



Measuring motivational relationship processes in experience sampling: A reliability model for moments, days, and persons nested in couples

Felix D. Schönbrodt¹ · Caroline Zygarr-Hoffmann¹ · Steffen Nestler² · Sebastian Pusch³ · Birk Hagemeyer³

Accepted: 5 September 2021
© The Author(s) 2021

Abstract

The investigation of within-person process models, often done in experience sampling designs, requires a reliable assessment of within-person change. In this paper, we focus on dyadic intensive longitudinal designs where both partners of a couple are assessed multiple times each day across several days. We introduce a statistical model for variance decomposition based on generalizability theory (extending P. E. Shrout & S. P. Lane, 2012), which can estimate the relative proportion of variability on four hierarchical levels: moments within a day, days, persons, and couples. Based on these variance estimates, four reliability coefficients are derived: between-couples, between-persons, within-persons/between-days, and within-persons/between-moments. We apply the model to two dyadic intensive experience sampling studies ($n_1 = 130$ persons, 5 surveys each day for 14 days, ≥ 7508 unique surveys; $n_2 = 508$ persons, 5 surveys each day for 28 days, ≥ 47764 unique surveys). Five different scales in the domain of motivational processes and relationship quality were assessed with 2 to 5 items: State relationship satisfaction, communal motivation, and agentic motivation; the latter consists of two subscales, namely power and independence motivation. Largest variance components were on the level of persons, moments, couples, and days, where within-day variance was generally larger than between-day variance. Reliabilities ranged from .32 to .76 (couple level), .93 to .98 (person level), .61 to .88 (day level), and .28 to .72 (moment level). Scale intercorrelations reveal differential structures between and within persons, which has consequences for theory building and statistical modeling.

Keywords Relationship · Motivation · Intensive longitudinal designs · Change reliability · Experience sampling · Ambulatory assessment

Variability is an inherent aspect of virtually all conceptualizations of the term motivation (e.g., Berridge, 2004; McClelland, 1987). Our momentary wishes and desires not only depend on past experiences (e.g., we get hungry when we have not eaten for some time), but also on situational cues, that signal the current availability of incentives, and the presence of competing desires. In motivation research, situational factors are usually manipulated in laboratory experiments to test causal hypotheses concerning the

conditions and consequences of motivational states (Heckhausen & Heckhausen, 2018; Schultheiss & Brunstein, 2010; Schultheiss & Köllner, 2021).

However, experimental studies tell us little about the time scale on which motivational states vary in everyday life. Is motivation waxing and waning from moment to moment within a day? Or is it a rather slow process that ramps up over several days, with little within-day fluctuation? Does it follow a weekly rhythm with some desires being stronger on weekends and weaker on workdays? Beyond these different time scales, motivational states might also vary between persons (Fleeson, 2001), which is a core assumption underlying research on motive dispositions (Hagemeyer, Neyer, Neberich, & Asendorpf, 2013; Schönbrodt & Gerstenberg, 2012; Schultheiss & Köllner, 2021). In addition, couples or even larger groups of people could be distinguishable in terms of their typical motivation, which adds additional potential levels of variability.

✉ Felix D. Schönbrodt
felix.schoenbrodt@psy.lmu.de

¹ Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstr. 13, 80802 München, Germany

² University of Münster, Münster, Germany

³ Friedrich Schiller University Jena, Jena, Germany

In our analyses of the time scale and levels of variability of several motivational constructs, we extend an existing statistical model for variance decomposition and reliability estimation (Cranford et al., 2006; Shrout & Lane, 2012) with an additional temporal level (moments within a day) and dyadic interdependence. For such statistical analyses, intensive longitudinal assessments of people's motivational states as they occur in their everyday lives are necessary (i.e., experience sampling studies; Hofmann, Finkel, & Fitzsimons, 2015; Hofmann, Vohs, & Baumeister, 2012; Zygar et al., 2018b; Bolger & Laurenceau, 2013; Laurenceau & Bolger, 2005). In this study, we focus on the dynamics of motivation in the life-domain of romantic relationships. Specifically, we investigate the variability and reliability of self-reported communal and agentic motivational states and relationship satisfaction as assessed in two intensive experience sampling studies. For this purpose, we propose a model for variance decomposition and reliability estimation that covers an ESM data structure where the order of multiple moments is crossed with days, days are crossed with persons, and persons are nested in couples.

Knowledge about the time scale and variability of motivational processes carries important information for the design of studies. For example, the frequency and time points of momentary assessment should match the time scale of variability, and limited resources call for a trade-off analysis whether short and intensive (within day) measurements, or longer (but less intensive) daily diaries, are more appropriate for the research question at hand. Furthermore, scale correlations on the between-person level usually do not reflect within-person processes (Molenaar, 2008). However, often within-person conclusions are drawn from between-person studies, which can result in an ecological fallacy such as the Simpson's paradox (Adolf & Fried, 2019; Medaglia, Jeronimus, & Fisher, 2019; Fisher, Medaglia, & Jeronimus, 2018; Kievit, Frankenhuis, Waldorp, & Borsboom, 2013). Consequently, scale intercorrelations can differ depending on the level of analysis. Just as reliability has to be considered on each level of analysis, construct validity also has to be analyzed on each level (Shrout & Lane, 2012; Horstmann & Ziegler, 2020).

In selecting motivational variables relevant for romantic relationships, we relied on the conceptualization of partner-related agentic and communal motives, as proposed by Hagemeyer and Neyer (2012). According to this view, agentic motivations focus on the individual self and strivings for independence and power in the relationship. Although independence and power are distinguishable classes of goals, both facets have in common that they entail a sense of psychological distance from one's relationship partner. In

terms of the hierarchy in a romantic relationship, independence strivings can be viewed as providing horizontal distance to one's partner, whereas power strivings provide vertical distance. Thus, independence and power are related to different behavioral strategies of motive implementation (independence strivings often lead to physical distance from the partner, whilst power might often be exerted in close proximity), but they share a common incentive, namely the experience of feeling as a capable and self-reliant individual. Communal strivings, on the other hand, are directed towards experiences of closeness and community with one's partner. According to Hagemeyer and Neyer (2012), they manifest in "enjoying joint activities and closeness, sharing of experiences and resources, sympathetic concern, efforts to improve the relationship, and feelings of loneliness in absence of the partner" (p. 116). These definitions were derived from Bakan's (1966) original concepts of agency and communion, and, accordingly, they are viewed as fundamental motivational dimensions in romantic relationships (Hagemeyer & Neyer, 2012; Hagemeyer, Neyer, Neberich, & Asendorpf, 2013). Previous studies mainly focused on partner-related agency and communion at the between-person level of motive dispositions and largely confirmed expected associations between the motives and measures of relationship quality (Hagemeyer, Neberich, Asendorpf, & Neyer, 2013; Hagemeyer & Neyer, 2012; Hagemeyer et al., 2013; Hagemeyer, Schönbrodt, Neyer, Neberich, & Asendorpf, 2015). Overall, self-reported (explicit) and indirectly assessed (implicit) agency motives showed negative associations, whereas communal motives showed positive associations with relationship quality.

There is increasing interest in the analysis of daily motivational processes in couples, for example focusing on helping motivation (Kindt, Vansteenkiste, Loeys, & Goubert, 2016), motives for sacrifice in intimate relationships (Impett, Gable, & Peplau, 2005), or sexual motivation (Muise, Impett, & Desmarais, 2013; Dewitte & Mayer, 2018). Concerning our focal constructs, we are only aware of three previous studies that addressed partner-related communion and agency motivation within partners of a couple in a longitudinal design. Hagemeyer, Schönbrodt, Neyer, Neberich, and Asendorpf (2015), Study 2, found in a two-week daily diary of 106 couples that daily relationship satisfaction in general was increased when partners spent more time together. However this positive effect of physical proximity was diminished in coresiding couples, when partners reported high state agency motivation. Further, in experience-sampling design with five assessments per day, momentary variations in self-reported partner-related communal and agentic motivation (over the course of a few hours) were positively related to variations in communal and agentic behavior, which corresponds to findings

on the between-person level (Zygar et al., 2018b; Zygar-Hoffmann, Pusch, Hagemeyer, & Schönbrodt, 2020), and communal motivation was predictive of relationship satisfaction in interaction with situational aspects (Zygar, Hagemeyer, Pusch, & Schönbrodt, 2018b). Thus, there is evidence that partner-related agentic and communal motivation are indeed relevant for the study of romantic relationships at a process level.

In addition to motivational variables, we included relationship satisfaction in our analyses of variability. On the one hand, relationship satisfaction as an indicator of partners' broad evaluations of their relationship quality is a primary outcome in many studies in couple research (Karney & Bradbury, 1995). Therefore, information on the time scale, levels of its variability, as well as reliability information will be of interest for relationship researchers. On the other hand, relationship satisfaction seems to display some motivational properties as well. In an experience sampling study with 115 couples (six daily assessments over one week), Hofmann, Finkel, and Fitzsimons (2015) found that day-to-day variations in goal progress were positively predicted by variations in relationship satisfaction. Moreover, this effect was mediated by positive affect, perceived partner support, perceived control, and goal focus. Thus, experiences of relationship satisfaction may support the successful implementation of motivational states by fostering a positive self-regulatory mindset.

In our analyses of the time scale and levels of variability regarding the three focal variables agency motivation, communion motivation, and relationship satisfaction, we pursued four research goals: (1) Extend an existing reliability model (Cranford et al., 2006; Shrout & Lane, 2012) with an additional temporal level (moments within a day) and dyadic interdependence (persons nested in couples). (2) Do a variance decomposition that informs on which level (between moments within a day, between days, between persons, between couples) the most variance of relationship motivations and satisfaction is located. (3) Estimate the reliability of relationship motivations and satisfaction on several levels of aggregation (within-person/between-moments, within-person/between-days, between-persons, and between-couples). (4) Evaluate one aspect of the scales' validity by inspecting scale intercorrelations at the four levels of aggregation.

Methods

Source code for all statistical models and reproducible analyses are available at the Open Science Framework (<https://osf.io/jmeaw/>). Raw data for both studies are available as scientific use files (Sample 1: Zygar, Hagemeyer, Pusch,

& Schönbrodt, 2018a; Sample 2: Zygar-Hoffmann, Hagemeyer, Pusch, & Schönbrodt, 2020).

Samples

Data from two intensive experience sampling studies were used. Sample 1 (henceforward, S1) uses a data set from Zygar, Hagemeyer, Pusch, & Schönbrodt, (2018b) which is available as a scientific use file (Zygar, Hagemeyer, Pusch, & Schönbrodt, 2018a). This data set includes ESM data from 130 German-speaking participants (52% women) nested in 68 heterosexual couples. Participants' mean age was 22.4 years, and the majority (78%) were students. Individuals were on average 2.35 years in a relationship, the majority was not married (97%), and only one participant had children. For a more detailed description of the data set, see Zygar, Hagemeyer, Pusch, and Schönbrodt (2018b).

Sample 2 (S2) includes ESM data from 508 German-speaking participants (50% women) nested in 258 heterosexual couples. Participants were mostly non-students (71%), but held a high school degree (German Abitur) or a higher educational degree (65%). Mean age was 31.4 years and individuals were on average 6.43 years in a relationship. The majority was not married (67%) and had no children (68%).

Procedure

In both studies, individuals completed an entry questionnaire (programmed with *formr*; Arslan, Walther, & Tata, 2019) on various measures. In the 14 days (S1) or 28 days (S2) that followed, they took part in an experience sampling study, where they answered questions five times a day on their own smartphones, summing up to 9100 scheduled surveys in S1 and 71400 scheduled surveys in S2. The surveys were scheduled semi-randomly across the day, at identical time points for both partners, but during a time-period which couples chose at study registration. Both studies used self-developed ESM apps. For technical reasons, in S1 only individuals with an Android device could participate. In S2 both Android and iOS users could participate. In S1, the first ESM day could be any day of the week. In S2, all participants started their ESM procedure on a Monday (although, due to a continuous enrollment, on several Mondays across a period of eight months).

The surveys were completed in a median time of 3.28 min (S1) and 2.70 min (S2). When notified, individuals had 45 min to complete the survey, which included the same questions at each assessment. An exception was the last survey in the evening in S2. This survey had a different set of items (e.g., did not include the motivation items that are investigated here), and could be completed within five hours, as individuals were instructed to answer it before

going to sleep. The average response rate before data exclusions was 84% (S1) and 88% (S2), incentivized by personalized feedback, course credit or money. For more detailed descriptions of the procedures including exclusions we refer the reader to Zygar et al. (2018b) and Zygar-Hoffmann, Pusch, Hagemeyer, and Schönbrodt (2020). Beyond the exclusions documented there, one additional couple was excluded as multiple flags for invalid responding showed up (see <https://osf.io/6v2rw/>).

Experience sampling items

State motivation At each measurement occasion in S1 and in the first four occasions in S2, three motivational state scales were assessed (see Tables 9 and 10 in the Appendix for all items, instructions and response scales). Communal motivation was assessed with four items at each moment (two Likert scale items and two slider items), for example “How emotionally close would you want to be to your partner at the moment?”. For independence motivation, two items were used, for example “Right now, do you wish: To solitarily pursue your own interests?”. Power motivation was assessed with two (S1) or three items (S2), for example “Right now, do you wish: To influence the feelings or behavior of your partner in any way?”. A fourth scale, referred to as state agency motivation, was computed by summing up independence and power motivation.

State relationship satisfaction State relationship satisfaction was assessed with two (S1) or three items (S2) at each moment (see Table 1). Exploratorily, we also constructed a more homogenous two-item scale in S2 by excluding the “annoyance” item, which showed the lowest correlations with the other items (this resulted in a different two-item set than in S1). All reported results concerning S2 refer to the full three-item scale, except the reliability and correlation analyses where results for the two-item scale are additionally reported.

Several other items were assessed during experience sampling, see the primary documentation of the data sets for a full list of items (see <https://osf.io/b8pu6> and <https://osf.io/psqx8>).

Statistical procedure

Different models for estimating reliability in intensive longitudinal measures have been proposed (Shrout & Lane, 2012; Cranford et al., 2006; Schoebi, 2008; Nezlek, 2016). Our model is based on generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Shavelson & Webb, 1991) and extends the Cranford et al. (2006) model with another level of measurement (the order of moments crossed with days) and dyadic interdependence (persons nested in

couples). We implemented the model as a random effects intercept-only model to decompose the variance of item responses, allowing to allocate the sources of variances to several temporal levels and multiple other factors. From the same variance decomposition, reliability estimates can be derived based on generalizability theory (Cranford et al., 2006; Shrout & Lane, 2012). Computationally, we estimated variance components using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) for linear mixed-effects models in the R environment for statistical computing (Core Team, 2020), where a random intercept variance was estimated for each factor in Eq. 1. We used maximum likelihood (instead of the default restricted ML) because the estimates were more stable (i.e., less dependent on starting values) for our current datasets. The specific function call is in the reproducible scripts on the OSF.

We defined dyad members as nested in couples, and we treated them as indistinguishable (Kenny, Kashy, & Cook, 2006). In our research on motivation in couples, we generally start with the presumption that motivational processes do not significantly differ between men and women (see, for example, Zygar et al., 2018b), and try to constrain effects to be equal for both genders. If partners are treated as indistinguishable in the variance decomposition, any systematic between-gender variance is captured by the person factor. From a personality perspective this makes sense to us, as gender differences *are* interindividual differences when looking at persons. Other research foci, however, might prefer to treat partners as distinguishable and to explicitly model a gender factor and its interactions.

Variance decomposition Conceptually, level 1 (L1) models the mean of the item responses, which are assessed at each moment (L2), which are crossed with days (L3), which are crossed with persons (L4), which are nested under couples (L5). Following generalizability theory, the full variance decomposition model is formalized as a four-way analysis of variance. For a person p nested in couple c , responding to item i in moment m on day d , the model for, say, communal motivation Y_{cpdmi} is

$$\begin{aligned}
 Y_{cpdmi} = & \mu + C_c + P_{p(c)} + D_d + M_m + I_i \\
 & + (CD)_{cd} + (CM)_{cm} + (CI)_{ci} + (PD)_{p(c)d} \\
 & + (PM)_{p(c)m} + (PI)_{p(c)i} + (DM)_{dm} + (DI)_{di} \\
 & + (MI)_{mi} + (CDM)_{cdm} + (CDI)_{cdi} + (CMI)_{cmi} \\
 & + (PDM)_{p(c)dm} + (PDI)_{p(c)di} + (PMI)_{p(c)mi} \\
 & + (DMI)_{dmi} + (CDMI)_{cdmi} \\
 & + e_{cpdmi}.
 \end{aligned} \tag{1}$$

Uppercase variables denote the factors *couple* (C), *person* (P), *day* (D), *moment* (M), and *item* (I). Subscripts with parentheses denote the nesting structure, for example $P_{p(c)}$

Table 1 Experience sampling items for the assessment of relationship satisfaction

Label	Instruction	Scale
Need satisfaction (RS-1)	Original: Wie fühlen Sie sich jetzt gerade in Ihrer Partnerschaft? [Englisch: How are you feeling at the moment in your relationship?	Continuous slider from 0 = <i>total frustriert</i> [totally frustrated], over 5 = <i>neutral</i> [neutral], to 10 = <i>total erfüllt</i> [totally satisfied]
Relationship mood (RS-3)	Original: Wie geht es Ihnen jetzt gerade mit Ihrer Beziehung? [Englisch: How do you feel about your relationship at the moment?	Continuous slider from 1 (S1) or 0 (S2) = <i>schlecht</i> [bad], over 3.5 (S1) or 5 (S2) = <i>neutral</i> [neutral], to 7 (S1) or 10 (S2) = <i>außergewöhnlich gut</i> [exceptionally good]
Annoyance (RS-4)	Original: Wie genervt sind Sie jetzt gerade von Ihrem Partner? [Englisch: How annoyed are you about your partner at the moment?	Continuous slider from 1 (S1) or 0 (S2) = <i>überhaupt nicht</i> [not at all], to 7 (S1) or 10 (S2) = <i>stark</i> [strongly]

S1 = Sample 1, S2 = Sample 2. The need satisfaction item (RS-1) was not assessed in S1. The annoyance item (RS-4) was reverse coded for scale calculation. These items can be reused under a CC-BY4.0 license

to indicate that persons are nested in couples. In our design, the four-way interaction $(PDMI)_{p(c)dmi}$ cannot be distinguished from the error term, because we have no replicate measurements for that interaction. Therefore, the term is subsumed under the error term and does not appear in Eq. 1. Compared to the full five-factorial model, seven terms that include a *couple x person* interaction are missing. As every person is nested under only one specific couple unit, there can be no interaction effect, and consequently such a model would not converge.

The indicator variable for *moment*, m , goes from 1 to 5 (S1) or 1 to 4 (S2), which means that $m = 1$, for example, denotes all morning surveys across all persons. The indicator variable for *day*, d , goes from 1 to 14 in S1, or 1 to 28 in S2. Hence, $d = 1$ denotes the first study day of all persons. Note that couples started the study on different calendar days in S1. Therefore factor D does not capture events which are specific to the calendar day or a specific weekday across participants, but rather systematic variance due to the onset and duration of the study. In S2, couples always started on a Monday, across multiple months. Here, factor D can additionally capture systematic weekend effects, as days 7, 14, 21, and 28 are Sundays for each participant. The person indicator p runs across couples to reflect the nested data structure (i.e., persons 1 and 2 belong to couple 1, persons 3 and 4 to couple 2, etc.). See Table 2 for an exemplary data structure.

The specific values for the number of items ($i = 1 \dots j$), the number of moments nested within each day ($m =$

$1 \dots l$), and the number of days ($d = 1 \dots k$) is given in Table 3.

A priori, we did not expect substantial systematic variation for some factors of the design. For example, we did not expect systematic effects for the *day* factor D in S1, as persons started on different calendar days. Day 7 of person A presumably has nothing specific in common with day 7 of person B, if these persons are from different couples (except that the same amount of time has passed in the study). In S2, in contrast, all participants started on a Monday. In this case, weekend effects would show up in this factor. Likewise, we did not expect that a certain item has a specific meaning on certain days (DI interaction), or on certain moments in general (MI interaction), or on certain days for certain persons (PDI interaction). Nonetheless, given that we have no empirical evidence for these guesses, we decided to run a factorial model which includes all possible (up to four-way) interactions. This maximal model allows to freely estimate all possible variance components in an explorative way and to see whether certain sources of variances indeed are (close to) zero.

Several conceptually meaningful units emerge in the model as interactions between factors. For example, the three-way interaction *person x day x moment*, PDM , refers to specific surveys of specific persons on specific days (e.g., $p = 5, d = 2, m = 1$, refers to the morning survey of person 5 on her second day). The variance of this component quantifies the variability between these specific surveys across all moments of all participants (averaging across all

Table 2 Exemplary data structure

couple_uid	person_uid	studyday_id	moment_id	item	value
1	1	1	1	1	4
1	1	1	1	2	5
1	1	1	2	1	5
1	1	1	2	2	4
1	1	2	1	1	5
1	1	2	1	2	1
1	1	2	2	1	6
1	1	2	2	2	4
1	2	1	1	1	5
1	2	1	1	2	4
1	2	1	2	1	3
1	2	1	2	2	5
1	2	2	1	1	2
1	2	2	1	2	5
1	2	2	2	1	5
1	2	2	2	2	2
2	3	1	1	1	5
2	3	1	1	2	6
2	3	1	2	1	3
2	3	1	2	2	4
2	3	2	1	1	5
2	3	2	1	2	4
2	3	2	2	1	4
2	3	2	2	2	2
2	4	1	1	1	2
2	4	1	1	2	5
2	4	1	2	1	5
2	4	1	2	2	3
2	4	2	1	1	6
2	4	2	1	2	5
2	4	2	2	1	4
2	4	2	2	2	1
3	5	1	1	1	4
3	5	1	1	2	5
3	5	1	2	1	5
3	5	1	2	2	4
3	5	2	1	1	5
...

Exemplary reduced data structure for persons in couples, answering two items on two moments on two days (couples could start on a different calendar date and moment 1 could be at different time points for each couple). Indexes for day are repeated within each person, and indexes for moment are repeated within each day of each person (as they are crossed). Indexes for couple and person, in contrast, are unique for each couple and each person in the sample (*uid* = unique id), as persons are nested in couples. The column *value* contains the item responses before standardization

items). The meaning of the other components together with an explanation of their respective variance components can be found in Table 4.

For estimating the model, several assumptions have to be made (Shrout & Lane, 2012): (a) Errors and true scores are independent, which also implies that no autoregressive

Table 3 Design settings

	RS2	RS3	Ind	Pow	A	C
Sample 1						
Number of items j	2	–	2	2	4	4
Number of days k	14	–	14	14	14	14
Number of moments l	5	–	5	5	5	5
Answered surveys	7545	–	7515	7508	7515	7544
Items	RS-3, RS-4	–	I-1, I-2	P-1, P-2	I-1, I-2, P-1, P-2	C-1, C-2, C-3, C-4
Sample 2						
Number of items j	2	3	2	3	5	4
Number of days k	28	28	28	28	28	28
Number of moments l	5	5	4	4	4	4
Answered surveys	60917	60917	47878	47871	47878	47913
Items	RS-1, RS-3	RS-1, RS-3, RS-4	I-1, I-2	P-1, P-2, P-3	I-1, I-2, P-1, P-2, P-3	C-1, C-2, C-3, C-4

j , k , and l are the numbers of scheduled items, days, and moments. Numbers of actually answered slightly differ within study when participants skipped a survey and only partial surveys were recorded. *RS2*, *RS3* = relationship satisfaction scale, measured with 2, resp. 3, items; *Ind* = independence motivation scale; *Pow* = power motivation scale; *A* = agentic motivation scale (pooled independence and power); *C* = communal motivation scale. *RS-1* = need satisfaction item, *RS-3* = relationship mood item, *RS-4* = annoyance item. For specific item wordings, see Tables 1, 9, and 10

effects are present, (b) the variances are fixed (i.e., the same for all units), (c) items have the same weight of the latent factor. There are good reasons why these assumptions do not reflect realistic properties of psychological data, and the consequences of violating them is discussed exemplarily for the current data in the limitation section.

Data preprocessing The items of our communion motivation scale were assessed on different response scales. The GT model covers differing mean levels of items with the item factor I . However, different scales can also pose (additional) problems for the assumption of equal item loadings and the assumption of fixed variances. In practice, items with different response options are typically averaged to a scale score by first standardizing them.¹ As we wanted to match our reliability analysis to the actually computed scale scores, we z -standardized all items across all measurement points of both genders. (The reliability estimates from unstandardized variables were virtually identical). Furthermore, we recoded one reversed item for relationship satisfaction (*RS-4*, see Table 1).

Reliability estimation Reliability estimation in the GT framework generally uses the formula

$$R_X = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_e^2} \quad (2)$$

¹We note that this practice makes the scale score sample-dependent, which is undesirable if the absolute value of a score should be interpreted. Alternatively, items could be rescaled to the same response scale.

where σ_T^2 is the variance of the true scores and σ_e^2 is the variance of the random measurement error, which is assumed to be constant across units and replications (Shrout & Lane, 2012).

Based on this general reliability approach, Cranford et al. (2006) and Shrout and Lane (2012) derived formulas that compute reliability on several levels in experience sampling designs. Here, we extend these formulas with an additional temporal level (moments crossed with days) and dyadic interdependence.

For all following reliability formula, we assume that days, D , are random (and not fixed), because participants started on different days across a period of several months, and the study period is not contingent on some common event. Moments, M , in contrast, were treated as fixed, as the moments each day (from morning to evening) were assumed to be comparable for each person.² Finally, the item factor, I , is treated as fixed (cf. Shrout & Lane, 2012), as no generalization beyond this specific item set is aimed for. Consequently, σ_M^2 , σ_I^2 , and σ_{MI}^2 play no role in the following reliability formulas.

Depending on the focal level for which reliability should be assessed, different terms contribute to the numerator (the true score variance) and the denominator (the observed variance). Generally, terms located on a higher level that do not vary within the focal level do not contribute to

²Specifically, the five surveys per day were pseudo-randomly distributed across the day. Start and end time could to some extent be personalized and some time spans of each day could be blocked in S2 because participants knew that they would not be able to answer in these periods.

Table 4 Variance decomposition of item responses: meaning of terms

Source	Explanation	Example/Comment
Theoretically relevant terms		
couple (C)	Variance between couples	
person (P)	Variance between persons	
day (D)	Variance between days 1 to 14/28 (pooled across all persons)	Time trends across the study, or effects of weekend vs. weekday.
moment (M)	Variance between time points 1 to 5 (moments are pooled within and across all persons)	Systematic effects of morning vs. evening
couple:day (CD)	Do specific days have different meanings for each couple? (days 1 to 14/28)	Shared daily characteristics (e.g. being together on a family gathering)
person:day (PD)	Variance between days (each day of each person is a unique day)	
couple:day:moment (CDM)	Event-level variance between couples	Shared momentary environment
person:day:moment (PDM)	Variance between moments (each moment of each person is unique)	
Nuisance terms		
item (I)	Do the mean level of items differ?	Items are z-standardized, therefore we expect only small values
couple:item (CI)	Do couples have a stable, differential understanding of items?	Couples agree on a common understanding of specific items
couple:moment (CM)	Systematic effects of moment for some couples	Systematic effects of morning vs. evening for some couples
moment:item (MI)	Do specific items have a specific meaning on specific time points of the day, pooled across all days of all persons?	All persons change the interpretation of some items in the evening.
person:item (PI)	Do persons have a stable, differential understanding of items?	Differential item functioning for men and women, or for specific persons
person:moment (PM)	Variance between time points of a day (pooled within each person)	Systematic effects of morning vs. evening for some persons
day:item (DI)	Do specific items have a specific meaning on specific days, across all persons?	All persons change the interpretation of some items on Fridays (assumed that all participant started on a Monday).
day:moment (DM)	Do certain events (e.g., moment 4 on day 9) have a special meaning across all persons?	All persons report higher values on all items on the first moment of the first ESM day.
couple:moment:item (CMI)	Do couples have a stable, differential understanding of items at specific time points across all days?	Couples differ in their shared understanding of items in the morning vs. in the evening.
couple:day:item (CDI)	Do couples have a stable, differential understanding of items at specific days?	Some couples change the interpretation of some items at specific days
person:moment:item (PMI)	Do person have a differential understanding of items on specific time points (1 to 5) across all days?	Some persons change the interpretation of some items in the evening
person:day:item (PDI)	Do persons have a differential understanding of items at specific days (1 to 14/28)?	Some persons change the interpretations of items on specific days
day:moment:item (DMI)	Do specific items have a specific meaning on specific moments of specific days (across all persons)?	All persons change the interpretation of an item on the evening of ESM day 6.
couple:day:moment:item (CDMI)	Do couples have a stable, differential	Different understanding of items after a

Table 4 (continued)

Source	Explanation	Example/Comment
	understanding of items at specific time points of specific days?	conflict between the partners
Error (e)	Residual error variance	

reliability estimation. For example, if we are interested in the measurement of purely within-person changes, the variance of the term *PI* (i.e., *person x item*), σ_{PI}^2 , neither contributes to systematic variances nor to the error term, as mean level biases in item understanding between persons are irrelevant for relative within-person assessments: Within each person, this is a constant mean level shift that does not contribute to variations within that person. Likewise, systematic variance between days of a person, σ_{PD}^2 , is an irrelevant source of variance if moment-to-moment change within a day is assessed, and between-couple variance, σ_C^2 , does not contribute to between-person reliability estimation or any other lower level.

For the *numerator*, one starts with a focal level for which reliability should be assessed, for example “between persons”. The numerator contains all sources of systematic variance for that level. In our example, this primarily is the person factor *P*, which contains all between-person variance. However, depending on which factors are a priori defined as fixed, some additional interaction terms also contribute to systematic between-person variance. Typically, just as it is in our case, the item factor *I* is considered fixed. Consequently, the *person x item interaction PI* contains idiosyncratic response patterns. If person A has on average higher scores on item 1 than would be expected by the main effects of the person mean and the item mean, this contributes to systematic between-person differences. As we assume moments *M* to be fixed, the *person x moment interaction PM* must be considered, too: If person B is not a morning person and always responds lower on all items in the morning survey, this variance component also contains systematic between-person variance. The same logic applies to the *PMI* interaction. Hence, the numerator contains the focal random factor, and all interactions of fixed factors with this random factor.

The *denominator* is the sum of systematic plus random (error) variance. Hence, along with all terms of the numerator, it contains all other random terms (including interactions with at least one random term) which are not on a higher level than the focal level. In the GT reliability computation, variance components are divided by the number of replications that are averaged when aggregating the scale scores, in order to account for the increased

precision when more measurements are available (see the explanations after the formula in the next paragraph).

Based on these assumptions, *between-couple reliability* (averaging all measurements of both persons of a couple across the entire study), R_{BC} , can be defined as:

$$R_{BC} = \frac{\sigma_C^2 + [\sigma_{CI}^2/j] + [\sigma_{CM}^2/l] + [\sigma_{CMI}^2/(l * j)]}{\sigma_C^2 + [\sigma_{CI}^2/j] + [\sigma_{CM}^2/l] + [\sigma_{CMI}^2/(l * j)] + [\sigma_{CD}^2/k] + [\sigma_{CDM}^2/(k * l)] + [\sigma_{CDI}^2/(k * j)] + [\sigma_{CDMI}^2/(k * l * j)] + [\sigma_P^2/2] + [\sigma_D^2/k] + [\sigma_{PD}^2/(2 * k)] + [\sigma_{PM}^2/(2 * l)] + [\sigma_{PI}^2/(2 * j)] + [\sigma_{DM}^2/(k * l)] + [\sigma_{DI}^2/(k * j)] + [\sigma_{PDM}^2/(2 * k * l)] + [\sigma_{PDI}^2/(2 * k * j)] + [\sigma_{PMI}^2/(2 * l * j)] + [\sigma_{DMI}^2/(k * l * j)] + [\sigma_e^2/(2 * k * l * j)]} \quad (3)$$

Constant *j* is the number of items, *k* is the number of days, *l* is the number of moments within each day (see Table 3 for the specific values). Each variance component is divided by the number of replications. For example, the *couple x item* variance is divided by *j*, as for each couple *j* estimates, for each level of the item factor, are considered. The residual error term σ_e^2 in R_{BC} is divided by $2 * k * l * j$ to take into account the increase in precision that results from averaging *j* items, assessed at *l* moments at each of the *k* days for both (i.e., two) persons in each couple. Finally, note that the variance components for I, M, and MI do not appear in the denominator as we assumed them to be fixed.

For computing *between-person reliability* (averaging all measurements of a person across the entire study), R_{BP} , we extend Equation (8) from Shrout and Lane (2012) by the new temporal level of moments, and all necessary interactions of the new moment factor with other factors:³

$$R_{BP} = \frac{\sigma_P^2 + [\sigma_{PI}^2/j] + [\sigma_{PM}^2/l] + [\sigma_{PMI}^2/(l * j)]}{\sigma_P^2 + [\sigma_{PI}^2/j] + [\sigma_{PM}^2/l] + [\sigma_{PMI}^2/(l * j)] + [\sigma_D^2/k] + [\sigma_{PD}^2/k] + [\sigma_{DM}^2/(k * l)] + [\sigma_{DI}^2/(k * j)] + [\sigma_{PDM}^2/(k * l)] + [\sigma_{PDI}^2/(k * j)] + [\sigma_{DMI}^2/(k * l * j)] + [\sigma_e^2/(k * l * j)]} \quad (4)$$

³Note that Shrout and Lane (2012) do not include $\sigma_{\text{TIME} * \text{ITEM}}^2$ in their Eq. (8), although time is considered random for this equation.

We computed *within-person change reliability from day to day*, R_{WPD} (averaging over l moments within a day), as:

$$R_{WPD} = \frac{\sigma_{PD}^2 + [\sigma_{PDI}^2/j] + [\sigma_{PDM}^2/l]}{\sigma_{PD}^2 + [\sigma_{PDI}^2/j] + [\sigma_{PDM}^2/l] + \sigma_D^2 + [\sigma_{DM}^2/l] + [\sigma_{DI}^2/j] + [\sigma_{DMI}^2/(l * j)] + [\sigma_e^2/(l * j)]} \quad (5)$$

On the lowest temporal level *within-person change reliability from moment to moment*, R_{WPM} , is computed as (cf. Shrout & Lane, 2012, Eq. 9):

$$R_{WPM} = \frac{\sigma_{PDM}^2}{\sigma_{PDM}^2 + [\sigma_e^2/j]} \quad (6)$$

The number of days within person, k , and the number of moments within day, l , is not constant if participants do not answer every single ESM survey. Therefore, for the actual computation of all reliabilities we inserted the average number of answered moments (i.e., response rate x maximum possible observations) and the average number of days into the formulas (see also Scott et al., 2018, footnote 5, and Shrout & Lane, 2012).

Application of reliability formulas to related data structures

The provided formulas can be adapted to related data structures. For measurement designs without a dyadic structure on the highest level, reliability formulas R_{BP} , R_{WPD} , and R_{WPM} are identical. In this case, the variance decomposition in Eq. 1 simply omits all terms including the factor C .

For measurement designs with a dyadic structure but with only a single daily measurement, the variance decomposition in Eq. 1 omits all terms including the factor M and the term PDI , as the latter cannot be distinguished from the error term, because no replicate measurements are present for that interaction. The between-person reliability formula simplifies to:

$$R_{BP_nondyadic} = \frac{\sigma_P^2 + [\sigma_{PI}^2/j]}{\sigma_P^2 + [\sigma_{PI}^2/j] + [\sigma_D^2/k] + [\sigma_{PD}^2/k] + [\sigma_{DI}^2/(k * j)] + [\sigma_e^2/(k * j)]} \quad (7)$$

Note that, in contrast to Eq. 8 in Shrout and Lane (2012), we added $\sigma_{DI}^2/(k * j)$ to the denominator, as time is considered to be random.

The within-person change reliability from day to day simplifies to:

$$R_{WPD_nondyadic} = \frac{\sigma_{PD}^2}{\sigma_{PD}^2 + \sigma_D^2 + [\sigma_{DI}^2/j] + [\sigma_e^2/j]} \quad (8)$$

Note that, in contrast to Eq. 9 in Shrout and Lane (2012) and Eq. 5 in Cranford et al. (2006), we added σ_D^2 and σ_{DI}^2/j to the denominator, as time is considered to be random.

Scale intercorrelation at four levels of aggregation

We computed correlation matrices of all scales on the four conceptual levels. Non-independence of data due to the hierarchical structure was handled by controlling for mean differences of all higher level units: (a) Scale scores on the *between-couple level* were computed by averaging all item responses of a scale across all measurements of both persons in a couple. (b) Scale scores on the *between-person level* were computed by subtracting the couple means from all answers and averaging the residuals across all measurements of each person. (c) Scale scores on the *within-person/between-days level* were computed by sequentially subtracting the couple and the person means and averaging the residuals across all measurements of each day of a person. (d) Scale scores on the *within-person/between-moments level* were computed by sequentially subtracting the couple, person, and day means and averaging the residuals within each moment of a person.

Centering the item responses to the mean of all higher units removes potential confounding effects. For example, for the between-days analysis, all potentially confounding between-couple and between-person effects are controlled for by removing the respective means from the item responses. After this preprocessing, correlations were computed across the full sample.

Results

Variance decomposition

Table 5 reports the absolute variance estimates and Table 6 reports a relative variance partitioning of the systematic (non-error) variances. For a better overview, we categorized sources of variance into “theoretically relevant terms” (i.e., of substantive interest) and “nuisance terms”, although some of the terms that we consider nuisance terms here might be centrally relevant for other research questions (e.g., for methodological and psychometric questions).

As a general pattern, four focal sources of variances had the largest share across scales and studies: persons (P ; around 19% of systematic variance), specific moments of persons (PDM ; around 15%), couple (C ; around 13%), and specific days of persons (PD ; around 8%). Beyond these general trends, however, specific variance components are more pronounced in some scales than others. For example, the largest share of couple-level variance is mostly present

Table 5 Variance decomposition of item responses: absolute variances

Variance component σ^2	Sample 1, absolute					Sample 2, absolute				
	RS*	Ind	Pow*	A*	C	RS*	Ind	Pow*	A*	C
Theoretically relevant terms										
couple (C)	.10	.04	.07	.05	.10	.20	.08	.06	.06	.16
person (P)	.08	.19	.11	.10	.18	.10	.16	.19	.12	.11
day (D)										
moment (M)		.01			.01					
couple:day (CD)	.09	.01		.01	.01	.08	.02	.01	.01	.02
person:day (PD)	.01	.11	.05	.02	.06	.06	.08	.07	.04	.07
couple:day:moment (CDM)	.11	.03	.01		.04	.09	.02			.03
person:day:moment (PDM)	.08	.17	.11	.04	.15	.09	.13	.13	.05	.12
Nuisance terms										
item (I)										
couple:item (CI)	.04	.02	.03	.03	.03	.04	.02	.02	.03	.04
couple:moment (CM)		.01								
moment:item (MI)				.01						
person:item (PI)	.12	.04	.20	.17	.08	.08	.08	.12	.16	.13
person:moment (PM)		.01			.01		.01			
day:item (DI)										
day:moment (DM)										
couple:moment:item (CMI)										
couple:day:item (CDI)	.02					.01			.01	.01
person:moment:item (PMI)				.01	.01					
person:day:item (PDI)	.05	.03	.08	.11	.05	.04	.05	.06	.10	.06
day:moment:item (DMI)										
couple:day:moment:item (CDMI)	.02			.02	.01	.01			.01	.01
Error (e)	.28	.34	.33	.43	.27	.19	.33	.32	.40	.24

RS = relationship satisfaction scale, Ind = independence motivation scale, Pow = power motivation scale, A = agentic motivation scale (pooled independence and power), C = communal motivation scale. Empty cells have values < .005. Scales marked with an asterisk do not contain the same items in S1 and S2

in relationship satisfaction and communal motivation. Furthermore, relationship satisfaction additionally has a unique large *couple x day* component (CD; around 11%) and *couple x day x moment* component (CDM; around 14%), which indicates that some days and some specific moments are more satisfying for some couples than other days or moments.

Concerning nuisance terms, two sources of variances had substantial contributions across scales and studies: After controlling for between-person variance, participants still had systematically different mean levels between item responses in general (PI; around 18% of variance), and on specific days (PDI; around 10%).

Reliability estimation

Table 7 reports reliability estimates for both studies on all levels. On couple level, reliabilities range from .32 to .76,

on person level from .93 to .98, on day level from .61 to .88, and on moment level from .28 to .72.⁴

Scale correlations on four levels of aggregation

The raw bivariate correlations are not corrected for unreliability of the scales, which has to be kept in mind when comparing the absolute sizes between the three levels. As reliability is lowest on the between-moment level, also

⁴If the maximum number of days and moments is inserted, instead of the average number of answered moments and days, reliabilities are virtually identical for R_{BC} (up to +.003) and R_{BP} (S1: +.005, S2: +.003), and slightly larger for R_{WPD} (S1: +.027, S2: +.021). In a previous publication based on S1 (Zygar et al., 2018b), a shorter two-item scale for communal motivation was employed, consisting of items C-1 and C-2 (see Table 9). This more homogenous scale demonstrated the following reliabilities in the larger S2 sample: $R_{BC} = .58$, $R_{BP} = .97$, $R_{WPD} = .88$, and $R_{WPM} = .70$.

Table 6 Variance decomposition of item responses: relative systematic variances

Source	Sample 1, relative in %					Sample 2, relative in %				
	RS*	Ind	Pow*	A*	C	RS*	Ind	Pow*	A*	C
Theoretically relevant terms										
couple (C)	13	6	10	8	13	25	12	9	9	21
person (P)	11	28	17	17	24	12	23	28	20	14
day (D)				1						
moment (M)		2			1					
couple:day (CD)	12	1		1	2	10	3	2	1	3
person:day (PD)	1	16	8	3	9	7	13	10	6	9
couple:day:moment (CDM)	16	5	1		5	11	3		1	3
person:day:moment (PDM)	11	25	16	7	21	11	19	19	8	15
Nuisance terms										
item (I)										
couple:item (CI)	5	3	4	6	4	5	3	3	6	5
couple:moment (CM)		1								
moment:item (MI)				1						
person:item (PI)	17	6	29	30	10	10	13	18	27	17
person:moment (PM)		1	1		2		1			
day:item (DI)										
day:moment (DM)										
couple:moment:item (CMI)										
couple:day:item (CDI)	3					2			2	1
person:moment:item (PMI)				1	1				1	
person:day:item (PDI)	7	4	12	20	7	5	8	9	16	8
day:moment:item (DMI)										
couple:day:moment:item (CDMI)	3			4	1	2			1	1

RS = relationship satisfaction scale, Ind = independence motivation scale, Pow = power motivation scale, A = agentic motivation scale (pooled independence and power), C = communal motivation scale. Empty cells have values <.05%. Scales marked with an asterisk do not contain the same items in S1 and S2

Table 7 Reliability estimates

Scale	Sample 1				Sample 2			
	R_{BC}	R_{BP}	R_{WPD}	R_{WPM}	R_{BC}	R_{BP}	R_{WPD}	R_{WPM}
RS2*	.58	.95	.61	.36	.74	.97	.88	.65
RS3					.76	.97	.86	.58
Ind	.32	.93	.79	.50	.46	.96	.74	.44
Pow*	.43	.95	.73	.40	.37	.98	.79	.54
A*	.44	.96	.66	.28	.44	.98	.73	.38
C2	.42	.95	.88	.72	.58	.97	.86	.70
C3	.47	.96	.85	.65	.66	.97	.85	.63
C4	.50	.95	.87	.70	.70	.97	.86	.67

R_{BC} = between-couples reliability, R_{BP} = between-persons reliability, R_{WPD} = within-person/between-days reliability, R_{WPM} = within-person/between-moments reliability. RS2, RS3 = relationship satisfaction scale, measured with 2, resp. 3, items, Ind = independence motivation scale, Pow = power motivation scale, A = agentic motivation scale (pooled independence and power), C = communal motivation scale. Scales marked with an asterisk do not contain the same items in S1 and S2. RS2 = relationship mood + annoyance items in S1 and relationship mood + need satisfaction items in S2 (see Table 1)

lower correlations are to be expected. Table 8 reports the correlations on each level of aggregation.

Generally, the matrices show largely similar patterns across aggregation levels. In particular, all differences between the day level correlations and the moment level correlations are less than .09, with an average absolute difference of .03. The correlations on person level, however, show some stronger differences to the day and moment level correlations. Specifically, the correlation between power and independence motivation is around .32 on the person level, but close to zero on the day and moment level. Furthermore, the negative correlation between independence motivation and communal motivation is stronger on the day and moment level (r between $-.30$ and $-.38$) compared to the person level ($r = -.15$).

Discussion

We presented a model for estimating the reliability of experience sampling measures which are assessed at multiple moments per day, across several days, for persons within dyads. This design allows researchers to estimate a variance decomposition and reliability on four levels of aggregation, (a) between-couples, (b) between-persons, (c) within-person/between-days, and (d) within-person/between-moments. The model was applied to estimate variance components and reliabilities of five scales that are central to the study of motivational dynamics and relationship satisfaction in couples: State relationship satisfaction, communal motivation, and agency motivation, which has been assessed with two subscales,

Table 8 Correlations on four levels of aggregation

	RS2	RS3	Ind	Pow	A	C
Between-couple correlations						
RS2			-.47	-.22	-.41	.40
RS3	.97					
Ind	-.16	-.20		.42	.84	-.18
Pow	.02	-.02	.46		.85	.45
A	-.06	-.10	.79	.91		.16
C	.60	.58	-.16	.40	.20	
Between-person correlations						
RS2			-.22	.05	-.11	.40
RS3	.94					
Ind	-.12	-.21		.29	.80	-.15
Pow	-.18	-.27	.35		.81	.58
A	-.19	-.30	.72	.90		.27
C	.30	.27	-.15	.32	.17	
Within-person, between-day correlations						
RS2			-.15	.05	-.08	.37
RS3	.94					
Ind	-.19	-.20		-.05	.74	-.38
Pow	.03	.01	.09		.63	.50
A	-.09	-.11	.65	.82		.04
C	.41	.42	-.30	.40	.13	
Within-person, between-moment correlations						
RS2			-.13	.00	-.10	.28
RS3	.90					
Ind	-.13	-.15		-.06	.75	-.36
Pow	.02	.00	.00		.62	.41
A	-.07	-.09	.61	.79		-.01
C	.33	.33	-.30	.34	.09	

Upper triangle in each matrix shows S1, lower triangle shows S2. RS2, RS3 = relationship satisfaction scale, measured with 2, resp. 3, items, Ind = independence motivation scale, Pow = power motivation scale, A = agentic motivation scale (pooled independence and power), C = communal motivation scale

independence motivation and power motivation. Two intensive longitudinal studies provided data on more than 7508 unique surveys in Sample 1 and more than 47764 unique surveys in Sample 2.

Variance decomposition and reliability estimation

One research question for this study was about on which temporal level (between moments within a day, between days, between persons, between couples) most variance of relationship motivations and satisfaction is located. This also allows the investigation of the time scale of variability of motivational processes and relationship satisfaction. Four theoretically relevant sources of variance had the largest share across scales and studies: persons, specific moments of persons, couples, and specific days of persons. That means, some persons and some couples are to some extent generally closer, more satisfied, or have more agentic motivation than other persons or couples. Furthermore, the investigated scales varied both from day to day and from moment to moment. The within-day variance, from moment to moment, was around twice as large as the between-day variance, and nearly as large as the between-person variance. Hence, the pattern of results shows (a) the existence of systematic inter-individual differences in self-reported motivational states and relationship satisfaction, (b) systematic inter-couple differences, that indicate some dyadic similarity in couples, and (c) that these scale values show more short-time variability within a day than variability between days.

Concerning nuisance terms, two sources of variance had substantial contributions across scales (in particular agency motivation) and studies. First, after controlling for between-person differences, participants still systematically demonstrated person-specific mean levels of item responses. This can be due to differential item functioning, which indicates that an item might be measuring different latent constructs for members of different subgroups. Follow-up analyses with explanatory variables, such as gender, marital status, or relationship duration, might reveal which specific subgroups have a differing understanding of items. Second, persons had a differential item understanding on specific days. This can happen, for example, if items are interpreted differently at weekends (vs. workdays) by some persons. From a psychometric point of view, these sources of variance should be as small as possible for a general-purpose questionnaire.

When item responses were aggregated on person level, all scales showed near perfect reliability $>.93$ (S1) and $>.97$ (S2). Aggregated on day level (across four or five moments per day), reliability of the more homogeneous scales fell between $.73$ and $.88$. The two items for state relationship satisfaction in S1 were quite inhomogenous, resulting

in a lower reliability of $.61$. Furthermore, combining independence and power motivation into a higher-order agency scale decreased reliability to $.66$ in S1.

On the lowest level of aggregation, at each moment, this trend was even stronger. Homogeneous scales showed (relatively) better reliabilities ranging from $.40$ to $.70$. The moment-level reliabilities of the combined agency scale ($.28$ in S1, $.38$ in S2) and the two heterogeneous relationship satisfaction items in S1 ($.36$) were unsatisfactory. Hence, concerning reliability, the two-item relationship satisfaction scale from S2 (with items RS-1 and RS-3) seems preferable to the two-item scale from S1 (with items RS-3 and RS-4). Although the full three-item scale in S2 does not improve reliability compared to the two-item scale, it covers a broader content range and might have better validity. It might thus be preferred, depending on the research question (see Zygar-Hoffmann & Schönbrodt, 2020, for validity considerations associated with this item).

Validity: Scale intercorrelations

The scale intercorrelations on the different temporal levels revealed some relevant insights into the underlying constructs. Generally, the correlation matrices were rather similar on all levels and did not show strong indicators of a Simpson's paradox, where associations between variables are very different between aggregation levels or even flip their sign. However, there were two notable exceptions where the person level correlations differed from the day and moment level correlations.

First, the independence and power motivation scales showed a positive correlation around $.32$ on the between-person level. Persons who generally had more independence motivation also generally had more power motivation, which can be interpreted that these scales are two facets of the overarching agency motive factor, which represents "a superordinate need to feel as a capable, self-reliant individual" (Hagemeyer & Neyer, 2012, p. 3). Within person, however, they were independent with correlations close to zero: On moments or days where persons experienced a strong motivation for independence, they did not necessarily experience a concurrent motivation for power. A theoretically consistent interpretation would be that independence and power are different implementation styles of enacting agency in relationships. Although they do not go together at each moment in time, both are different (and to some extent exchangeable) ways to express a superordinate need for agency.

This correlation structure of the agency subscales has implications both for assessment and theory building. Zero correlations on a momentary level lead to low reliabilities of the combined agency scale. Consequently, unless one is explicitly treating agentic motivation as a formative

construct on the day or moment level, we generally recommend not to use that combined scale, but rather to treat both subscales as separate on the day or moment level. On the between-person level, in contrast, the subscales showed a substantial positive correlation, which was also reflected in higher reliabilities of the combined agency scale.

Dissociations of motivational processes and domains at different conceptual levels should also get more attention in theory building. Within-person processes do not necessarily reflect between-person structures, and vice versa (Molenaar, 2008). Consequently, theory building in motivation ideally covers both levels, and researchers should be careful when inferences and implications are transferred from one level (e.g., within-person experimental manipulations in the lab) to the other level (e.g., between-person structures of motivational domains). This call is in line with previous research that demonstrated differences in between-person and within-person structures of the Big Five personality traits (e.g., Borkenau & Ostendorf, 1998; Grice, Jackson, & McDaniel, 2006) or positive and negative affect (e.g., Brose, Voelkle, Lövdén, Lindenberger, & Schmiedek, 2015).

Second, independence and communal motivation were, to some extent, mutually exclusive on the daily and momentary level (with an r around $-.34$), but not so much on the between-person level ($r = -.15$). On a behavioral level this makes immediate sense, as it is difficult to be (emotionally) close to the partner, and at the same time to independently follow your own interest. On the motivational level, in contrast, such an ambiguity is imaginable, where persons simultaneously want to be close and distant from the partner. Empirically, however, the negative correlation shows that such ambiguous motivational states were rather rare. On the person level, in contrast, the correlation is only slightly negative, indicating that a person's general level of communal motivation was largely independent of the general level of independence motivation.

When the agency and the communion motive have been assessed as stable dispositions, they typically have shown negative correlations around $-.40$, both on an explicit level, assessed with self-report questionnaires (Hagemeyer et al., 2013), and on an implicit level, assessed with indirect methods (Hagemeyer & Neyer, 2012). In contrast to these previous results, we found slightly positive correlations of agentic and communal motivation on person level between $.17$ and $.27$, and correlations in the range of $-.01$ to $.13$ on the moment or day level. This deviation from previous results can partly be explained by the specific conceptualization of the combined agency scale in the current ESM studies. Inspecting the two agency subscales reveals that the independence subscale showed the expected negative correlation to communal motivation on day and moment level, and a weak negative correlation on person level. The explicit agency (dispositional) motive in the

studies cited above has been assessed with the ABC scales (Hagemeyer et al., 2013), which focus on the agentic aspect of “forming separations” (Bakan, 1966). Hence, items such as “I like to be completely alone” from the ABC scales are most closely related to the independence motivation items in the current study, which did show the expected negative correlation (albeit, with a smaller effect size).

The positive correlation between power motivation and communal motivation on all levels of aggregation might be due to two different factors. First, some of our ESM power items were inspired by prosocial aspects of the power motive as described in Winter (1994) and Hagemeyer and Neyer (2012), where power motivation includes supportive behaviors/motivations within the relationship as well as a positive influence on the partner. Therefore, our ESM power items focus on prosocial aspects of power and do not address aspects that are usually valued negatively, such as dominance in the relationship. Thus, the power and the communion scale share a common positive connotation. Second, in contrast to independence, the power aspect of agency often requires contact to the partner. Thus, the power and communion items share a common mode of implementation, namely seeking proximity to the partner. One way to further disentangle different facets of agency would be to separately investigate coercive or aggressive dominance as another facet of agency (Suessenbach, Loughnan, Schönbrodt, & Moore, 2019). Dominance motivated instrumental behavior in that sense also requires proximity to the partner, but does not share the same positive connotation as our operationalization of power motivation does.

Implications for Future Research

The results have some direct implications for the design and the statistical analysis of studies using these scales. First, a considerable amount of variance was located on the between-couple level. Hence, the dyadic structure should not be ignored in statistical analyses. Second, all scales showed more variance between moments (within a day) than between days. Hence, a daily diary, which has only a single measurement per day, probably misses large parts of the fluctuations in these constructs. Third, the analyses revealed an unexpected large amount of differential item functioning between persons, but also between days within persons. This underscores the importance of proper psychometric analyses and intensive pilot testing of the ESM item wordings and how participants understand them. In the current two studies, we did multiple pilot studies where we refined items and asked participants in S1 in a post-ESM-questionnaire how they interpreted the items, using open ended questions. Additionally, in both studies before starting the ESM part, all participants received instructions (written in S1, video-recorded in S2) on how to interpret

the items, and could look up the instructions for each item during the study. Despite these efforts, not all persons had the same understanding of items, and we suppose that this source of variance might be even larger in studies that do not have the same amount of pretesting.

Fourth, change reliability on the moment-to-moment level was mostly not satisfactory. When such unreliable scale scores are used as predictors or outcomes in follow-up statistical models, two aspects influence statistical power, working in opposite directions: As reliability is lowest on the most fine-grained moment level, statistical power is lowered. At the same time, this level also has the largest number of measurement points, which in turn increases the statistical power to detect existing effects. For example, despite the low reliability of .36 in the two-item relationship satisfaction scale used in S1, Zygar et al. (2018b) found reliable evidence for hypothesized effects on this outcome variable (see robustness check, footnote 10).

When designing an ESM study, specifically the frequency, timing, and length of measurements, several factors must be considered. The expected rate of change of a construct determines the frequency of sampling, and reliability and burden of participants must be balanced (for further aspects regarding the sampling plans of state relationship satisfaction, see Zygar-Hoffmann and Schönbrodt, 2020). For planning a study, power analyses are needed to investigate the relative impact of these determinants on statistical power.

Limitations

Several limitations follow from the assumptions that have to be made for computing the variance components (Shrout & Lane, 2012). Most importantly, the components of Eq. 1 are assumed to be independent, which is most likely violated in multiple ways. Although the random intercept for *couple* accounts for some of the dyadic interdependence, it does not model covariances between dyad members. This ignorance of dyadic covariances is acceptable if covariances are positive, as was the case in our data sets. In this case, variances are shifted towards a higher level (e.g., between-person variance gets reallocated to the couple level if persons within a couple are more alike to each other), which makes sense. However, if dyadic covariances are negative, this can lead to estimation problems and/or biased variance estimates. Another likely violation of the independence assumption is that consecutive time points in an ESM presumably have some autoregressive effect, which is ignored in the GT model. Finally, the model assumes equal item loadings. Simulations by Lane and Shrout (2010) showed that the GT method underestimates the reliability to the extent that the assumption of equal item loadings is violated. One way to get closer to equal items loadings can be the standardization of items before calculating the scale, in particular when

items do not have the same response scale. This, however, is not always desirable in terms of interpretation.

Bearing these limitations in mind, we think that this model is an acceptable approximation for our current research question. We note, however, that this does not necessarily generalize to other data sets, in particular when negative dyadic covariances are present.

Further, the analysis relates only to our specific operationalization of motivation and relationship satisfaction. The statelikeness of a phenomenon is also a feature of the specific item wording, and a different phrasing might shift the variance components more towards the person or couple level. Furthermore, we only used two to four items per scale. This gave only few possibilities to do item selection. Scale development for ESM studies can benefit from a larger item pool in a pilot study that allows to choose an item set that balances homogeneity and content width. Joint efforts to collect ESM items and curate the documentation of their psychometric quality are another important step that helps to achieve reliable and valid ESM scales (see e.g., <http://www.esmitemrepository.com>; Kirtley et al., 2020).

Conclusions

Creating items and scales for ESM has some special challenges. Many ESM studies use ad-hoc scales with very few items, and proper psychometric analyses are rarely seen. Here we extend the psychometric toolbox by proposing a variance decomposition and reliability model for data sets where constructs are assessed with multiple items at multiple moments each day in couples. Applying this model to four motivation scales and different scales for state relationship satisfaction showed substantial variability on state level, different reliabilities depending on the level of aggregation, and theoretically interesting patterns of scale intercorrelations. The model can also easily applied to data set where persons are not nested in couples (by removing all terms related to the couple), and we encourage researchers to use the provided R-Scripts on the OSF to calculate reliability analyses of their ESM scales for individual as well as dyadic study designs.

Open Practices Statement

Due to the dyadic nature of the data set, we cannot make the data fully openly available. The data and materials for Sample 1 (<https://doi.org/10.5160/psychdata.zrce16dy99>) and Sample 2 (<https://doi.org/10.5160/psychdata.zrce18mo99>) are published as scientific use files, which restricts access to scientific users. The reliability analyses presented here were not preregistered. Reproducible scripts for all data analyses reported in this paper are available at the Open Science Framework (<https://osf.io/jmeaw/>).

Appendix A: Item wordings

Table 9 Original experience sampling items for the assessment of motivation

Domain & Label	Instruction		Scale
	Partner present	Partner not present	
		Stellen Sie sich nun vor, Sie bekämen jetzt ca. 30 Minuten Zeit aufgrund von Leerlauf geschenkt (freie Zeit, in der Sie keine Verpflichtungen erledigen können), die Sie mit Ihrem Partner verbringen könnten aber nicht müssten (d.h. Ihr Partner hätte gerade auch 30 Minuten Zeit und könnte in Ihrer Nähe sein).	
		Wünschen Sie sich jetzt gerade:	Wünschen Sie sich in dieser Zeit:
Communion	C-1	Erfahrungen, Gedanken oder Gefühle mit Ihrem Partner gemeinsam zu teilen? (z.B. von einem Erlebnis, einer Idee, Vorfreude, Sorgen erzählen)	Likert scale: S1:
Communion	C-2	Emotionale Zuneigung von Ihrem Partner zu erhalten? (z.B. eine liebevolle Geste oder liebevolle Worte)	4 = ja, sehr stark, 3 = ja, stark,
Independence	I-1	Unabhängig von Ihrem Partner zu handeln und Entscheidungen zu treffen? (z.B. nicht auf Ihren Partner angewiesen zu sein, nicht die Meinung Ihres Partners zu einem Thema einzuholen oder ohne die Unterstützung Ihres Partners ein Problem zu lösen.)	2 = ja, mittelmäßig, 1 = ja, aber nur schwach, 0 = nein, danach habe ich gerade kein , Bedürfnis -1 = nein, das würde mich gerade sogar ein bisschen stören -2 = nein, das würde mich gerade sehr stören
Independence	I-2	Alleine Ihren eigenen Interessen nachzugehen? (z.B. ein eigenes Hobby ausüben; an einem eigenen Projekt weiterzuarbeiten.)	S2: 4 = ja, sehr stark, 3 = ja, stark,
Power	P-1	Die Gefühle oder das Verhalten von Ihrem Partner in irgendeiner Weise zu beeinflussen? (z.B. Ihren Partner zum Lachen zu bringen oder zu überraschen; Ihren Partner von einer Meinung zu überzeugen; Begehren bei Ihrem Partner auszulösen.)	2 = ja, mittelmäßig, 1 = ja, aber nur schwach, 0 = nein, danach habe ich gerade kein , Bedürfnis
Power	P-2	Dass es einen Austausch mit Ihrem Partner gibt, in dem es um Sie geht, Sie im Mittelpunkt stehen? (z.B. dass <u>Ihr Partner Ihre Bedürfnisse über die Eigenen stellt; Ihr Partner Ihnen volle Aufmerksamkeit schenkt; Sie Ihren Partner beeindrucken.</u>)	-1 = nein, das würde mich gerade stören
Power	P-3	Dass Ihr Partner sich nach Ihnen richtet? (z.B. dass <u>Ihr Partner Ihre Bedürfnisse über die Eigenen stellt; Ihre Wünsche erfüllt.</u>)	
Communion	C-3	Wie emotional nahe wären Sie Ihrem Partner jetzt gerade gerne?	Discrete slider from 1 = mit etwas Abstand to 7 = maximal nah in S1 and 1 = Abstand to 7 = maximale Nähe in S2, each position showing one picture of the Inclusion of Other in the Self Scale ^a
Communion	C-4	Stellen Sie sich vor, Sie bekämen jetzt zwei Stunden Zeit aufgrund von Leerlauf geschenkt (freie Zeit, in der Sie keine Verpflichtungen erledigen können), die Sie mit Ihrem Partner verbringen könnten aber nicht müssten (d.h. Ihr Partner hätte gerade auch 2 Stunden Zeit und könnte in Ihrer Nähe sein). Wie würden Sie diese 2h (Leerlauf-)Zeit gerade gerne nutzen?	Continuous slider from 1 (S1) or 0 (S2) = <i>Komplett ohne Ihren Partner (als Zeit für mich)</i> to 7 (S1) or 10 (S2) = <i>Komplett mit Ihrem Partner (als gemeinsame Zeit)</i>

Dashed underlined black text was not part of the items in S1, dashed underlined gray text was only part of the item in S1

^aAron, Aron, & Aron, (1992). We consider item C4 as bipolar item that covers both the communal and agentic domain with regard to independence. In the current analyses, we only use the item as a communion item. These items can be reused under a CC-BY4.0 license

Table 10 English translation of experience sampling items for the assessment of motivation

Domain & Label	Instruction		Scale
	Partner present	Partner not present	
		Imagine you would now <u>get to spend have approx. 30 minutes of time</u> (free time, in which you cannot take care of duties), time which you could spend with your partner, but wouldn't have to (i.e. <u>your partner would also have those two hours, and could be by your side</u>).	
	Right now, do you wish:	In this given time, do you wish:	
Communion	C-1	To share experiences, thoughts or feelings with your partner? (e.g. to tell your partner about an experience, an idea, a pleasant anticipation or worries.)	Likert scale:
Communion	C-2	To receive emotional affection from your partner? (e.g. a loving gesture or loving words.)	S1: 4 = yes, very strongly, 3 = yes, strongly, 2 = yes, moderately, 1 = yes, but only weakly, 0 = no, I don't need this right now,
Independence	I-1	To act and decide independent of your partner? (e.g. <u>don't have to rely on your partner, don't obtain the opinion of your partner on a topic or to solve a problem without the support of your partner.</u>)	-1 = no, that would rather bother me a little bit -2 = no, that would bother me quite a lot
Independence	I-2	To solitarily pursue your own interests? (e.g. to pursue an own hobby; to work on an own project.)	
Power	P-1	To influence the feelings or behavior of your partner in any way? (e.g. to make your partner laugh or to surprise your partner; to convince your partner of an opinion; to cause your partner to desire you.)	S2: 4 = yes, very strongly, 3 = yes, strongly, 2 = yes, moderately, 1 = yes, but only weakly, 0 = no, I don't need this right now, -1 = no, that would bother me right now
Power	P-2	That there is an exchange with your partner, which is about you, where you are the center of attention? (e.g. that <u>your partner puts your needs above his/her own; your partner gives you full attention; that you impress your partner.</u>)	
Power	P-3	That your partner fits in with your wishes? (e.g. that <u>your partner puts your needs above his/her own; satisfies your wishes.</u>)	
Communion	C-3	How emotionally close would you want to be to your partner at the moment?	Discrete slider from 1 = with some distance to 7 = maximally close in S1 and 1 = distance to 7 = maximal closeness in S2, each position showing one picture of the Inclusion of Other in the Self Scale ^a
Communion	C-4	Imagine you would now <u>get to spend have two hours of time</u> (free time, in which you cannot take care of duties), time which you could spend with your partner (i.e. <u>your partner would also have those two hours, and could be by your side</u>).	Continuous slider from 1 (S1) or 0 (S2) = Entirely without your partner (as me-time) to 7 (S1) or 10 (S2) = Entirely with your partner (as shared time)
		How would you like to spend this time right now?	

S1 = Sample 1, S2 = Sample 2. Dashed underlined black text was not part of the items in S1, dashed underlined gray text was only part of the item in S1

^aAron, Aron, and Smollan (1992). We consider item C4 as bipolar item that covers both the communal and agentic domain with regard to independence. In the current analyses, we only use the item as a communion item. These items can be reused under a CC-BY4.0 license

Author Note We embrace the values of openness and transparency in science (<http://www.researchtransparency.org/>). The data of both studies are available as scientific use files (Zygar et al. 2018a for Study 1; Zygar-Hoffmann et al. 2020 for Study 2). The data of Study 1 have previously been used by Zygar et al. (2018b),

Pusch, Schönbrodt, Zygar-Hoffmann, & Hagemeyer, (2020), Zygar-Hoffmann et al. (2020). The data of Study 1 and Study 2 have previously been used by Zygar-Hoffmann and Schönbrodt (2020). Reproducible scripts for all data analyses reported in this paper are available at the Open Science Framework (<https://osf.io/jmeaw/>).

Acknowledgements We thank David Kenny for helpful comments regarding the generalizability model. We thank Helen Baumann for assistance during data collection of Study 1, as well as Tobias Kächele, Lukas Müller, Ludwig Zellner, and Ronan Feig for app development.

Funding Open Access funding enabled and organized by Projekt DEAL. This research was funded by the German Research Foundation (DFG SCHO 1334/5-1, Felix Schönbrodt; HA 6884/2-1 Birk Hagemeyer).

Author Contributions B.H.: Funding acquisition, Investigation, Writing - original draft, and Writing - review & editing. C.Z.-H.: Conceptualization, Data curation, Investigation, Project administration, Writing - original draft, and Writing - review & editing. F.S.: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Supervision, Writing - original draft, and Writing - review & editing. S.N.: Formal analysis, Methodology, Software, Validation, and Writing - review & editing. S.P.: Investigation, Writing - original draft, and Writing - review & editing.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adolf, J. K., & Fried, E. I. (2019). Ergodicity is sufficient but not necessary for group-to-individual generalizability. *Proceedings of the National Academy of Sciences*, 201818675. <https://doi.org/10.1073/pnas.1818675116>.
- Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63, 596–612.
- Arslan, R. C., Walther, M. P., & Tata, C. S. (2019). Formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*. <https://doi.org/10.3758/s13428-019-01236-y>.
- Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion*. Chicago: Rand McNally.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & Behavior*, 81, 179–209. <https://doi.org/10.1016/j.physbeh.2004.02.004>.
- Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York: Guilford Press.
- Borkenau, P., & Ostendorf, F. (1998). The big five as states: How useful is the five-factor model to describe intraindividual variations over time? *Journal of Research in Personality*, 32, 202–221. <https://doi.org/10.1006/jrpe.1997.2206>.
- Brose, A., Voelkle, M. C., Lövdén, M., Lindenberg, U., & Schmiedek, F. (2015). Differences in the between-person and within-person structures of affect are a matter of degree. *European Journal of Personality*, 29, 55–71. <https://doi.org/10.1002/per.1961>.
- Cranford, J. A., Shrout, P. E., Iida, M., Rafaeli, E., Yip, T., & Bolger, N. (2006). A procedure for evaluating sensitivity to within-person change: Can mood measures in diary studies detect change reliably? *Personality and Social Psychology Bulletin*, 32, 917–929. <https://doi.org/10.1177/0146167206287721>.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (Eds.) (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York: Wiley.
- Dewitte, M., & Mayer, A. (2018). Exploring the link between daily relationship quality, sexual desire, and sexual activity in couples. *Archives of Sexual Behavior*, 47, 1675–1686. <https://doi.org/10.1007/s10508-018-1175-x>.
- Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, 115, E6106–E6115. <https://doi.org/10.1073/pnas.1711978115>.
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80, 1011–1027. <https://doi.org/10.1037/0022-3514.80.6.1011>.
- Grice, J. W., Jackson, B. J., & McDaniel, B. L. (2006). Bridging the Idiographic-Nomothetic divide: A Follow-up study. *Journal of Personality*, 74, 1191–1218. <https://doi.org/10.1111/j.1467-6494.2006.00407.x>.
- Hagemeyer, B., Neberich, W., Asendorpf, J. B., & Neyer, F. J. (2013). (In)Congruence of implicit and explicit communal motives predicts the quality and stability of couple relationships. *Journal of Personality*, 81, 390–402. <https://doi.org/10.1111/jopy.12016>.
- Hagemeyer, B., & Neyer, F. J. (2012). Assessing implicit motivational orientations in couple relationships: The Partner-Related Agency and Communion Test (PACT). *Psychological Assessment*, 24, 114–128. <https://doi.org/10.1037/a0024822>.
- Hagemeyer, B., Neyer, F. J., Neberich, W., & Asendorpf, J. B. (2013). The ABC of social desires: Affiliation, being alone, and closeness to partner. *European Journal of Personality*, 27, 442–457. <https://doi.org/10.1002/per.1857>.
- Hagemeyer, B., Schönbrodt, F. D., Neyer, F. J., Neberich, W., & Asendorpf, J. B. (2015). When ‘together’ means ‘too close’: Agency motives and relationship functioning in coresident and living-apart-together couples. *Journal of Personality and Social Psychology*, 109, 813–835. <https://doi.org/10.1037/pspi0000031>.
- Heckhausen, J., & Heckhausen, H. (Eds.) (2018). *Motivation and Action* (3rd ed.). Springer International Publishing. Retrieved April 26, 2019, from <https://www.springer.com/de/book/9783319650937>.
- Hofmann, W., Finkel, E. J., & Fitzsimons, G. M. (2015). Close relationships and self-regulation: How relationship satisfaction facilitates momentary goal pursuit. *Journal of Personality and Social Psychology*, 109, 434–452. <https://doi.org/10.1037/pspi0000020>.
- Hofmann, W., Vohs, K. D., & Baumeister, R. F. (2012). What people desire, feel conflicted about, and try to resist in everyday life. *Psychological Science*, 23, 582–588. <https://doi.org/10.1177/0956797612437426>.
- Horstmann, K. T., & Ziegler, M. (2020). Assessing personality states: What to consider when constructing personality state measures. *European Journal of Personality*, 34, 1037–1059. <https://doi.org/10.1002/per.2266>.
- Impett, E. A., Gable, S. L., & Peplau, L. A. (2005). Giving up and giving in: The costs and benefits of daily sacrifice in intimate relationships. *Journal of Personality and Social Psychology*, 89, 327–344. <https://doi.org/10.1037/0022-3514.89.3.327>.

- Karney, B., & Bradbury, T. (1995). The longitudinal course of marital quality and stability: A review of theory, method, and research. *Psychological Bulletin*, *118*, 3–34. <http://www.sciencedirect.com/science/article/B6WY5-46R0XFP-12/2/704bf89457e42e1d3f6cbee359f5feb2>.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford.
- Kievit, R., Frankenhuus, W. E., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*. <https://doi.org/10.3389/fpsyg.2013.00513>.
- Kindt, S., Vansteenkiste, M., Loeys, T., & Goubert, L. (2016). Helping motivation and well-being of chronic pain couples: A daily diary study. *Pain*, *157*, 1551–1562. <https://doi.org/10.1097/j.pain.0000000000000550>.
- Kirtley, O. J., Hiekkaranta, A. P., Kunkels, Y. K., Eisele, G., Verhoeven, D., Van Nierop, M., & Myin-Germeys, I. (2020). The experience sampling method (ESM) item repository. *OSF*. <https://doi.org/10.17605/OSF.IO/KG376>.
- Lane, S. P., & Shrout, P. E. (2010). Abstract: Assessing the reliability of Within-Person change over time: A dynamic factor analysis approach. *Multivariate Behavioral Research*, *45*, 1027–1027. <https://doi.org/10.1080/00273171.2010.534380>.
- Laurenceau, J.-P., & Bolger, N. (2005). Using diary methods to study marital and family processes. *Journal of Family Psychology*, *19*, 86–97. <https://doi.org/10.1037/0893-3200.19.1.86>.
- McClelland, D. C. (1987). *Human motivation*. New York: Cambridge University Press.
- Medaglia, J. D., Jeronimus, B. F., & Fisher, A. J. (2019). Reply to Adolf and Fried: Conditional equivalence and imperatives for person-level science. *Proceedings of the National Academy of Sciences*, 201820221. <https://doi.org/10.1073/pnas.1820221116>.
- Molenaar, P. C. (2008). On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation. *Developmental Psychobiology*, *50*, 60–69. <https://doi.org/10.1002/dev.20262>.
- Muise, A., Impett, E. A., & Desmarais, S. (2013). Getting it on versus getting it over with: Sexual motivation, desire, and satisfaction in intimate bonds. *Personality and Social Psychology Bulletin*, *39*, 1320–1332. <https://doi.org/10.1177/0146167213490963>.
- Nezlek, J. B. (2016). A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*, 1–7. <https://doi.org/10.1016/j.jrp.2016.06.020>.
- Pusch, S., Schönbrodt, F. D., Zygar-Hoffmann, C., & Hagemeyer, B. (2020). Truth and wishful thinking: How interindividual differences in communal motives manifest in momentary partner perceptions. *European Journal of Personality*, *34*, 115–134. <https://doi.org/10.1002/per.2227>.
- Core Team, R. (2020). *R: A language and environment for statistical computing. manual R Foundation for Statistical Computing*. Austria: Vienna. <https://www.R-project.org/>.
- Schoebi, D. (2008). The coregulation of daily affect in marital relationships. *Journal of Family Psychology*, *22*, 595–604. <https://doi.org/10.1037/0893-3200.22.3.595>.
- Schönbrodt, F. D., & Gerstenberg, F. X. R. (2012). An IRT analysis of motive questionnaires: The Unified Motive Scales. *Journal of Research in Personality*, *46*, 725–742. <https://doi.org/10.1016/j.jrp.2012.08.010>.
- Schultheiss, O. C., & Brunstein, J. C. (2010). *Implicit motives*. Oxford: Oxford University Press.
- Schultheiss, O. C., & Köllner, M. G. (2021). Implicit motives. In O. P. John, & R. W. Robins (Eds.) *Handbook of personality psychology: Theory and research (4th ed.)*, (pp. 385–410). New York: Guilford Press.
- Scott, S. B., Sliwinski, M. J., Zawadzki, M., Stawski, R. S., Kim, J., Marcussen-Clavertz, D., . . . , Smyth, J. M. (2018). A coordinated analysis of variance in affect in daily life. *Assessment*, *107319111879946*. <https://doi.org/10.1177/1073191118799460>.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. London: SAGE.
- Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In Mehl, M. R., & Conner, T. S. (Eds.) *Handbook of research methods for studying daily life*. Retrieved November 23, 2018, from <https://www.guilford.com/books/Handbook-of-Research-Methods-for-Studying-Daily-Life/Mehl-Conner/9781462513055/contents>, (p. New York): Guilford Press.
- Suessenbach, F., Loughnan, S., Schönbrodt, F. D., & Moore, A. B. (2019). The dominance, prestige, and leadership account of social power motives. *European Journal of Personality*, *33*, 7–33. <https://doi.org/10.1002/per.2184>.
- Winter, D. G. (1994). *Manual for scoring motive imagery in running text (4th edn.)* Ann Arbor: University of Michigan.
- Zygar, C., Hagemeyer, B., Pusch, S., & Schönbrodt, F. D. (2018a). From motive dispositions to states to outcomes: An intensive experience sampling study on communal motivational dynamics in couples. *European Journal of Personality*, *32*, 306–324. <https://doi.org/10.1002/per.2145>.
- Zygar, C., Hagemeyer, B., Pusch, S., & Schönbrodt, F. D. (2018b). *From motive dispositions to states to outcomes: Research data of an intensive experience sampling study on communal motivational dynamics in couples (Version 2.1.0) [data and documentation]*. Trier: Psychologisches Datenarchiv PsychData des Leibniz-Zentrums für Psychologische Information und Dokumentation ZPID. Retrieved June 13, 2018, from <https://doi.org/10.5160/psychdata.zrce16dy99.v20100>.
- Zygar-Hoffmann, C., Hagemeyer, B., Pusch, S., & Schönbrodt, F. D. (2020). *Eine große Längsschnittstudie zu Motivation Verhalten und Zufriedenheit von Paaren: Forschungsdaten einer vierwöchigen Experience-Sampling-Studie mit einer Vor- Nach- und einjährigen Follow-up-Befragung. [A large longitudinal study on motivation, behavior and satisfaction in couples. Research data from a four-week experience sampling study with a pre-, post-, and one-year follow-up assessment.]* Trier: Psychologisches Datenarchiv PsychData des Leibniz-Institut für Psychologie ZPID. <https://doi.org/10.5160/psychdata.zrce18mo99>.
- Zygar-Hoffmann, C., Pusch, S., Hagemeyer, B., & Schönbrodt, F. D. (2020). Motivated behavior in intimate relationships: Comparing the predictive value of motivational variables. *Social Psychological Bulletin*, *15*. <https://doi.org/10.32872/spb.2873>.
- Zygar-Hoffmann, C., & Schönbrodt, F. D. (2020). Recalling experiences: Looking at momentary, retrospective and global assessments of relationship satisfaction. *Collabra: Psychology*, *6*, 7. <https://doi.org/10.1525/collabra.278>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.