Capturing situational dynamics and processes: Strengths and pitfalls of the experience sampling method

Philipp K. Masur

University of Hohenheim

- PREPRINT -

Prepared for the conference proceedings of the 2017 annual meeting of the DGPK Methods section, “Dynamische Prozesse der öffentlichen Kommunikation: Methodische Herausforderungen”, edited by Stefan Geiß, Christian Schemer, Teresa K. Naab, Christina Peter, and Philipp Müller.

Correspondence concerning this article should be addressed to Philipp Masur, Department of Media Psychology, University of Hohenheim, Stuttgart, Germany. E-mail: philipp.masur@uni-hohenheim.de.

Abstract

Communication research has often either used self-report questionnaires or experimental designs to study communication phenomena. Scholars thereby either draw inferences about these phenomena through people's retrospection in surveys or by isolating particular factors and behaviors to identify causal mechanisms. Although both approaches have strengths, they nonetheless fail to capture dynamics and processes that results from different situationally-varying environmental and personal factors. This chapter therefore proposes a situational framework that allows to investigate non-situational and situational antecedents of perceptions or behaviors. It further describes the merits of using the experience sampling method (ESM) for testing such processes. Different types of data collection techniques and sampling strategies are presented and discussed. Using a real-world data set of 164 participants who completed 1,104 situational questionnaires, this chapter further illustrates how data obtained in ESM studies can be analyzed using a multilevel approach. Several aspects of this analytical approach such as power issues, centering decisions, random intercept and random slope models will be presented and discussed. Finally, an analytical strategy to test whether the potential burden of having to complete repeated measurements in ESM designs produce biases in participant's response patterns will be presented.

Keywords: Situational processes, experience sampling method, multilevel modeling, sampling procedures, bias in response patterns

1. Introduction

What we know about communication phenomena is often based on what people have told us in standardized, retrospective surveys. With the aim of measuring people's feelings, thoughts, attitudes, or behaviors, we often ask participants to remember past experiences and to aggregate them into a single set of coherent responses in a self-report questionnaire (cf. BARRETT/BARRETT 2001: 175). For example, scholars have asked people to indicate what they expect from using a certain medium and have tried to correlate these expected gratifications with participants' self-reported usage frequency (cf. uses and gratification approach; e.g., RUBIN 1973; LEUNG/WEI 2000; SMOCK/ELLISON/LAMPE/WOHN 2011). Similarly, another field of research has studied the frequency of online communication and its relation with measures of overall well-being (e.g., KRAUT ET AL. 1998; 2002; VALENZUELA/PARK/KEE 2009, DIENLIN/MASUR/TREPTE 2017). Survey designs are particularly adopted whenever scholars aim at identifying differences in personality factors as the source for differences in perceptions or behaviors (e.g., MOORE/MCELROY 2012; SEIDMAN 2013). In all fields of research, communication scholars have embraced and used survey methods because they are comparably easy to draft and implement, yet allow for interesting insights into interpersonal differences. Although findings from these studies are invaluable, there is an inherent caveat: Whenever we want to measure a certain behavior (that may vary across different situations), we have no choice but to ask participants to estimate their *general* behavior in retrospective surveys (typical survey question such as "How often do you usually...?", "On average, how willing are you to...?", "Please estimate how many times you use... per day?"). Such questions require individuals to aggregate several, potentially varying experiences of different situations (in which they remember having engaged in that particular behavior) into a coherent set of responses to predefined items. These aggregated measures flatten situational variance into one broad measure that may be indicative of a general tendency (although potentially biased due to difficulties in estimation and recollection), but at the same time, do not contain any information about potential deviations that may result from situational factors.

With regard to the examples mentioned above, we could ask: In how far can findings obtained in the tradition of the uses and gratification approach predict whether a person will watch TV in the evening on a particular day? We potentially know what this person generally expects from watching TV, but we also need to know what she currently feels and in particular which of the many needs that may be satisfied by watching TV are currently not satisfied. Going further, we might realize that we also would need to know if other persons are present, if the person routinely watches the show that runs on that particular evening, how she currently feels and so on. All these latter factors are inherently situational and vary from evening to evening, from day to day, or even from one situation to the next. With regard to the second example, the rationale is no different: Well-being (as many other concepts) has even been theorized as having state and trait components (DIENER/SUH/LUCAS/SMITH 1999). It is easily imagined that there are situations in which a lot of communication does not contribute to our well-being and others in which too little communication likewise negatively affects our well-being (e.g., whether communication positively influences our well-being surely depends on who we are communicating with).

Of course, not all research uses survey designs: Scholars have also studied communication phenomena by systematically varying the characteristics of a situation and observing changes in a person's affective, cognitive, or behavioral reactions. Experiments are thus valuable for measuring situational behavior, particularly because they allow to prove causality. However,

they likewise have several caveats (e.g., CZIENSKOWSKI 1996; BROSIUS/HAAS/KOSCHEL, 2016): Experiments are designed to maximize internal validity by isolating single effects and thereby ruling out alternative explanations. But they also reduce the external validity and thereby the generalizability of the findings. Experiments in laboratory settings generally represent highly artificial situations. More importantly, they only capture a small amount of possible intraindividual variance and cannot provide information about the complexity and diversity of situational dynamics. Experiments (whether between- or within-designs) mostly just represent one to three situation(s) among a multitude of similar or diverging situations encountered in daily life. Given that an individual is always influenced by a changing immediate external and social environment (cf. STONE/SHIFFMAN/ATIENZA/NEBELING, 2007: 7), this may result in findings that are difficult to generalize to everyday life situations.

We hence know that several environmental and personal factors are essential in determining whether a person shows a particular behavior or experiences a certain feeling (e.g., decides to use a certain medium, feels happy or sad...). In light of this, I argue that it is important to adopt a situational perspective in order to gain a more comprehensive understanding of many communication phenomena. That is not to say, however, that we should neglect dispositional factors. Instead, I suggest that we should acknowledge and consider *dispositional* (or trait-like personal factors), *contextual* and *situational* factors when trying to understand behavioral processes. Indeed, some environmental and personal factors are stable and thus non-situational, others are consistent across some, but not all situations, and some factors are different in each situation.

That being said, it currently seems that we are tempted to explain exclusively in dispositional terms (i.e. non-situational person-related factors) what ought to be understood largely in situational terms. In light of this, this paper discusses the experience sampling method (ESM), which can be regarded as an extensive longitudinal research methodology, designed to capture people's momentary feelings, thoughts, behaviors across different situations and as they are encountered in participant's daily lives. It thereby allows researchers to sample a broad range of variables in different situations (BARRETT/BARRETT 2001: 176). This chapter thus represents a primer, illustrating not only the promise and opportunities of ESMs in capturing situational dynamics, but also the challenges and difficulties associated with conducting such research. In the following, I will first provide a situational framework that allows to capture both non-situational and situational antecedents of perceptions or behaviors. I will then provide a more nuanced rationale for the experience sampling method. Third, I will describe different types of experience sampling methods in more detail. Fourth, I will discuss aspects of collecting and analyzing experience sampling data. Multilevel modelling is presented as a standard analytical procedure using a data set that was collected to investigate both personal and environmental factors of self-disclosure in mobile communication situations (MASUR 2018). In doing so, I will also discuss whether the somewhat intensive burden of responding to several prompts per day could have an influence on the data quality. Finally, I will summarize the promises and pitfalls of the method and provide some perspectives on how future research may benefit from adopting a situational perspective and implementing experience sampling designs.

2. Adopting a situational perspective and sampling experiences

Combining the two predominant approaches towards analyzing communication phenomena (drawing inferences from retrospective survey measures or studying causal mechanisms with experiments) requires an in-depth examination of the situation as the unit of analysis. Although the literature is full of attempts to define the situation, we must at least temporarily settle on a

working definition that allows us to make the manifold variety of situations amenable to theoretical and empirical investigation.¹

For the purpose of this article, it is important to note that there is a certain agreement between scholars of different backgrounds that both stable and situationally activated personal factors and environmental cues determine a person's psychological representation of a situation (e.g., LEWIN 1951; MAGNUSSON 1981a; SAUCIER ET AL. 2007; ROSS/NISBETT, 2001; RAUTHMANN/SHERMAN/FUNDER, 2015a). The term situation thereby has been used to represent those parts of the „total world“ that an individual can experience through sensory perception, and interprets as having reference to himself and his behavior (MAGNUSSON 1981b: 15). These combinations of environmental and personal factors can be positive (or driving) forces and negative (or inhibiting) forces (ROSS/NISBETT 2011: 14) of situationally exhibited behavior.

Based on this work, I argue that we can understand a situation “as the entirety of circumstances that affect people's thoughts, feelings, and behaviors at a given time” (MASUR 2018: chap. 7.1). These circumstances include both *personal* and *environmental* factors (cf. MASUR 2018: chap. 7.1.3) that vary with regard to their stability. Personal factors, on the one hand, refer to characteristics of the person itself, such as personality traits and trait-like qualities (e.g., opinions, attitudes, skills, long-term goals...) that are rather stable and thus do not vary across situations. On the other hand, they also refer to internal affective and cognitive processes (e.g., perceived duties, motives, goals, or feelings) that are situationally activated and thus less stable. Environmental factors can be differentiated into interpersonal (e.g., assessments of other persons) and external factors (physical or virtual cues of the environment) that generally differ across situations. That being said, they sometimes also show a certain stability across contexts. For example, in the context of my family, my assessments of those persons that are typically present (e.g., my parents or siblings) will be fairly similar across all potential situations that may occur within this context. Similarly, and unequally more important to studying behavior in media environments, certain characteristics of a medium (e.g., norms in certain media environments) are environmental factors that should to some extent always influence a person's behavior when he or she uses this particular medium. They thus have a certain stability across some situations, yet are not as stable as person-related factors (e.g., a person can use different media). There is an on-going debate about how long a situation lasts and how the stability of factors can be conceptualized (e.g., RAUTHMANN/SHERMAN/FUNDER 2015b). For example, scholars have argued that a situational framework needs to distinguish between (1) occurrences, (2) situations, (3) episodes, (4) life events, (5) typical situations, and (6) contexts. As Rauthmann and colleagues denote, “one may think of these concepts as being “nested” within each other like layers of an onion” (2015b, p. 5). For our purpose, it is sufficient to accept that a person experiences a “new” situation as soon as any of the situational factors changes. With regard to the stability of certain factors, a broad differentiation between non-situational and situational factors may often be sufficient. However, if environmental factors are studied that are clearly related to contexts, certain media, or cultures, their stability should be appropriately addressed (see also section 4).

In sum, this situational framework allows to theoretically examine general behavioral patterns while still acknowledging situational variance. Specifically, we can investigate the influence of stable person characteristics on situational behavior while controlling for situational influences both of personal or environmental nature. Additionally, we can study the influence of situational personal and environmental factors on situational behavior while controlling for trait characteristics. Finally, we can also investigate in how far situational influences are affected by more stable context or person characteristics. In many contexts, the effects of situational factors on behavior might differ between persons. We can hence analyze

¹ The literature is too vast to be summarized here. For comprehensive overviews and reviews, see e.g., Magnusson, 1981; Rauthmann, Sherman & Funder, 2015a; 2015b; Masur, 2018.

what person-related characteristics moderate potential situational processes. Yet, to investigate this situational framework empirically, we cannot rely on traditional survey methods or experimental designs. Instead, we have to assess situational factors across a variety of differing situations and additionally measure more stable person characteristics. In the following, I will show how a combination of experience sampling and survey methods allows to capture such situational processes in more detail.

3. Sampling from everyday life: Assessing actual experiences

The experience sampling method (ESM) is an intensive longitudinal data collection methodology that can be understood as a specific form of ecological momentary assessment (EMA). The basic idea of such methods is to measure thoughts, opinions, feelings, or behavior directly or slightly after they occurred. Other forms of EMA include diary studies, behavioral observations, continuous physiological assessments, or self-monitoring. Within ESM studies, participants are prompted to answer several short self-report questionnaires per day over a longer period of time. While the exact methods, technologies, and sampling strategies can differ, the overall goal is to collect data about momentary states in real-time, in natural environments, with multiple repeated assessments over time (SCOLLON/KIM-PEITRO/DIENER 2003: 5).

Although several groups of scholars developed techniques to measure behavior and momentary states *in situ* independently, the method that closest resembles its current form was developed in the 1970s by psychologists from the University of Chicago (CSIKSZENTMIHALYI & KUBEY 1981; CSIKSZENTMIHALYI/LARSON 1987; KUBEY/LARSON/CSIKSZENTMIHALYI 1996; HEKTNER/SCHMIDT/CSIKSZENTMIHALYI 2007; LARSON/CSIKSZENTMIHALYI 2014). These researchers aimed at measuring “internal (how people think and feel) as well as external (time, location, and social context) dimensions of experience” (cf. VANDEWATER/LEE 2009: 8). Since then, it has become an integral part of psychological methodology (SCOLLON/KIM-PEITRO/DIENER 2003: 5). In communication science, however, the method has not gained as much popularity so far. Few scholars, for example, have applied the ESM to investigate TV use (CSIKSZENTMIHALYI & KUBEY 1981; KUBEY/LARSON 1990; VANDEWATER/LEE 2009; KARNOWSKI/VON PAPE 2009), video games (KUBEY & LARSON 1990), music listening (SLOBODA/O’NEILL/IVALDI 2001; GREASLEY/LAMONT 2011; RANDALL/RICKARD 2013), gratifications of media use (SCHLÜTZ 2002; SCHERER & SCHLÜTZ 2002), media use for procrastination (REINECKE/HOFMANN 2016), mood and situational Facebook usage (BAYER ET AL. 2017), and privacy and self-disclosure processes (SCHENK ET AL. 2012; MASUR 2018). These studies all obtained fine-grained – and often surprising – insights in the studied phenomena. Yet, what is the precise rationale for conducting ESM studies?

3.1 Rationale for the experience sampling method

In general, three advantages of the ESM can be brought forward (cf. STONE ET AL. 2007: 3): (1) ESM minimizes the recalling of information and thus reduces potential bias in self-reports; (2) has a high ecological validity as it collects data of momentary states in real world situations, and (3) achieves a high temporal resolution by repeated assessments and thereby allows for the analysis of dynamic processes over time. In light of the proposed necessity to focus more on situational dynamics of communication phenomena, particularly the second and the third characteristic make a compelling case for using the ESM. In the following, I will briefly discuss each rationale.

3.1.1 *Momentary data: Reducing recall bias*

Traditional surveys provide reliable measures of variables of interest when these are reasonably stable over a longer period of time, represent historical information, or refer to rare or at least regular behaviors (SCHWARTZ 2007: 12; BROSIUS/HAAS/KOSCHEL 2016: 133). Only in those cases, participants are able to provide accurate information.² If the variables of interest refer to more frequent, yet irregular and mundane behaviors or indications of intensity in past experience, retrospective self-reports are likely to be biased (SCHWARZ 2007: 12). A growing body of research has shown that the recollection of such information is influenced by many different factors including the salience of recent experiences, heuristics, implicit theories, or survey-related motives such as social desirability (e.g., BRADBURN ET AL. 1987; NISBETT/ROSS 1980; TOURANGEAU ET AL. 2000). Human beings are generally not very good at retrieving certain information and thus rather try to reconstruct what must have happened based on heuristics and other mental-shortcuts (GORIN/STONE 2002).

ESMs reduce the methodological disadvantages of surveys because instead of having to remember *past* experiences, people are able to report their *current* experiences and observations directly while or shortly after they occurred. It has to be noted, however, that depending on the type of sampling strategy (see section 2.2.1), self-reports in ESM studies can also be biased. That being said, the time between the actual experiences and the assessment is often considerably shorter compared to traditional survey designs.

3.1.2 *Ecological validity: Assessing situational factors in real-world environments*

As noted already earlier, we are often interested in individuals' behavior under certain conditions or in specific environments. In other words, we want to investigate within-person processes. In survey studies, environmental influences are barely investigable and can only be achieved by confronting participants with scenarios (also called vignettes) in order to compare their responses after they are prompted to imagine being in these particular scenarios. In laboratory settings, researchers generally try to model real-world scenarios, but findings from such artificial settings may not necessarily transfer into the real world (STONE ET AL., 2007: 7).

By implementing the ESM, participants answer questionnaires in real-world environments and during or shortly after situations of interest. Researchers are hence able to assess environmental influences on their behavior while still controlling for person-related factors. In context of psychotherapeutical assessments, for example, researchers are often interested in what causes pain. Using an ESM design allows to assess pain whenever it occurs while simultaneously enquiring about potential situational causes. In communication science, assessing communication whenever it occurs allows to investigate what aspects of the environment (e.g., who is communicating, the type of relationship between the communicators, the location...) as well as which internal personal processes (e.g., motives, feelings, perceived norms or duties...) influence how much people disclose of themselves, what type of language they use, how they react to certain messages, or any other aspects relevant to the research goal respectively.

3.1.3 *Temporal dynamics: Analyzing processes over time*

Finally, ESM provide longitudinal data that can be used to analyze temporal processes. By sampling several situations (whether time-based, random, or event-based, see further below)

² Depending on the type of information that has to be recalled, survey measures of historical information or rare behavior can still be biased. However, experience sampling methods offer no advantage towards measuring these types of constructs.

throughout a person's day for one or two weeks, researchers have the possibility to study intra-individual change of behaviors, moods or other states over time and independent of situational factors. Research on media use, for example, could benefit heavily from ESM because they could try to identify a variety of temporal patterns in media use (e.g., are there diurnal or weekly patterns in TV use? Does Internet use differ between weekdays and weekends?). Although such question can sometimes also be studied with time-budget or diary studies, ESM often provides a higher temporal resolution of the studied processes.

For example, Hasebrink and colleagues (2013; 2017) argued that we need to investigate people's media repertoires. Assessing which media people use over the course of each day over several weeks might not only provide better assessments of people's media repertoires, but also add a temporal resolution to the analysis and thus providing insights into how media repertoires are connected to different times of the day and how they evolve over time as situational conditions change.

3.2 Classification of experience sampling methods

ESMs can be differentiated according to their data collection technique and their respective sampling strategy. Each particular method has its own strengths and weaknesses which should be evaluated with regard to its fit with the aims and objectives of respective studies. Particularly the sampling strategy profoundly affects in what situations participants will be prompted to answer the short questionnaires and thus how close these assessments will be to the relevant real-world experiences. Investigators have to consider their research questions and hypotheses, their knowledge about the studied phenomenon, practical and logistical considerations, and also need to develop a-priori defined plans for the data analysis in order to find the appropriate data collection and sampling technique (cf. SHIFFMAN 2007: 27).

3.2.1 Data collection techniques

In early ESM studies, participants were given small hand-held devices such as pagers or alarm watches (personal digital assistants = PDA) that signaled them when to answer short paper-pencil questionnaires. Today, three types of data collection approaches can be distinguished (HOFMANN/PATEL 2015: 238): (1) the original *PDA approach*, which gives the researcher much freedom in programming the necessary sampling strategy, but requires participants to carry extra devices at all times; (2) the *short message system (SMS) survey distribution approach*, which refers to sending links to mobile versions of a web-based survey via SMS to individual's smartphones and thus requires that participants' devices have internet access and (3) the *local application approach* which requires participants to install a specifically-programmed application on their own smartphones that both signals and collects data and only requires internet access in order to upload data every once in a while. The proliferation of smartphones in recent years as well as constant improvements in their technical capabilities have made the SMS survey distribution and the local app approach much more feasible. Furthermore, the time needed to answer ESM questionnaires is thereby significantly reduced. As much contemporary communication practices occur via smartphones, sampling smartphone-based communication situations or smartphone usage episodes becomes even more unobtrusive. The specific benefits of the mobile experience sampling method (M-ESM; KARNOWSKI/DOEDENS 2010), however, will be discussed further below.

3.2.2 Sampling designs

As the goal of any ESM study is to capture experiences in real-world settings using multiple repeated assessments, researchers have to carefully evaluate what kind of data collection

arrangement they want to implement. Most importantly, they have to ask themselves what situations they want to sample and how they will be sampled and measured. Most studies will be either focusing on assessing the flow of behavior and experience over time (and thereby receive reliable assessments of varying variables with minimal retrospective bias) or on assessing particular events in which situational processes will be analyzed (e.g., measuring a variety of situational factors including personal and environmental characteristics that may be used to predict situationally varying behavior). A first question is hence whether a study aims at sampling discrete situations or rather a continuous flow of experiences. Whereas the former requires to create a sampling scheme that allows to get as close to the relevant situations as possible, the latter can only be accomplished by intensive time-based or random sampling procedures. Table 1 provides an overview of typical sampling designs and combinations (cf. also BARRETT/BARRET 2001: 176; SCOLLON ET AL. 2003: 7-8; SHIFFMAN 2007: 30-31).

We can generally distinguish between (1) time-based sampling, (2) random sampling, (3) event-based sampling, and (4) combinations thereof. *Time-based sampling* refers to all sampling designs that prompt participants to answer questionnaires at regular or irregular, yet set times throughout the day. We can further distinguish between daily sampling, interval sampling, and intensive sampling. Daily sampling is particularly valuable, if the investigator is interested in summative measures for each day. Similar to panel studies, such designs allow to investigate day-to-day changes or how day-specific characteristics influence aggregative measures for each day. Interval sampling refers to prompting participants at pre-defined times per day (both at set intervals or at systematically varying intervals). It is hence particularly valuable when the study focuses on phenomena that vary considerably in intensity or character at different times during the day, e.g., feelings of hunger or tiredness (SHIFFMAN 2007: 35). However, if the phenomenon under investigation is considered to be continuously present (yet with varying intensity, e.g., measures of well-being or feelings) and a continuous analysis is wanted, interval sampling may not provide a representative sample of all possible experiences or situations. In this case, intensive sampling (or a random sampling procedure; see further below) can provide higher data quality. Intensive sampling refers to prompting people in short intervals (e.g., every 30 minutes or every hour). That being said, such a procedure is very burdensome for participants and might even interfere and influence the phenomenon under investigation.

Random sampling generally refers to prompting participants to answer questionnaires at irregular, random intervals. In comparison to time-based interval sampling designs, such a procedure is often implemented to avoid potential bias and ensure a representative sample of a person's overall experience (SHIFFMAN 2007: 39). Most ESM studies implemented such a design in order to obtain a representative sample of all situations their participants experienced during the time of the study.

Finally, *event-based sampling* designs are most important for studying situational dynamics and processes. The general idea is to prompt participants to answer a questionnaire whenever the studied phenomena occurs (e.g., the person watches TV, feels pain, is happy, has communicated, uses the smartphone...). Event-based sampling strategies can be categorized into self-sampling and automatic sampling. Self-sampling relies on participants to identify target events themselves. Such a design hence requires a clear a-priori definition of the events that should prompt participants to answer a questionnaire. If the event is objectively measurable (e.g., every time a person has watched TV), such a sampling procedure can be very useful. However, if the event is defined as a subjective experience (e.g., pain), idiosyncratic biases may arise. Furthermore, compliance in event-based self-sampling designs is hard to document. If the goal is to measure the frequency of certain events, such a design may provide less valid results. Automatic event-based sampling refers to all procedures that allow to trigger a questionnaire

Table 1: Overview of different sampling designs

Sampling-Approach	Definition	Target phenomena	Advantages	Issues
Time-based sampling				
Daily sampling	Participants are prompted to answer the questionnaire once a day (usually at the end of the day)	Summary of day's events or experiences (e.g., media consumption per day ...)	Coverage of whole day via recall and summary	Recall-based assessment may lead to bias; Time of completion may influence responses
Interval sampling	Participants are prompted at, pre-defined times a day (e.g., 8.00 am, 2.00 pm, and 8.00 pm)	Summary of certain episodes throughout the day (e.g., phase-related variables such as stress, level of tiredness)	Coverage of whole day via recall and summary of smaller episodes	Recall-based assessment may lead to bias; Time of completion may influence responses
Intensive sampling	Frequent prompts (e.g., once per hour or even more)	Fast-changing or continuous phenomena (e.g., feelings, stress...)	Almost continuous assessment of the phenomena of interest	Significant subject burden due to intensive prompting
Random sampling	Participants are prompted to answer the questionnaire at random, variable intervals (e.g., five times a day)	Continuous phenomena (e.g., levels of well-being, stress...)	Aims to achieve a representative sample of occasions; allows for stratified or weighted sampling schemes	If certain events are of interest (e.g., media use), measures might be recall-based; requires devices for prompting and implantation of sampling strategy
Event-based sampling				
Self-sampling	Participants identify the pre-defined events or episodes and answer the questionnaire autonomously	Pre-defined events or episodes that are easily accessible to individuals' perception, but impossible to track (e.g., eating, smoking...)	Multiple events can be tracked; no extra signaling devices necessary (e.g., by using printed questionnaires)	Events must be well defined; compliance per questionnaire hard to document
Automatic sampling	Participants are prompted to answer the questionnaire whenever a pre-defined event or episode occurs (requires the trackability of the event)	Pre-defined events or episodes that can be tracked (e.g., TV use, application use, smartphone use, using a particular website, watching a particular show...)	Multiple events can be tracked; compliance can be documented	Tracking and prompting device necessary (e.g., mobile phone); requires a complex sampling strategy (particularly when the pre-defined events occur often); willingness to participate may be low due to privacy concerns
Combinations				
Event-based, time-based or random sampling and daily sampling	Combining within-day event or time-based sampling with end of the day summary	Daily summary after within-day assessments (e.g., assessing evaluations of daily media use and assessing situational factors of actual media use)	Daily and within-day assessments can be sampled independently	Daily assessment is recall-biased and should be analyzed separately
Event-based and random sampling	Combining event-based and random sampling	Correlates or antecedents of events (e.g., motives to watching TV and gratifications while or after watching TV)	Potentially analyses of retrospective or prospective event antecedents	Overlaps between event-based and randomly sampled events possible

automatically after a certain event occurred. Schenk et al. (2012), for example, provided participants with a browser extension that logged participants Internet use. Every time a person engaged in specific Facebook activities (e.g., posted a status update) and a corresponding link was saved as log datum, the software automatically triggered a questionnaire that popped up as a new window in the browser (a similar approach was also implemented by BAYER ET AL. 2017). Masur (2018) used the local application approach. In his two-week study, participants were required to install the application *movisensXS* on their smartphones which tracked their application use and provided the possibility to automatically trigger situational questionnaires after participants had used certain pre-defined applications (e.g., WhatsApp, Twitter, Facebook, ...). The advantage of such a sampling design is that it allows to sample directly from pre-defined situations. If the goal of a study is to sample such specific, clearly defined situations, time-based or random sampling procedures are less useful as they sample independent of whether the event of interest occurred. Although there is the chance that the event of interest is sampled, the probability of missing it is very high. In those cases, we can only ask participants to remember the last time the event occurred which again introduces bias due to difficulties in recollection.

Finally, it is important to note that all sampling designs may be combined to achieve an even greater temporal resolution. Particularly combining time-based, random, or event-based sampling schedules with daily assessments can be useful to get a more detailed picture of the studied phenomena. It should be noted that the literature provides no single recommendation on how many times participants should be prompted per day. According to Christensen et al. (2007, p. 61), decisions in this regard (independent of the used sampling method) should be based on four criteria: a) the number of observations needed for stable estimations on the within-person level (see also power consideration in section 4.1), b) the naturalistic occurrence of the studied situations, c) the burden to participants, and d) anticipated compliance.

3.2.3 *Focus: The mobile experience sampling method*

Some scholars have recently introduced the term *mobile experience sampling method* (M-ESM) for research designs in which mobile devices are used to collect data (KARNOWSKI/DOEDENS 2010; KARNOWSKI 2013; RANDALL/RICKARD 2013). In light of the growing proliferation of smartphones as daily companions of individuals around the world, communication scholars thereby have the unique opportunity to study mediated communication on individuals' own devices (GÖRLAND 2017, p. 23). This has several beneficial implications for conducting mobile communication research (cf. KARNOWSKI/DOEDENS 2010: 214-215): First, participants do not have to carry an extra device (e.g., electronic pager) which makes sampling less obtrusive and less burdensome. Second, participants do not have to carry printed questionnaires as web- or application-based surveys can be specifically programmed and assessed right from participants' smartphones. M-ESM often allow to record time stamps both for the trigger and for the completion of the survey. Researchers hence obtain data on whether and when a questionnaire was answered and thereby document compliance more easily.

With continuously growing technical abilities of regular smartphones, M-ESM studies may furthermore allow to integrate other methods of data and content collection (e.g., multimedia content, smartphone sensor data such as GPS, etc.). This way, the local application approach allows to implement automatic event-based sampling procedures (e.g., using *movisensXS*) and participants' smartphone use can be objectively assessed by recording log files. Finally, data can be collected, saved, and uploaded during the study reducing the probability of data loss.

4. Collecting and analyzing experience sampling data

When conducting an ESM study, the researcher is confronted with many decisions that involve (1) deciding for an appropriate strategy to recruit participants (including power analysis and incentive strategies), (2) investigating which of the sampling strategies above is most appropriate to study the phenomenon of interest (including decisions with regard to filtering and potential sampling breaks), (3) evaluating the necessity of conducting several larger pretests, (4) deciding about necessary data preparation steps (including missing value analyses and centering decisions), (5) choosing an appropriate data analysis method, and (6) evaluating the quality of the data. At any rate, researchers should clearly describe and report their design and decisions (for first reporting guidelines, see STONE/SHIFFMAN 2007: 371-386).

I will exemplify these decisions with the help of a multi-method study on privacy and self-disclosure processes in smartphone-based communication situations (MASUR 2018)³. This study combined survey (to assess non-situational personal factors), tracking (logging smartphone and application use), and experience sampling methods (to assess situational personal and environmental factors). Participants answered in an initial online survey and subsequently installed a specifically programmed application (movisensXS) on their own smartphone that tracked their smartphone use for 14 days. Following an automatic event-based sampling design, the application automatically triggered situational questionnaires (roughly a maximum of 6 questionnaires per day depending on the individual applications use) after participants used pre-defined applications (i.e. whenever the log data revealed that participants used communication applications such as WhatsApp, Threema, Email, Facebook, Google+, Twitter, Instagram...). Overall, $N = 164$ smartphone users took part in the study ($M = 30$ years, 67% female). Overall, $N = 3,802$ questionnaires were prompted. The number of prompts per day per person varied considerably due to individual differences in application use over the course of the study ($M = 20.56$ per person on average, $SD = 15.04$; range = 1-87). The following analyses are based on a subset of $n = 1,104$ situations in which participants indicated having communicated to one or several other persons.

4.1 Preparation and data collection

It is important to note that conducting ESM studies is disproportionately more complex than data collection methods with one or only a limited number of repeated observations.⁴ First, several pretests are recommended for different reasons. On the one hand, investigators should take care to develop reliable short scales or items for the situational assessments. Larger scale scales should be avoided due to the repeated measurements and the related burden for participants. A preliminary survey can help to identify those items of a larger scale that best represent the latent variable (e.g., the item with the highest factor loading) or to develop more generic items (e.g., by correlating these new items with the latent construct of an existing scale). On the other hand, the technical side of ESM designs should be tested carefully prior to the actual data collection. Final simulations of the whole design should be undertaken with several subjects in order to evaluate the burden for future participants and to identify potential problems that may arise during the study. It is furthermore important to ensure that participants are able to contact the researcher at any time during the study.

A second a-priori consideration refers to estimating the power needed to test situational and/or personal effects. In ESM studies, it is important to note that effects of non-situational,

³ For more information about the methods used in the overall study, see Part II in Masur (2018). For the subset analyzed in this paper and the corresponding R code, see the online supplement: <https://osf.io/jpdvc/>

⁴ Of course, there are also very complex surveys and experimental designs. Yet, compared to most of such studies, ESM studies are nonetheless rather complex and involve many unusual decisions.

person-related characteristics on behavior are estimated as interpersonal effects and are thus based on the number of participants. Effects of environmental or situationally varying personal factors on behavior, in contrast, are tested based on the number of situations obtained by the respective sampling strategy. For example, if we want to test *small* non-situational effects (i.e. an effect size of $r = .10$) while assuming an alpha-error of 5%, we would need to recruit $N = 779$ participants to achieve a power of 80%. In the example study ($N = 164$ participants), we are only able to test non-situational effects with moderate effect sizes ($r > .21$) with a power of 80%. On the situational level, however, it is much easier to achieve enough power to test small effects. If $N = 164$ participants provide on average $M = 6$ situational questionnaires ($= 164 * 6 = 984$ situations), we can test situational effects with a power of 88%. In the present study, participants on average disclosed private information in $M = 6.78$ ($SD = 5.34$) situations during the time of the study (total number of situations that can be used in the analysis: $N = 1,104$). Small situational effects can hence be tested with a power of 92%.

Finally, it should be noted that ESM studies require participants to be engaged with the study for a longer period of time (usually 1 to 2 weeks) and involves being prompted to answer questionnaires several times a day. Thus, the willingness to participate is considerably lower compared to traditional survey studies. If additional data is collected (e.g., log data), privacy concerns might also influence the willingness to participate. Researchers aiming at recruiting large sample sizes (particularly on the person level) should consider using higher incentives for participation.

4.2 Analyzing the data

Analyzing data obtained from ESM studies can be tricky. On the one hand, such data are generally organized in a hierarchical structure that we have to account for in our analyses. On the other hand, they often require preparative steps before we can actually analyze them with statistical models. Of course, the specific types of data preparations and the analytical approach depends on the research questions or hypotheses. Depending on the study's goals, methods may include aggregative approaches, time series analysis, structural equation modelling, or multilevel approaches. However, when studying situational dynamics of human behavior as outlined above, multilevel modelling often represents the most valuable approach. In the following, I will discuss important aspects and potential problems that scholars typically encounter when adopting such a data analytical approach.

4.2.1 Hierarchical data structures

In ESM studies, it is important to remember that we sample data in two consecutive steps. First, we take a sample of individuals, and next we take a sample of situations that these individuals experience in their daily lives (cf. HOX 2010: 2). The data structure is called hierarchical because several observations are *nested* within persons (or: several repeated measurements are *clustered* within individuals). Sometimes, it could be meaningful to think about other nested structures in such data sets as well. For example, situations could also be nested in media environments (e.g., situations that occurred while using Facebook compared to situations that occurred while using WhatsApp), contexts (family, friends...) or certain time-based structures (days, weeks, mornings, evenings...). Even the participants could be nested in countries or cultures. Whether to account for these hierarchies or not depends on the variance in the respective factor and the research question at hand. In the following, I will only discuss the typical data structure that mostly results from conducting ESM studies: measurements nested in persons.

Depending on the study's design, variables may be defined both at the personal (Level 2) and at the situational level (Level 1). In the example study, I measured socio-demographics,

trait-like characteristics such as personality and other non-situational factors such as privacy concerns directly at the personal level (in a preliminary survey). Additionally, I measured situationally varying factors such as interpersonal assessments and people's behavior (e.g., the depth of self-disclosure) at the situational level (in the subsequent ESM study). A trivial, yet important first step is to merge both types of data sets. For being able to do so, it is important to provide each of the data sets with unique person identifiers (IDs). Only with such IDs, one is able to determine which situational measures belong to which personal measures. There are generally two ways of merging these two data sets which will be discussed in the next section. To exemplify the procedure, I used a subset of the original data set which only includes two variables on the person level (age and sex) and three variables on the situational level (app used, interpersonal trust, and self-disclosure).

4.2.1.1 Data formats

We can generally distinguish between a so-called *wide* and a *long* format. In the wide format, each row refers to one subject. A subject's repeated responses is hence in a single row and each response is in a separate column (see Table 2). Repeated measures are indicated by the respective time stamp in front of the variable name (e.g., T1_, T2_, etc.). The wide format allows to conduct descriptive analyses of person-related variables. Similar to data formats obtained from classical survey studies, for example, we can estimate population means. Using simple descriptive analyses, we can conclude that our example sample of participants is $M = 30$ years ($SD = 10$ years) in average and 67% of them are female.

In the long format, each row refers to one situation in which the situational variables were assessed (see Table 3). Each response per situational questionnaire is hence in a separate column. In contrast to the wide format, several rows now belong to one person (as indicated by the ID variable).

Table 2: Long format

id	age	sex	time	it	sd
25	24	1	T1	9,7	5,8
25	24	1	T2	9,7	7,2
35	24	1	T1	7,6	8,3
35	24	1	T2	9,3	8,7
35	24	1	T3	9,1	6,6
35	24	1	T4	5,8	3,6
35	24	1	T5	9,4	0,5
...	...	...	...	...	...

Note: Each row represents measures in one situational questionnaire.
it = interpersonal trust, sd = self-disclosure

Table 3: Wide format

id	age	sex	T1 it	T1 sd	T2 it	T2 sd	...
25	24	1	9,7	5,8	9,7	7,2	...
35	24	1	7,6	8,3	9,3	8,7	...
37	40	1	7	6,1	8,1	6,4	...
41	47	1	6,2	3,2	6,6	5,8	...
48	33	1	8,2	6,4	8,5	5,6	...
49	27	1	4,9	3,2	9,6	0,2	...
54	26	1	9,8	1,1	9,9	5,8	...
...	...	...	...	...	...	...	...

Note: Each row represents one participant.
it = interpersonal trust, sd = self-disclosure

The long format is typically used to estimate multilevel models (see further below). It is generally possible to move a variable from the situational level to the personal level by aggregation (Hox 2010: 2). For example, I could assign to each person his or her mean self-disclosure score. Disaggregation, in contrast, refers to moving a variable from a higher to a lower level. In the wide format, age and sex are person-related characteristics that occupy one column each. In the long format, we need to assign such person-related characteristics to each situation (similar to assigning the mean score of situational variables to each situation). As Hox (2010: 3) denoted, multilevel problems have historically led to approaches that moved all

variables by aggregation or disaggregation to one single level. The resulting data set was then used to estimate ordinary multiple regression models, to conduct analysis of variance, or to use other standard analysis methods. However, using aggregation or disaggregation in this manner, i.e. moving variables that belong to different levels to one single level (in ESM designs, this was often the personal level), is inadequate for several reasons (HOX 2010: 3): First, aggregation results in a loss of information and also in a loss of power. Second, disaggregation is similarly problematic as “using the larger number of disaggregated cases for the sample size leads to significance tests that reject the null-hypothesis far more often than the nominal alpha level suggests” (HOX 2010: 3). Third, aggregation and disaggregation may easily lead to wrong conclusions (e.g., when disaggregated data is wrongly interpreted on the situational level).

Multilevel models avoid these statistical problems and treat each variable appropriately with regard to its generic level of measurement. As we will see, however, aggregation and disaggregation is nonetheless useful if appropriately implemented.

4.2.1.2 Missing values

In data analyses of traditional survey or experimental data, we often treat missing values resulting from item nonresponse with listwise deletion. Particularly if the amount of missing values is small, such an approach can be appropriate. In ESM studies, in contrast, missing values can occur on both levels and thereby amount to a cause of concern. A first type of missing values results from participants not providing the same number of situations. Depending on the type of sampling strategy and participants’ compliance, participants can differ significantly regarding the number of completed situational questionnaires. Although systematic missingness may result in bias, a varying number of measurements per person is not necessarily worrisome for the data analysis because even one or two observations per person can be enough to fit a multilevel model (HOX 2010: 106). That being said, the (random) intercepts may not be estimated precisely for persons with few observations. Yet, even few observations per person can still provide partial information for the estimation of the regression coefficients and variance parameters on both levels (GELMAN/HILL 2009: 276).

A second and more concerning type of missing values refers to participants’ item nonresponse both in the preliminary survey and in the situational questionnaires. If we implement listwise deletion, nonresponse to one item in a situational questionnaire would cause the deletion of that situation and would thus only leave that particular participant with fewer situational measurements. Missing values in the preliminary survey, however, are more worrisome as they would eliminate the participant from the analysis. Given that it is more difficult to recruit participants for ESM studies, it can be disastrous if a person who provided high quality data in each situational questionnaire is deleted only due to one missing item in the preliminary survey. In sum, treating missing values with case-wise deletion may severely decrease the power to detect effects on both levels. I hence highly recommend to conduct a missing value analysis before engaging in the actual data analysis and consider imputation strategies if the missing values can be considered missing at random (for an overview, see SCHAFFER/GRAHAM 2002).

4.2.1.3 Data preparation and centering decisions

A final step of data preparation refers to centering. Many communication phenomena and particularly psychological constructs are expressed on arbitrary metrics (e.g., Likert-type scales) that lack a clearly interpretable or meaningful zero point (BLANTON/JACCARD 2006). Centering can be used to establish more meaningful zero points on these scales (e.g., the population’s mean) resulting in interpretable intercepts. In the context of multilevel modelling, however, centering decisions are more complex as situational measures (Level 1) can be centered around the grand mean (i.e., the population’s mean) or they can be centered around a person’s individual mean (i.e., the mean of a person’s repeated measurements). Whereas the

first has been termed centering at the grand mean (CGM) or grand-mean centering, the latter is referred to as centering within cluster (CWC) or group-mean centering (ENDERS/TOFIGHI 2007: 121). Although both centering approaches can be used to establish meaningful zero points on level 1 variables, it is important to note that they produce different parameters that must be interpreted differently (p. 121). It can be reasonable to use both in the context of one study when different research question need to be answered (ENDERS/TOFIGHI, 2007: 136). Enders and Tofighi (2007) generally propose the following guidelines: (1) CWC is appropriate when one is primarily interested in relationships between two Level 1 variables, (2) CGM is appropriate when one is interested in Level 2 predictors while simultaneously controlling for Level 1 covariates; and (3) CWC is preferable when one is interested in investigating cross-level interactions and interactions that involve a pair of Level 1 variables, while CGM is appropriate for interactions between Level 2 variables (p. 136).

In the context of ESM studies which aim at investigating situational processes, the following procedure, however, is recommended: First, all level 2 predictors that lack a meaningful zero point should be centered using CGM. This way, we account for (2) and (3) of the guidelines and get an interpretable intercept for Level 2 predictors (the value of the dependent variable when the predictor equals the population's mean). Second, level 1 predictors should be separated into between-person (non-situational) and within-person (situational) variance. To do so, one first needs to aggregate the situational measure to the (higher) person level by estimating the mean per person (e.g., estimating the average perception of interpersonal trust for each participant in the study, see online supplement). Including this aggregate measure allows to identify influences of stable variance parts in situational measures on the stable variance parts in the dependent variable. Second, the situational measure should be centered around this aggregated person's mean thus accounting for (1) and (3) of the guidelines. This centered situational measure represents situational deviations from a person's average ESM measure and hence allows to identify truly situational effects. In the following multilevel analysis, we will include both the person's mean (the aggregated measure) and the person-mean centered situational variable in the model.

4.2.2 Multilevel analysis

The multilevel regression model (also known as random coefficients model, variance component model, hierarchical linear model, or mixed-effects model) generally assumes a hierarchical data set with a single dependent variable at the lowest level (HOX 2010: 11). In our example, we will use a multilevel model to predict the depth of self-disclosure (variable code = sd) using both Level 1 (interpersonal trust; variable code = it) and Level 2 predictors (age and sex).⁵

We could choose to ignore the multilevel structure of the data and estimate a simple regression model (no pooling approach) which would result in a moderate, positive relationship; $b = 0.20$, $se = 0.03$, $\beta = .17$. Or we could decide to estimate a regression model for each person (complete pooling approach). In this case, the mean of all intrapersonal effects even suggests a negative relationship; $M_b = -0.18$, $SD_b = 6.06$ (For more information on no pooling and complete pooling approaches and their caveats, see online supplementary material). However, complete pooling ignores variation between persons and no pooling overstates it (cf. GELLMAN/HILL 2009: 253). Only in very limited cases (i.e. when there is very little between-person variance), the multilevel regression model is similar to the simple regression model. The multilevel model is designed as a compromise between these two extremes. It is thus often referred to as a form of *partial pooling* (GELLMAN/HILL 2009: 258). It can be thought of as a generalization of the linear regression model in which intercepts and possibly also slopes are allowed to vary per

⁵ Please note that all analytical steps described in the following sections can be reproduced with the online supplementary material: <https://osf.io/jpdvc/>.

group (in the case of ESM data: per person).⁶ The aim is to estimate effects that reflect both the relative information we have about within-person variance and between-person variance (p. 253). We therefore estimate *fixed* and *random* effects. Fixed effects can be thought of as weighted estimates of the effects that pertain to all persons. Random effects represent how much each person deviates from these fixed effects.

The most basic multilevel model is the *random-* or *varying-intercept* model represented in the following formula:

$$y_i = \alpha_{j[i]} + \beta x_i + \varepsilon_i$$

In this model, we predict a situational outcome variable (with $i = 1, \dots, n$ referring to situations as the lowest level of measurement) with another situational (or disaggregated personal) variable while allowing the intercept α to vary accross $j = 1, \dots, J$ persons. We hence assume a fixed effect for the relationship between x and y , but acknowledge that each person's intercept may deviate from the fixed intercept (see also Figure 3).

The varying-intercept and varying-slope model further allows slopes to vary between persons.⁷ It is represented with the following formula:

$$y_i = \alpha_{j[i]} + \beta x_{j[i]} + \varepsilon_i$$

In this model, we still estimate fixed effects for the intercept and the slope, yet both are allowed to vary. We get additional random effects for both the intercept and the slopes which tell us how much persons deviate from the weighted fixed effects.

4.2.2.1 Assessing personal and situational variance

A first step in conducting multilevel analyses of ESM data should be the estimation of a varying-intercept model that includes no predictors. This null-model provides one fixed effect that is equivalent to weighed mean across all situations and persons (i.e., persons with more situational measures provide more information than persons with less situational measures). We further get random effects, which, in this case, represent the variance that is attributable to between-person differences ($\sigma^2_{person|intercept} = 1.86$) and the unexplained residual variance (i.e. situational and error variance; $\sigma^2_{residual} = 1.45$). This is helpful, because it allows us to compute the *intraclass correlation coefficient* (ICC) with the following formula:

$$ICC = \frac{\sigma^2_{person|intercept}}{\sigma^2_{person|intercept} + \sigma^2_{residual}}$$

In our example, the ICC for depth of self-disclosure is .29, meaning that 29% of the variance in depth of self-disclosure is attributable to stable between-person differences. In other words, person-related, non-situational factors can explain one third of the variance of the depth of self-disclosure. Conversely, this suggest that a larger part of the variance may be due to varying situational factors. Hence, self-disclosure (as a communication behavior) is inherently situational.

⁶ Formal statistical description can be found in de Leeuw and Meijer (2007), Gelman and Hill (2009), and Hox (2010).

⁷ Although random slope models are theoretically possible, they are seldom found in reality. For this reason, I refrained from discussing them here.

4.2.2.2 Estimating fixed and random effects

In a next step, we add predictors to the varying-intercept model. Table 4 shows an overview of all estimated models. Model M0 represents the null-model described above. Model M1 includes the situational predictor interpersonal trust (both the person's mean (it.m) and the person-mean centered situational deviance (it.c)). The results reveal that both stable variance ($b = .23, se = .11$) and situational variance components ($b = .23, se = 0.03$) of interpersonal trust are positively related to depth of self-disclosure. This means, that people, who generally perceive other the recipients in smartphone-based communication situations as trustworthy, are also generally more likely to self-disclose. More importantly, the results also suggest that if a person perceives receivers of their disclosures as more trustworthy than normally, he or she will also be more inclined to self-disclose more in that situation. Both findings are visualized in Figure 3 which shows that the same slope was estimated for each person, but intercepts were allowed to differ.

Figure 1: Visualization of the random-intercept model: Only intercepts are allowed to vary. The slopes of all participants are hence the same.

In model M2, age and sex were included as Level 2 predictors. Only age emerged as a positive predictor ($b = 0.03, se = 0.01$) indicating that older people are generally more willing to self-disclose, independent of situational trust evaluations. Interestingly, the small effect of trait trust on self-disclosure thereby became non-significant. In model M3, the situational predictor was allowed to vary across situations. We are hence estimating a varying intercept, varying slope model. Although the fixed estimate for the relationship between situational trust and self-disclosure barely changes ($b = 0.22, se = 0.05$), we now also get random effect estimates (i.e., the variance of the deviations from this fixed effect: $\sigma^2_{person|trust} = 0.12$). In general, allowing situational effects to vary across persons should result in a better model fit. However, if the increase in model fit is negligible, we could stick with the varying-intercept model which is easier to interpret. As a general rule, it is recommended to estimate both models and compare their model fits. If the increase in model fit is not significant, one can interpret the varying-intercept model appropriately. As the models are nested, they can be compared using the likelihood ratio test (SNIJDERS & BOSKER 2011: 97). This test is based on a comparison of the models' deviances. Models with a lower deviance show a better fit than models with a

higher deviance (Hox 2010: 47). As the difference in deviances has a χ^2 -distribution, we can use a standard χ^2 -test to assess whether varying slopes significantly improved model fit. In our case, inclusion of random effects for the situational predictor variable did increase model fit ($\chi^2(2) = 38.35, p < .001$). Looking at the distribution of the random effects (Figure 4), we can see that some persons deviate considerably from the overall fixed effect. For some individuals, the relationship between situational trust and depth of self-disclosure is even negative.

Figure 2: Visualization of the random intercept, random slope model: Both intercepts and slopes are allowed to vary.

Finally, model M4 includes cross-level interactions. Whenever a situational effect varies, cross-level interactions can help to understand why they vary. In our example, we analyze whether including the interaction between (1) age and situational trust and (2) sex and situational trust improves the model fit. It is important to bear in mind that after inclusion of interaction terms, the effects of the moderated relationships become conditional. Due to our centering decision, the effect of situational trust on self-disclosure only applies to men who have the population's average age. However, results show that age does not moderate the situational effect of trust on self-disclosure. The relationship between trust and self-disclosure is only slightly stronger for women compared to men. However, the interaction effect is not significant ($b = 0.12, se = 0.11$) and thus does not result in an improved model fit ($\chi^2(2) = 1.23, p = .540$).

Based on the likelihood ratio test, we reached the best model fit already with model M3. Yet, we do not have any information about its absolute fit. Computing the amount of explained variance is not as straightforward as in standard linear regression models. The simplest approach is to compute how much the model with predictors reduces the residual variance on both levels compared to the null-model. This variance reduction approach can only be used with random-intercept models. In our example, it reveals that model M2 explains 5% of the variance on Level 1, but no variance on Level 2. This is not too surprising as the non-situational variables (trait trust, age, and sex) barely influence self-disclosure. Nakagawa and colleagues (2013) recently proposed a marginal and a conditional R^2 that are simple measures of the variance explained by fixed effects only and by both fixed and random factors. Johnson (2014) further extended this approach to random-slope models. In this regard, model M3 explains 5% ($R^2_{\text{marginal}} = .049$) by fixed effects only, but 42% ($R^2_{\text{conditional}} = .417$) by fixed and random effects.

Table 4: Results from the multilevel models predicting self-disclosure

	M0	M1	M2	M3	M4
	Fixed Effects				
(Intercept)	4.65 (0.13) ***	4.63 (0.13) ***	4.52 (0.23) ***	4.55 (0.23) ***	4.53 (0.23) ***
Level 1					
Interpersonal trust (CWC)		0.23 (0.03) ***	0.23 (0.03) ***	0.22 (0.05) ***	0.14 (0.09)
Level 2					
Average interpersonal trust (GMC)		0.23 (0.11) *	0.17 (0.11)	0.16 (0.11)	0.16 (0.11)
Age (GMC)			0.03 (0.01) *	0.03 (0.01) *	0.03 (0.01) *
Sex (1 = female)			0.15 (0.28)	0.11 (0.28)	0.13 (0.28)
Cross-level interactions					
Age x Interpersonal trust					−0.00 (0.01)
Sex x Interpersonal trust					0.12 (0.11)
	Random Effects				
$\sigma^2_{intercept person}$	1.86	1.87	1.88	1.97	1.97
$\sigma^2_{residual}$	4.45	4.23	4.22	3.76	3.76
$\sigma^2_{trust person}$				0.12	0.12
	Goodness-of-Fit Statistics				
AIC	4982.42	4936.22	4934.77	4900.42	4903.19
BIC	4997.44	4961.25	4969.82	4945.48	4958.26
log-likelihood	−2488.21	−2463.11	−2460.38	−2441.21	−2440.59
deviance	4976.42	4926.22	4920.77	4882.42	4881.19
Δ deviance		50.20	5.45	38.35	1.23
df		2	2	2	2
<i>p</i>		< .001	.066	< .001	.540

Note: Fixed effects are estimated by restricted maximum likelihood (REML). Numbers in parentheses represent standard errors.

Basis: $n_{participants} = 164$, $n_{situations} = 1,104$.

*** $p < .001$, ** $p < .01$, * $p < .05$.

4.2.3 *Assessing potential intervention effects*

An important, yet seemingly neglected aspect of analyzing ESM data refers to possible biases resulting from the considerable burden that participants might experience during the study. If individuals are prompted to answer questionnaires several times a day for a longer period of time, it seems likely that this somehow affects how they respond. Stone, Kessler, and Haythomthwatte (1991) denoted that diary studies (and thus ESM studies as well due to their similar assessment) require respondents to maintain interest over a considerable period of time. Otherwise, responses in situational questionnaires might be of poor quality because people might answer more quickly and with less reflection over the course of the study. To the best of my knowledge, there is almost no literature on such potential intervention effects in ESM study. Nonetheless, I believe that several effects can be postulated and should be investigated.

First, we can simply investigate whether the repetition of the same questions somehow affects our variables of interest. We can test this assumption by computing a variable that indicates the current number of repeated measures for each situational questionnaire (i.e., the first questionnaire per person received the number 1, the second number 2, and so on). This variable reflects the temporal position of each questionnaire within the study. We can now use this variable to predict our variable of interest using multilevel modeling. In our example, depth of self-disclosure increased by 0.05 points with each additional questionnaire. However, the effect was not significant ($b = 0.05$, $se = 0.03$).

Second, we could assume that participants' increasingly deviate from the scale's midpoint. In this particular study, the situational measures were assessed using a slider. A pretest had shown that this answering tool has a high usability. Yet, a slider could also lead to less thoughtful answers by just sliding to right or left extremes. I hence computed another variable, that measures the deviation of the response to the scale's midpoint in each situational questionnaire. Results again suggest that people did not deviate more from the scale's midpoint over the course of the study ($b = -0.02$, $se = 0.01$).

Third, we could assume that the situational variance decreases with each additional situational assessment). However, the difference between two consecutive situations did not decrease significantly over the course of the study ($b = -0.01$, $se = 0.02$).

Finally, we could also test whether the between-person variance decreases over time as response patterns could become more consistent. Therefore, we only have to compute the variance of self-disclosure per situational questionnaire (across all persons). Results suggest that between-person variance did decrease over time ($b = -0.13$, $se = 0.01$).

In this study, these analyses suggest that only small biases exist that can be considered negligible (see also Figure 5 in the online supplement). From a general perspective, this can be regarded as good news. At least in this study, it seems that the sampling design did not affect the way people responded. However, as the last analysis has shown, ESM designs seem to be able to influence the balance between situational and personal variance in the situationally measured variables. In light of this influence on people's response patterns, particular intensive longitudinal ESM studies should test whether such biases exist and if so, control for them in the main analyses.

5. Summary and future perspectives

In this chapter, I have argued that we should often adopt a situational perspective when investigating communication phenomena. Many theories in communication science provide assumptions about situational processes that have been neglected in empirical research. In many

fields of research, it seems that we too often overestimate the power of dispositional factors to predict human behavior and simultaneously neglect situational factors. Much of contemporary research either draws inferences about communication phenomena through people's retrospection in surveys or by isolating particular factors and behaviors to identify causal mechanisms. It thereby misses the opportunity to bridge both perspectives. If the goal is to arrive at a comprehensive understanding of human behavior (and I believe it should be), we should not ignore the situation and its potential to alter behavior or behavioral intentions. Based on a long tradition of investigating person-situation interactions, I first proposed a general situational framework which allows to investigate both *non-situational* and *situational* factors and their interactions as antecedents of behavior (cf. MASUR, 2018).

Second, I have introduced the experience sampling method (ESM) as a promising data collection method to test such a situational framework. The appeal of ESM thereby rests upon three general rationales (cf. STONE ET AL. 2007: 3): (1) ESM studies avoid the disadvantages of retrospective self-reports; (2) they increase the ecological validity of self-reports as people are surveyed in their natural environments; and (3) they allow researchers to investigate temporal patterns in people's daily experiences. In sum, ESM studies provide the opportunity to study people's behavior immediately when it occurs and while participants are in their real-world settings. This way, analyzing the influences of the environment and other situational factors becomes possible. With a variety of sampling designs, it provides a promising data collection approach for a variety of research questions.

Third, I have introduced multilevel modelling as a best practice for analyzing situational processes with ESM data. The motivation for conducting multilevel analysis is that it accounts for *both* situational and non-situational variation in estimating individual-level and situation-level regression coefficients. In contrast to complete pooling or no pooling approaches, it does not under- or overestimate between-person variance. That being said, a drawback of implementing multilevel modelling is its complexity. Several a-priori decisions concerning data transformations and centering decisions have to be made and researchers have to carefully plan both the data collection process and the corresponding data analyses. With this chapter, I have outlined several important steps in this regard. It is my hope that this chapter (and its corresponding online supplement) will help scholars, who are interested in adopting a situational perspective and in conducting ESM studies, to overcome potential methodological issues.

In the future, ESM studies will undoubtedly contribute to existing communication research. Particularly the multifunctionality of smartphones and their embeddedness in the daily lives of individuals provides communication scholars with novel opportunities to capture mediated communication and how it is enacted in real-world settings. With the increasing data processing power of handheld devices, scholars should aim at combining ESM data with other (potentially automatic) data collection approaches. For example, ESM data in combination with physiological data obtained from smartwatches can help to study unconscious reactions in communication situations or when people engage with different media content. Combining log data of people's media use with ESM sampling strategies can provide more information about what situations have been sampled and in how far the sampling strategy resulted in a representative sample of a person's experiences. Within the ESM sampling framework, scholars can also ask participants to capture their immediate environments by taking pictures or recording videos and thereby providing an actual window into their world. In this regard, the ESM could also bring interesting data for qualitative approaches. With the growing mobility of devices, scholars become more and more interested in examining mobile media use and mobile communication. The ESM provides a possibility to study these situations. Combining with data obtained from the smartphone's ability to use GPS and geo-positioning technologies, it is even possible to sample "on the go"-situations automatically and representatively.

In sum, we can conclude that the ESM is a useful and innovative data collection approach that can be used for a variety of research questions. Combined with the situational perspective outline in this chapter, I strongly believe that it can provide novel insights into long-researched as well as new communication phenomena. Yet, as Scollon and colleagues (2003) denote: “like any other tool, [the ESM’s] utility can only be measured by the care with which the researcher plans and conducts the research, analyzes the findings, and interprets the results” (p. 28). If done with care, the ESM can be a powerful tool that provides researchers with fine-grained and rare insights into people’s behavior – or more generally: into people’s lives.

References

- BARRETT, LISA F.; BARRETT, DANIEL J.: An introduction to computerized experience sampling in psychology. In: *Social Science Computer Review*, 19(2), 2001, pp. 175–185
- BAYER, JOSEPH.; ELLISON, NICOLE B.; SCHOENEBECK, SARITA; BRADY, ERIN; FALK, EMILY B.: Facebook in context(s): Measuring emotional responses across time and space. In: *New Media & Society*, 20(3), 2017, pp. 1047 - 1067
- BLANTON, HART; JACCARD, JAMES: Arbitrary metrics in psychology. In: *American Psychologist*, 61, 2006, pp. 27–41
- BRADBURN, NORMAN M.; RIPS, LANCE, & SHEVELL, STEVEN: Answering autobiographical questions: The impact of memory and inference on surveys. In: *Science*, 236(4798), 1987, pp. 157–161
- BROSIUS, HANS-BERND; HAAS, ALEXANDER; KOSCHEL, FRIEDERICKE: *Methoden der empirischen Kommunikationsforschung: Eine Einführung* [Methods of empirical communication science: An introduction]. Wiesbaden [Springer], 2006
- CHRISTENSEN, TAMLIN C.; BARRETT, LISA F.; BLISS-MOREAU, ELIZA ET AL.: A practical guide to experience-sampling procedures. In: *Journal of Happiness Studies*, 4, 2003, pp. 53-78.
- CSIKSZENTMIHALYI, MIHALY; KUBEY, ROBERT: Television and the rest of life: A systematic comparison of subjective experience. In: *Public Opinion Quarterly*, 45(3), 1981, pp. 317-328
- CSIKSZENTMIHALYI, MIHALY; LARSON, REED: Validity and reliability of the Experience-Sampling Method. In: *The Journal of Nervous and Mental Disease*, 175(9), 1987, pp. 526–536
- CZIENSKOWSKI, UWE: *Wissenschaftliche Experimente: Planung, Auswertung, Interpretation* [Scientific experiments: Planning, analysis, interpretation]. Weinheim [Beltz, Psychologie-Verl.-Union.], 1996
- DIENER, ED; SUH, EUNKOOK M.; LUCAS, RICHARD E.; SMITH, HEIDI L: Subjective well-being: Three decades of progress. In: *Psychological Bulletin*, 125(2), 1999, pp. 276–302
- DIENLIN, TOBIAS, MASUR, PHILIPP K. & TREPTE, SABINE: Displacement or reinforcement? The reciprocity of FtF, IM, and SNS communication and their effects on loneliness and life satisfaction. In: *Journal of Computer-Mediated Communication*, 22(2), 2017, pp. 71-87
- ELLISON, NICOLE B.; STEINFELD, CHARLES; LAMPE, CLIFF: The benefits of Facebook “friends”: Social capital and college students’ use of online social network sites. In: *Journal of Computer-Mediated Communication*, 12(4), 2007, pp. 1143–1168.
- ENDERS, CRAIG K.; TOFIGHI, DAVOOD: Centering predictor variables in crosssectional multilevel models: A new look at an old issue. In: *Psychological Methods*, 12(2), 2007, pp. 121–138
- GELMAN, ANDREW; HILL, JENNIFER: *Data analysis using regression and multilevel/hierarchical models: Analytical methods for social research* (11th edition). Cambridge [Cambridge University Press], 2009
- GÖRLAND, STEPHAN: Surveys go mobile: Experience Sampling Method als methodischer Ansatz für mobile Rezeptionssituationen [Surveys go mobile: Experience sampling method as methodological approach to mobile reception situations]. In: HOOFFACKER, G; WOLF, C. (Eds.): *Technische Innovationen - Medieninnovationen?* Wiesbaden [Springer], 2017, pp. 247–258
- GORIN, AMY A.; STONE, ARTHUR A.: Recall biases and cognitive errors in retrospective self-reports a call for momentary assessments. In: BAUM. A.; REVENSON, T.;

- SINGER, J. (Eds.): *Handbook of Health Psychology*. Mahwah, NJ [Lawrence Erlbaum Associates], 2002, pp. 405-413
- GREASLEY, ALINKA E.; LAMONT, ALEXANDRA: Exploring engagement with music in everyday life using experience sampling methodology. In: *Musicae Scientiae*, 15(1), 2011, pp. 45–71
- HASEBRINK, UWE; HEPP, ANDREAS: How to research cross-media practices? Investigating media repertoires and media ensembles. In: *Convergence*, 23(4), 2017, 362-377
- HASEBRINK, UWE; SCHMIDT, JAN-HINDRIK: Medienübergreifende Informationsrepertoires. In: *Media Perspektiven*, 1, 2013, pp. 2–12.
- HEKTNER, JOEL M., SCHMIDT, JENNIGER A., & CSIKSZENTMIHALYI, MIHALY: *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks Calif. u.a. [Sage Publications], 2007
- HOFMANN, WILHELM; PATEL, PARESH V.: SurveySignal: A convenient solution for experience sampling research using participants' own smartphones. In: *Social Science Computer Review*, 33(2), 2015, pp. 235–253
- HOX, JOOP J.: *Multilevel analysis: Techniques and applications* (2nd ed.). Quantitative methodology series. New York [Routledge], 2010
- KARNOWSKI, VERONIKA: Befragung in situ: Die Mobile Experience Sampling Method (MESM) [Survey in situ: The mobile experience sampling method (MESM)]. In: Möhring, W.; Schlütz, D. (Eds.): *Handbuch standardisierte Erhebungsverfahren in der Kommunikationswissenschaft*. Wiesbaden [Springer], 2013, pp. 235–247
- KARNOWSKI, VERONIKA; DOEDENS, SEBASTIAN: Mobile Experience Sampling: Eine Methode zur Untersuchung mobilen Mediennutzungsverhaltens [Mobile experience sampling: A method to investigate mobile media use]. In: Jakob, N.; Zerback, T.; Jandura, O.; Maurer, M. (Eds.): *Das Internet als Forschungsinstrument und -gegenstand in der Kommunikationswissenschaft* (Vol. 6). Methoden und Forschungslogik der Kommunikationswissenschaft. Köln [Herbert von Halem], 2010, pp. 235–247
- KARNOWSKI, VERONIKA; VON PAPE, THILO: Mobile TV im Alltag der Nutzer: Ergebnisse einer dreimonatigen Panelstudie [Mobile TV in the everyday life of the user: Results from a three-month panel study]. In: Krone, J. (Ed.): *Fernsehen im Wandel*. Baden-Baden [Nomos], 2009, pp. 241–255
- KRAUT, ROBERT; KIESLER, SARA; BONEVA, BONKA; CUMMINGS, JONATHAN; HELGESON, VICKI; CRAWFORD, ANNE: Internet paradox revisited. In: *Journal of Social Issues*, 58(1), 2002, pp. 49–74
- KRAUT, ROBERT; PATTERSON, MICHAEL; LUNDMARK, VICKI; KIESLER, SARA; MUKOPADHYAY, TRIDAS; SCHERLIS, WILLIAM: Internet paradox: A social technology that reduces social involvement and psychological well-being? In: *American Psychologist*, 53(9), 1998, pp. 1017–1031.
- KUBEY, ROBERT; LARSON, REED: The use and experience of the new video media among children and young adolescents. In: *Communication Research*, 17(1), 1990, pp. 107–130.
- KUBEY, ROBERT; LARSON, REED; CSIKSZENTMIHALYI, MIHALY: Experience sampling method applications to communication research questions. In: *Journal of Communication*, 46(2), 1996, pp. 99–120
- LARSON, REED; CSIKSZENTMIHALYI, MIHALY: The experience sampling method. In: CSIKSZENTMIHALYI, M. (Ed.): *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi*. Dordrecht [Springer], 2014, pp. 21–34
- LEUNG, LOUIS; WEI, RAN: More than just talk on the move: Uses and gratifications of the cellular phone. In: *J&MC Quarterly*, 77(2), 2000, pp. 308-320

- LEWIN, KURT: *Field theory in social science: Selected theoretical papers*. New York [Harper & Brothers], 1951
- MAGNUSSON, DAVID (Ed.): *Toward a psychology of situations: An international perspective*. Hillsdale, NJ [Erlbaum], 1981a
- MAGNUSSON, DAVID: Problems in environmental analyses – An introduction. In: MAGNUSSON, D. (Ed.): *Toward a psychology of situations: An international perspective*. Hillsdale, NJ [Erlbaum], 1981a, pp. 3-7
- MASUR, PHILIPP K.: *Situational privacy and self-disclosure: Communication processes in online environments*. Cham, Switzerland [Springer International Publishing], 2018
- MOORE, KELLY; MCELROY, JAMES C.: The influence of personality on Facebook usage, wall postings, and regret. In: *Computers in Human Behavior*, 28, 2012, pp. 267-274
- NISBETT, RICHARD E.; ROSS, LEE: *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ [Prentice-Hall, Inc.], 1980
- RANDALL, WILLIAM M.; RICKARD, NIKKI S.: Development and trial of a mobile experience sampling method (m-ESM) for personal music listening. In: *Music Perception: An Interdisciplinary Journal*, 31(2), 2013, pp. 157–170
- RAUTHMANN, JOHN F.; SHERMAN, RYNE A.; FUNDER, DAVID C.: Principles of situation research: Towards a better understanding of psychological situations. In: *European Journal of Personality*, 29(3), 2015a, 363–381
- RAUTHMANN, JOHN F.; SHERMAN, RYNE A.; FUNDER, DAVID C.: New Horizons in Research on Psychological Situations and Environments. In: *European Journal of Personality*, 29(3), 2015b, pp. 382-432
- REINECKE, LEONARD; HOFMANN, WILHELM: Slacking off or winding down? An experience sampling study on the drivers and consequences of media use for recovery versus procrastination. In: *Human Communication Research*, 43(3), 2016, 441–461.
- ROSS, LEE: The intuitive psychologist and its shortcomings. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York [Academic Press], 1977
- ROSS, LEE; NISBETT, RICHARD E.: *The person and the situation: Perspectives of social psychology*. London [Pinter & Martin], 2011
- RUBIN, ALLAN, M.: Television uses and gratifications: The interactions of viewing patterns and motivations. In: *Journal of Broadcasting*, 27(1), 1983, pp. 37-47.
- SAUCIER, GERARD; BEL-BAHAR, TARIK; FERNANDEZ, CYNTHIA: What modifies the expression of personality tendencies? Defining basic domains of situation variables. In: *Journal of Personality*, 75(3), 2007, pp. 479–503.
- SCHAFER, JOSEPH L.; GRAHAM, JOHN W.: Missing data: Our view of the state of the art. In: *Psychological Methods*, 7(2), 2002, pp. 147-177.
- SCHENK, MICHAEL; NIEMANN, JULIA; REINMANN, GABI; ROßNAGEL, ALEXANDER: *Digitale Privatsphäre: Heranwachsende und Datenschutz auf sozialen Netzwerkplattformen* [Digital privacy: Adolescents and data protection on social network sites]. Schriftenreihe Medienforschung der Landesanstalt für Medien Nordrhein-Westfalen. Berlin [Vistas Verlag], 2012
- SCHERER, HELMUTH; SCHLÜTZ, DANIELA: Gratifikation à la minute: Die zeitnahe Erfassung von Gratifikationen [Gratification à la minute: The realtime measurement of gratifications]. In: RÖSSLER, P; KUBISCH, S.; GEHRAU, V. (Eds.), *Empirische Perspektiven der Rezeptionsforschung* (Vol. 23). Angewandte Medienforschung. München [Fischer], 2002, pp. 133–152
- SCHLÜTZ, DANIELA: *Bildschirmspiele und ihre Faszination: Zuwendungsmotive, Gratifikationen und Erleben interaktiver Medienangebote* [Video games and their fascination: Motives, gratifications, and experiencing interactive media content]. München [Fischer], 2002

- SCOLLON, CHRISTIE N.; KIM-PRIETO, CHU; & DIENER, ED: Experience sampling: Promises and pitfalls, strengths and weaknesses. In: *Journal of Happiness Studies*, 4(1), 2003, 5–34
- SEIDMAN, GWENDOLYN: Self-presentation and belonging on Facebook: How personality influences social media use and motivations. In: *Personality and Individual Differences*, 54, 2013, pp. 402-407
- SHIFFMAN, SAUL: Designing protocols for ecological momentary assessment. In: STONE, A.A.; SHIFFMAN, S; ATIENZA, A. A.; NEBELING, L. (Eds.): *The science of real-time data capture*. Oxford [Oxford University Press], 2007, pp. 27–53
- SLOBODA, JOHN A., O'NEILL, SUSAN A.; IVALDI, ANTONIA: Functions of music in everyday life: An exploratory study using the experience sampling method. In: *Musicae Scientiae*, 5(1), 2001, pp. 9–32.
- SNIJDERS, TOM A. B. & BOSKER, ROEL J.: *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Los Angeles, Calif. [Sage Publications], 2011
- STONE, ARTHUR. A.; KESSLER, RONALD C.; HAYTHOMTHWATTE, JENNIFER A.: Measuring daily events and experiences: Decisions for the researcher. In: *Journal of Personality*, 59(3), 1991, pp- 575–607.
- STONE, ARTHUR A.; SHIFFMAN, SAUL: Appendix: Capturing momentary, self-report data: A proposal for reporting guidelines. In: STONE, A.A.; SHIFFMAN, S; ATIENZA, A. A.; NEBELING, L. (Eds.): *The science of real-time data capture*. Oxford: Oxford University Press, 2007, pp. 371-386
- STONE, ARTHUR A.; SHIFFMAN, SAUL; ATIENZA, AUDIE, A.; NEBELING, LINDA: Historical roots and rationale of ecological momentary assessments (EMA). In: STONE, A.A.; SHIFFMAN, S; ATIENZA, A. A.; NEBELING, L. (Eds.): *The science of real-time data capture*. Oxford: Oxford University Press, 2007, pp. 3–10
- TOURANGEAU, ROGER; RIPS, LANCE J.; RASINSKI, KENNETH: *The psychology of survey response*. Cambridge u.a. [Cambridge University Press], 2000
- VALENZUELA, SEBASTIÁN; PARK, NAMSU; KEE, KERK F.: Is there social capital in a social networking site? Facebook use and college students' life satisfaction, trust, and participation. In: *Journal of Computer-Mediated Communication*, 14, 2009, pp. 875–901.
- VANDEWATER, ELIZABETH A.; LEE, SOOK-JONG: Measuring children's media use in the digital age: Issues and challenges. In: *The American Behavioral Scientist*, 52(8), 2009, pp. 1152–1176