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

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, n.d.).

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, variability in motivation could also be situated at the between person level. In this case, motivational states might not only vary within persons, but also between persons (Fleeson, 2001), which is a core assumption underlying research on motive dispositions (Birk Hagemeyer et al., 2013; Schönbrodt & Gerstenberg, 2012; Schultheiss & Köllner, n.d.). 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; ). 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 the data structure where ESM surveys covers multiple moments per day, and persons are nested in couples. Knowledge about the time scale and variability of motivational processes carries important information for the design of future studies, pertaining to the frequency and time points of momentary assessment, and trade-off analyses whether limited resources should be rather spent on short and intensive (within day) measurements or rather longer (but less intensive) daily diaries.

In selecting motivational variables relevant for couple relationship, we relied on the conceptualization of partner-related agentic and communal motives, as proposed by Hagemeyer & 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 & 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 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 (B. Hagemeyer et al., 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 association with relationship quality. To date, only one previous study addressed motivational processes within partners of a couple, using self-reports of the partner-related communion motive in a two-week experience-sampling design with five assessments per day (Zygar et al., 2018a). 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 agency and communion motives 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 and levels of its variability 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, & 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 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 (Study 1: <https://doi.org/10.5160/psychdata.zrce16dy99>, Study 2: We currently prepare the submission of a scientific use file to a repository. Data available for review upon request).

## Samples

Two samples from intensive longitudinal assessment are used. Study 1 (henceforward, S1) uses a data set from Zygar et al. (2018a) which is available as a scientific use file (Zygar et al., 2018b). This data set includes ESM data from 130 persons nested in 68 heterosexual couples. Fifty-two percept were women, mostly students (78%), and mean age was 22.39 years. 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).

Study 2 (S2) includes ESM data from 510 persons nested in 259 heterosexual couples. Fifty percent were women, mostly not students (71%), but with a German Abitur (high school degree) or a higher educational degree (65%), and mean age was 31.40 years. 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 a preliminary questionnaire (programmed with *formr*; ) on various measures. Subsequentely, 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 coudl 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 took a median of 3.28 minutes (S1) and 2.70 minutes (S2) to answer. When notified, individuals had 45 minutes to complete the survey, which included the same questions at each assessment, except for the last survey in the evening in S2. This survey had a different set of items, for example did it 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 compliance before data exclusions was 84% (S1) and 88% (S2), incentivized by personalized feedback, course credit or money. For a more detailed procedure including exclusions, see Zygar et al. (2018a) and <https://osf.io/v2uxs/>.

## Experience sampling items

### State motivation

At each measurement occasion, three motives were assessed (see Tables [[tab:motitems]](#tab:motitems) and [[tab:motitems2]](#tab:motitems2) in the Appendix for all items, instructions and response scales). Closeness 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, state agency motivation, is a superordinate scale combining independence and power motivation.

### State relationship satisfaction

State relationship satisfaction was assessed with two (S1) or three items (S2) at each moment (see Table [[tab:rsitems]](#tab:rsitems)). Exploratorily, we also constructed a homogenous two-item scale in S2 by excluding the “annoyance” item. 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.

|  |  |  |
| --- | --- | --- |
| Label | Instruction | Scale |
| Relationship mood | Original: **Wie geht es Ihnen *jetzt gerade* mit Ihrer Beziehung?** | 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 | Original: **Wie genervt sind Sie *jetzt gerade* von Ihrem Partner?** | Continuous slider from 1 (S1) or 0 (S2) = *überhaupt nicht [not at all]*, to 7 (S1) or 10 (S2) = *stark [strongly]* |
| Need satisfaction | Original: **Wie fühlen Sie sich *jetzt gerade* in Ihrer Partnerschaft?** | Continuous slider from 0 = *total frustriert [totally frustrated]*, over 5 = *neutral [neutral]*, to 10 = *total erfüllt [totally satisfied]* |

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

## 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]](#footnote-30)

### Variance decomposition

Following generalizability theory, the full variance decomposition model is formalized as a four-way analysis of variance. For a person in couple , responding to motivation item in moment on day , the model for closeness motivation is

Uppercase variables denote the factors *couple (C)*, *person (P)*, *day (D)*, *moment (M)*, and *item (I)*. In our design, the four-way interaction 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. [[eq:GT]](#eq:GT).

The indicator variable for *moment*, , goes from 1 to 5 (S1) or 1 to 4 (S2), which means that , for example, denotes all morning surveys across all persons. Likewise, the indicator variable for *day*, , goes from 1 to 14 in S1, or 1 to 28 in S2. Hence, denotes the first study day of all persons. But note that most participants started the study on a completely different calendar day in S1, or on a different Monday in S2. The specific values for the number of items (), the number of moments nested within each day (), and the number of days () is given in Table [[tab:itemStats]](#tab:itemStats).

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
| (l2ptr2pt)2-5 (l2ptr2pt)6-9  Scale | items | days | moments | surveys | items | days | moments | surveys |
| Relationship satisfaction (RS) | 2 | 14 | 5 | 7545 | 3 | 28 | 5 | 60917 |
| Independence motivation (Ind) | 2 | 14 | 5 | 7515 | 2 | 28 | 4 | 47878 |
| Power motivation (Pow) | 2 | 14 | 5 | 7508 | 3 | 28 | 4 | 47871 |
| Agency motivation (A) | 4 | 14 | 5 | 7515 | 5 | 28 | 4 | 47878 |
| Closeness motivation (C) | 4 | 14 | 5 | 7544 | 4 | 28 | 4 | 47913 |

*Note.* , , and 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.

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 *day x moment* term, indicates a specific survey and whether this survey has received a different rating on a specific day (e.g., *5:2* is the second survey on day 5 of all persons). The three-way interaction *person x day x moment*, , 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 [[tab:varDecompExplanation]](#tab:varDecompExplanation).

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

Items of our scales were assessed on different response scales. The GT model covers differing mean levels of items with the item factor . 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]](#footnote-33) As we wanted to match our reliability analysis to the actually computed scale scores, we -standardized all items across all measurement points of both genders. (The reliability estimates from unstandardized variables were virtually identical, with a maximum difference of ).

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. ([[eq:GT]](#eq:GT)) using the *lmer* package (Bates, Mächler, Bolker, & Walker, 2015). The specific function call is in the reproducible scripts on OSF.

### Reliability estimation

Reliability estimation in the GT framework generally uses the formula

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

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 (i.e., person x item), , 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, , 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), , 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 or moments each day (from morning to evening) were assumed to be comparable for each person.[[3]](#footnote-35) We use Equation (8) from Shrout & Lane (2012), including the *person x day x item* interaction, , as an additional source of error, but not moment-to-moment variance, as all participants are assumed to have the same effects of moments:

$$\label{eq:R\_BP}
\begin{split}
R\_{BP} = \frac{
\sigma^2\_{P} + [\sigma^2\_{PI} / j]
}{
\splitdfrac{
\sigma^2\_{P} + [\sigma^2\_{PI} / j] + [\sigma^2\_{D} / k] + [\sigma^2\_{PD} / k] }
{ + [\sigma^2\_{PDI} / (k\*j)] + [\sigma^2\_{e} / (k\*l\*j)]}
}
\end{split}$$

The constant is the number of items, is the number of days, is the number of moments within each day (see Table [[tab:itemStats]](#tab:itemStats)). 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 , the number of items; the residual error term in is divided by to take into account the increase in precision that results from averaging items, assessed at moments at each of the days of a person.

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

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*, , is computed as (cf. , Eq. 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. ([[eq:GT]](#eq:GT)) simply omits all terms including the factor .

The number of days within person, , and the number of moments within day, , is not constant if participants do not answer every single ESM survey. Therefore, we inserted the average number of answered moments (i.e., compliance x maximum possible observations) and the average number of days into the formulas (see also , footnote 5, and ).

### 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 be 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).

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 -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 [[tab:varDecompAbs]](#tab:varDecompAbs) reports the absolute variance estimates, and Table [[tab:varDecompRel]](#tab:varDecompRel) 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”. (Note that terms we consider as nuisance terms here might be centrally relevant for other research questions, e.g., for methodological and psychometric questions).

|  |  |  |
| --- | --- | --- |
| Source | Explanation | Example/Comment |
|  |  |  |
| couple (C) | Variance between couples |  |
| person (P) | Variance between persons |  |
| day (D) | Variance between days 1 to 14/28 (pooled acrossed 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) |  |
|  |  |  |
| 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 |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Variance component | RS | Ind | Pow | A | C | RS | Ind | Pow | A | C |
|  |  |  |  |  |  |  |  |  |  |  |
| 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 |
|  |  |  |  |  |  |  |  |  |  |  |
| 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, *Ind* = independence motivation, *Pow* = power motivation, *A* = agentic motivation (pooled independence and power), *C* = closeness motivation.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Source | RS | Ind | Pow | A | C | RS | Ind | Pow | A | C |
|  |  |  |  |  |  |  |  |  |  |  |
| 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% |
|  |  |  |  |  |  |  |  |  |  |  |
| 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.

As a general pattern, four focal sources of variances had the largest share across scales and studies: persons (; around 19% of stable variance), specific moments of persons (; around 15%), couple (; around 12%), and specific days of persons (; 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 closeness motivation. Furthermore, relationship satisfaction additionally has a unique large *couple x day* component (; 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: Participants had systematically different mean levels between item responses in general (; around 16% of variance), and on specific days (; around 9%).

## Reliability estimation

Table [[tab:reliability]](#tab:reliability) 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]](#footnote-40)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| (l2ptr2pt)2-4 (l2ptr2pt)5-7  Scale |  |  |  |  |  |  |
| 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.* = between-person reliability, = within-person/between-days reliability, = within-person/between-moments reliability. *RS2, RS3* = relationship satisfaction, measured with 2, resp. 3, items, *Ind* = independence motivation, *Pow* = power motivation, *A* = agentic motivation (pooled independence and power), *C* = closeness motivation.

## 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 [[tab:cor.all]](#tab:cor.all) reports the uncorrected and disattenuated correlations on each level of aggregation.

ldddddddddd & &  
(l2ptr2pt)2-6 (l2ptr2pt)7-11

& RS & Ind & Pow & A & C & RS & Ind & Pow & A & C  
  
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 &  
  
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 &  
  
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, *Ind* = independence motivation, *Pow* = power motivation, *A* = agentic motivation (pooled independence and power), *C* = closeness motivation. Disattenuation can result in correlations > 1, these were set to 1.

Disattenuated correlation matrices show largely similar patterns across aggregation levels. In particular, all differences between day level and moment level correlations are less than , with an average absolute difference of .05. The correlations on person level show some stronger differences to the day and moment level correlations. In particular, the correlation between *Power* and *Independence* motivation is around .40 on person level, but close to zero on moment level. Furthermore, the negative correlation between *Independence* and *Closeness* motivation is stronger on a day and moment level (around ) compared to the person level (around )

# Discussion

We presented a model for estimating reliability when in an experience sampling design multiple items are assessed at multiple moments per day, across several days, in dyadic data. This design allows 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 focus on motivational dyanamics on relationship satisfaction in couples: state relationship satisfaction, closeness motivation, and agency motivation, which has been assessed with two subscales, independence motivation and power motivation. Two intensive longitudinal studies provided data

## Variance decomposition and reliability estimation

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 investigates scales do vary 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. This shows that the employed items are able to capture state variance.

Concerning nuisance terms, two sources of variances had substantial contributions across scales and studies. First, participants had 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 what 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 are 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 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 two heterogeneous relationship satisfaction items in S1 (.36) and the combined agency scale (.28 in S1, .38 in S2) were unsatisfactory.

## Validity: Scale intercorrelations

The scale intercorrelations on the different temporal levels reveal some relevant insights about the underlying constructs. In the following, we base our interpretation on the disattenuated correlations. Generally, the correlation matrices were rather similar on all levels and do 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 are two notable exceptions where the person level correlations differ from the day and moment level.

First, the *Independence* and *Power* motivation scales show a positive correlation around .40 on the between-person level. Persons who generally have more independence motivation also generally have more power motivation, which can be interpreted that these scales are two facets of the overarching agency motive factor, as “a superordinate need to feel as a capable, self-reliant individual” (Hagemeyer & Neyer, 2012, p. 3). Within person, however, they are independent with correlations close to zero: On moments or days where persons experience a strong motivation for independence, they do 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, we recommend not to use that scale on the day or moment level. On the between-person level, in contrast, the subscales show a substantial positive correlation, which is also reflected in much better reliabilities. Hence, this conceptualization of agency as a multifaceted construct only makes sense on a between-person level; for analyses and theoretical models on a momentary level, *Independence* and *Power* should be treated as separate motivational subscales. This finding 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 *Closeness* motivation are, to some extent, mutually exclusive on the daily and momentary 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 are rather rare. On the person level, in contrast, the correlation is only slightly negative, indicating that a person’s general level of closeness motivation is largely independent of the general level of independence motivation.

When the agency and the communion motive are assessed as stable dispositions, they typically show (uncorrected) 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 closeness 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 shows the expected negative correlation to closeness 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 do show the expected negative correlation (albeit, with a smaller effect size).

The positive correlation between power motivation and closeness motivation could be, at least partly, explained by the fact that exerting influence on the partner typically requires physical closeness. Furthermore, our ESM power items were inspired by prosocial aspects of the power motive as described in Winter (1994) and Hagemeyer & Neyer (2012), where power motivation also includes dominant or supportive behaviors within the relationship, and also includes a positive influence on the partner. To summarize, power motivation, as one aspect of a broader agency factor, serves to satisfy agentic needs within the relationship – which typically requires contact and physical closeness. Independence, in contrast, refers to agency needs that are satisfied outside of the relationship, which typically requires distance from the partner.

## 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-couples 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. 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 concerning statistical power are counterbalanced. Reliability is lowest on the most fine-grained moment level, which lowers statistical power. 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).

The design of an ESM study, specifically the frequency, timing, and length of measurements, must consider several factors. The expected speed of change of a construct determines the frequency of sampling, and reliability and burden of participants must be balanced. 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. ([[eq:GT]](#eq:GT)) 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. 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 & Shrout (2010) showed that the GT method underestimates the reliability to the extent that this assumption 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 phenomenen 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 to choose 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 decompostion 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 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 Study 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 Study 2 to a repository. The reliability analyses presented here were not preregistered.

# Item wordings

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| (r)1-1 (r)2-3 (r)4-4 |  |  |  |
|  |  | Stellen Sie sich nun vor, Sie bekämen jetzt **ca. 30 Minuten Zeit aufgrund von Leerlauf** (freie Zeit, in der Sie keine Verpflichtungen erledigen können), **die Sie mit Ihrem Partner verbringen könnten aber nicht müssten** . |  |
|  | ... Wünschen Sie sich jetzt gerade: | ... Wünschen Sie sich in dieser Zeit: |  |
| (r)1-3  Closeness, Communion |  |  |  |
| (r)1-3  Closeness, Communion |  |  |  |
| (r)1-3  Independence |  |  |  |
| (r)1-3  Independence |  |  |  |
| (r)1-3  Power |  |  |  |
| (r)1-3  Power |  |  |  |
| (r)1-3 |  |  |  |
| Closeness |  | 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 IOS picture |  |
| Closeness, (Independence) |  | 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.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| (r)1-1 (r)2-3 (r)4-4 |  |  |  |
|  |  | Imagine you would now **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** . |  |
|  | ... Right now, do you wish: | ... In this given time, do you wish: |  |
| (r)1-3  Closeness, Communion |  |  |  |
| (r)1-3  Closeness, Communion |  |  |  |
| (r)1-3  Independence |  |  |  |
| (r)1-3  Independence |  |  |  |
| (r)1-3  Power |  |  |  |
| (r)1-3  Power |  |  |  |
| (r)1-3 |  |  |  |
| Closeness |  | 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 IOS picture |  |
| Closeness, (Independence) |  | 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 = Study 1, S2 = Study 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.

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](https://doi.org/10.1073/pnas.1818675116)

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](https://doi.org/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](https://doi.org/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](https://doi.org/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](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. doi:[10.1177/0146167206287721](https://doi.org/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](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. doi:[10.1037/0022-3514.80.6.1011](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. doi:[10.1111/j.1467-6494.2006.00407.x](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. doi:[10.1111/jopy.12016](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. doi:[10.1037/a0024822](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. doi:[10.1002/per.1857](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. doi:[10.1037/pspi0000031](https://doi.org/10.1037/pspi0000031)

Heckhausen, J., & Heckhausen, H. (Eds.). (2018). *Motivation and Action* (3rd ed.). Springer International Publishing. Retrieved 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](https://doi.org/10.1037/pspi0000020)

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](https://doi.org/10.3389/fpsyg.2013.00513)

Kreft, I., & de Leeuw, J. (1998). *Introducing Multilevel Modeling*. London, United Kingdom: SAGE Publications, Ltd. doi:[10.4135/9781849209366](https://doi.org/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](https://doi.org/10.1080/00273171.2010.534380)

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](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. doi:[10.1002/dev.20262](https://doi.org/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](https://doi.org/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](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. doi:[10.1016/j.jrp.2012.08.010](https://doi.org/10.1016/j.jrp.2012.08.010)

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

Schultheiss, O. C., & Köllner, M. G. (n.d.). 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.

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 from <https://www.guilford.com/books/Handbook-of-Research-Methods-for-Studying-Daily-Life/Mehl-Conner/9781462513055/contents>

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. *European Journal of Personality*, *32*, 306–324. doi:[10.1002/per.2145](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*. Trier: Psychologisches Datenarchiv PsychData des Leibniz-Zentrums für Psychologische Information und Dokumentation ZPID. Retrieved from <https://doi.org/10.5160/psychdata.zrce16dy99>

1. Exploratively, we also ran a model with distinguishable dyads, which yields comparable results. [↑](#footnote-ref-30)
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 same response scale. [↑](#footnote-ref-33)
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. [↑](#footnote-ref-35)
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 (S1: +.002, S2: +.001), and slightly larger for (S1: +.03, S2: +.02). [↑](#footnote-ref-40)