraction Strategy**

To extract cues corresponding to participants' psychological situations, we had to match the smartphone sensing data with the

<sup>5</sup> <https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionApi>.

<sup>6</sup> <https://developer.spotify.com/documentation/web-api/reference/#/operations/get-several-audio-features>.

<sup>7</sup> <https://developers.google.com/location-context/fused-location-provider/>.

<sup>8</sup> <https://developer.here.com>.

<sup>9</sup> [https://developer.here.com/documentation/places/dev\\_guide/topics/place\\_categories/places-category-system.html](https://developer.here.com/documentation/places/dev_guide/topics/place_categories/places-category-system.html).

respective ES instances within one situational snapshot. This combination of smartphone sensing and ES data at the level of the ES is a novel contribution to the current literature, so we created two different feature extraction modes.

**Inter-Situm.** First, we extracted inter-situm features, that is, cues measured at the exact point of time when participants opened an ES questionnaire (see Step 2 in Figure 3). This feature extraction mode, however, was only available for sensing data types running in parallel while participants actively interacted with the ES questionnaire in the PhoneStudy app (e.g., music player logs, Bluetooth connectivity, power plug status). In contrast, sensing data types resulting from active engagement with the phone (e.g., phone and app usage logs) could not be produced while filling out the ES questionnaire. Because inter-situm features referred to a status query at a specific point in time, most of them were binary (e.g., power plug status as plugged-in or not plugged-in), and only a few were continuous (e.g., the tempo score of the song currently playing).

**Circa-Situm.** Second, we extracted circa-situm features by aggregating sensing data over the time window (i.e., the situation) surrounding the point of time the ES questionnaire was opened. While it is generally difficult to determine the temporal scope of situations from continuous real-life data (Blake et al., 2020; Rauthmann & Sherman, 2016b), we defined this window based on practical considerations of our data structure. We chose a 60-min timeframe (30 min before and after an ES instance) because this was sufficiently large to extract meaningful features (i.e., to capture rare events like texting) but small enough to avoid overlap between consecutive ES instances (which could occur within a minimum of 1 hr apart). We used different quantification metrics (i.e., sum, average, variation, minimum, maximum) to aggregate the sensing data per data type within these windows. For the GPS-based circa-situm features, we additionally applied some more sophisticated quantification metrics suggested in previous literature, such as the standard deviation of displacements or the radius of gyration (see Stachl et al., 2020). In contrast to inter-situm features, this extraction mode was available for all data types (except time).

### Extracted Cues

Finally, our feature preprocessing pipeline resulted in the extraction of a total of 1,356 cues (58 inter-situm, 1,298 circa-situm). The complete list of cues, including detailed definitions of the respective measures, can be found in the codebook and in Table S2 in the online supplemental materials. For further analysis, we categorized the 1,356 features into theory-driven cue groups based on a hierarchical taxonomy: In a first step, we assigned each feature to one of the three Level-2 groups of situation cues, smartphone cues, or environment cues as defined in the Introduction. In a second step, each situation cue feature received a second label (Level-1 cue groups). We adopted the widely used approach from the previous research to categorize situation cues as interactions/persons, objects, activities/events, location, or time (Rauthmann, Sherman, & Funder, 2015). We summarize the distribution of features among Level-1 and Level-2 cue groups and give some examples in Table 1.

**Table 1**  
*Overview of Cue Groups and Examples*

Cue category	No.	Example
Situation cues	212	
Interactions/persons	47	Total number of outgoing calls during the situation window (circa-situm)
Objects	55	Headphone status is currently “plugged” (inter-situm)
Activities/events	89	Total number of usage sessions during the situation window (circa-situm)
Location	16	Current location is “at home” (inter-situm)
Time	5	Current timestamp is “in the evening” (inter-situm)
Smartphone cues	1,130	Variation in the ringing of incoming calls during the situation window (circa-situm)
Environment cues	14	Density of inhabitants of the current location’s city (inter-situm)

*Note.* The numbers indicate how many features were assigned to the respective cue group. The brackets indicate whether the exemplary features were extracted at the exact moment of the ES (inter-situm) or for the 60-min window surrounding the ES (circa-situm). ES = experience sampling.

### Analytic Strategy

#### *Validation Analysis of Self-Reported DIAMONDS*

As indicated above, we changed the response format for the DIAMONDS dimensions measure from a Likert-type to a binary rating scale for maximum economy during our intensive study. To confirm the validity of this adapted measure, we followed the example of Horstmann and Ziegler (2020) and proceeded as follows: First, we examined the intraclass correlations (ICCs) of the DIAMONDS dimensions to see if they still had a substantial amount of within-person variance despite the binary rating scale. Second, we examined nomological associations to check whether the patterns of association among DIAMONDS dimensions and with other nomological constructs (mood states, Big Five personality traits) were similar to those reported in the previous literature. In particular, for (DIAMONDS dimensions and mood) state measures, we computed person-mean scores across all ES questionnaires and then calculated Pearson correlations at the person level. To compare our association patterns with those from previous research, we searched the literature and found  $n = 4$  studies reporting intercorrelations between DIAMONDS dimensions,  $n = 3$  studies reporting correlations between DIAMONDS dimensions and mood states, and  $n = 7$  studies reporting correlations between DIAMONDS dimensions and personality traits. The list of included references is available in Table A1 in the Appendix A and the full coding scheme in our OSF project. We only selected studies that reported empirical correlations based on in situ person-mean DIAMONDS ratings or that provided the raw data sets so that we were able to calculate these correlations on our own (Horstmann & Ziegler, 2019; Kritzler et al., 2020). If articles included more than one study, we included each study as a separate observation. We pooled the

correlations by means of the *meta* package (Balduzzi et al., 2019) and used a random effects-model to account for between-study heterogeneity (Harrer, 2022).

### Machine Learning

The 1,356 situation features (136 after target-independent preprocessing) served us to predict their corresponding situation characteristics in a machine learning approach. Thereby, we treated each of the self-reported DIAMONDS dimensions as the target of a binary classification task. All machine learning analyses were conducted within the *mlr3verse* (Lang & Schratz, 2022). For our visualizations, we used the packages *fmsb* and *ggplot2* (Nakazawa, 2023; Wickham, 2016).

**Preprocessing.** First, we performed target-independent preprocessing of the smartphone sensing features using the *caret* package (Kuhn, 2022). We scanned the data set for outliers. Thereby, we replaced only extreme outliers with more than four standard deviations from the mean to exclude anomalies in the data caused by technical recording errors, while preserving extreme expressions of situation features in the data. In addition, we removed all features with more than 90% missing values across all observations and features with only few unique values (Kuhn, 2022). Although our chosen models can handle multicollinearity, we removed features with high intercorrelations ( $r > .75$ ) to avoid ambiguities when applying interpretable machine learning techniques (Kuhn & Johnson, 2013). Second, to avoid overfitting our models to our specific data set, we integrated all remaining preprocessing steps into the resampling. These steps included the imputation of missing values using the median, scaling, and weighting for imbalanced data as outlined below.

**Models.** We trained regularized logistic linear regression models (least absolute shrinkage and selection operator [LASSO]; Tibshirani, 1996) and nonlinear tree-based random forest models (Breiman, 2001). We chose these two models because they have a model-inherent selection of relevant features and can, thus, cope with high-dimensional and intercorrelated predictor spaces. We used the default hyperparameter settings of the models' implementations in the *mlr3* package (Lang et al., 2019).

We ran these models with class-dependent costs to account for imbalanced class distributions (e.g., *Deception* was only "present" in 3% of all sampled situations). More specifically, we assigned a class-dependent theoretical weight to each observation to increase the effect of the minority class and decrease the effect of the majority class observations (Stern et al., 2021). The weighting factor was determined separately for each target dimension based on its distribution in the sample.

For benchmark purposes, we additionally trained featureless baseline models, which predicted the most common class of the training data set's target variable for all observations in the respective test set without considering any features.

**Performance Evaluation.** For all target variables, we conducted one benchmark experiment comparing the predictive performance of the two models against the baseline. To estimate the predictive performance for unseen data points, we separated training and test data and ran all benchmarks with 10-times repeated 10-fold cross-validation ( $10 \times 10$  CV) as the resampling procedure. However, as our target variables resided at the level of observations (i.e., each ES questionnaire) instead of persons (i.e., participants), our data structure was nested, and within-person observations were dependent. To still

obtain realistic performance estimates for previously unseen persons, we applied blocking, whereby all observations of the same participant are considered to belong together and are assigned to either the training or the testing set but never split between both (Dragicevic & Casalicchio, 2020).

We evaluated model performances with the area under the receiver operating characteristic curve (AUC). In binary classification problems, the receiver operating characteristic (ROC) curve plots a model's true positive rate (sensitivity) against its false positive rate at various threshold settings (1—specificity; Majnik & Bosnić, 2013; Maloof, 2003). Thus, the ROC curves show the model's predictive ability resulting from sensitivity and specificity as a function of different discrimination thresholds for a binary classification task. The AUC performance measure, in turn, integrates the tradeoff between both error measures across various discrimination thresholds into a single parameter (i.e., the area under the curve). Thereby, the metric is insensitive to imbalanced class distributions. The AUC indicates the probability of correctly ranking two random situations, one from each class (e.g., a situation high in Duty is ranked higher than a low-Duty situation when such a pair is randomly selected; Viaene & Dedene, 2005). The baseline model exhibits a linear relationship between sensitivity and specificity which manifests in an AUC of .50. In prediction tasks, model performance can be better (AUC > .50) or worse (AUC < .50) than the baseline with AUC values ranging between 0 and 1. For each prediction model, we computed the AUC performance metric for each of the  $10 \times 10$  CV iterations and averaged it across all iterations.

Furthermore, we compared the AUCs obtained in each of the 100 resampling iterations between prediction and baseline models, which is similar to comparing, for example, pairs of persons (in our case resampling iterations), which are not independent of each other. We applied pairwise student *t* tests (one-sided), which were variance-corrected to account for the dependency structure introduced by cross-validation (Bouckaert & Frank, 2004; Nadeau & Bengio, 1999; Stachl et al., 2020). For each prediction outcome, we adjusted for multiple comparisons (2 models  $\times$  8 DIAMONDS dimensions = 16 tests) via Bonferroni correction. Models whose AUC was significantly ( $p < .001$ ) above the baseline were considered predictive as they were consistently successful across resampling iterations.

Finally, we also report phi coefficients  $r_\phi$ —also known as Matthews Correlation Coefficient in machine learning—as performance measures (Chicco & Jurman, 2020). They indicate the association between the actual (self-reported) and the predicted situation characteristics. The higher the phi coefficient, the better the model predicted the respective situation characteristic.

**Feature Transformation.** As far as we know, there are currently no machine learning algorithms implemented off-the-shelf and proven in our application setting that could accommodate nested data structures. Therefore, as already mentioned, we trained our models at the observational level, which raises the question of whether the models learned within-person or between-person patterns. Because our models do not take the nested structure into account, the model parameters cannot directly give an answer to this question. To still differentiate within-person and between-person effects, we adopted an approach commonly used in multilevel modeling, namely centering within context with reintroduction of the mean; CWC(*M*), see Zhang et al., 2009 for more details from the multilevel setting. That means we ran a second benchmark experiment ( $10 \times 10$  CV) and included different feature sets with identical cues but different

transformations: (a) our nontransformed original features; (b) features that were person-mean centered (CWC) and, therefore, only contained information on within-person differences; (c) person mean (M) features that only contained information on between-person differences; and (d) both person-mean centered features and the person means (CWC(M)) that contained information on both within- and between-person differences. Comparing the predictive performances between the original and transformed feature sets provided insight into whether our models learned within-person or between-person patterns or both.

### Model Interpretations

To gain further insights into the successful DIAMONDS dimensions prediction models, we ran several interpretable machine learning analyses. Giving a sneak peek into our results, LASSO and random forest models performed equally. Therefore, we focused on the LASSO models because they are easier to interpret due to their linearity and sparsity.

**Single Feature Importance.** To understand which individual features were most relevant for predicting the respective DIAMONDS dimensions, we refit the LASSO models for which we achieved significant predictive performances on the full data set. The LASSO automatically performs feature selection by integrating only the most useful features in its predictions. For all features with nonzero coefficients, we retrieved the model-inherent, standardized  $\beta$  regression weights. Similar to standard logistic regression, the magnitude of the standardized regression coefficients is an indicator of the importance of single features for the models' predictions.

**Grouped Feature Importance.** In a second step, we were also interested in the unique relevance of our theory-driven feature groups (see Table 1). Therefore, we assigned our features to the different cue groups and ran another benchmark experiment ( $10 \times 10$  CV) with different subsets of features for each predicted situation characteristic:

For inspecting the importance of features assigned to Level-2 cue groups, we included the LASSO model trained on the full set of features, and three further LASSO models, each trained on a subset of the features, where one Level-2 cue group (i.e., situation cues *or* smartphone cues *or* environment cues) was excluded. For each of these subsets, we extracted the mean prediction performances over the resampling iterations and calculated the mean  $AUC_{Loss}$  as the pairwise difference between the AUC achieved by the full feature set and the feature set reduced by the respective Level-2 cue group:  $AUC_{Loss\ L2i} = M_{AUC\ (full\ feature\ set)} - M_{AUC\ (full\ feature\ set - feature\ set\ L2i)}$ , where L2i symbolizes one of the Level-2 groups. The  $AUC_{Loss\ L2i}$  reflects the importance of the excluded Level-2 cue group. The more negative it was, the more important the corresponding cue group was for the prediction of the respective situation characteristic.

To inspect the importance of the features assigned to the different Level-1 cue groups, we applied the same procedure, only on one hierarchical level below. That means, for each situation characteristic, we trained the LASSO model on the full set of situation cues, and we ran five further LASSO models, each on a subset of the situation cues, where one Level-1 cue group (interactions/persons *or* objects *or* activities/events *or* location *or* time) was excluded. Again, for each of the Level-1 feature subsets, we extracted the mean prediction performances over the resampling iterations and calculated the mean  $AUC_{Loss}$  as the pairwise difference between

the AUC achieved by the full situation cue feature set and the feature set reduced by the respective Level-1 cue category:  $AUC_{Loss\ L1i} = M_{AUC\ (situation\ cues)} - M_{AUC\ (situation\ cues - feature\ set\ L1i)}$ , where L1i is one of the Level-1 groups.

## Results

Across 455 participants, we sampled a total of 9,790 situational snapshots, of which 47.1% were perceived as containing Duty, 26.8% as Intellect, 4.6% as Adversity, 28.2% as Mating, 72.8% as pOsitivity, 18.6% as Negativity, 3.3% as Deception, and 56.8% as Sociality. Detailed descriptive statistics for self-reported DIAMONDS dimensions are reported in Table 2, while those for cues, including Pearson correlations with self-reported DIAMONDS dimensions, can be found in our online supplemental materials (Table S2).

### Validity Analysis of Self-Reported Situation Characteristics

As a basis for all further analyses, we first investigated the validity of our adapted DIAMONDS measures, whose rating scale had been changed from the original 7-point Likert scale to a binary format. First, we considered the ICCs of our DIAMONDS dimensions, assuming that situational state measures should vary not only between but also within persons (Horstmann & Ziegler, 2020). Indeed, Table 2 exhibits ICCs between 0.20 (for Duty) and 0.41 (for Mating), indicating a sufficient amount of within-person variance across situational snapshots to confirm our expectation.

Second, we considered nomological associations and calculated (a) person-level intercorrelations of the DIAMONDS dimensions and (b) person-level correlations between the DIAMONDS dimensions and mood states as well as Big Five personality traits to compare our association patterns with those reported in previous research. As an obligatory first step, we pooled correlations reported in past literature and added the resulting coefficients to the second row of each cell in Table 2. Based on their 95% intervals, we marked the substantial pooled correlations in Table 2, creating an overview of expected *convergent* (highlighted in black) and *divergent* (highlighted in gray) correlations. Next, we checked whether our empirically determined correlation coefficients lay within the 95% confidence interval of the pooled correlation coefficients from previous literature. Please note that this procedure was explorative and fulfilled purely descriptive purposes, in particular, because our pooled confidence intervals were rather wide due to the small sample of past studies reporting the respective associations (see Table A1 in the Appendix A). Table 2 shows that our data were largely consistent with the pattern of expected convergent and divergent correlations from the literature. Out of the six external validation constructs, only Emotional Stability exhibited a pattern deviating from past studies with stronger correlations for Mating, pOsitivity, Negativity, and Sociality. However, because previous studies assessed the construct in reverse as Neuroticism (which we re-coded for Table 2), these discrepancies could be related to the slight differences in conceptualization. Beyond that, there seemed to be systematic deviations from the previous literature only for the DIAMONDS dimension of Mating, whose correlations with Big Five personality traits (except for Openness) were higher and whose intercorrelations with Adversity and Deception were lower than in the past. Interestingly, our ICC for Mating was also relatively high

**Table 2**  
*Descriptive Statistics and (Inter)Correlations for Self-Reported Situation Characteristics and Nomologic Net Constructs in Comparison to Pooled Effect Sizes From Previous Literature*

Variable	M	SD	ICC	D	I	A	M	O	N	De	S
DIAMONDS	0.47	0.50	0.20	—							
Duty	0.27	0.44	0.28	0.57	—						
Intellect	0.05	0.21	0.39	0.17 [0.04, 0.35]	0.37 [0.04, 0.48]	—					
Adversity	0.28	0.45	0.41	0.08 [−0.16, 0.19]	0.09 [0.02, 0.31]	0.16 [0.19, 0.45]	—				
Mating	0.73	0.45	0.33	−0.01 [−0.23, 0.41]	−0.04 [−0.12, 0.59]	−0.15 [−0.18, −0.52, 0.21]	0.29 [−0.12, 0.45]	—			
pOstivity	0.19	0.39	0.31	0.20 [−0.01, 0.48]	0.33 [−0.12, 0.59]	0.52 [0.39, 0.70]	−0.02 [−0.29, 0.43, −0.14]	−0.52 [−0.43, −0.14]	—		
Negativity	0.03	0.18	0.39	0.17 [−0.22, 0.31]	0.30 [−0.15, 0.54]	0.65 [0.41, 0.94]	0.32 [0.22, 0.42]	−0.20 [−0.62, 0.31]	0.44 [0.41, 0.66]	—	
Deception	0.57	0.50	0.31	0.33 [0.02, 0.53]	0.27 [0.04, 0.46]	0.10 [−0.51, 0.58]	0.35 [−0.15, 0.71]	0.51 [0.26, 0.70]	0.08 [−0.30, 0.32]	0.03 [−0.53, 0.57]	—
Sociality											
Nomologic constructs											
Mood valence	4.61	0.74	0.45	−0.04 [−0.38, 0.69]	−0.06 [−0.22, 0.72]	−0.15 [−0.65, 0.51]	0.21 [−0.27, 0.51]	0.71 [0.17, 0.92]	−0.58 [−0.68, 0.23]	−0.09 [−0.79, 0.68]	0.15 [0.06, 0.71]
Openness	−0.23	0.77		0.02 [−0.01, 0.08]	0.07 [0.05, 0.17]	−0.03 [−0.19, 0.05]	0.07 [−0.13, 0.09]	0.10 [−0.03, 0.24]	−0.02 [−0.07, 0.02]	−0.01 [−0.21, 0.12]	0.14 [0.10, 0.23]
Conscientiousness	−0.02	0.76		0.08 [0.02, 0.17]	0.07 [0.00, 0.12]	−0.10 [−0.33, −0.05]	0.17 [−0.17, 0.11]	0.16 [−0.13, 0.28]	−0.14 [−0.19, −0.01]	−0.10 [−0.38, −0.01]	0.13 [0.10, 0.28]
Extraversion	−0.24	0.78		−0.01 [−0.03, 0.15]	0.01 [−0.01, 0.11]	−0.04 [−0.15, 0.03]	0.19 [0.01, 0.14]	0.23 [0.04, 0.29]	−0.16 [−0.22, −0.03]	−0.03 [−0.14, −0.01]	0.15 [0.11, 0.20]
Agreeableness	0.02	0.81		−0.02 [−0.05, 0.10]	−0.07 [−0.04, 0.13]	−0.16 [−0.26, −0.06]	0.11 [−0.15, 0.03]	0.13 [−0.03, 0.29]	−0.09 [−0.17, −0.02]	−0.12 [−0.27, −0.06]	0.07 [−0.07, 0.19]
Emotional Stability <sup>a</sup>	0.07	0.79		0.02 [−0.13, 0.07]	−0.02 [−0.08, 0.04]	−0.04 [−0.20, 0.13]	0.01 [−0.05, 0.07]	0.03 [−0.11, 0.17]	−0.10 [−0.28, 0.08]	0.00 [−0.13, 0.13]	0.12 [−0.07, 0.03]

*Note.* Pearson correlations were calculated at the between-person level. For this purpose, all state measures (i.e., DIAMONDS dimensions and mood valence) were first aggregated to person-mean scores. The elements in the second row of each cell depict pooled correlations and their 95% confidence intervals, which we aggregated from effect sizes reported in the previous literature. White text highlighted in black indicates substantial pooled correlations that one would expect to find again and that, thus, can be used to explore convergent validity. All other pooled correlations (gray background) represent nonsubstantial correlations that one would not expect to find again and that, thus, can be used to explore divergent validity. Due to the differential availability of survey data, the sample size ( $n = 455$ ) was reduced to  $n = 380$  for all statistics related to personality traits. ICC = intraclass correlation coefficient; DIAMONDS = Duty, Intellect, Adversity, Mating, pOstivity, Negativity, Deception, Sociality.

<sup>a</sup> Our study assessed Emotional Stability, while previous studies assessed the reverse construct of Neuroticism. For ease of interpretation, we reversed the reported pooled correlations from the past literature so it matches our directionality.

compared to previous studies, which found values between 0.19 and 0.29 (Horstmann et al., 2021; Kritzler et al., 2020; Sherman et al., 2015). This indicates that our Mating responses contained comparatively more variance because of between-person differences, which, in turn, could be responsible for the differences in the nomological correlations. One reason for our elevated levels of between-person variance could be our sample composition, which was representative in terms of age and gender, and was, thus, more heterogeneous than the student samples used in the previous literature (Horstmann et al., 2021; Kritzler et al., 2020; Sherman et al., 2015). Alternatively, the pandemic setting could be related to the greater differences between individuals' Mating perceptions. Despite these few deviations, the majority of our nomological correlations align well with the patterns of association in the previous literature, so we consider our analyses as empirical support for the validity of our adjusted DIAMONDS dimensions measures.

**Prediction of Situation Characteristics**

***How Much Situational Information Do Smartphone Sensing Data Contain?***

In our preregistered analyses, we investigated whether self-reported situation characteristics can be predicted by objective cues

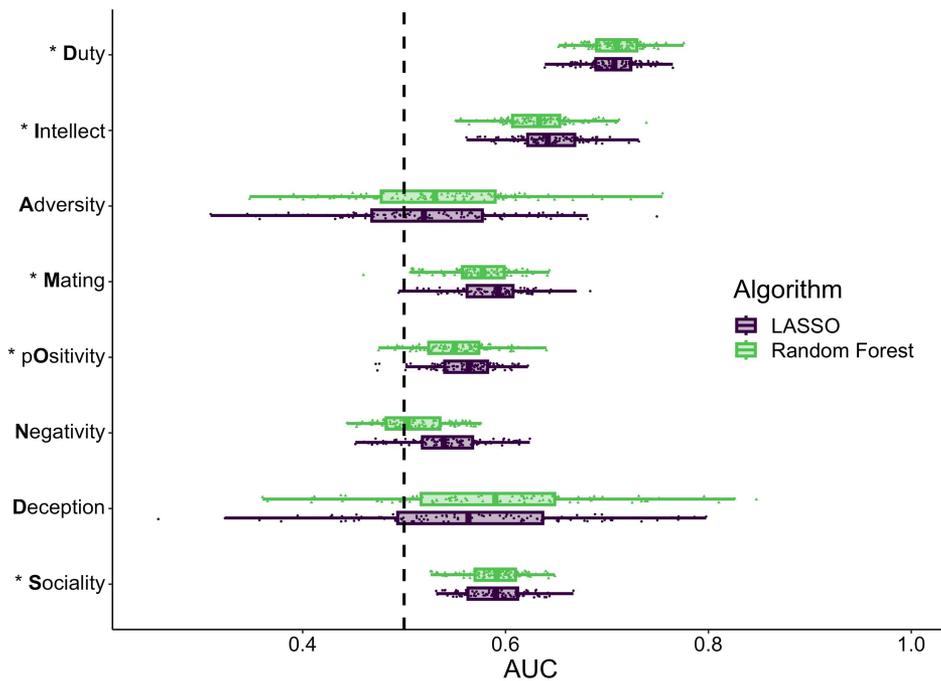
assessed via smartphone sensing. All descriptive and test statistics of the respective prediction analyses can be found in Table B1 in the Appendix B, and the corresponding ROC curves are available in our online supplemental materials (see Figure S1).

The linear LASSO and the nonlinear random forest models achieved comparable predictive performances across all DIAMONDS dimensions (see boxplots in Figure 4). Following the principle of parsimony, we, therefore, focus our further report and exploratory analysis on the less complex and more interpretable LASSO models.

Figure 4 and Table B1, Appendix B show that Duty, Intellect, Sociality and Mating, and pOsitivity were successfully predicted across the resampling iterations. That means their LASSO models' average prediction performances (i.e., their mean AUCs) were each significantly higher than the prediction performance of the featureless baseline model. In contrast, the prediction performances for the dimensions of Adversity, Negativity, and Deception were not significantly different from the baseline performance. Thus, our LASSO models could not grasp systematic variance in Adversity, Negativity, and Deception based on the cues extracted from our sensing data.

Among the significant results, not every DIAMONDS dimension was predicted equally well from the sensing data. The correlations between the self-reported and predicted DIAMONDS dimensions,

**Figure 4**  
*Distribution of Prediction Performances Across the Iterations of Repeated Cross-Validation by Situation Characteristics and Models*



*Note.* Distribution of the AUC across the 100 resampling iterations of the applied 10x10 CV scheme for LASSO and random forest models. The black dotted line at an AUC of 0.50 represents the prediction performance of baseline model against which both algorithms were compared in a pairwise manner. AUCs of the single iterations are represented by the single dots. The boxes contain all values between the 25% and 75% quantiles and their middle line indicates the median. For clarity, the AUC scale was cutoff at .25. Significantly predictive models ( $p < .001$ ) are marked by an asterisk (\*). More detailed descriptive information and test results can be found in Table B1 in the Appendix B. AUC = area under the curve; 10 × 10 CV = 10-fold cross-validation; LASSO = least absolute shrinkage and selection operator. See the online article for the color version of this figure.

which we inspected as an additional performance measure, confirmed this differential pattern of predictability: LASSO models worked best for Duty ( $r_\phi = 0.30$ ), and Intellect ( $r_\phi = 0.18$ ), followed by Sociality and Mating (both  $r_\phi = 0.13$ ). In comparison, the prediction performance for pOstivity ( $r_\phi = 0.08$ ) was clearly lower, despite a significant test result (see Table B1).

We wanted to limit our follow-up analyses to those DIAMONDS dimensions that were successfully predicted from sensed cues and selected all outcomes that were predicted significantly better by the LASSO models than by the baseline models. Thus, despite its comparatively low predictive performance, we included pOstivity in our selection for further analyses (along with Duty, Intellect, Mating, and Sociality).

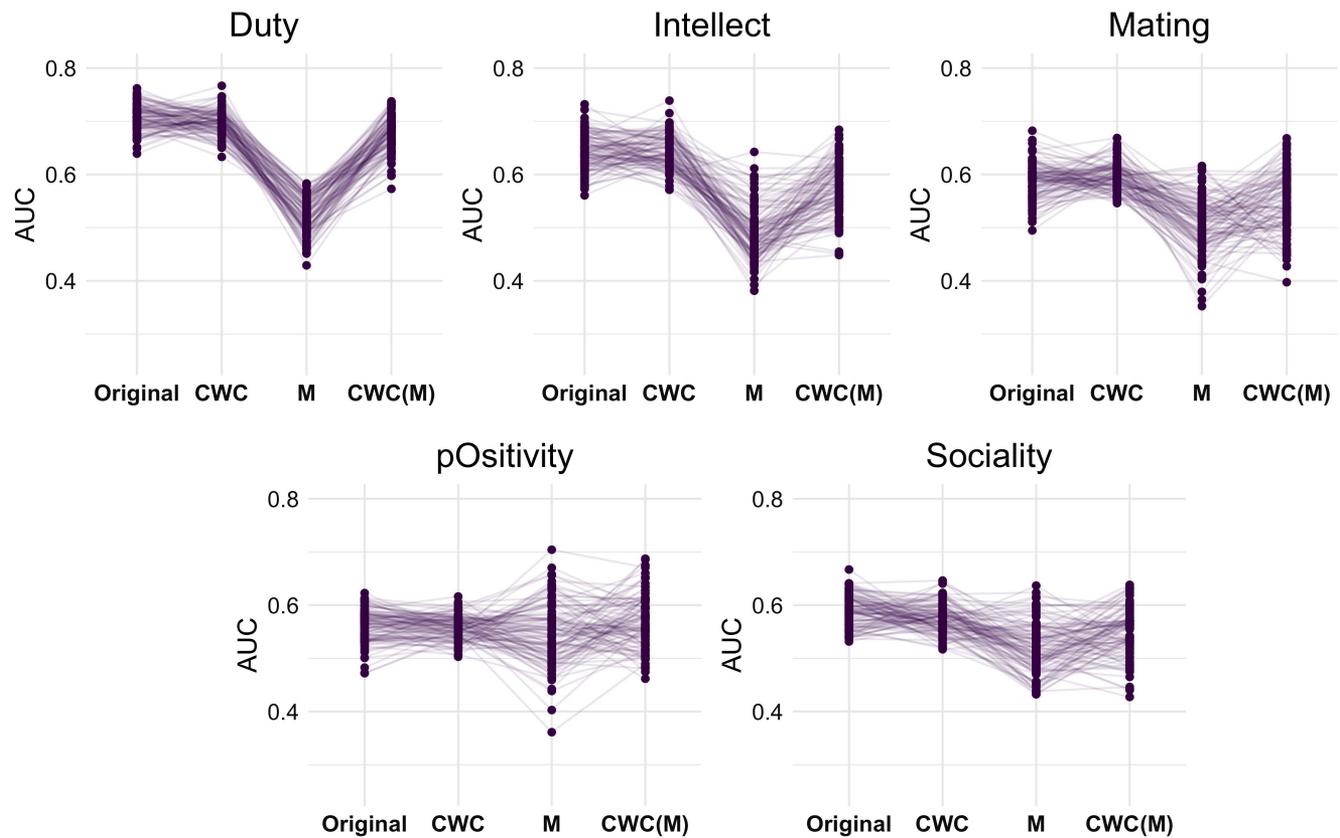
### Do Predictions Rely on Within- or Between-Person Patterns?

In an exploratory next step, we investigated if our LASSO models learned within- or between-person patterns in the smartphone sensing data when predicting situation characteristics. Consider this highly simplified example illustrating the idea behind this analysis:

Let us assume that the cue *smartphone usage time* was the only feature in a LASSO model predicting Sociality. Which association patterns could the model learn? On the one hand, the LASSO model could learn that many individuals' smartphone usage time was higher in situations perceived as not containing Sociality, compared to the same individuals' smartphone usage time in other situations (i.e., within-person patterns). On the other hand, the LASSO model could learn that the smartphone usage time was higher for some individuals compared to others in general (i.e., across sampled situations) and that among those with generally higher smartphone usage times, more individuals perceived the situation as not containing Sociality than among those with generally lower smartphone usage times (i.e., between-person patterns).

To explore if our models learned within- or/and between-person patterns, we compared the prediction performances of (a) our original LASSO models (using a nontransformed feature set) with LASSO models trained on a set of (b) person-mean centered features (CWC; containing within-person information only), (c) person-mean features ( $M$ ; containing between-person information only), and (d) a combination thereof (CWC( $M$ )). Figure 5 displays the results of this comparison. The lines represent the 100 iterations of

**Figure 5**  
Prediction Performance Across Resampling Iterations by Situation Characteristics and Feature Transformation Approach



*Note.* The AUC was determined for each of the 100 resampling iterations of the  $10 \times 10$  CV scheme. The lines represent the 100 iterations and the first column of dots (Original) represent the AUCs of our original LASSO models using nontransformed features, which is compared to the models using the person-mean-centered features (CWC), the person mean of features ( $M$ ), and a combination of both (CWC( $M$ )). These analyses were only conducted for situation characteristics successfully predicted from the original feature set in our main analyses. AUC = area under the curve;  $10 \times 10$  CV = 10-fold cross-validation; LASSO = least absolute shrinkage and selection operator; CWC = person-mean centered. See the online article for the color version of this figure.

the applied  $10 \times 10$  CV scheme, and the points indicate the AUCs of our initial LASSO models compared to the models using the CWC, ( $M$ ), and CWC( $M$ ) feature sets, respectively. Figure 5 shows a similar order across all situation characteristics: The performance of our original LASSO models was, on average, equally good as for models using CWC features, slightly better than for models using CWC( $M$ ) features, and considerably better than for models using the ( $M$ ) features. While present for all situation characteristics, this pattern was most evident for the dimension of Duty, probably because Duty had the highest level of prediction performance overall, regardless of the transformation approach used. Please note that, when interpreting the results in Figure 5, poorer performance of a machine learning model is expressed not only by a decreased AUC (i.e., downward shift in the points) but also by a higher dispersion of the AUC (i.e., wider spread of points). Descriptive statistics of the performance results across transformation approaches are displayed in Table B2.

To sum up, because the LASSO models using CWC features, in contrast to ( $M$ ) features, achieved comparable prediction performances as our original LASSO models, we conclude that our original models mostly learned within-person association patterns. LASSO

models using CWC( $M$ ) features performed slightly worse than our original LASSO models, probably because the ( $M$ ) features introduced some “noise” in the within-person information contained in the CWC features.

### Interpretation of Situation Characteristics Predictions

Having provided some initial insight into *how much* information on the different dimensions of situation experience is available in the cues captured by smartphone sensing, we turned to the second question of *what* information drives the models’ predictions.

### Importance of Individual Cues

First, we explored which cues—considered individually among the full feature set—were most important for predicting each of the DIAMONDS dimensions. We present the top five features for each situation characteristic based on the magnitude of the standardized regression coefficients (i.e.,  $\beta$  weights) extracted from the LASSO models in Table 3. The coefficients may be interpreted the same way

**Table 3**  
Top Five Important Predictors per LASSO Model Predicting Situation Characteristics

Rank	Cues	$\beta$	Cue category	
			Level-2	Level-1
<b>Duty (29 cues)</b>				
1	Situation snapshot is on the <i>weekend</i>	-0.38	Situation cue	Time
2	Situation snapshot is in the <i>evening</i>	-0.34	Situation cue	Time
3	Minimum distance to the workplace during situation window	-0.23	Smartphone cue	
4	Situation snapshot is at <i>home</i>	-0.20	Situation cue	Location
5	Situation snapshot is at the <i>workplace</i>	0.13	Situation cue	Location
<b>Intellect (27 cues)</b>				
1	Situation snapshot is on the <i>weekend</i>	-0.28	Situation cue	Time
2	Situation snapshot is in the <i>evening</i>	-0.18	Situation cue	Time
3	Situation snapshot is at <i>home</i>	-0.17	Situation cue	Location
4	Minimum distance to the workplace during situation window	-0.16	Smartphone cue	
5	Mobility (standard deviation of displacements during situation window)	-0.06	Smartphone cue	
<b>Mating (22 cues)</b>				
1	Situation snapshot is in the <i>evening</i>	0.16	Situation cue	Time
2	Situation snapshot is on the <i>weekend</i>	0.14	Situation cue	Time
3	Ratio between total duration of smartphone checks and sessions during situation window	-0.08	Smartphone cue	
4	Minimum distance to the workplace during situation window	0.07	Smartphone cue	
5	Popularity (number of a participant’s visits) of the situation snapshot’s Geohash	0.06	Environment cue	
<b>pOsitivity (46 cues)</b>				
1	Situation snapshot is on the <i>weekend</i>	0.12	Situation cue	Time
2	Situation snapshot is at the <i>workplace</i>	-0.11	Situation cue	Location
3	Minimum distance to the workplace during situation window	0.10	Smartphone cue	
4	Situation snapshot’s city’s categorized density of inhabitants	-0.09	Environment cue	
5	Situation snapshot was at a service place	-0.07	Situation cue	Location
<b>Sociality (46 cues)</b>				
1	Situation snapshot is at <i>home</i>	-0.14	Situation cue	Location
2	Music usage was tracked during situation window	0.11	Environment cue	
3	Popularity (duration of a participant’s visits) of the situation snapshot’s Geohash	-0.08	Environment cue	
4	Variation of the distance to home during situation window	-0.07	Smartphone cue	
5	Minimum total duration of screen sessions during situation window	-0.07	Smartphone cue	

*Note.* We present the top five most important features in decreasing order for each successfully predicted situation characteristic. Features were selected and ranked based on the standardized regression coefficients  $\beta$  extracted from the respective dimension’s LASSO models. Numbers in brackets indicate how many features were not shrunk to zero by the LASSO model’s inherent feature selection. Italic text (colors) illustrates which features (cue groups) repeatedly made it into the top five ranking across the DIAMONDS dimensions. DIAMONDS = Duty, Intellect, Adversity, Mating, pOsitivity, Negativity, Deception, Sociality; LASSO = least absolute shrinkage and selection operator. See the online article for the color version of this table.

as those from standard logistic regression that most readers are likely familiar with. As example, for a *binary* feature with a *negative* coefficient ( $\beta = -0.38$ ), the LASSO model's estimated probability that a situation snapshot contained Duty was, on average, higher if the snapshot did *not* occur on the weekend (when all other cues remained constant). As an example, for a *continuous* feature with a *positive* coefficient ( $\beta = 0.12$ ), the LASSO model's estimated probability that a situation snapshot contained pOstivity was, on average, higher if the minimum distance from the workplace in the hour around the snapshot was higher than the sample mean (when all other cues remained constant).

Table 3 shows that across all situation characteristics, most of the top informative features belonged to the Level-2 group of situation cues (13; smartphone cues: 8, environment cues: 4). Among these situation cues most features were part of the Level-1 groups location and time. Thereby, four situation cues were repeatedly part of the top five features, independent of the target: home, workplace, weekend, and evening (see *italic* text in Table 3).

We chose these four cues to descriptively explore how our trained models worked. Figure 6 illustrates the resulting prototypical situation profiles. The purple profiles represent the cue constellation when the LASSO models predicted the respective situation characteristic as *present*, and the gray profile indicates the cue constellation when the respective situation characteristic was predicted as *absent*. For example, among all situations predicted as containing Duty, 0.4% took place on weekends, 0.5% in the evening, 49.3% at home, and 57.7% at the workplace. In contrast, among all situations predicted as *not* containing Duty, 46.0% took place on weekends, 44.6% in the evening, 66.7% were at home, and 47.3% were at the workplace.

When comparing the predictions of Duty and Intellect, the profile plots exhibit a similar picture for the constellation of the selected location and time cues: Of all situations predicted as containing Intellect, 0.0% occurred on weekends, 1.6% in the evening, 45.8% at home, and 54.7% at work and of all situations predicted as *not* containing Intellect, 44.9% occurred on weekends, 41.8% in the evening, 70.3% at home, and 51.2% at work.

For Mating we found a reversed profile in comparison to Duty and Intellect: Of all situations predicted as containing Mating, 51.2% were on the weekend, 51.5% in the evening, 67.9% at home, and 52.2% at the workplace and of all situations predicted as *not* containing Mating, 2.1% took place on weekends, 1.0% in the evening, 50.2% at home, and 53.7% at work.

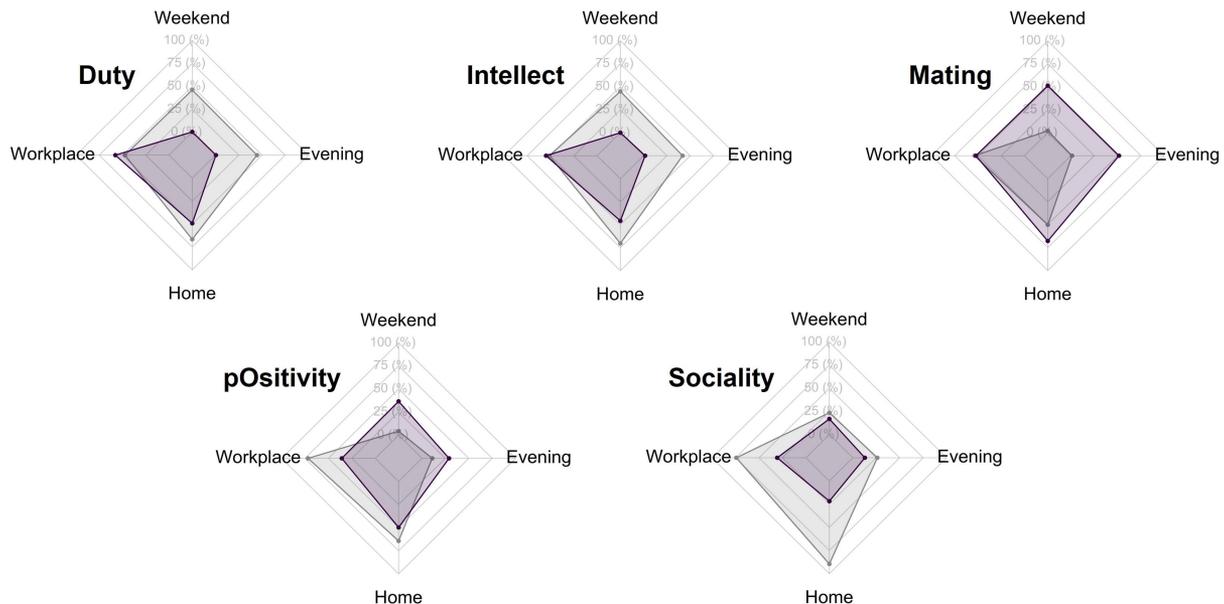
For the remaining DIAMONDS dimensions, the profiles looked more distinct. For example, of all situations predicted as containing pOstivity, 36.6% happened on the weekend, 28.9% in the evening, 49.8% at home, and 35.8% at the workplace, and of all situations predicted as *not* containing pOstivity, 4.4% occurred on weekends, 10.9% in the evening, 64.4% at home, and 71.8% at work.

Finally, of all situations that were predicted as containing Sociality, 16.9% were on the weekend, 12.9% in the evening, 21.8% at home, and 30.6% at the workplace. In addition, of all situations predicted as *not* containing Sociality, 23.6% took place on weekends, 26.2% in the evening, 89.6% at home, and 74.2% at work.

### Importance of Cue Groups

Finally, we explored which cues groups—as categorized by our theory-driven two-level grouping structure—were most important

**Figure 6**  
Prototypical Profiles of Selected Situation Cues for Predicted Situation Characteristics



*Note.* The profile lines illustrate the percentage of situations predicted to contain (purple) or to not contain (gray) the respective DIAMONDS dimensions that took place on weekends, in the evening, at home, and at the workplace. Prediction profiles are only presented for situation characteristics successfully predicted from the full feature set in our main analyses. The depicted selection of situation cues was based on their frequency within the top five features in Table 3. DIAMONDS = Duty, Intellect, Adversity, Mating, pOstivity, Negativity, Deception, Sociality. See the online article for the color version of this figure.

for predicting the respective DIAMONDS dimensions. Following our hierarchical cue taxonomy, we first inspected the importance of the three higher-level groups. Figure 7 shows that for predicting Duty, Intellect, Mating, and pOsitivity, the group of situation cues (range  $AUC_{Loss} = -0.04$  to  $-0.11$ ) was most important for the LASSO models, followed by the group of smartphone cues (range  $AUC_{Loss} = 0.00$  to  $-0.01$ ). In contrast, environment cues were not relevant for predicting these four situation characteristics ( $AUC_{Loss} = 0.00$  for all groups).

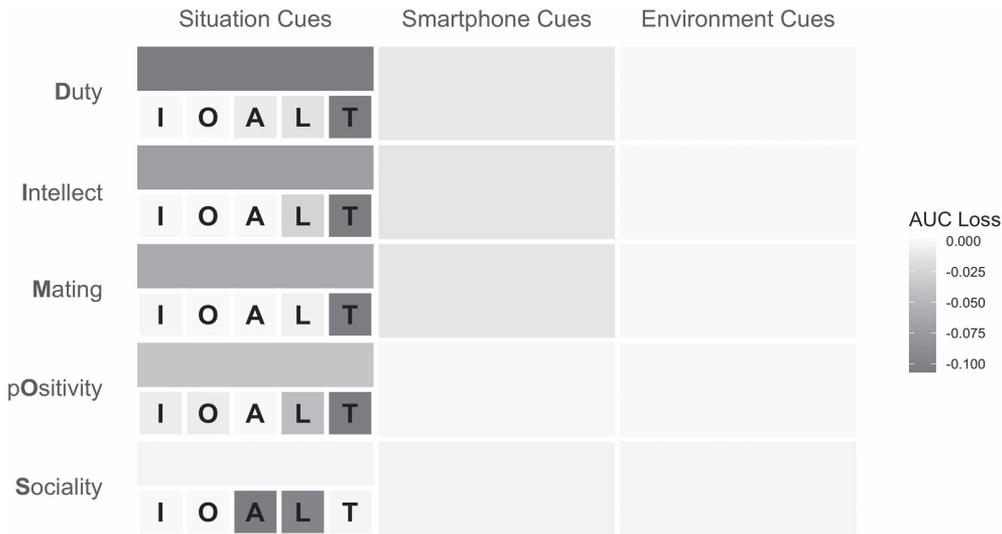
Across situation characteristics, situation cues had the greatest relevance for predicting Duty ( $AUC_{Loss} = -0.11$ ), followed by Intellect ( $AUC_{Loss} = -0.07$ ), Mating ( $AUC_{Loss} = -0.06$ ), and, finally, pOsitivity ( $AUC_{Loss} = -0.04$ ). For Sociality, situation cues, smartphone cues, and environment cues ( $AUC_{Loss} = 0.00$  for all) were equally important. That means if one of the three groups was omitted, the LASSO models still performed equally well with the two remaining cue groups as they did with all three groups. Thus, the situational information regarding Sociality that was contained in one cue group was apparently compensated by the other two groups.

In a second step, we explored the importance of the Level-1 cue groups among the features categorized as situation cues on Level-2. Figure 7 highlights that for predicting Duty, Intellect, and pOsitivity, time cues (range  $AUC_{Loss} = -0.02$  to  $-0.10$ ), followed by location cues (range  $AUC_{Loss} = -0.01$  to  $-0.02$  for), were most important for the models' predictions. For predicting Mating, only time cues were particularly important ( $AUC_{Loss} = -0.05$ ), and for predicting Sociality, activity and location cues were both relevant (both  $AUC_{Loss} = -0.02$ ).

### Similarity Between Prediction Models

In summary, both our single and grouped importance analyses exhibited considerable overlap in the most relevant features across the situation characteristic-specific prediction models. More specifically, the different models had in common that situation cues and, among situation cues, primarily cues representing time and location were most important. While we found some differences in the relative constellation of single cues (see Figure 6) and the relative importance of the different cue groups (see Figure 7) between the situation characteristics, these differences seemed rather small. Therefore, we wanted to conclude by checking whether our DIAMONDS dimensions-specific models may have all learned the same patterns in the sensing data, making “one-size fits all” predictions instead of predicting the specific situation characteristics. For this purpose, we re-trained the LASSO models on the full data set, extracted the predicted situation characteristic scores, and inspected their inter-correlations (see Table B3 in the Appendix B). We expected that if all LASSO models worked the same across situation characteristics, the intercorrelations should (a) be (almost) perfect, that is, close to 1.00 and (b) be all the same. We found that intercorrelations were high among Duty, Intellect, and Mating (range of  $|r_{\phi}| = 0.70-0.82$ ), but considerably lower among the remaining DIAMONDS dimensions (range of  $|r_{\phi}| = 0.14-0.41$ ). In line with the intercorrelations between our self-reported DIAMONDS dimensions and, therefore, conceptually plausible (see Table B3), we found the highest intercorrelation between predicted Duty and Intellect ( $r_{\phi} = 0.82$ ) and the lowest

**Figure 7**  
Heatmap of Prediction Performance Loss by Cue Groups for the Illustration of Grouped Feature Importance



*Note.* For each significantly predicted situation characteristic (represented by rows), the  $AUC_{Loss}$  was determined as an indicator of grouped feature importance: For Level-2 cue groups (situation cues, smartphone cues, environment cues), the  $AUC_{Loss}$  indicates the difference between the mean AUC across resampling iterations using the full feature set minus the mean AUC of a feature set without the respective Level-1 cue group of interest (see columns). For Level-1 cue groups (I: interactions/people, O: objects, A: activities/events, L: location, T: time), the  $AUC_{Loss}$  indicates the difference between the mean AUC over the resampling iterations when the full situation cue feature set was used minus the AUC when the situation cue feature set without the Level-1 cue category of interest was used (see squares). AUC = area under the curve.

between pOsitivity and Sociality ( $r_\phi = 0.14$ ). Based on this pattern of intercorrelations between both self-reported and predicted DIAMONDS scores, we believe that the LASSO models for the different DIAMONDS dimensions did learn situation characteristic-specific association patterns in smartphone data. Therefore, it is likely that the important cues (see Table 3 and Figure 7) in our models overlap primarily because the targets (i.e., situation characteristics) for which the models were trained also exhibit conceptual overlap, and, thus, it is only reasonable that they *share* important cues.

## Discussion

In the present study, we employed a longitudinal, multimethod approach to assess the objective realities and perceived characteristics of situations encountered in everyday life by an age- and gender-representative sample. To account for the complexity of daily situations, we collected a large variety of cues extracted from smartphone sensing data, which we used to predict ESs of situation experiences in terms of the DIAMONDS dimensions in a machine learning framework. In a series of follow-up analyses, we opened the black box of machine learning algorithms and explored which patterns of association between situation cues and characteristics prove robust across situations (see Appendix C, for a compact overview of our main findings).

Our results demonstrate that objective sensing data contain predictive information about several dimensions of self-reported situation perception, namely Duty, Intellect, Mating, pOsitivity, and Sociality. Thereby, different groups of smartphone features varied in their contribution, with situation cues, and, among them, time and location cues, being particularly predictive for the perception of everyday situations. While the psychological experience of situations was, across dimensions, associated with highly similar types and constellations of cues, this overlap only reflects the construct similarity between the situation characteristics, so our models did learn characteristic-specific association patterns. In the following sections, we discuss these main results in the context of literature, and we highlight the empirical evidence our study contributes to the mapping of cues and characteristics of situations in daily life.

### Differential Predictability Across Situation Characteristics

To get an impression of *how much* situational information smartphone sensing data generally provide with regard to situation experience, we used a broad range of cues from different sensing modalities to predict each of the situational eight DIAMONDS dimensions.

Overall, the magnitudes of our successful prediction performances ( $r_\phi = 0.08$ – $0.30$ ) were, on average, below those found when predicting personality traits from smartphone sensing data aggregated over several weeks ( $r = 0.20$ – $0.40$ ; Schoedel et al., 2018; Stachl et al., 2020), but comparable to those found when predicting personality states from smartphone sensing data aggregated over shorter time windows (i.e., 30 min;  $r = 0.12$ – $0.26$ ; Rügger et al., 2020). Given the closer proximity in design, the latter study may provide a more meaningful comparison to our findings.

Also in line with previous research applying machine learning techniques in psychology (e.g., Jacobucci et al., 2021; Pargent & Albert-von der Gönna, 2018; Rügger et al., 2020), the simpler

linear models (LASSO) performed no worse than the complex random forest models in our study.

Focusing on the easier-to-interpret linear models, we found that smartphone sensing data were not equally informative about the different DIAMONDS dimensions—a finding previously reported in a study using Tweets as digital traces of situations (Serfass & Sherman, 2015). In more detail, Duty obtained the highest prediction performance in our study, followed by Intellect, Sociality and Mating, and, finally, pOsitivity. Adversity, followed by Negativity and Deception, obtained the worst performances. For the latter DIAMONDS dimensions, the models were, thus, not able to capture any systematic information in the cues to make predictions. Of the five successfully predicted dimensions, pOsitivity was predicted the worst, indicating that the LASSO model also had some difficulty in capturing systematic variance in pOsitivity based on the smartphone data. Thus, whether a situation characteristic can be predicted may not be a clear-cut decision but rather judgment based on a continuum of predictability.

### Consensual Versus Idiosyncratic Variance of Situation Characteristics

One reason for the differential predictability patterns of situation characteristics may be that some DIAMONDS dimensions are more strongly based on the consensual interpretation of cues, while others depend more on the perceivers' subjective (i.e., idiosyncratic) interpretation of cues (Rauthmann et al., 2014; Rauthmann & Sherman, 2021). In other words, when perceiving certain situation characteristics, the objective cues from the situational reality may undergo a higher degree of idiosyncratic interpretation than when perceiving other characteristics (see Block & Block, 1981; Rauthmann, 2012; Rauthmann, Sherman, & Funder, 2015; Rauthmann & Sherman, 2021; Serfass & Sherman, 2013). For example, participants' shared meaning systems (e.g., cultural norms; Block & Block, 1981; Kenny, 1988; Rauthmann, Sherman, & Nave, 2015) may have led them to consensually interpret situations taking place *at the workplace* as dutiful. In contrast, participants may have perceived the cue *at the workplace* differentially when forming mental representations of the pOsitivity or Negativity of a situation. Thereby, person factors such as participants' level of Conscientiousness might have played a role in cue perception. In conclusion, we assume that different DIAMONDS dimensions achieved different levels of prediction performances because our models learned only the consensual interpretation of cues across participants and situations, while situation perceptions also—to different degrees—contain idiosyncratic variance components.

To underpin this assumption, we combined different machine learning analyses to find out which variance components our models had learned. By their methodological design, our models were trained on the level of observations. They neither knew which observations belonged to the same participants nor did they have any external information on stable person factors (e.g., by including personality traits as features). Thus, our models might have confounded patterns on the within- and between-person level when making predictions. In particular, the models may have learned patterns on the person-level to predict (interindividual) idiosyncratic variations in situation perception that are related to participants' stable dispositions (e.g., personality traits). However, our follow-up CWC(*M*) analyses ruled out this possibility by

showing that our models learned mostly within- but no between-person patterns. Idiosyncratic variance related to fluctuating factors represented noise to our models and could not be learned. That would, for example, be the case if there is one instance where a situation *at the workplace*, which would be normatively considered to contain Duty, is not experienced as dutiful (e.g., because the perceiver is for once in a very good mood). Thus, based on the models' functionality and on our additional analyses, it seems plausible that our models learned only the consensual meaning of cues across situations and, therefore, predicted only the consensual variance component of situation characteristics. This would suggest that the well-predictable dimensions are those with higher consensual variance proportions, while the not-so-well-predictable dimensions contain more idiosyncratic variance. This reasoning is consistent with previous findings that the interrater agreement between in situ raters (i.e., those directly experiencing the situation) and ex-situ raters (i.e., those not directly involved in a situation) is higher for certain DIAMONDS dimensions (e.g., Duty, Sociality) than for others (e.g., Adversity, Deception; Rauthmann et al., 2014; Rauthmann & Sherman, 2021). In this sense, our cue-based predictions and ex-situ ratings of situation characteristics both represent how situations are consensually and not how they are idiosyncratically interpreted (Rauthmann, 2012; Rauthmann, Sherman, & Funder, 2015).

### ***Differential Coverage of Situation Characteristics***

As an alternative explanation, one may argue that the differential predictability patterns of the DIAMONDS dimensions are a methodological artifact arising from the coverage of characteristics and cues across situation snapshots. On the one hand, the base rates varied considerably between the eight situation characteristics, whereby those with the lowest base rates were not or were less well-predicted. That could be particularly true for Adversity and Deception, both of which were rated as absent in the great majority of situational snapshots in our sample, confirming the low variances repeatedly found for these dimensions in past studies (see Jonason & Sherman, 2020; Rauthmann et al., 2014; Sherman et al., 2015). This absence of situations high in Deception and Adversity may reflect an actual (fortunate) lack of such situations in everyday life or may be a methodological artifact introduced by a sampling bias if participants do not answer ESs in these situations across studies. Either way, we countered this underrepresentation by applying specific weighting techniques for imbalanced class distributions in our data analyses, as explained in our methods section (see also Sterner et al., 2021). Thus, we can rule out that the differential predictability of the eight DIAMONDS dimensions was merely a methodological artifact of the algorithms applied in our study. However, the low coverage of some characteristics in everyday situations is, nevertheless, a common methodological issue in current situation research, causing a lack of knowledge on these characteristics.

On the other hand, cue coverage might have also contributed to the differential pattern of predictability across the DIAMONDS dimensions. That was the case if some situation characteristics were predicted better than others due to the types of cues inspected in this study. Even though we extracted an extensive set of situation, smartphone, and environment cues, they may not have fully covered the entire objective parameter space of daily situations and, therefore, may have lacked information relevant to

the not well-predicted DIAMONDS dimensions. For example, the situation cue *presence of a mate or spouse* was associated with less Negativity in situation experiences in previous studies (Rauthmann et al., 2014). While our smartphone data captured social interactions, they only covered smartphone-mediated types of interactions (e.g., social media or communication app use or calls) and lacked further details on the interaction, such as the type of interaction partner (e.g., a friend vs. colleague).

### **The Role of Different Cues**

As of today, situation research has mainly focused on situation characteristics but less on the physical reality of situations (e.g., Horstmann et al., 2021; Kritzler et al., 2020; Rauthmann et al., 2014). This imbalance may have been caused—at least in part—by difficulties in the objective assessment of cues. However, with the advent of smartphone-based data collection, the tables have now started to turn (Harari, Müller, & Gosling, 2020; Harari, Vaid, et al., 2020). Even though cues have no psychological meaning per se, they lay the foundation for forming psychological situation representations and, thus, could (indirectly) help to explain individual differences in thoughts, feelings, and behaviors (Rauthmann & Sherman, 2021). Therefore, our study applied a series of interpretable machine learning techniques to explore *what* situation information in smartphone sensing data (i.e., cues) was informative for predicting situation experience.

### ***The Relevance of Situation Cues***

We found that the successful prediction of situation characteristics relied on a broad range of different smartphone-sensed cues ranging between 22 and 46 per DIAMONDS dimension (see Table 3). Thus, analogous to personality prediction research (Schoedel et al., 2018; Stachl et al., 2020; Sust et al., 2022), single cues contributed little information on their own, while their holistic constellation was much more informative.

For a better overview of the importance of cues, we assigned them to theory-driven groups to understand if certain types of cues and their constellations were generally more informative for situation characteristics. When considering individual feature importance, we found that cues of all three groups (situation, smartphone, and environment cues) were among the most informative predictors across the DIAMONDS dimensions. However, when considering groups directly, the most dominant features were situation cues.

This finding is good news for situation research because it serves as a data-intensive proof of concept for the long-help assumption that objective information in a given situation has no intrinsic psychological meaning unless it is being processed by the perceiver (Miller, 2007; Rauthmann & Sherman, 2021). In more detail, situation researchers postulate that any situational information must be processed by the human perceptual system in order to contribute to the psychological representation of the situation and, consequentially, to play a role in thoughts, feelings, and behaviors (Rauthmann et al., 2014). In our study, such conscious processing was only certain for the group of situation cues (e.g., *being at the workplace or on weekend time*)—which are traditionally defined as highly salient short-term aspects of the situation that are likely being processed (Rauthmann, Sherman, & Nave, 2015; Rauthmann, 2021; Rauthmann & Sherman, 2021). In contrast to these well-established

features, our two remaining cue groups contained technically more complex features that our study had newly introduced to situation research based on the availability of smartphone sensing data. Features from these two groups of smartphone and environment cues represented less salient information, such as the *number of changes in the Wi-Fi status* during a given situation or the *density of inhabitants of the city* a situation occurred in and were likely not consciously processed by participants. Thus, our finding that these more subtle groups of situation cues were less informative for the prediction of situation characteristics points to the relevance of salience and conscious processing of cues when forming situation experiences.

However, our analyses also show that these two cue groups were not completely uninformative. This is best illustrated by the example of Sociality, whose most relevant individual features contained both smartphone and environment cues. In addition, grouped feature importance showed that all three cue groups were equally informative about Sociality. Overall, the low, albeit nonnegligible relevance of smartphone and environment cues could indicate that some of their features, despite not being perceptible per se, may contain information about other more salient (situation) cues. For example, smartphone cues such as the *minimum total duration of screen sessions* during a given situation might be a proxy for using the smartphone less when other persons are present. Persons, such as family or friends, in turn, have been found to be salient situation cues for perceiving Sociality (Blake et al., 2020). To summarize, within smartphone sensing data, salient, and perceptible situation cues proved to be most meaningful for predicting situation characteristics, while the remaining, more technically complex cue groups contained far less relevant information.

### ***Time and Location as Strong Cues***

When further inspecting the relevance of different features from the group of situation cues, our results show that time and location cues were most informative for several situation characteristics. This is in line with previous research consistently relating cues of time and location, such as evening, home, or office/university, to situation characteristics (Blake et al., 2020; Rauthmann et al., 2014; Rauthmann & Sherman, 2016a).

One reason for the superiority of these two groups of situation cues across studies could be that they are particularly prone to consensual perception. As discussed earlier, we argue that DIAMONDS dimensions were not equally well-predicted because our models learned primarily consensual variance components—which vary between characteristics. This line of interpretation, in turn, suggests that situation cues were particularly informative if they provoked a stronger consensual perception. Put more simply: Individuals perceive and interpret time and location cues consensually (e.g., as Duty) because they share common mental situation schemata, arising, for example, from cultural norms (Block & Block, 1981; Kenny, 1988; Rauthmann, Sherman, & Funder, 2015). As an example, most people go to work to make a living, so they strongly associate locations such as the workplace with the perception of Duty.

This reasoning also aligns well with the situational strength concept (Mischel, 1977; Snyder & Ickes, 1985): Strong situation cues override individual differences in people's perception and lead to agreement in the ratings for situation characteristics, causing consensual variance. In contrast, weak cues leave room for person

parameters to manifest in subjective situation interpretations causing idiosyncratic variance. Accordingly, based on our results, we consider time and location as strong cues.

Having the concept of situational strength and our results in mind, we believe that research in personality and social psychology could greatly benefit from establishing the collection of time and location cues in future studies. The differentiation between strong and weak situations (enabled by collecting time and location as covariates), in turn, might foster our understanding of person-situation interactions and resulting behaviors (Funder, 2006; Lewin, 1936). Given the increasing number of commercial research apps, which can readily log time and are beginning to integrate GPS logging, the assessment of these variables via passive sensing has become easier than ever.

### ***The Absence of Interaction Cues***

With regard to the role of cues, we were somewhat surprised that interaction cues were not informative for Sociality while activity and location cues were. Based on the DIAMONDS' theoretical conception (Rauthmann et al., 2014), we would have expected interaction cues to be most informative about whether a situation enables or requires social interactions. Accordingly, past findings repeatedly related the perception of a situation's Sociality to the presence of others and communication activities (Blake et al., 2020; Rauthmann et al., 2014; Rauthmann & Sherman, 2016a). In light of these seemingly contradictory findings, Sociality may depend on real-world but not smartphone-mediated social interactions. In line with this reasoning, situations were rated as containing Sociality if they occurred on the way (i.e., not at home), at a rather popular place (i.e., where participants potentially meet others), or if the phone's minimum screen time was low (i.e., so that participants were not using their phone), that is, when they were not using their phone for communication. Note that this post hoc explanation should not be generalized without further confirmation. Future research could, for example, include smartphone-sensed situation cues indicating real-life social interactions, such as conversations detected via microphone sensors (Harari, Müller, & Stachl, 2020).

### ***Cue Overlap Between Models***

Finally, we found that different situation characteristics "shared" similar informative cues. Especially Duty and Intellect showed great overlap in the type of cues and their constellation relevant to predictions. Compared to the duo of Duty and Intellect, the degree of cue similarity decreased from Mating to pOsitivity and, finally, Sociality. In line with our findings, past studies also found correlational patterns relating pairs of situation characteristics to common situation cues (e.g., Blake et al., 2020; Rauthmann et al., 2014; Rauthmann & Sherman, 2016a). For example, Blake et al. (2020) also reported similarities between Duty and Intellect. Based on these past findings and additional analysis to check if our models learned situation characteristic-unspecific patterns in the smartphone data, we decided that the common cues and constellations were not an artifact of our method. Because the corresponding pairs of self-reported situations characteristics also exhibit moderate-to-high intercorrelations, both in our data and in the literature (e.g., Abrahams et al., 2021; Horstmann & Ziegler, 2019; Rauthmann et al., 2014; Rauthmann & Sherman, 2016a), it

seems most likely that the overlap between situation characteristics in terms of their “shared” cues reflects the conceptual similarity between different dimensions of situation experience. Therefore, we conclude that the psychological representation of situations might be based on similar situation cues and their constellation but still in a (more or less) specific manner for each dimension of situation perception. This finding, again, highlights the complexity of mapping characteristics and cues of situations in daily life. As an outlook, more nuanced differences between specific prediction models may occur when considering the full space of situation cues and their constellations, which, however, is beyond the scope of this report.

## Implications for Situation Research

### *Predictive Value of Psychological Situation Theories*

Psychological theories have recently been criticized for their exclusive focus on the development of mechanistic and complex models for explaining and understanding psychological phenomena that have little (or unknown) predictive abilities (Yarkoni & Westfall, 2017). Our study contributes to this debate by testing the predictive ability of the theoretical assumptions underlying the Situational Eight DIAMONDS taxonomy by Rauthmann et al. (2014; Rauthmann, Sherman, & Funder, 2015). Our findings provide evidence that the theories on the psychological situation do not only hold explanative but also predictive value. Of course, based on our prediction findings, we cannot make any causal claims or rule out the role of any third variables. But still, we demonstrated that perceived situation characteristics are indeed associated with the physical realities of a given situation.

### *Situation Characteristics Prediction Models for Situation Research*

Our study provides pretrained and evaluated models that future situation researchers may apply to new smartphone sensing data to make automated predictions on situation characteristics without repeatedly asking participants for their ratings. However, users of our models should be aware of some considerations. First, as discussed in detail above, our models learned consensual variance components. Thus, they can only predict consensual situation perceptions (i.e., how individuals generally interpret situations) when applied to new data but *not* the subjective meaning contained in idiosyncratic variance shares (e.g., Rauthmann, 2012; Rauthmann, Sherman, & Funder, 2015; Serfass & Sherman, 2013). Consequently, the application of our models may be particularly interesting for social psychological research questions but also for personality psychologists studying the personality triad. For example, one could use our models to predict the consensual Duty of a situation and assess how an individual’s self-reported perception of Duty differs. These discrepancies reflect the idiosyncratic meaning of the situation and may, in turn, relate to behavioral and person variables. In contrast to consensual predictions, our models should not be used for single-case diagnostics of situation perceptions because they cannot account for the subjective components of situation characteristics.

In addition, even though we rigorously trained and evaluated our prediction models on separate data sets, for quality assurance, we recommend that future researchers collect their own validation data set. They should compare predictions from our models to the self-reported

DIAMONDS ratings in new studies to see whether the prediction performances on new samples are comparable to the ones reported in this article. That may be particularly relevant if the new sample deviates in composition from our age- and gender-representative German sample. Moreover, as discussed in the methods section, our data collection took place during the COVID-19 pandemic so we cannot exclude that the scope of situations participants encountered in daily life was biased by the pandemic. Future research should, therefore, carefully check if our models generalize to different phases of the pandemic and the “new normal.” Finally, researchers should also test how well our models generalize to study designs with different situation sampling schedules (e.g., predicting situation characteristics every hour instead of only two to four times a day).

### *Smartphone Sensing in Situation Research*

Recently, scholars have highlighted the potential of smartphone sensing for situation research (Harari et al., 2015; Harari, Müller, & Gosling, 2020; Wrzus & Mehl, 2015, 2020). In light of our findings, we agree that smartphone sensing offers great opportunities to assess situation cues in an objective and ecologically valid manner. Thereby, smartphone sensing provides an immense variety of situation cues and other novel cues, which cannot only serve for predictive modeling but also as a starting point for confirmatory research in the future. We made some suggestions in the previous sections, such as focusing on the situation strength concept (Mischel, 1977; Snyder & Ickes, 1985).

At the same time, we think that smartphone sensing data can only tell a part of the story of everyday situations because situation cues per se have no psychological meaning (Rauthmann, Sherman, & Funder, 2015). Therefore, in our opinion, the combination of smartphone sensing and self-report-based data collection tools such as ES can provide the greatest added value for situation research, as it allows for the collection of both objective *and* subjective situation parameters. In summary, we argue that smartphone sensing is a very fruitful addition to, but not a replacement for, previously established data collection approaches in situation research.

### Limitations and Outlook

Our study faced some limitations that may be a good starting point for future research. First, we cannot rule out that our study suffered from biases in the sampling of persons and their daily situations. Regarding person-related biases, our sample is representative of the German population and, thus, avoids typical sampling biases (e.g., in favor of young age) in personality science. Nevertheless, it is still drawn from a westernized, educated, industrialized, rich, and democratic (i.e., Henrich et al., 2010) society, so our findings may not generalize well to other societies. Moreover, our approach of employing smartphones to study psychological situations is generally limited in its applicability to less developed countries due to a lack of smartphone penetration (see Newzoo, 2022). A second person-related bias is that our sample comprised only users of Android smartphones, excluding those owning iOS devices. However, following the past literature comparing users of different operating systems, we believe that possible selection biases for demographic and personality traits are negligible (at least for the German population, Götz et al., 2017; Keusch et al., 2020). Concerning biases in the situation sampling, it should be noted that participants might have carried their smartphones

in many but not every situation of daily life and, thus, might have provided sensing data and answered ES questionnaires only in selected situations. By implementing a sophisticated ES presentation algorithm proposed by van Berkel et al. (2019), we minimized the risk of selectively sampling only certain situations in our study (e.g., only during working hours). Nevertheless, we may have missed situational realities (e.g., at a church service or in the sauna) and psychological situations (e.g., those high in Negativity or Deception) that were incompatible with carrying a phone and/or answering ES questionnaires. However, this situation selection bias generally applies to all studies sampling natural situations from smartphones and, thus, is not specific to the methodological innovations of our study.

A second limitation of our study relates to our definition of the time window around ES questionnaires to aggregate circa-situm features. The temporal scope and position of this window resulted “naturally” for us from the selected ES schedule. With this approach, we investigated (representative) snapshots of situations but, strictly speaking, no delimited situations. As the length of situations can vary widely (Rauthmann & Sherman, 2016b), we cannot rule out that our circa-situm features aggregated situation cues over different situations or missed the beginning or end of situations. Future studies could overcome this issue by explicitly asking participants to report the beginning and end of their current situation in the respective ES questionnaires.

Third, while our study covered a very wide range of situation cues, we still see room for improving the scope and resolution of extracted situation cues with the advancement of smartphone sensing technology in the future. For example, grooming and eating were identified as associated cues across different situation characteristics in previous studies (Blake et al., 2020; Rauthmann et al., 2014). These situational activity cues could not be operationalized in our study—except via summarizing the use of thematically related apps (e.g., shopping apps for drugstores, food service apps). In this context, image processing techniques could be applied to extract high-resolution situation cues from the information visible on persons’ screens (e.g., the exact food order within an app) or photos taken (e.g., the big pizza just photographed; Brinberg et al., 2021). In addition, more sophisticated analyses of audio recordings, such as the identification of speakers or classification of ambient sounds (see Harari, Müller, & Gosling, 2020), could produce more features from person/interaction cues which, in turn, might improve predictive performances in some dimensions (e.g., Sociality). Moreover, situation cues from smartphones could be complemented by data from other sources, such as smartwatches, to create a wider range of situation cues on events or activities (e.g., doing sports or sleeping). Expanding the smartphone-sensed situation cues may eventually provide information relevant to DIAMONDS dimensions that are not predictable in our study.

Fourth, we predicted each dimension of the Situational Eight DIAMONDS separately in one binary classification task. Even though we carefully checked the validity of our binary items, this artificial separation may still not do full justice to the dimensional conception of situation perceptions (e.g., Rauthmann et al., 2014; Rauthmann & Sherman, 2016a). In addition, as situation characteristics are interrelated, it will be an interesting extension of our research to predict all DIAMONDS at once while accounting for their intercorrelation structures (i.e., multitarget prediction, Au et al., 2019) once the corresponding statistical machine learning techniques are ready for implementation.

Finally, our data set was collected at a time when legislative measures were in place in response to the COVID-19 pandemic, so we cannot rule out that participants’ daily situations (in terms of cues and characteristics) were affected by the pandemic (see Kuper et al., 2021). Based on a descriptive analysis, we showed that our data collection at least took place during a period of loosened restrictions in Germany and that the restrictions still in place were comparable across German federal states, that is, across our participants. Nevertheless, people may have encountered social situations primarily at work but not during leisure time due to social distancing. Also, the restriction measures may have altered their mobility behaviors, and these adaptations probably depended on individual differences in personality traits such as Conscientiousness (Chan et al., 2021; Elarde et al., 2021). Therefore, our findings should be interpreted against the background of a very special global situation, namely the COVID-19 pandemic, and future research may want to replicate our findings when life is back to the “new normal.”

## Conclusion

The present work employed a longitudinal multimethod approach to demonstrate that everyday situation’s objective reality captured via smartphone sensing can predict different dimensions of psychological situation experience (Duty, Intellect, Mating, pOsitivity, and Sociality). Thereby, different groups of smartphone features varied in their level of informativeness: Salient situation cues, particularly time and location, were most informative for the perception of everyday situations. Even though different dimensions of situation perception overlapped in terms of their types and constellations of associated cues, they still showed differential association patterns with these shared situation cues.

In sum, our findings provide new insights into the mapping of situation cues and situation characteristics across real-life situations. Our suggested methodological framework accounts for the complex nature of daily situations and contributes to advance future research on psychological situations.

## References

- Abrahams, L., Rauthmann, J. F., & Fruyt, F. D. (2021). Person-situation dynamics in educational contexts: A self- and other-rated experience sampling study of teachers’ states, traits, and situations. *European Journal of Personality*, 35(4), 598–622. <https://doi.org/10.1177/08902070211005621>
- 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), 3–25. <https://doi.org/10.1037/amp0000191>
- Arendasy, M. (2009). *BFSI: Big-Five Struktur-Inventar* [BFSI: Big-Five Structure Inventory] (Test & Manual). SCHUHFRIED GmbH.
- Au, Q., Schalk, D., Casalicchio, G., Schoedel, R., Stachl, C., & Bischl, B. (2019). *Component-wise boosting of targets for multi-output prediction*. arXiv. <https://doi.org/10.48550/ARXIV.1904.03943>
- Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health*, 22(4), 153–160. <https://doi.org/10.1136/ebmental-2019-300117>
- 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(EICS), 1–24. <https://doi.org/10.1145/3397872>

- 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>
- Block, J., & Block, J. H. (1981). Studying situational dimensions: A grand perspective and some limited empiricism. In D. Magnusson (Ed.), *Toward a psychology of situations: An interactional perspective* (pp. 85–106). Lawrence Erlbaum.
- Böhrmer, M., Hecht, B., Schöning, J., Krüger, A., & Bauer, G. (2011). *Falling asleep with Angry Birds, Facebook and Kindle: A large scale study on mobile application usage* [Conference session]. 13th International Conference on Human Computer Interaction with mobile devices and services—MobileHCI '11, Stockholm, Sweden. <https://doi.org/10.1145/2037373.2037383>
- Bouckaert, R. R., & Frank, E. (2004). Evaluating the replicability of significance tests for comparing learning algorithms. In H. Dai, R. Srikant, & C. Zhang (Eds.), *Advances in knowledge discovery and data mining* (Vol. 3056, pp. 3–12). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-24775-3\\_3](https://doi.org/10.1007/978-3-540-24775-3_3)
- 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>
- Brinberg, M., Ram, N., Yang, X., Cho, M.-J., Sundar, S. S., Robinson, T. N., & Reeves, B. (2021). The idiosyncrasies of everyday digital lives: Using the human screenome project to study user behavior on smartphones. *Computers in Human Behavior*, 114, Article 106570. <https://doi.org/10.1016/j.chb.2020.106570>
- 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 & 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(3). <https://doi.org/10.1177/1474704915593662>
- Buss, D. M. (1987). Selection, evocation, and manipulation. *Journal of Personality and Social Psychology*, 53(6), 1214–1221. <https://doi.org/10.1037/0022-3514.53.6.1214>
- Buss, D. M. (2009). An evolutionary formulation of person–situation interactions. *Journal of Research in Personality*, 43(2), 241–242. <https://doi.org/10.1016/j.jrp.2008.12.019>
- Canzian, L., & Musolesi, M. (2015). *Trajectories of depression: Unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis* [Conference session]. ACM International Joint Conference on Pervasive and Ubiquitous Computing—UbiComp '15, Osaka, Japan. <https://doi.org/10.1145/2750858.2805845>
- Caspi, A., & Roberts, B. W. (2001). Personality development across the life course: The argument for change and continuity. *Psychological Inquiry*, 12(2), 49–66. [https://doi.org/10.1207/S15327965PLI1202\\_01](https://doi.org/10.1207/S15327965PLI1202_01)
- 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 & Personality Science*, 12(6), 1018–1029. <https://doi.org/10.1177/1948550620952572>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), Article 6. <https://doi.org/10.1186/s12864-019-6413-7>
- Chirico, M. (2020). *Package 'geohashTools'*. <https://cran.r-project.org/web/packages/geohashTools/index.html>
- Conner, T. S., & Mehl, M. R. (2015). Ambulatory assessment: Methods for studying everyday life. In R. A. Scott & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences* (1st ed., pp. 1–15). Wiley. <https://doi.org/10.1002/9781118900772.etrds0010>
- de Raad, B. (2004). Situations that matter to personality. In A. Elias, S. E. Hampson, & B. de Raad (Eds.), *Advances in personality psychology* (Vol. 2, pp. 179–204). Psychology Press.
- Destatis. (2021). *Städte (Alle Gemeinden mit Stadtrecht) nach Fläche, Bevölkerung und Bevölkerungsdichte am 31.12.2020*: © Daten (im Auftrag der Herausbergemeinschaft Statistische Ämter des Bundes und der Länder) [Data set]. Statistisches Bundesamt (Destatis). <https://www.destatis.de/DE/Themen/Laender-Regionen/Regionales/Gemeindeverzeichnis/Administrativ/05-staedte.html>
- Destatis. (2023). *Erwerbstätige im Home Office*. Statistisches Bundesamt. [https://www.destatis.de/DE/Presse/Pressemitteilungen/Grafiken/Arbeitsmarkt/2022/\\_Interaktiv/20220614-zdw-homeoffice-erwerbstaetige.html](https://www.destatis.de/DE/Presse/Pressemitteilungen/Grafiken/Arbeitsmarkt/2022/_Interaktiv/20220614-zdw-homeoffice-erwerbstaetige.html)
- 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 smartphones* [Conference session]. 13th International Conference on Ubiquitous Computing—UbiComp '11, Beijing, China. <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>
- Dragevic, M., & Casalicchio, G. (2020). *Resampling—Stratified, blocked and predefined*. Mlr-Org. <https://mlr-org.com/gallery/basic/2020-03-30-stratification-blocking/>
- Edwards, J. A., & Templeton, A. (2005). The structure of perceived qualities of situations. *European Journal of Social Psychology*, 35(6), 705–723. <https://doi.org/10.1002/ejsp.271>
- 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), Article e0259031. <https://doi.org/10.1371/journal.pone.0259031>
- Ender, N. S. (1981). Situational aspects of interactional psychology. In D. Magnusson (Ed.), *Toward a psychology of situations: An interactional perspective* (pp. 361–373). Lawrence Erlbaum.
- Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, 13(2), 83–87. <https://doi.org/10.1111/j.0963-7214.2004.00280.x>
- Fleeson, W. (2007). Situation-based contingencies underlying trait-content manifestation in behavior. *Journal of Personality*, 75(4), 825–861. <https://doi.org/10.1111/j.1467-6494.2007.00458.x>
- Funder, D. C. (2001). Personality. *Annual Review of Psychology*, 52(1), 197–221. <https://doi.org/10.1146/annurev.psych.52.1.197>
- 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>
- Funder, D. C. (2016). Taking situations seriously: The situation construal model and the riverside situational Q-sort. *Current Directions in Psychological Science*, 25(3), 203–208. <https://doi.org/10.1177/0963721416635552>
- 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(4), 716–742. <https://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), Article e0176921. <https://doi.org/10.1371/journal.pone.0176921>
- Griner, L. A., & Smith, C. A. (2000). Contributions of motivational orientation to appraisal and emotion. *Personality and Social Psychology Bulletin*, 26(6), 727–740. <https://doi.org/10.1177/0146167200268007>

- Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (oxford COVID-19 government response tracker). *Nature Human Behaviour*, 5(4), 529–538. <https://doi.org/10.1038/s41562-021-01079-8>
- 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., Müller, S. R., & Gosling, S. D. (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. 298–311). 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., Rentfrow, P. J., Campbell, A. T., & 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., Bühner, M., & 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>
- Harrer, M. (2022). *Doing meta-analysis with R: A hands-on guide* (1st ed.). CRC Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- 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. (2018). Measurement of situational influences. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *The Sage handbook of personality and individual differences* (Vol. 3, pp. 465–484). SAGE Publications.
- 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>
- Ickes, W., Snyder, M., & Garcia, S. (1997). Personality influences on the choice of situations. In J. A. Johnson & S. R. Briggs (Eds.), *Handbook of personality psychology* (pp. 165–195). Elsevier. <https://doi.org/10.1016/B978-012134645-4/50008-1>
- 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>
- Janke, S., & Glöckner-Rist, A. (2012). *Deutsche Version der Positive and Negative Affect Schedule (PANAS)*. Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS. <https://doi.org/10.6102/ZIS146>
- Jonason, P. K., & Sherman, R. A. (2020). Personality and the perception of situations: The big five and dark triad traits. *Personality and Individual Differences*, 163, Article 110081. <https://doi.org/10.1016/j.paid.2020.110081>
- Kemp, S. (2022, January 27). *DIGITAL 2021: Global overview report*. DATAREPORTAL. <https://datareportal.com/reports/digital-2021-global-overview-report>
- Kenny, D. A. (1988). Interpersonal perception: A social relations analysis. *Journal of Social and Personal Relationships*, 5(2), 247–261. <https://doi.org/10.1177/026540758800500207>
- 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*. Advance online publication. <https://doi.org/10.1177/0049124120914924>
- 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, Article 103929. <https://doi.org/10.1016/j.jrp.2020.103929>
- Kuhn, M. (2022). *caret: Classification and regression training* (package Version 6.0-93). <https://CRAN.R-project.org/package=caret>
- Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer New York. <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), Article 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>
- Kushlev, K., & Heintzelman, S. J. (2018). Put the phone down: Testing a complement–interfere model of computer-mediated communication in the context of face-to-face interactions. *Social Psychological & Personality Science*, 9(6), 702–710. <https://doi.org/10.1177/1948550617722199>
- 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), Article 1903. <https://doi.org/10.21105/joss.01903>
- Lang, M., & Schratz, P. (2022). *mlr3verse: Easily install and load the “mlr3” package family* (package Version 0.2.7). <https://CRAN.R-project.org/package=mlr3verse>
- Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In M. Csikszentmihalyi (Ed.), *Flow and the foundations of positive psychology* (pp. 21–34). Springer Dordrecht. [https://doi.org/10.1007/978-94-017-9088-8\\_2](https://doi.org/10.1007/978-94-017-9088-8_2)
- Lewin, K. (1936). *Principles of topological psychology*. McGraw-Hill. <https://doi.org/10.1037/10019-000>
- Magnusson, D. (1981). *Toward a psychology of situations: An interactional perspective*. Lawrence Erlbaum. [http://www.123library.org/book\\_details/?id=107209](http://www.123library.org/book_details/?id=107209)
- Majnik, M., & Bosnić, Z. (2013). ROC analysis of classifiers in machine learning: A survey. *Intelligent Data Analysis*, 17(3), 531–558. <https://doi.org/10.3233/IDA-130592>
- Maloof, M. A. (2003). *Learning When data sets are imbalanced and when costs are unequal and unknown* [Conference session]. ICML-2003 workshop on learning from imbalanced data sets II. <http://www.eiti.uotta.wa.ca/~nat/Workshop2003/maloof-icml03-wids.pdf>
- Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., & Roser, M. (2020). *Coronavirus pandemic (COVID-19)*. OurWorldInData.Org. <https://ourworldindata.org/coronavirus>
- 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), 1367–1385. <https://doi.org/10.1037/pspp0000297>
- 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), 862–877. <https://doi.org/10.1037/0022-3514.90.5.862>

- Mehl, M. R., & Robbins, M. L. (2012). Naturalistic observation sampling: The Electronically Activated Recorder (EAR). In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 176–192). Guilford Press.
- Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B. (2019). “LIWC auf Deutsch”: The development, psychometrics, and introduction of DE-LIWC2015. PsyArXiv. <https://doi.org/10.31234/osf.io/uq8zt>
- Miller, G. (2007). Reconciling evolutionary psychology and ecological psychology: How to perceive fitness affordances. *Acta Psychologica Sinica*, 39(3), 546–555.
- Mischel, W. (1977). The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology* (pp. 333–352). Lawrence Erlbaum.
- Mischel, W., & Shoda, Y. (1995). A cognitive–affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, 102(2), 246–268. <https://doi.org/10.1037/0033-295X.102.2.246>
- Morse, P. J., Sauerberger, K. S., Todd, E., & Funder, D. (2015). Relationships among personality, situational construal and social outcomes. *European Journal of Personality*, 29(2), 97–106. <https://doi.org/10.1002/per.1987>
- 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>
- Murray, H. A. (1938). *Explorations in personality: A clinical and experimental study of fifty men of college age*. Oxford University Press.
- Nadeau, C., & Bengio, Y. (1999). Inference for the generalization error. *Advances in Neural Information Processing Systems*, 12, 239–281. [https://proceedings.neurips.cc/paper\\_files/paper/1999/file/7d12b66d3df6af8d429c1a357d8b9e1a-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/1999/file/7d12b66d3df6af8d429c1a357d8b9e1a-Paper.pdf)
- Nakazawa, M. (2023). *fmsb: Functions for medical statistics book with some demographic data* (R package Version 0.7.5). <https://CRAN.R-project.org/package=fmsb>
- Newzoo. (2022). *Global games market report*. <https://newzoo.com/insights/rankings/top-countries-by-smartphone-penetration-and-users>
- Niemeyer, G. (2008). *Geohash*. <https://en.wikipedia.org/wiki/Geohash>
- Nystedt, L. (1981). A model for studying the interaction between the objective situation and a person’s construction of the situation. In D. Magnusson (Ed.), *Toward a psychology of situations: An interactional perspective* (pp. 375–391). Lawrence Erlbaum.
- 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>
- 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>
- Pervin, L. A. (1978). Definitions, measurements, and classifications of stimuli, situations, and environments. *Human Ecology: An Interdisciplinary Journal*, 6(1), 71–105. <https://doi.org/10.1007/BF00888567>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ramanujam, E., Perumal, T., & Padmavathi, S. (2021). Human activity recognition with smartphone and wearable sensors using deep learning techniques: A review. *IEEE Sensors Journal*, 21(12), 13029–13040. <https://doi.org/10.1109/JSEN.2021.3069927>
- Rauthmann, J. F. (2012). You say the party is dull, I say it is lively: A componential approach to how situations are perceived to disentangle perceiver, situation, and Perceiver × Situation variance. *Social Psychological & Personality Science*, 3(5), 519–528. <https://doi.org/10.1177/1948550611427609>
- Rauthmann, J. F. (2021). Capturing interactions, correlations, fits, and transactions: A person–environment relations model. In J. F. Rauthmann (Ed.), *The handbook of personality dynamics and processes* (pp. 427–522). Elsevier. <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., 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., & Sherman, R. A. (2016a). Measuring the Situational Eight DIAMONDS characteristics of situations: An optimization of the RSQ-8 to the S8\*. *European Journal of Psychological Assessment*, 32(2), 155–164. <https://doi.org/10.1027/1015-5759/a000246>
- Rauthmann, J. F., & Sherman, R. A. (2016b). Situation change: Stability and change of situation variables between and within persons. *Frontiers in Psychology*, 6, Article 1938. <https://doi.org/10.3389/fpsyg.2015.01938>
- Rauthmann, J. F., & Sherman, R. A. (2016c). 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. (2018a). S8-I. Situational Eight DIAMONDS - deutsche Fassung [Verfahrensdokumentation und Fragebogen]. *Leibniz-Institut für Psychologie (ZPID) (Hrsg.), Open Test Archive. ZPID*. <https://doi.org/10.23668/psycharchives.6568>
- Rauthmann, J. F., & Sherman, R. A. (2018b). The description of situations: Towards replicable domains of psychological situation characteristics. *Journal of Personality and Social Psychology*, 114(3), 482–488. <https://doi.org/10.1037/pspp0000162>
- Rauthmann, J. F., & Sherman, R. A. (2021). Conceptualizing and measuring the psychological situation. In D. Wood, S. J. Read, P. D. Harms, & A. Slaughter (Eds.), *Measuring and modeling persons and situations* (pp. 427–463). Elsevier. <https://doi.org/10.1016/B978-0-12-819200-9.00009-0>
- 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. (2008). Reinvigorating the concept of situation in social psychology. *Personality and Social Psychology Review*, 12(4), 311–329. <https://doi.org/10.1177/1088868308321721>
- Remus, R., Quasthoff, U., & Heyer, G. (2010). *SentiWS—A publicly available German-language resource for sentiment analysis* [Conference session]. Seventh International Conference on Language Resources and Evaluation (LREC’10), Valletta, Malta.
- Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7(4), 331–363. <https://doi.org/10.1037/1089-2680.7.4.331>
- Roy, A., & Pebesma, E. (2017). *A machine learning approach to demographic prediction using Geohashes* [Conference session]. Proceedings of the 2nd International Workshop on Social Sensing, Pittsburgh, PA, USA. <https://doi.org/10.1145/3055601.3055603>
- 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>
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178. <https://doi.org/10.1037/h0077714>
- Saucier, G., Bel-Bahar, T., & Fernandez, C. (2007). What modifies the expression of personality tendencies? Defining basic domains of situation

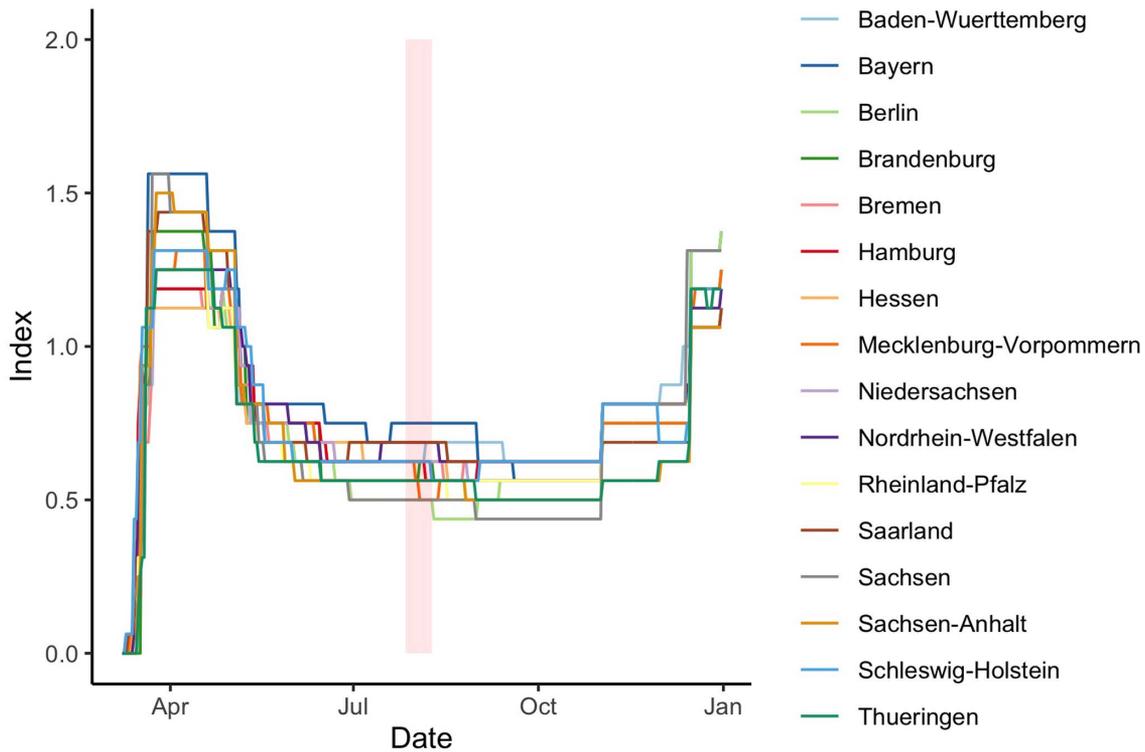
- variables. *Journal of Personality*, 75(3), 479–503. <https://doi.org/10.1111/j.1467-6494.2007.00446.x>
- Schoedel, R., Au, Q., Völkel, S. T., Lehmann, F., Becker, D., Bühner, M., Bischl, B., Hussmann, H., & Stachl, C. (2018). Digital footprints of sensation seeking: A traditional concept in the Big Data Era. *Zeitschrift für Psychologie*, 226(4), 232–245. <https://doi.org/10.1027/2151-2604/a000342>
- Schoedel, R., & 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). Systematic categorisation of 3,091 smartphone applications from a large-scale smartphone sensing dataset. *Journal of Open Psychology Data*, 10(1), Article 7. <https://doi.org/10.5334/jopd.59>
- Serfass, D. G., & Sherman, R. A. (2013). Personality and perceptions of situations from the thematic apperception test. *Journal of Research in Personality*, 47(6), 708–718. <https://doi.org/10.1016/j.jrp.2013.06.007>
- Serfass, D. G., & Sherman, R. A. (2015). Situations in 140 characters: Assessing real-world situations on Twitter. *PLOS ONE*, 10(11), Article 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* [Conference session]. 26th International Conference on World Wide Web Steering Committee, Perth, Australia. <https://doi.org/10.1145/3038912.3052618>
- 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>
- Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 3, pp. 883–947). Random House.
- 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 of the United States of America*, 117(30), 17680–17687. <https://doi.org/10.1073/pnas.1920484117>
- 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>
- Sterner, P., Goretzko, D., & Pargent, F. (2021). *Everything has its price: Foundations of cost-sensitive learning and its application in psychology*. PsyArXiv. <https://doi.org/10.31234/osf.io/7asgz>
- Sust, L., Stachl, C., Kudchadker, G., Bühner, M., & Schoedel, R. (2022). *Personality computing with naturalistic music listening behavior: Comparing audio and lyrics preferences*. PsyArXiv. <https://doi.org/10.31234/osf.io/xnv2q>
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Ushay, K. (2022). *renv: Project environments* (R package Version 0.15.5). <https://rstudio.github.io/renv/>
- van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys*, 50(6), 1–40. <https://doi.org/10.1145/3123988>
- 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>
- Viaene, S., & Dedene, G. (2005). Cost-sensitive learning and decision making revisited. *European Journal of Operational Research*, 166(1), 212–220. <https://doi.org/10.1016/j.ejor.2004.03.031>
- 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, R., Aung, M. S. H., Abdullah, S., Brian, R., Campbell, A. T., Choudhury, T., Hauser, M., Kane, J., Merrill, M., Scherer, E. A., Tseng, V. W. S., & Ben-Zeev, D. (2016). *CrossCheck: Toward passive sensing and detection of mental health changes in people with schizophrenia* [Conference session]. International Joint Conference on Pervasive and Ubiquitous Computing, New York, NY, United States. <https://doi.org/10.1145/2971648.2971740>
- 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), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>
- Wilt, J., & Revelle, W. (2019). The Big Five, everyday contexts and activities, and affective experience. *Personality and Individual Differences*, 136, 140–147. <https://doi.org/10.1016/j.paid.2017.12.032>
- Wrzus, C., & Mehl, M. R. (2015). Lab and/or field? Measuring personality processes and their social consequences. *European Journal of Personality*, 29(2), 250–271. <https://doi.org/10.1002/per.1986>
- 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. 311–327). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190263348.013.13>
- Wrzus, C., Wagner, G. G., & Riediger, M. (2016). Personality-situation transactions from adolescence to old age. *Journal of Personality and Social Psychology*, 110(5), 782–799. <https://doi.org/10.1037/pspp0000054>
- Yang, Y.-H., & Teng, Y.-C. (2015). Quantitative study of music listening behavior in a smartphone context. *ACM Transactions on Interactive Intelligent Systems*, 5(3), 1–30. <https://doi.org/10.1145/2738220>
- 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>
- Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models: Problems and solutions. *Organizational Research Methods*, 12(4), 695–719. <https://doi.org/10.1177/1094428108327450>
- 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>

(Appendices follow)

## Appendix A

### Supplementary Methods

**Figure A1**  
Index of COVID-19 Restrictions in 2020 Across Federal States in Germany



*Note.* Based on an open-access data set by Steinmetz et al. (2022), we calculated a daily index composed of 16 different kinds of governmental restrictions during the pandemic in 2020 for each federal state in Germany. The index scale ranges from 0 (*no restriction*) to 2 (*full restriction*). The red box marks the time period inspected in our study. See the online article for the color version of this figure.

**Table A1**  
*Overview of Previous Studies Used for Pooling Correlations*

References	Number of samples	Correlations between		
		DIAMONDS—DIAMONDS	DIAMONDS—Valence	DIAMONDS—Big Five traits
Abrahams et al. (2021)	1			x
Horstmann et al. (2021)	1	x	x	x
Horstmann and Ziegler (2019)	1	x	x	
Jonason and Sherman (2020)	1			x
Kritzler et al. (2020)	1	x	x	x
Rauthmann, Sherman, and Nave (2015)	2			x
Sherman et al. (2015)	1	x		x

*Note.* We only selected studies with empirical correlations based on in situ person-mean DIAMONDS ratings or studies that provided the raw data sets, so we were able to calculate these correlations on our own. If articles included more than one study, we included each study as single observation. DIAMONDS = Duty, Intellect, Adversity, Mating, pOsitivity, Negativity, Deception, Sociality.

(Appendices continue)

## Appendix B

### Supplementary Results

**Table B1**

*Prediction Performance Across Resampling Iterations by Situation Characteristics Targets and Models*

Target	LASSO					Random forest				
	$M_{AUC}$	$SD_{AUC}$	$t(99)$	$p_{corr}$	$r_{\phi}$	$M_{AUC}$	$SD_{AUC}$	$t(99)$	$p_{corr}$	$r_{\phi}$
D	<b>0.706</b>	<b>0.026</b>	<b>23.039</b>	<b>&lt;.001</b>	<b>0.302</b>	<b>0.709</b>	<b>0.025</b>	<b>23.888</b>	<b>&lt;.001</b>	<b>0.299</b>
I	<b>0.644</b>	<b>0.034</b>	<b>12.212</b>	<b>&lt;.001</b>	<b>0.177</b>	<b>0.633</b>	<b>0.035</b>	<b>11.089</b>	<b>&lt;.001</b>	<b>0.168</b>
A	0.522	0.085	0.765	1.000	0.015	0.531	0.079	1.142	1.000	-0.001
M	<b>0.588</b>	<b>0.035</b>	<b>7.283</b>	<b>&lt;.001</b>	<b>0.132</b>	<b>0.574</b>	<b>0.035</b>	<b>6.081</b>	<b>&lt;.001</b>	<b>0.052</b>
O	<b>0.560</b>	<b>0.031</b>	<b>5.550</b>	<b>&lt;.001</b>	<b>0.074</b>	<b>0.550</b>	<b>0.034</b>	<b>4.174</b>	<b>&lt;.001</b>	<b>0.031</b>
N	0.541	0.039	2.990	.028	0.044	0.508	0.034	0.714	1.000	-0.005
De	0.565	0.109	1.701	.736	0.040	0.587	0.101	2.480	.119	-0.001
S	<b>0.590</b>	<b>0.031</b>	<b>8.355</b>	<b>&lt;.001</b>	<b>0.129</b>	<b>0.588</b>	<b>0.028</b>	<b>9.005</b>	<b>&lt;.001</b>	<b>0.113</b>

*Note.* The area under the curve (AUC) was determined for each of the 100 resampling iterations of the  $10 \times 10$  CV scheme. This table provides means and standard deviations of the AUC across iterations. Per target variable, we applied variance-corrected one-sided  $t$  tests to compare each model (LASSO and random forest) against the baseline (AUC = 0.50) over the 100 iterations ( $df = 99$ ). We applied Bonferroni correction ( $n = 16$ ) to account for multiple testing of 8 targets  $\times$  2 models. Significantly predictive models ( $p < .001$ ) are marked in bold. We report the phi coefficient  $r_{\phi}$  as a measure of association between two binary variables (i.e., between self-reported and predicted situation characteristics).  $10 \times 10$  CV = 10-fold cross-validation; LASSO = least absolute shrinkage and selection operator.

**Table B2**

*Prediction Performance Across Resampling Iterations by Situation Characteristics Targets and Feature Transformation Approach*

Target	Feature transformation approach			
	Original	CWC	( $M$ )	CWC( $M$ )
Duty	0.71 (0.03)	0.70 (0.02)	0.52 (0.04)	0.68 (0.03)
Intellect	0.64 (0.03)	0.64 (0.03)	0.49 (0.05)	0.58 (0.05)
Mating	0.59 (0.04)	0.60 (0.03)	0.50 (0.05)	0.55 (0.06)
pOstivity	0.56 (0.03)	0.56 (0.02)	0.55 (0.06)	0.57 (0.05)
Sociality	0.59 (0.03)	0.58 (0.03)	0.52 (0.05)	0.55 (0.05)

*Note.* This table provides means and standard deviations (in brackets) of the AUC across resampling iterations. The AUC was determined for each of the 100 resampling iterations of the  $10 \times 10$  CV scheme. Different feature transformation approaches were applied: Original means that the feature set was not transformed (see also Table B1); CWC means that features were person-mean centered; ( $M$ ) means that the person mean of the features was used; CWC( $M$ ) is a set of combined CWC and ( $M$ ) features; AUC = area under the curve;  $10 \times 10$  CV = 10-fold cross-validation.

**Table B3**

*Intercorrelations of the Predicted Situation Characteristics*

Variable	Intellect	Mating	pOstivity	Sociality
Duty	0.82 (0.49)	-0.74 (-0.06)	-0.41 (-0.12)	0.19 (0.22)
Intellect	—	-0.70 (-0.03)	-0.38 (-0.12)	0.22 (0.21)
Mating		—	0.40 (0.20)	-0.18 (0.26)
pOstivity			—	0.14 (0.08)

*Note.* We trained the LASSO models on the full data set and extracted predicted scores for situation characteristics. We then computed phi correlations  $r_{\phi}$  as association measure between binary outcome variables at the observational level. For comparison, the phi correlations between the self-reported situation characteristics (also at the observational level) are presented in brackets. All coefficients presented ignore the multilevel data structure of observations nested within persons. LASSO = least absolute shrinkage and selection operator.

(Appendices continue)

## Appendix C

### Summary of Research Questions, Analytical Strategies, and Key Findings

Analytical strategy	Key finding
Research Question 1: Validity analysis of self-reported situation characteristics	
Are the self-reported situation characteristics valid despite the binary rating scale used?	
We compared our empirical association patterns among DIAMONDS and with nomological constructs with association patterns (in terms of pooled correlations) identified in the previous literature.	The association patterns generally matched and can therefore be considered as validity support for our adjusted DIAMONDS measure.
Research Question 2: Prediction of situation characteristics	
How much situational information do smartphone sensing data contain with regard to situation experience?	
We conducted a benchmark experiment, in which we used smartphone sensing data to predict self-reported DIAMONDS characteristics. Specifically, we compared different machine learning models against a baseline model and against each other.	The linear and nonlinear models performed equally well. We found a differential pattern of predictability for the different dimensions (in decreasing order): Duty, Intellect, Mating and Sociality, pOsitivity. Adversity, Deception, and Negativity could not be predicted.
We ran another benchmark to compare the LASSO models' performances when trained on not-transformed and transformed feature sets to get insights into whether within-person or between-person differences were learned by the LASSO models.	Our models learned within-person patterns for predicting the situation characteristics.
Research Question 3: Interpretation of situation characteristics predictions	
What situational information contained in smartphone sensing data drive the models' predictions?	
We inspected standardized $\beta$ regression coefficients extracted from the LASSO models to explore which single cues were important for the respective LASSO models' predictions.	Considered individually, features that were relevant to the different situation characteristics showed considerable overlap in terms of the type of the cues (situation cues: time and location).
We ran another benchmark with the full feature set and the feature sets each reduced by a Level-1/Level-2 cue category to explore the importance of cue groups (i.e., performance loss) for the respective LASSO models' predictions.	Considered as theory-driven cue groups, situation cues, and among situation cues, time and location were most important for predictions across all situation characteristics.
We inspected intercorrelations among the predicted situation characteristics to explore whether LASSO models learned distinct patterns to make predictions for each specific situation characteristic.	The magnitude of the intercorrelations varied pairwise across the different predicted situation characteristics. We concluded that the models shared similar cues to come to their predictions but learned situation characteristic-specific association patterns.

*Note.* DIAMONDS = Duty, Intellect, Adversity, Mating, pOsitivity, Negativity, Deception, Sociality; LASSO = least absolute shrinkage and selection operator.

Received August 9, 2022  
Revision received May 2, 2023  
Accepted May 5, 2023 ■