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

Felix D. Schönbrodt
Ludwig-Maximilians-University Munich

Caroline Zygar-Hoffmann
Ludwig-Maximilians-University Munich

Steffen Nestler
University of Münster

Sebastian Pusch
Friedrich Schiller University Jena

Birk Hagemeyer
Friedrich Schiller University Jena

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, three reliability coefficients are derived: 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, $\geq$ 7508 unique surveys; n_2 = 510 persons, 5 surveys each day for 28 days, $\geq$ 47871 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, which consist 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 .95 to .98 (person level), .52 to .86 (day level), and .28 to .70 (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, in press).

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, in press). 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.

To investigate the time scale and levels of motivation, 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, Hagemeyer, Pusch, & Schönbrodt, 2018a). In this study, we focus on the dynamics of motivation in the life-domain of

couple 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 a ESM data structure where multiple moments are nested in a day, 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 analyses whether short and intensive (within day) measurements, or longer (but less intensive) daily diaries, are more appropriate for the research question at hand.

In selecting motivational variables relevant for couple relationship, 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 couple 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 couple relationships (Hagemeyer & Neyer, 2012; Hagemeyer, Neyer, et al., 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, Neyer, 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.

To date, only two previous studies addressed motivational processes within partners of a couple. Hagemeyer et al. (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 agency motivation. Zygar et al. (2018a) used self-reports of the partner-related communal motivation in a two-week experience-sampling design with five assessments per day. Corresponding with findings on the between-person level, momentary variations (over the course of a few hours) in communal motivation were positively related to variations in communal behavior and relationship satisfaction. Thus, there is evidence that partner-related agentic and communal motivation are indeed relevant for the study of couple relationships at a process level.

In addition, 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 et al. (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 statistical model for variance decomposition and reliability estimation (Cranford et al., 2006) 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 (between-person, within-person/between-days, and within-person/between-moments). (4) Evaluate one aspect of the scales' validity by inspecting scale intercorrelations at three levels (between-person, within-person/between-days, and within-person/between-moments).

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: https://doi.org/10.5160/psychdata.zrce16dy99, Sample 2: We currently prepare the submission of a scientific use file to a repository. Data available for review upon request).

Samples

Data from two intensive experience sampling studies were used. Sample 1 (henceforward, S1) uses a data set from Zygar et al. (2018a) which is available as a scientific use file (Zygar, Hagemeyer, Pusch, & Schönbrodt, 2018b). This data set includes ESM data from 130 persons (52% women) nested in 68 heterosexual couples. Participants' mean age was 22.39 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 et al. (2018a).

Sample 2 (S2) includes ESM data from 510 persons (50% women) nested in 259 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 31.40 years and individuals were on average 7.10 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 two (four) weeks that followed, they took part in an experience sampling study, where they answered questions five times a day on their own smartphones for 14 days (S1) or 28 days (S2), 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 the period of eight months).

The surveys were completed in a median time of 3.28 minutes (S1) and 2.70 minutes (S2). When notified, individuals had 45 minutes to complete the survey, which included the same questions at each assessment. The 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), 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. (2018a) and https://osf.io/v2uxs/.

Experience sampling items

State motivation. At each measurement occasion, three five motivational states were assessed (see Tables A1 and A2 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. All reported results refer to the full

three-item scale, except the reliability 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.

Statistical procedure

Different models for estimating reliability in intensive longitudinal measures have been proposed (Cranford et al., 2006; Nezlek, 2016; Schoebi, 2008; Shrout & Lane, 2012). Our model is based on the Cranford et al. (2006) model, which we extended to include another level of measurement (moments crossed with days) and the 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 other factors. From the same variance decomposition, reliability estimates can be derived based on generalizability theory (Cranford et al., 2006; Shrout & Lane, 2012).

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). We treated dyad members as indistinguishable, as gender-specific effects can better be modeled as fixed effects in a follow-up model.[1]

Variance decomposition. Following generalizability theory, the full variance decomposition model is formalized as a four-way analysis of variance. For a person p in couple c, responding to motivation item i in moment m on day d, the model for communal motivation Y_{cpdmi} is

$$
\begin{aligned}
Y_{cpdmi} = &\mu + C_c + P_p + D_d + M_m + I_i + \\
& (CD)_{cd} + (CM)_{cm} + (CI)_{ci} + (PD)_{pd} + \\
& (PM)_{pm} + (PI)_{pi} + (DM)_{dm} + (DI)_{di} + (MI)_{mi} + \\
& (CDM)_{cdm} + (CDI)_{cdi} + (CMI)_{cmi} + (PDM)_{pdm} + \\
& (PDI)_{pdi} + (PMI_{pmi}) + (DMI)_{dmi} + \\
& (CDMI)_{cdmi} + \\
& e_{cpdmi}.
\end{aligned}
\tag{1}
$$

Uppercase variables denote the factors *couple (C)*, *person (P)*, *day (D)*, *moment (M)*, and *item (I)*. In our design, the four-way interaction $(PDMI)_{pdmi}$ cannot be distinguished from the error term, because we have no replicate measurements for that specific moment x day interaction. Therefore, the term is subsumed under the error term and does not appear in Eq. (1).

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. Likewise,

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 mostly started the study on different calendar days in S1, or on a different Mondays in S2. Therefore factor D does not capture events which are specific to the calendar day across participants, but rather systematic variance due to the onset and duration of the study. The specific values for the number of items ($i = 1 \ldots j$), the number of moments nested within each day ($m = 1 \ldots l$), and the number of days ($d = 1 \ldots k$) is given in Table 2.

A priori, we did not expect substantial systematic variation for some factors of the design. Specifically, we conceptualize days as being nested under persons, as persons started on different calendar days and we did not expect systematic effects of the passage of time. That means, we expected for example that day 7 of person A has nothing in common with day 7 of person B, if these persons are from different couples. Likewise, we would not expect that a certain item has a specific meaning on certain days or on certain moments in general, or on certain days for certain persons. 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_{pdm}, denotes a specific survey of a specific person on a specific day. It indicates whether this person has responded differently to this survey on the day (across the items). The meaning of the other components together with an explanation of their respective variance components can be found in Table 3.

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 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.

Data preprocessing. The items of our scales 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.[2] As we wanted to match our reliability analysis

---

[1]Exploratively, we also ran a model with distinguishable dyads, which yields comparable results.

[2]We 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

Table 1

Experience Sampling Items for the Assessment of Relationship Satisfaction.

Label	Instruction	Scale
Relationship mood (RS-1)	Original: **Wie geht es Ihnen *jetzt gerade* mit Ihrer Beziehung?** [English: **How do you feel about your relationship *at the moment*?**]	Continuous slider from 1 (S1) or 0 (S2) = *schlecht [bad]*, over 3.5 (S1) or 5 (S2) = *neutral [neutral]*, to 7 (S1) or 10 (S2) = *außergewöhnlich gut [exceptionally good]*
Annoyance (RS-2)	Original: **Wie genervt sind Sie *jetzt gerade* von Ihrem Partner?** [English: **How annoyed are you about your partner *at the moment*?**]	Continuous slider from 1 (S1) or 0 (S2) = *überhaupt nicht [not at all]*, to 7 (S1) or 10 (S2) = *stark [strongly]*
Need satisfaction (RS-3)	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 = *total frustriert [totally frustrated]*, over 5 = *neutral [neutral]*, to 10 = *total erfüllt [totally satisfied]*

Note. S1 = Sample 1, S2 = Sample 2. The need satisfaction item was not assessed in S1. The annoyance item was reverse coded for scale calculation.

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, with a maximum difference of ±.02).

Furthermore, we recoded one reversed item for relationship satisfaction and reformatted the data into the long format, where each row represents the value of one item answer. We estimated the variance components from Eq. (1) using the *lmer* package (Bates, Mächler, Bolker, & Walker, 2015). The specific function call is in the reproducible scripts on the OSF.

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

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

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 for an additional temporal level (moments nested in days). Dyadic interdependence is taken into account in the step of variance decomposition, but terms including the couple factor C do not contribute to reliability formulas.

Depending on the level where reliability should be assessed, different terms contribute to the numerator (the true

score variance) and the denominator (the observed variance). For example, if we are interested in the measurement of purely within-person changes, the variance of the term PI_{pi} (i.e., person x item), σ_{PI}^2, neither contributes to stable variances nor to the error term, as mean level biases in item understanding between persons are irrelevant for relative within-person assessments. Likewise, stable 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.

For computing *between-person reliability* (averaging all measurements of a person across the entire study), R_{BP}, we assumed that days are random (and not fixed), because participant started on different days across a period of several months, and the study period is not contingent on some common event. Moments, in contrast, were treated as fixed, as the $l = 4$ or $l = 5$ moments each day (from morning to evening) were assumed to be comparable for each person.[3] We use Equation (8) from Shrout and Lane (2012), including the *person x day x item* interaction, *PDI*, as an additional source of error, but not moment-to-moment variance, as all participants are assumed to have the same effects of moments:

---

same response scale.

[3]Specifically, 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 because participants knew that they would not be able to answer in these periods.

$$R_{BP} = \frac{\sigma_P^2 + [\sigma_{PI}^2/j]}{\begin{array}{c}\sigma_P^2 + [\sigma_{PI}^2/j] + [\sigma_D^2/k] + [\sigma_{PD}^2/k] \\ + [\sigma_{PDI}^2/(k*j)] + [\sigma_e^2/(k*l*j)]\end{array}} \quad (3)$$

The constant j is the number of items, k is the number of days, l is the number of moments within each day (see Table 2). In the GT reliability computation, variance components are divided by the number of replications that are averaged when aggregating the scale scores. For example, the *person x item* variance is divided by j, the number of items; the residual error term in R_{BP} is divided by $k*l*j$ to take into account the increase in precision that results from averaging j items, assessed at l moments at each of the k days of a person.

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

$$R_{WPD} = \frac{\sigma_{PD}^2 + [\sigma_{PDI}^2/j]}{\sigma_{PD}^2 + \sigma_D^2 + [\sigma_{PDI}^2/j] + [\sigma_e^2/(l*j)]} \quad (4)$$

Again we treat moments as fixed, because we do not randomly sample moments in each day, but rather average across all (fixed) available moments.

On the lowest temporal level *within-person change reliability from moment to moment*, R_{WPM}, is computed as (cf. Cranford et al., 2006, Eq. 5):

$$R_{WPM} = \frac{\sigma_{PDM}^2}{\sigma_{PDM}^2 + [\sigma_e^2/j]} \quad (5)$$

Note that all reliability formulas are identical for measurement designs without a dyadic structure on the highest level. In this case, the variance decomposition in Eq. (1) simply omits all terms including the factor C.

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, 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).

Scale intercorrelation at three levels of aggregation. 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 (Adolf & Fried, 2019; Fisher, Medaglia, & Jeronimus, 2018; Kievit, Frankenhuis, Waldorp, & Borsboom, 2013; Medaglia, Jeronimus, & Fisher, 2019). Consequently, scale intercorrelations can differ depending on the level of analysis. Just as reliability has to be considered on each level of

Table 2
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
surveys	7545	-	7515	7508	7515	7544
items	RS-1, RS-2	-	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
surveys	60917	60917	47878	47871	47878	47913
items	RS-1, RS-3	RS-1, RS-2, RS-3	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

Note. j, k, and l are the numbers of scheduled items, days, and moments. survey is the number of actually answered surveys. Numbers slightly differ within study when participants skipped a survey in between 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. For item wordings, see Tables 1, A1, and A2.

analysis, construct validity also has to be analyzed on each level (Shrout & Lane, 2012).

We computed correlation matrices on three levels. (a) Correlations on the *between-person level* were computed by averaging all item responses of a scale across all moments of a person. These person means of all scales were then correlated across the sample. (b) Correlations on the *within-person/between-days level* were computed by averaging all item responses of a scale across all moments of each day of a person. Then the between-day correlation matrix was computed for each person, and these matrices were averaged across persons by first Fisher's Z-transforming the correlations, computing the mean, and then back-transforming them into the correlation metric. (c) Correlations on the *within-person/between-moments level* were computed by averaging all item responses of a scale within each moment of a person. However, in order to remove potential confounding with between-day effects, we first centered the item responses within each day (Kreft & de Leeuw, 1998). Then the between-moments correlation matrix was computed for each person, and these matrices were averaged across persons. Note that we ignored the nested couple structure for computing these correlation matrices.

Results

Variance decomposition

Table 4 reports the absolute variance estimates and Table 5 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 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 stable variance), specific moments of persons (*PDM*; around 15%), couple (*C*; around 12%), 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 large share of couple-level variance is mostly present in relationship satisfaction and communal motivation. Furthermore, relationship satisfaction additionally has a unique large *couple x day* component (*CD*; around 11%), which indicates that some days are more satisfying for couples than other days.

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 16% of variance), and on specific days (*PDI*; around 9%).

Reliability estimation

Table 6 reports reliability estimates for both studies on all levels. Generally, the more measurements are aggregated, the higher is the reliability. On person level, reliabilities range from .95 to .98, on day level from .52 to .86, and on moment level from .28 to .70.[4]

Scale correlations on several levels

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 lower correlations are to be expected. Table 7 reports the uncorrected and disattenuated correlations on each level of aggregation.

In the following description we focus exclusively on the disattenuated correlations. 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 .16, with an average absolute difference of .05. 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 .40 on the person level, but close to zero on the moment level. Furthermore, the negative correlation between independence and communal motivation is stronger on the day and moment level (around $r = -.50$) compared to the person level (around $r = -.15$).

Discussion

We presented a model for estimating reliability of experience sampling measures which are assessed at multiple moments per day, across several days, and within in dyads. This design allows researchers to estimate a variance decomposition and reliability on three levels of aggregation, (a) between-persons, (b) within-person/between-days, and (c) 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

---

[4]If 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_{BP} (S1: +.002, S2: +.001), and slightly larger for R_{WPD} (S1: +.03, S2: +.02). In a previous publication based on S1 (Zygar et al., 2018a), a shorter two-item scale for communal motivation was employed, consisting of items C-1 and C-2 (see Table A1). This more homogenous scale demonstrated the following reliabilities: $R_{BP} = .96$, $R_{WPD} = .83$, and $R_{WPM} = .72$.

Table 3
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 accrossed 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)	couple:moment	Systematic effects of morning vs. evening for some couples
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.
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
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
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.
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.
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
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
couple:day:moment:item (CDMI)	Do couples have a stable, differential understanding of items at specific time points of specific days?	Different understanding of items after a conflict between the partners
Error (e)	Residual error variance	

has been assessed with two subscales, independence motivation and power motivation. Two intensive longitudinal studies provided data on more than 7,508 unique surveys in Sample 1 and more than 47,871 unique surveys in Sample 2.

Variance decomposition and reliability estimation

One research question for this study was 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, couples and persons are to some extent generally closer, more satisfied, or have more agentic motivation than other couples or persons. 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 stable interindividual differences in self-reported motivational states and relationship satisfaction, (b) 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 and studies. First, participants demonstrated systematically 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 > .95. Aggregated on day level (across four or five moments per day), reliability of the more homogeneous scales fell between .68 and .86. The two items for state relationship satisfaction in S1 were quite

Table 4

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)	0.10	0.04	0.10	0.03	0.11	0.21	0.08	0.06	0.06	0.15
person (P)	0.08	0.19	0.21	0.12	0.17	0.09	0.16	0.17	0.12	0.15
day (D)	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
moment (M)	0.00	0.01	0.00	0.00	0.01	0.01	0.00	0.00	0.07	0.03
couple:day (CD)	0.08	0.01	0.00	0.00	0.02	0.08	0.02	0.01	0.01	0.02
person:day (PD)	0.01	0.11	0.09	0.02	0.06	0.06	0.08	0.07	0.04	0.07
couple:day:moment (CDM)	0.11	0.03	0.01	0.00	0.04	0.09	0.02	0.00	0.00	0.03
person:day:moment (PDM)	0.08	0.17	0.19	0.04	0.16	0.09	0.13	0.13	0.05	0.12
Nuisance terms										
item (I)	0.00	0.00	0.11	0.00	0.00	0.01	0.01	0.11	0.00	0.06
couple:item (CI)	0.04	0.02	0.08	0.04	0.03	0.04	0.02	0.02	0.04	0.05
couple:moment (CM)	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
day:item (DI)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
day:moment (DM)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
moment:item (MI)	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00
person:item (PI)	0.12	0.04	0.35	0.16	0.08	0.08	0.08	0.12	0.16	0.12
person:moment (PM)	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00
couple:day:item (CDI)	0.02	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.01
couple:moment:item (CMI)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
day:moment:item (DMI)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
person:day:item (PDI)	0.05	0.03	0.14	0.11	0.05	0.04	0.05	0.06	0.10	0.06
person:moment:item (PMI)	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
couple:day:moment:item (CDMI)	0.02	0.00	0.01	0.02	0.01	0.01	0.00	0.00	0.01	0.01
Error (e)	0.28	0.34	0.56	0.43	0.27	0.19	0.33	0.32	0.40	0.24

Note. *day* runs from 1 to 14 in S1, and from 1 to 28 in S2. *moment* runs from 1 to 5 in S1 and from 1 to 4 in S2. *RS* = relationship satisfaction scale, *Ind* = independence motivation scale, *Pow* = power motivation scale, *A* = agentic motivation scale (pooled independence and power), *C* = communal motivation scale.

inhomogenous, resulting in a low reliability of .52. Furthermore, combining independence and power motivation into a higher-order agency scale decreased reliability to .63 in S1.

On the lowest level of aggregation, at each moment, this trend was even stronger. Homogeneous scales showed (relatively) better reliabilities ranging from .41 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-1 and RS-2). 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, depending on the research question.

Validity: Scale intercorrelations

The scale intercorrelations on the different temporal levels revealed some relevant insights into the underlying con-

structs. In the following, we base our interpretation on the disattenuated correlations. 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 .40 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 dif-

Table 5
Variance Decomposition of Item Responses: Relative stable variances.

	Sample 1, relative					Sample 2, relative				
Source	RS	Ind	Pow	A	C	RS	Ind	Pow	A	C
Theoretically relevant terms										
couple (C)	13%	6%	8%	5%	14%	25%	12%	8%	9%	17%
person (P)	11%	28%	16%	21%	23%	11%	23%	22%	18%	17%
day (D)	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%
moment (M)	0%	2%	0%	0%	1%	1%	0%	0%	10%	3%
couple:day (CD)	12%	1%	0%	0%	2%	10%	3%	2%	1%	3%
person:day (PD)	2%	16%	7%	4%	8%	7%	12%	9%	5%	8%
couple:day:moment (CDM)	15%	5%	1%	1%	5%	11%	3%	0%	1%	3%
person:day:moment (PDM)	11%	25%	14%	7%	21%	11%	19%	16%	7%	13%
Nuisance terms										
item (I)	0%	0%	8%	0%	0%	1%	1%	15%	0%	6%
couple:item (CI)	5%	3%	6%	8%	4%	5%	3%	3%	5%	5%
couple:moment (CM)	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%
day:item (DI)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
day:moment (DM)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
moment:item (MI)	0%	0%	0%	2%	0%	0%	0%	0%	0%	0%
person:item (PI)	16%	6%	26%	27%	10%	9%	12%	16%	24%	13%
person:moment (PM)	0%	1%	0%	0%	2%	0%	1%	0%	0%	0%
couple:day:item (CDI)	3%	0%	0%	1%	1%	1%	0%	0%	2%	1%
couple:moment:item (CMI)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
day:moment:item (DMI)	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
person:day:item (PDI)	7%	4%	11%	19%	7%	5%	8%	8%	14%	7%
person:moment:item (PMI)	0%	0%	0%	2%	1%	0%	0%	0%	1%	0%
couple:day:moment:item (CDMI)	3%	0%	1%	4%	1%	2%	1%	0%	1%	1%

Note. day runs from 1 to 14 in S1, and from 1 to 28 in S2. moment runs from 1 to 5 in S1 and from 1 to 4 in S2. *RS* = relationship satisfaction scale, *Ind* = independence motivation scale, *Pow* = power motivation scale, *A* = agentic motivation scale (pooled independence and power), *C* = communal motivation scale.

Table 6
Reliability Estimates.

	Sample 1			Sample 2		
Scale	R_{BP}	R_{WPD}	R_{WPM}	R_{BP}	R_{WPD}	R_{WPM}
RS2*	.96	.52	.36	.98	.86	.64
RS3				.97	.83	.58
Ind	.95	.74	.50	.97	.68	.44
Pow*	.96	.69	.41	.98	.73	.54
A*	.96	.63	.28	.98	.69	.38
C	.97	.83	.70	.98	.82	.67

Note. R_{BP} = between-person 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.

Table 7
Correlations on Three Levels of Aggregation

	Plain correlations					Disattenuated correlations				
	RS	Ind	Pow	A	C	RS	Ind	Pow	A	C
Between-person correlations										
RS		−.36	−.12	−.29	.38		−.38	−.12	−.30	.39
Ind	−.18		.38	.82	−.12	−.18		.40	.86	−.13
Pow	−.06	.41		.84	.52	−.06	.42		.87	.54
A	−.13	.76	.91		.25	−.13	.78	.92		.26
C	.53	−.16	.37	.19		.55	−.16	.38	.19	
Within-person, between-day correlations										
RS		−.18	−.01	−.14	.40		−.29	−.01	−.25	.61
Ind	−.20		.03	.82	−.38	−.27		.04	1.00	−.49
Pow	.01	.14		.70	.47	.01	.19		1.00	.62
A	−.12	.74	.84		.01	−.17	1.00	1.00		.01
C	.41	−.31	.38	.08		.50	−.42	.49	.11	
Within-person, between-moment correlations										
RS		−.13	.00	−.11	.27		−.30	−.01	−.34	.54
Ind	−.13		−.05	.78	−.34	−.25		−.12	1.00	−.58
Pow	−.02	.03		.63	.39	−.03	.06		1.00	.73
A	−.09	.69	.80		−.03	−.20	1.00	1.00		−.08
C	.30	−.27	.30	.05		.48	−.50	.50	.10	

Note. Upper triangle in each matrix shows S1, lower triangle shows S2. *RS* = relationship satisfaction scale, *Ind* = independence motivation scale, *Pow* = power motivation scale, *A* = agentic motivation scale (pooled independence and power), *C* = communal motivation scale. Disattenuation can result in correlations > 1, these were set to 1.

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

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 re-search 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, but not on the between-person level. On a behavioral level this makes immediate sense, as it is difficult to be 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 (uncorrected) negative correlations around -.40, both on an explicit level, assessed with self-report questionnaires

(Hagemeyer, Neyer, 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 around .22, and virtually zero correlations 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, Neyer, 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 might be due to two different factors. First, 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 add dominance as another facet of agency (Suessenbach, Loughnan, Schönbrodt, & Moore, 2018). Dominance motivated instrumental behavior also benefits from 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 must 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 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 in S1, Zygar et al. (2018a) 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 state relationship satisfaction, see Zygar & Schönbrodt, 2019). For planning a study, power analyses are needed to investigate the relative impact of these determinants on 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 presum-

ably have some autoregressive effect, which is ignored in the GT model. Finally, the model assumes equal item loadings. This can be tackled by standardizing the items before calculating the scale, which is however not always desirable in terms of interpretation. 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.

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.

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 no real room to do item selection. Scale development for ESM studies can benefit from a larger item pool in a pilot study that allows choosing more homogeneous items for scales.

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. The model can also easily applied to data set where persons are not nested in couples. Applying this model to four motivation scales and 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.

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 are published as a scientific use file (Zygar et al., 2018b), available at https://doi.org/10.5160/psychdata. zrce16dy99, which restricts access to scientific users. We currently are in the process to submit the data of Sample 2 to a repository. The reliability analyses presented here were not preregistered.

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. doi: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*. doi: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. doi:10.18637/jss.v067.i01

Berridge, K. C. (2004). Motivation concepts in behavioral neuroscience. *Physiology & Behavior*, *81*, 179–209. doi:10.1016/j.physbeh.2004.02.004

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. doi:10.1006/jrpe.1997.2206

Brose, A., Voelkle, M. C., Lövdén, M., Lindenberger, 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. doi: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. doi:10.1177/0146167206287721

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. doi: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. doi: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. doi: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. doi: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. doi:10.1037/a0024822. pmid:21787092

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. doi: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. doi:10.1037/pspi0000031

Heckhausen, J., & Heckhausen, H. (Eds.). (2018). *Motivation and Action*. 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. doi: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. doi:10.1177/0956797612437426

Karney, B., & Bradbury, T. (1995). The longitudinal course of marital quality and stability: A review of theory, method, and mesearch. *Psychological Bulletin*, *118*, 3–34. Retrieved from http://www.sciencedirect.com/science/article/B6WY5-46R0XFP-12/2/704bf89457e42e1d3f6cbee359f5feb2

Kievit, R., Frankenhuis, W. E., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00513

Kreft, I., & de Leeuw, J. (1998). *Introducing Multilevel Modeling*. doi:10.4135/9781849209366

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. doi:10.1080/00273171.2010.534380. pmid:26760733

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. doi: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. doi:10.1002/dev.20262

Nezlek, J. B. (2016). A practical guide to understanding reliability in studies of within-person variability. *Journal of Research in Personality*, 1–7. doi:10.1016/j.jrp.2016.06.020

Schoebi, D. (2008). The coregulation of daily affect in marital relationships. *Journal of Family Psychology*, *22*, 595–604. doi: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. doi:10.1016/j.jrp.2012.08.010

Schultheiss, O. C., & Brunstein, J. (2010). *Implicit Motives*. Oxford University Press.

Schultheiss, O. C., & Köllner, M. G. (in press). Implicit motives. In O. P. John & R. W. Robins (Eds.), *Handbook of personality psychology: Theory and research (4th ed.)*. (pp. 603–633). New York, NY US: Guilford Press.

Scott, S. B., Sliwinski, M. J., Zawadzki, M., Stawski, R. S., Kim, J., Marcusson-Clavertz, D., ... Smyth, J. M. (2018). A Coordinated Analysis of Variance in Affect in Daily Life. *Assessment*, 107319111879946. doi:10.1177/1073191118799460

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of Research Methods for Studying Daily Life* (pp. 302–320). New York, NY, US: Guilford Press. Retrieved November 23, 2018, from https://www.guilford.com/books/Handbook-of-Research-Methods-for-Studying-Daily-Life/Mehl-Conner/9781462513055/contents

Suessenbach, F., Loughnan, S., Schönbrodt, F. D., & Moore, A. B. (2018). The Dominance, Prestige, and Leadership Account of Social Power Motives: Dominance, prestige, and leadership motives. *European Journal of Personality*. doi:10.1002/per.2184

Winter, D. G. (1994). *Manual for scoring motive imagery in running text (4th edition)*. 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. *Eu-*

ropean Journal of Personality, *32*, 306–324. doi: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*. Trier: Psychologisches Datenarchiv Psych-Data des Leibniz-Zentrums für Psychologische Information und Dokumentation ZPID. Retrieved June 13, 2018, from https : / / doi . org / 10 . 5160 / psychdata . zrce16dy99

Zygar, C., & Schönbrodt, F. D. (2019). *Recalling experiences: Looking at momentary, retrospective and global assessments of relationship satisfaction*. Retrieved May 13, 2019, from https : / / psyarxiv . com / mvynu/

Appendix
Item wordings

Table A1

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:
Communion	C-2	**Emotionale Zuneigung von Ihrem Partner zu erhalten?** *(z.B. eine liebevolle Geste oder liebevolle Worte)*		S1: 4 = *ja, sehr stark*, 3 = *ja, stark*, 2 = *ja, mittelmäßig*,
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.)*		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 auszuüben; an einem eigenen Projekt weiterzuarbeiten.)*		S2: 4 = *ja, sehr stark*, 3 = *ja, stark*, 2 = *ja, mittelmäßig*,
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.)*		1 = *ja, aber nur schwach*, 0 = *nein, danach habe ich gerade kein Bedürfnis*, -1 = *nein, das würde mich gerade stören*,
Power	P-2	**Dass es einen Austausch mit Ihrem Partner gibt, in dem es um Sie geht, Sie im Mittelpunkt stehen?** *(z.B. dass Ihr Partner Ihre Bedürfnisse über die Eigenen stellt; Ihr Partner Ihnen volle Aufmerksamkeit schenkt; Sie Ihren Partner beeindrucken.)*		
Power	P-3	**Dass Ihr Partner sich nach Ihnen richtet?** *(z.B. dass Ihr Partner Ihre Bedürfnisse über die Eigenen stellt; Ihre Wünsche erfüllt.)*		
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) = *Komplett ohne Ihren Partner (als Zeit für mich)* to 7 (S1) or 10 (S2) = *Komplett mit Ihrem Partner (als gemeinsame Zeit)*

Note. Dashed underlined black text was not part of the items in S1, dashed underlined gray text was only part of the item in S1. [a] Aron, 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.

Table A2

English Translation of Experience Sampling Items for the Assessment of Motivation.

Domain & Label		Instruction		Scale
		Partner present	Partner not present	
			Imagine you would now ~~get to spend~~ have **approx. 30 minutes of time** (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. your partner would also have those two hours, and could be by your side).	
		... 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,*
Independence	I-1	**To act and decide independent of your partner?** *(e.g. 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.)*		1 = *yes, but only weakly,* 0 = *no, I don't need this right now,* -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.)*		S2: 4 = *yes, very strongly,* 3 = *yes, strongly,*
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.)*		2 = *yes, moderately,* 1 = *yes, but only weakly,* 0 = *no, I don't need this 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 your partner puts your needs above his/her own; your partner gives you full attention; that you impress your partner.)*		-1 = *no, that would bother me right now*
Power	P-3	**That your partner fits in with your wishes?** *(e.g. that your partner puts your needs above his/her own; satisfies your wishes.)*		
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 ~~get to spend~~ have **two hours of time** (free time, in which you cannot take care of duties), **time which you could spend with your partner** (i.e. your partner would also have those two hours, and could be by your side). **How would you like to spend this time right now?**		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)*

Note. 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. [a] Aron, 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.