

Mix It to Fix It: Emotion Regulation Variability in Daily Life

Elisabeth S. Blanke

Humboldt-Universität zu Berlin, Germany

Annette Brose

Humboldt-Universität zu Berlin, Germany; KU Leuven, Belgium

Elise K. Kalokerinos

University of Newcastle, Australia

Yasemin Erbas

KU Leuven, Belgium

Michaela Riediger

Heisenberg Research Group Socio-emotional Development and Health Across the Lifespan,
Freie Universität Berlin and Max Planck Institute for Human Development, Berlin, Germany

Peter Kuppens

KU Leuven, Belgium

©American Psychological Association, 2020. This paper is not the copy of record and may not exactly replicate the authoritative document published in the APA journal. Please do not copy or cite without author's permission. The final article is available, upon publication, at: Blanke, E.S., Brose, A., Kalokerinos, E.K., Erbas, Y., Riediger, M., & Kuppens, P. (2020). Mix it to fix it: Emotion regulation variability in daily life. *Emotion*, 20(3), 473–485. <https://psycnet.apa.org/doi/10.1037/emo0000566>

Author Note

The research leading to the results reported in this paper was supported in part by a grant awarded to Annette Brose by the German Research Foundation [Deutsche Forschungsgemeinschaft, DFG], BR 3782/3-1, as well as by the Research Fund of KU Leuven (GOA/15/003). Elise Kalokerinos was supported by a Marie Skłodowska-Curie

individual fellowship (No. 704298) under the European Union's Horizon 2020 research and innovation programme. Elise Kalokerinos is now supported by an Australian Research Council Discovery Early Career Researcher Award (DE180100352). Yasemin Erbas was supported by a postdoctoral fellowship from the Flemish Fund for Scientific Research (FWO). Michaela Riediger was supported by a Heisenberg stipend of the German Research Foundation [Deutsche Forschungsgemeinschaft, DFG], RI 1797/3-1. Michaela Riediger is now at the Institute for Psychology of the University of Jena.

We thank our student research assistants and interns for their help with the data collection. We would also like to express our gratitude to our participants.

This is a study containing multiple data sets from different laboratories. Results from these datasets were previously published to test different research questions. A list of the publications pertaining to each data set is provided in the *Method* section of the manuscript. We report how we determined our sample sizes, all exclusions of participants, and all measures as relevant for the research questions. There were no manipulations. Results reported in this manuscript were previously presented as a flash talk at the Annual Conference of the Society for Affective Science 2017 in Boston, MA, USA.

Correspondence concerning this article should be addressed to Elisabeth S. Blanke, Humboldt-Universität zu Berlin, Institut für Psychologie, Unter den Linden 6, 10099 Berlin, Germany. Email: elisabeth.blanke@hu-berlin.de

Abstract

Emotion regulation (ER) strategies are often categorized as universally adaptive or maladaptive. However, it has recently been proposed that this view is overly simplistic: instead, adaptive ER involves applying strategies variably to meet contextual demands. Using data from four experience-sampling studies ($N_s=70, 95, 200, \text{ and } 179$), we tested the relationship between ER variability and negative affect (NA) in everyday life. The constantly changing demands of daily life provide a more ecologically valid context in which to test the role of variability. We calculated two global indicators of variability: within-strategy variability (of particular strategies across time) and between-strategy variability (across strategies at one time-point). Associations between within-strategy variability and NA were inconsistent. In contrast, when controlling for mean strategy endorsement, between-strategy variability was associated with reduced NA across both individuals and measurement occasions. This is the first evidence that variably choosing between different strategies within a situation may be adaptive in daily life.

152 words

Keywords: *emotion regulation, variability, flexibility, experience sampling (4/5)*

Word count: 7,172

Mix It to Fix It: Emotion Regulation Variability in Daily Life

In daily life, situations and emotions change dynamically, and the ability to respond flexibly to these changes has been proposed as an essential building block for psychological health (Hollenstein, Lichtwarck-Aschoff, & Potworowski, 2013; Kashdan & Rottenberg, 2010). In response to these dynamic emotions and situations, people use strategies to influence their emotions, a process which is called emotion regulation (ER; Gross, 1998). Ineffective ER is a risk factor for both psychological (e.g., Aldao & Nolen-Hoeksema, 2010) and physical problems (e.g., cardiovascular diseases; Appleton & Kubzansky, 2014). Much of the past research investigated how effective different ER strategies are in altering feelings, outward expressions, and physiological processes, or their cognitive or interpersonal costs (e.g., Gross, 2002). Based on the results of this work, strategies have been characterized as either adaptive (e.g., the reinterpretation of emotional stimuli referred to as reappraisal) or maladaptive (e.g., expressive suppression).

More recently, in line with work emphasizing the benefits of psychological flexibility, it has been proposed that this characterization of strategies is a fallacy (Bonanno & Burton, 2013). Contemporary ER theory suggests that effective ER does not involve inflexibly using the same “adaptive” strategy. Instead, it is not only important *how* people regulate their emotions, but also *how variably or flexibly* they choose ER strategies in response to situational demands. Yet, to date available empirical evidence for this proposition is limited for two primary reasons: first, different studies have used diverging operationalizations of ER variability and flexibility. And second, most of the existing research has been conducted in the laboratory. Following recommendations by Aldao, Sheppes, and Gross (2015), we addressed these two issues in the present research: We utilized experience-sampling data from four studies obtained in daily life (in Belgium and Germany), and investigated the adaptiveness of two global indicators of ER variability.

Aldao et al. (2015) suggested that ER variability is a superordinate construct that encompasses flexibility: flexibility occurs when variability is synchronized with situational changes in a way that is congruent with an individual's goals. They proposed that variability can be divided into *within-strategy variability* and *between-strategy variability*. Within-strategy variability is the variation in the intensity of usage of single strategies over different contexts and time. More precisely, within-strategy variability occurs when a person uses strategies in some occasions, but not in others. Between-strategy variability refers to the selection of particular strategies from a pool of strategies at one moment in time, possibly reflecting a search for the best strategy, or a prioritization of certain strategies in accordance with contextual demands. High between-strategy variability indicates that an individual neither tries to use all strategies simultaneously to a similar extent nor strongly prioritizes only one strategy, but chooses few strategies and uses these.

Research examining within-strategy variability, operationalized as the self-reported ability to flexibly use different ER (or coping)¹ strategies across situations, has demonstrated that it is associated with positive adjustment (Bonanno, Pat-Horenczyk, & Noll, 2011; Cheng, 2001; Kato, 2012). Relatedly, laboratory work has also demonstrated that the ability to variably modulate emotional expressiveness is associated with positive long-term outcomes (Bonanno, Papa, Lalande, Westphal, & Coifman, 2004). Aldao and Nolen-Hoeksema (2012) showed that, for two types of strategies (acceptance and problem-solving, but not the other strategies studied), within-strategy variability was associated with lower levels of psychopathology. Furthermore, Troy, Shallcross, and Mauss (2013) demonstrated that reappraisal, a putatively adaptive strategy, was not adaptive in controllable situations. Research on between-strategy variability is scarce, but in a study by Birk and Bonanno (2016), the ability to flexibly switch from a suboptimal to an optimal strategy was associated with higher satisfaction with life.

Thus, an emerging body of lab work suggests that different forms of ER variability may be adaptive. However, lab research suffers from some important limitations, particularly when applied to the study of variability. First, studies of self-reported flexibility may not be reflective of actual everyday behavior due to memory biases (e.g., Bolger, Davis, & Rafaeli, 2003). Second, the ability to display flexible ER strategy use in the laboratory may not translate outside the lab, where demands and situations are constantly changing on different dimensions (such as controllability, sociality, or importance). Relatedly, there are non-emotional (neutral) situations in daily life which may not warrant any regulation at all. In daily life, individuals thus need to identify situations in which regulation is warranted, and then choose an appropriate strategy, whereas in laboratory experiments, they usually only need to choose a strategy. To get a clear picture of the functionality of variability, it is therefore necessary to measure people at many different time-points in changing environments. To address these issues, Aldao et al. (2015) suggested that the experience-sampling method (ESM) could provide an ideal lens through which to study ER variability.

Preliminary evidence from ESM and diary studies suggests that within-strategy variability is indeed adaptive. In a diary study by Cheng (2001), flexible coping was determined using hierarchical cluster analysis. The flexible coping group consisted of participants who reported high variability in problem-focused and emotion-focused coping over time, as well as high variability in whether situations were perceived as controllable or uncontrollable. Compared to other clusters, this group had the lowest depression scores, and also scored favorably in other domains. A recent ESM study (Haines et al., 2016) replicated the finding from the laboratory study by Troy et al. (2013; see above) in daily life showing that the use of reappraisal is adaptive when a situation is perceived as uncontrollable, but not when it is perceived as controllable.

However, this small body of research is thus far inconclusive, as these studies focus only on the variability of one or two strategies (e.g., reappraisal), and often use retrospective

reports (diary studies). This single-strategy focus makes it impossible to test the role of between-strategy variability. Thus, to the best of our knowledge, no ESM study has been published to date that has systematically investigated the adaptiveness of within- and between-strategy variability in daily life. One recent study has looked at age differences in emotion regulation variability, using a daily diary approach, but did not focus on adaptiveness (Eldesouky & English, 2018).

To summarize, strong and consistent evidence supporting the presumably adaptive nature of ER variability and flexibility—one of the central propositions of modern ER theory—is thus far lacking. Previous research was primarily conducted in the lab, which provides only a small window to examine variability in limited context, without the many time-points necessary to study everyday variability. In addition, the few studies conducted outside the lab have focused on only one or two strategies, meaning that the role of between-strategy variability has not been tested.

To address these issues, we used pre-existing ESM data from four studies which assessed different emotion regulation strategies in daily life. Since these studies were not designed for the investigation of ER flexibility, we focused on ER variability in the current research. As proposed by Aldao et al. (2015), we used standard deviations (*SDs*) as indicators of variability. We hypothesized that, on average, greater ER variability would be adaptive, as ER variability is necessary (though not sufficient) for flexibility. We investigated whether ER variability relates to lower levels of negative affect (NA). We chose NA as our key dependent variable because people are usually motivated to experience low levels of NA (Riediger, Schmiedek, Wagner, & Lindenberger, 2009). We thus hypothesized that greater ER variability would be associated with lower NA, and interpreted associations between variability and reduced NA as adaptive regulation.

We analyzed between-strategy and within-strategy variability as person-level characteristics. More precisely, we had one value for between-strategy variability for each

occasion (i.e., state information), and this was averaged across occasions to obtain a person-level characteristic. For within-strategy variability, we observed state information on strategy use at each occasion. For each strategy, the distribution of these observations provided a basis to estimate the standard deviation. These within-strategy *SDs* for each strategy were then averaged across strategies to obtain a person-level indicator of participants' average within-strategy variability.

We were also interested in how time-varying aspects of between-strategy variability (i.e., state variance components) covaried with NA across time. For analyses, we used the time series of between-strategy variability values. For between-strategy variability, we thus examined our hypothesis at both the between- and within-person level (i.e., on average and at the level of within-person dynamics). In sum, we expected that individuals who show greater ER variability on average (between- and within-strategy variability) would report lower NA across the measurement period. Furthermore, we expected that occasions at which individuals prioritize some strategies (i.e., show more between-strategy variability) are occasions at which they experience lower levels of NA.

We expected ER variability only to be adaptive when we controlled for mean strategy endorsement (for a similar approach, see e.g., Koval, Pe, Meers, & Kuppens, 2013). High ER strategy endorsement has been related to unfavorable outcomes, including greater NA (e.g., Dixon-Gordon, Aldao, & De Los Reyes, 2015), possibly because of failed regulation efforts (Aldao & Nolen-Hoeksema, 2013). However, it is not possible to have high levels of variability at very low or very high levels of mean strategy endorsement. This means that variability (as assessed with the *SD*) can be confounded with mean strategy endorsement. We therefore separate the effect of mean ER endorsement from the effect of variability in our prediction and analyses.

Given that we used pre-existing data that were not initially designed to answer our research questions and hypotheses, we consider this work to be a first step in the investigation

of the adaptiveness of ER variability. Thus, despite our hypotheses, this work is somewhat exploratory in nature. The use of four studies allowed us to apply meta-analytic tools to determine whether and to what extent ER variability may be adaptive for reducing NA.

Method

Participants and Procedure

All datasets reported here were parts of larger studies. To answer our present research questions, we used meta-analytic techniques to get an overall estimate of the effect sizes. Sample sizes for each individual study were determined by each respective principal investigator before data collection on the basis of previous experiences with experience-sampling. There was no optional stopping in any of the studies. In Study 3, the sample size was determined to detect small to medium effects ($r = .30$, $\alpha = .05$).

The items used in the four studies and the data for Studies 1-3 are available in the supplementary material on the Open Science Framework. The release of data from Study 4 to the public is regulated by contract and will happen after data collection for this longitudinal study is finished. The supplementary material is available here:

https://osf.io/mxjfh/?view_only=5118406e8780402c8230a278eea0a502

Study 1. This convenience sample consisted of 70 students from various disciplines ($n = 35$ female) aged between 20 and 30 years ($M = 25.55$, $SD = 2.74$ years; see Blanke & Brose, 2017; Blanke, Riediger, & Brose, 2018). They were recruited via posters, online advertisement, and university mailing lists in the Berlin area, Germany. The participants took part in two laboratory sessions, with the ESM phase falling in between sessions. In the two sessions, they gave informed consent to participate, and filled out questionnaires including a German version of the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977; German version by Hautzinger & Bailer, 1993). They received smartphones (Huawei Ascend G330), which were programmed with an ESM technology that was developed and applied in previous studies (e.g., Rauers, Blanke, & Riediger, 2013; Riediger et al., 2009).

The ESM phase started the following day and lasted nine days, during which six ESM prompts (beeps) occurred semi-randomly each day in a fixed 12-hour time frame (selected by the participants). The students were given the opportunity to prolong the study by up to three days if they missed more than one assessment a day. They received a fixed reimbursement for the laboratory sessions and an additional reimbursement according to the number of ESM questionnaires they had completed with a bonus of 10 Euros for 45 or more completed beeps; however, it was communicated to the participants that we aimed for 54 answered beeps (9 days times 6 beeps). Participants answered 54.41 beeps on average ($SD = 3.25$; range: 48–65). Due to the extra days, participants were able to answer more than the target of 54 beeps. However, if 54 beeps or more are considered a response rate of 100%, the average response rate was 98.3% ($SD = 2.7%$, range 89–100%). For the present analyses, no participant was excluded from this sample. In total, participants received 65 Euros on average. The ethics committee of the Humboldt-Universität zu Berlin approved of the study.

Study 2. The final sample consisted of 95 undergraduate students ($n = 59$ female) aged between 18 and 24 years ($M = 19.06$, $SD = 1.28$ years). They were recruited from a pool of 439 undergraduates at the University of Leuven, Belgium, who completed a Dutch translation of the CES-D, and who were selected to maximize variation in depression scores (for a more detailed description, see study 2 in Brans, Koval, Verduyn, Lim, & Kuppens, 2013; for other publications with this data, see Erbas, Ceulemans, Koval, & Kuppens, 2015; Koval, Ogrinz, Kuppens, Van den Bergh, Tuerlinckx, & Sütterlin, 2013; Koval, Pe, et al., 2013; Pe, Koval, & Kuppens, 2013). As participants in psychological studies often report relatively low levels of depressive symptoms, participants were selected to also represent higher depression levels. The participants took part in an introductory session in the laboratory, in which they gave informed consent to participate, filled out questionnaires, and received palmtops (Tungsten E2 PalmOne, Mankato, MN), which were programmed with the Experience-Sampling Program (Barrett & Barrett, 2010). The ESM phase started the

following day and lasted seven days, during which 10 beeps occurred semi-randomly each day in a 12-hour time frame. Participants answered 91.5 % of the beeps ($SD = 6.2\%$, range: 67–100% of all beeps). From the initial sample ($N = 100$), one participant withdrew from the study, and four participants were excluded from data analysis after data collection because of equipment malfunction ($n = 3$), and poor compliance ($n = 1$; $> 40\%$ missing data; see Brans et al., 2013). The students were reimbursed with 70 Euros for the entire study. The ethics committee of the University of Leuven approved of the study.

Study 3. The final sample consisted of 200 first-year students ($n = 110$ female) aged between 17 and 24 years ($M = 18.32$, $SD = 0.96$ years). The majority of the participants were recruited from a pool of 686 undergraduates at the University of Leuven, Belgium, who completed a Dutch translation of the CES-D. Like in Study 2, individuals were selected to maximize variation in depression scores (for a more detailed description, see Koval et al., 2015; this sample was Wave 1 of a longitudinal study; for other publications with this data, see Bastian, Koval, Erbas, Houben, Pe, & Kuppens, 2015; Brose, Wichers, & Kuppens, 2017; Erbas, et al., in press; Dejonckheere et al., 2018; Pe, Brose, Gotlib, & Kuppens, 2016; Pe, Koval, Houben, Erbas, Champagne, & Kuppens, 2015).

The participants took part in an introductory session in the laboratory, in which they gave informed consent to participate, filled out questionnaires, and received smartphones (Motorola Defy Plus), which were programmed with custom-built software. The ESM phase started the following day and lasted seven days, during which 10 beeps occurred semi-randomly each day in a 12-hour time frame (10 a.m. to 10 p.m.). Participants answered 87.27% of the beeps on average ($SD = 9.05\%$, range: 55–100% of all beeps). The target sample size was 200; two participants were oversampled, but later excluded from the initial sample ($N = 202$) because they answered less than 50% of the beeps (see Koval et al., 2015). The students were reimbursed with 60 Euros for their participation in this wave of the study. The ethics committee of the University of Leuven approved of the study.

Study 4. The sample consisted of 179 adults ($n = 94$ female) aged between 38 and 61 years ($M = 50.93$, $SD = 5.76$ years). Participants came from the innovation sample of the German Socio-Economic Panel (SOEP-IS), a longitudinal survey in which participants are visited yearly in their private households in Germany (Richter & Schupp, 2015). Participants from the SOEP-IS between 38 and 61 years of age were contacted and invited to participate in our psychological study if they had participated in the panel for at least two waves of data collection, and if they participated in 2014. This sample was Wave 1 of a longitudinal study.

For the introductory session, participants were visited at their homes by interviewers from the Humboldt-Universität zu Berlin. They gave informed consent to participate, filled out questionnaires including a 10-item short version of the German CES-D (Irwin, Artin, & Oxman, 1999; German translation by Hautzinger & Bailer, 1993). Participants received smartphones (Huawei Ascend G330), which were programmed with the same program used in Study 1 with one difference: starting the day after the visit, the ESM phase included three assessment phases of four sampling days, which were followed by four pause days. If participants missed more than one assessment a day, they had the opportunity to prolong each assessment wave by up to two days (leaving only two pause days). The target sample size was $N = 180$. When data collection was finished, we realized that one participant did not meet the inclusion criteria. This participant was not considered to be part of the sample and was thus excluded. Participants were reimbursed with 20 Euros for the session and 60 Euros for participation in the ESM. Participants were told that their target was 60 beeps, and they received a bonus of 10 Euros if they completed 60 beeps or more. Participants were able to answer more than the 60 beep target, and answered 69.33 beeps on average ($SD = 7.59$, range: 30–85). However, if 60 beeps or more are considered a response rate of 100%, the average response rate was 98.7% ($SD = 7.1\%$, range 50–100%). The ethics committee of the Humboldt-Universität zu Berlin approved of the study.

Measures

All four studies assessed various ER strategies at each beep (ESM), NA at each beep (ESM), and as depressive symptoms once (before the ESM phase). As reported in Table 1, the four studies used different items to assess these constructs. The studies that we used to address our research questions were not collected for the purpose of investigating ER variability. Thus, the studies feature different NA items and ER strategies, and the assessments were not aiming to be comprehensive. However, differences between the studies should not be problematic for our research questions, as the principle of ER variability is not linked to specific strategies. In addition, replicating findings across different sets of items and strategies would support the robustness of our conclusions.

In Studies 1 and 4, we selected NA items from the well-known PANAS scales (Watson, Clark, & Tellegen, 1988) that showed sufficient variability in previous studies with intense longitudinal designs (Study 1: nervous, distressed; Study 4: nervous, distressed, jittery, upset; Röcke, Li, & Smith, 2009). We also added the item downhearted to capture sad/depressed mood, and because this item has been used successfully to assess this construct in German experience sampling studies (e.g., Riediger et al., 2009). In Studies 2 and 3, we selected NA items based on Russell's core affect model (Russell, 2003). Items were selected to measure low arousal negative emotion (sad, depressed) and high arousal negative emotion (anxious, angry). The high arousal item *angry* was also assessed in Study 4.

Items that measured ER strategies were selected to fit the rationale of the studies. For example, Study 1 was primarily designed to measure mindfulness. As one of the main components of mindfulness is attention to the present moment, we selected ER strategies that also focused on attentional deployment. In the other studies, we selected well-researched strategies from different stages of the process model of emotion regulation (Gross, 1998): attentional deployment (reflection, rumination, distraction), cognitive change (reappraisal, acceptance), and—in Studies 2 and 3—also response modulation (expressive suppression, social sharing).

Besides the measures used to answer our research questions, the studies featured other items and measures. These other measures depended on the focus of the study, and included things like personality, well-being, critical life events, executive functioning, and reactivity to emotional film clips. For all studies, we applied the procedure we describe below. Descriptive information on all measures used in our analyses are displayed in Table 2.

ESM measures. The ESM items were used to calculate indicators of NA, ER mean strategy endorsement, and ER variability at the moment-level and at the person-level.

Moment-level aggregation. Each individual i has a number of N_i measurement occasions (or time points) t . At each measurement occasion t , there are a number of $s = 1$ to L emotion regulation strategies that the individual can use to a certain degree. Thus, x_{sti} is the value of strategy s at measurement occasion t of individual i . L is the same for all individuals at each measurement occasion in one study (e.g., for Study 2: $L = 6$; rumination, distraction, reflection, reappraisal, suppression, and social sharing).

For the between-strategy variability index, the SD was calculated per measurement occasion. The average intensity with which the strategies were employed at each measurement occasion (the between-strategy mean or mean endorsement) was calculated as follows.

Between-strategy mean (moment-level) / mean endorsement:

$$M_{(between)ti} = \frac{1}{L} \sum_{s=1}^L x_{sti}$$

For mean strategy endorsement (moment-level), we calculated within-person omega scores as measures of reliability based on multilevel confirmatory factor analyses (Geldhof, Preacher, & Zyphur, 2014) conducted in Mplus. These were as follows: Study 1: .54; Study 2: .53; Study 3: .52; Study 4: .56.

Between-strategy index (moment level):

$$SD_{(between)ti} = \sqrt{\frac{1}{L-1} \sum_{s=1}^L (x_{sti} - M_{(between)ti})^2}$$

Figure 1 illustrates these calculations. At each beep, we also calculated the mean of the respective NA items. For mean NA, we calculated within-person omega scores as measures of reliability (Geldhof et al., 2014). These were as follows: Study 1: .67; Study 2: .76; Study 3: .76; Study 4: .81. These measures for NA were based on two-factor models that also incorporated positive affect (see Footnote 2).

Average NA and ER strategy use across the study (person-level aggregation). For each individual, we calculated the mean NA level across all beeps. As a person-level measure of mean ER strategy endorsement, we calculated the mean across all ER strategies across all beeps. Between-person reliabilities were calculated in the same models as the within-person reliabilities (Geldhof et al., 2014). Between-person omegas were as follows for mean ER strategy endorsement: Study 1: .83; Study 2: .81; Study 3: .85; Study 4: .85. Between-person omegas were as follows for NA: Study 1: .93; Study 2: .94; Study 3: .96; Study 4: .96.

Average ER within-strategy and between-strategy variability (person-level aggregation). As an indicator of within-strategy variability, Aldao et al. (2015) proposed the *SD* of a given ER strategy endorsement across different contexts. A high *SD* indicates that a person does not apply a strategy in question in every situation to a similar extent, but is able to inhibit a strategy.

The average intensity with which *one strategy* is employed *across all measurement occasions* t for each individual i was calculated as follows.

$$M_{(within)i} = \frac{1}{N_i} \sum_{t=1}^{N_i} x_{ti}$$

The within-strategy *SD* for each individual i for *one strategy* was then calculated as follows.

$$SD_{(within)i} = \sqrt{\frac{1}{N_i - 1} \sum_{t=1}^{N_i} (x_{ti} - M_i)^2}$$

For the mean within-strategy variability index, the *SDs* of the strategies were averaged across strategies for each individual. Because within-strategy variability depends on the assessment of multiple measurement occasions, this measure could only be obtained at the person level. High global within-strategy variability indicates that a person uses all considered strategies variably across time.

Within-strategy variability index (person level):

$$M_{SD(wichin)i} = \frac{1}{L} \sum_{s=1}^L SD_{(within)si}$$

Reliabilities calculated using Cronbach's α for each study were as follows: Study 1: .86; Study 2: .88; Study 3: .86; Study 4: .88). As a person-level measure of between-strategy variability, the between-strategy *SD* (calculated at the moment-level) was averaged across all measurement occasions for each person.

Between-strategy index (person level):

$$M_{SD(between)i} = \frac{1}{N_i} \sum_{t=1}^{N_i} SD_{(between)ti}$$

Figure 1 illustrates examples of different patterns of higher and lower *SD* values, indicating the pathways through which higher and lower scores on between-strategy variability can be obtained. A higher *SD* between the strategies at each beep can be obtained by several patterns, as illustrated in Figure 1. Higher values are obtained when prioritizing few strategies strongly. Lower values are obtained when either endorsing multiple strategies to a similar extent or endorsing few strategies, but only weakly.

Depressive symptoms. Depressive symptoms were assessed with variations of the Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977). Mean scores for these scales were computed. Reliabilities calculated using Cronbach's α for each Study were as follows: Study 1: .92; Study 2: .91; Study 3: .88; Study 4: .84)

Data Analysis

For the between-person analyses (person level), we used multiple regression models computed in IBM SPSS version 22 for Windows (2013). To obtain the mean effect size of the associations between ER variability and NA across the four studies, we performed fixed effect meta-analyses on the results using Comprehensive Meta-Analysis version 2 (Borenstein, Hedges, Higgins, & Rothstein, 2005).² We chose to use fixed, rather than random-effect meta-analyses because all heterogeneity statistics [Q] were non-significant (Shadish & Haddock, 1994), however we should note that the results using random-effects meta-analyses were not substantively different from the fixed-effect results reported here. In these analyses, we controlled for depressive symptoms, because in two of the studies, the within-study variability of depressive symptoms was increased by the sampling technique (i.e., Study 2 and 3 oversampled individuals with particularly high and low levels of depressive symptoms).³

For the within-person association (moment-level) relating between-strategy variability to NA at particular moments, we used multilevel models. In these models, beeps (Level-1) were nested within persons (Level-2). NA at each beep was predicted by moment-level between-strategy variability and mean strategy endorsement (all Level-1). The predictors were person-mean centered and modeled as fixed and as random effects, with the random intercept and slopes being allowed to covary. In the following, the multilevel equations are presented. Again, x_{ti} (e.g., NA_{ti}) refers to measurement occasion t of individual i .

$$NA_{ti} = \beta_{0i} + \beta_{1i} \times (\text{between-strategy } SD_{ti}) + \beta_{2i} \times (\text{mean ER endorsement}_{ti}) + r_{ti} \quad (\text{L-1})$$

$$\beta_{0i} = \gamma_{00} + \mu_{0i} \quad (\text{L-2})$$

$$\beta_{1i} = \gamma_{10} + \mu_{1i}$$

$$\beta_{2i} = \gamma_{20} + \mu_{2i}$$

Random effects were tested using the deviance statistic (Singer & Willett, 2003). Models were run using the PROC MIXED procedure in SAS version 9.3. A spatial power error structure accounted for the autocorrelation of the unevenly spaced measurement occasions.

Results

Associations Between Mean Endorsement and the Two Types of Variability

We first examined the relationship between the ER variability measures and mean ER strategy endorsement using correlations (Table 3). As expected, there were generally significant positive associations between the variability measures and the mean ER strategy endorsement. An exception was the non-significant association between the within-strategy variability and mean ER endorsement in Study 4. A positive association between mean endorsement and variability indicates that individuals who endorsed strategies to a higher degree also used the strategies more variably across different situations. Within-strategy and between-strategy variability were also positively related, indicating that individuals who did not use particular strategies to the same degree over time also prioritized some strategies over others at any given moment. This association was also significant for all four studies when partialing out the mean ER endorsement (study 1: $r=.480$; study 2: $r=.549$; study 3: $r=.413$; study 4: $r=.490$; all $p < .01$; meta-analytic result: $r=.472$, $p < .001$; 95% CI [.403, .535]).

Associations between variability indicators and NA (person level)

Next, we examined associations between the averaged ER variability indicators and NA (Tables 4 and 5). We used multiple regression analyses. In step 1, we controlled for mean ER strategy endorsement. This allowed us to obtain a measure of ER variability that was not confounded with the mean. As expected, mean strategy endorsement was positively related to

NA, indicating that when participants endorsed many strategies intensively they also have high NA levels. In step 2, we entered variability measures, and in step 3, we controlled for depressive symptoms.

Within-strategy variability. Within-strategy variability was not significantly related to NA in all four studies in step 2 (Table 4). When we controlled for depressive symptoms in step 3, three of the four negative associations between within-strategy variability and NA became significant (Studies 2, 3, 4, $p < .05$). A meta-analysis conducted with the semi-partial correlations (i.e., controlling for mean ER strategy endorsement and depressive symptoms) yielded a significant effect ($r = -.136$, $p = .002$, 95% CI $[-.218, -.052]$). These results suggest that within-strategy variability is associated with lower NA levels independent of depressive symptoms, although the magnitude of this relationship was small.⁵

Between-strategy variability. For between-strategy variability, the results were more consistent (Table 5): In three of the four studies, higher levels of between-strategy variability were significantly related to lower NA levels. Moreover, when controlling for depressive symptoms in step 3, all associations became significant. A meta-analysis conducted with the semi-partial correlations (controlling for mean ER strategy endorsement and depressive symptoms) yielded a significant effect of medium size ($r = -.316$, $p < .001$, 95% CI $[-.390, -.238]$). These results suggest that individuals who, on average, prioritized some strategies over others, experienced less NA during the study.

Given the relationship between within- and between-strategy variability, we also ran regressions with both variability indicators as predictors. When controlling for within-strategy variability (as well as mean ER and depressive symptoms), between-strategy variability stayed significant in three out of the four studies. The semi-partial correlations taken from the regression model were as follows: Study 1: $-.403$, Study 3: $-.245$, Study 4: $-.387$ (all $p < .01$); Study 2: $-.074$ ($p = .255$). These results suggest that within-strategy variability and

between-strategy variability shared predictive variance in NA, but between-strategy variability was predictive above and beyond within-strategy variability.

Associations between between-strategy variability and NA (moment-level)

Finally, we examined whether between-strategy variability and NA were also associated within individuals (at the level of within-person dynamics), controlling for mean strategy endorsement (see Table 6). In these analyses, we tested whether occasions at which individuals prioritized some ER strategies over others (i.e., did not endorse all strategies at the same time or prioritized only one strategy, evidenced by a high *SD* across strategies at that moment) were also occasions at which they experienced less NA. These within-person results yielded similar results as the between-person analyses: again, we found a negative association between NA and between-strategy variability in all four studies. This indicates that at times when individuals used some strategies more than others, they felt less NA. Overall, in the different studies, 14-22% of the variance of NA within individuals was explained by mean ER endorsement and between-strategy variability.

Discussion

Despite an ongoing theoretical discussion centering on the adaptiveness of the variable and flexible use of ER strategies (e.g., Aldao et al., 2015; Bonanno & Burton, 2013; Cheng, Lau, & Chan, 2014; Kashdan & Rottenberg, 2010), there has yet to be a comprehensive study conducted in daily life. Studying variability in daily life is critical, as it provides the opportunity to test the role of variability across many time-points with changing situational demands. Here, we differentiated between two global indicators: within- and between-strategy ER variability. We examined whether they were related to reduced NA—indicating adaptive strategy use—in data from four ESM studies. We included a variety of ER strategies, allowing us to meaningfully estimate these two types of variability.

We found that average within-strategy variability was only weakly associated with NA across studies, and only when controlling for depressive symptoms. This seems to be

indicative of a suppression effect originating from the slightly positive association between within-strategy ER variability and depressive symptoms (see Footnote 3). That is, a small, but apparently not negligible, proportion of the within-strategy ER variance was not adaptive variability but was instead associated with depressive symptoms. If we controlled for this proportion of the variance, within-strategy variability and NA were negatively related. In other words, greater within-strategy variability predicted lower NA (and was potentially adaptive), only when the variance in variability that relates to depression was not considered. This finding can be linked to research on affective variability which suggests that there may be an optimal level of variability, particularly for NA: Very low variability may indicate inflexibility but high variability may indicate instability (Houben, Van Den Noortgate, & Kuppens, 2015). The fact that there was a small but significant correlation between within-strategy variability and depressive symptoms may indicate that similar dynamics are at play for emotion regulation variability. A better understanding of context might help us pull apart instability and flexibility.

When controlling for mean strategy endorsement, between-strategy variability was consistently associated with lower NA, both at the between-person level and the within-person level. These results suggest that (a) individuals who prioritize certain ER strategies, rather than using all strategies to the same degree, experience less NA on average, and (b) situations with more variable between-strategy ER use are situations in which people experience less NA. These results offer support for the idea that between-strategy ER variability has adaptive value, at least when adaptiveness is conceptualized as lower NA.

We did not examine the mechanisms underlying ER variability, but theory suggests that prioritizing some strategies over others might be indicative of a successful search for the best strategy, or of some knowledge of how each strategy best fits the situation. In line with this view, between-strategy variability, which thus far remains understudied, was of key importance in our data. Our results suggest that recruiting all available strategies

simultaneously is not an adaptive way to manage negative emotions. Moreover, between-strategy and within-strategy variability were relatively strongly associated with each other, indicating that individuals high in between-strategy variability also varied the strategies that they chose across time. This suggests that future research should place greater emphasis on studying how selection from a broader repertoire of strategies relates to affective well-being.

This work represents an important initial step in establishing whether ER variability is adaptive. However, we used existing data to address our research questions, meaning that the conditions for studying ER variability and flexibility were not always optimal. We relied on the assumption that contexts in everyday life change frequently. This meant that our variability indices represented, at least in part, context-dependent flexible behavior. However, it will be critical for future research to directly capture contextual change. This will be particularly important because some individuals experience more diverse contexts than others. Measuring context directly will allow us to capture how these individual differences are implicated in ER variability. For example, older adults experience less diverse contexts than younger adults (Brose, Scheibe, & Schmiedek, 2013) and are less variable in their NA (Röcke & Brose, 2013), which may be a reason why they show less variability in daily ER strategies (Eldesouky & English, 2018). That is, reports of contexts are not random, but likely tied to person-level characteristics. In turn, some individuals may need to regulate their emotions more variably than others, and this may pose a threat to the validity of our findings. As three of our four samples were student samples, our ability to investigate questions like interindividual differences in contextual diversity may be limited. In future studies, this problem could be addressed with an event-contingent experience sampling design with more predictable and comparable events across participants.

Measuring context and variability together will also be important in understanding effective ER in several other ways. First, previous research has demonstrated the importance of subjective evaluations of the context in determining which specific ER strategies are

adaptive (e.g., perceived controllability of the situation: Cheng, 2001; Haines et al., 2016, Troy et al., 2013). These subjective context evaluations have proven important for interventions (e.g., Cheng, Kogan, & Chio, 2012). However, this line of research has yet to examine how this contextual information may influence the selection of strategies from a repertoire. Second, we do not know how contextual intensity matters for ER variability. In our data, we focused on normal daily life, and thus we primarily captured more mundane situations, instead of emotionally more intense situations such as major life events. Finally, research on context-dependent variability in ER would inform the debate about whether certain ER strategies are inherently (mal)adaptive. Such research could investigate whether strategies are adaptive in a broader versus a narrower range of contexts. For example, rumination was related to maladaptive outcomes across several studies and contexts (e.g., Aldao & Nolen-Hoeksema, 2010; Brans et al., 2011), indicating that rumination may indeed be maladaptive most of the time.

In the current investigation, we did not include goals or a direct marker of regulatory success. That is, we did not ask participants whether they felt that their regulation efforts were successful in achieving their goals. As people are usually motivated to experience low levels of NA, we used low NA as a marker of regulatory success. In future, we believe that it will be important to test the role of goals in driving ER variability, thus directly testing the concept of emotion regulation flexibility (Aldao et al., 2015). This could be achieved by combining controlled laboratory research (which may foster an understanding of an individual's capacity to use strategies flexibly) with ESM (which may foster an understanding of actual flexible use in daily life).

Since our data is correlational, the temporal order of events remains unclear. It is thus possible that individuals either successfully reduced their NA using variable ER strategies, or that lower NA levels prompted more variable strategy use. It is entirely possible that variability in ER is a consequence, not an antecedent, of lower NA in daily life. Indeed, using

the data set we used in Study 2, Brans et al., (2011) found that, for single strategies, affect and ER may influence each other dynamically. This may also be the case for ER variability, with higher levels of variability resulting in lower levels of NA, which in turn enables higher variability. However, it may also be that high levels of NA make high levels of variability somewhat less likely, since high NA may prompt individuals to try different strategies with a relatively strong intensity.

Relatedly, we did not account for the temporal order of strategy use, which would foster the understanding of how successful ER is achieved (Kalokerinos, Résibois, Verduyn, & Kuppens, 2016). This is especially important to consider in the interpretation of between-strategy variability, as higher levels of between-strategy variability are obtained when few strategies are strongly prioritized. In our data, we do not know whether strategies were used successively or in combination. If individuals try one strategy only briefly, and then switch to another strategy (as investigated, for example, by Birk & Bonanno, 2016), it is also possible that the first strategy may not even be reported in an ESM. Thus, the global index of between-strategy variability as suggested by Aldao et al. (2015) remains somewhat ambiguous in our data. In addition, each study used different strategies, and in Study 1, strategies were paired. Variability may have differed depending on whether individuals are particularly good at selecting similar versus different strategies.

Moreover, we used standard deviations to examine emotion regulation variability because these measures were proposed specifically to assess this construct (Aldao et al., 2015). We controlled for mean ER endorsement, as in previous research. However, in research on intraindividual variability (for example, NA variability), other indicators have been proposed (cf. Ram & Gerstorf, 2009; Röcke & Brose, 2013), including time-structured measures (e.g., the mean squared successive difference; Jahng, Wood, & Trull, 2008) as well as other indicators of net variability (e.g., entropy measures; Benson, Ram, Almeida, Zautra, & Ong, 2018). The degree to which these different variability measures converge and have

shared predictive validity is not clear yet. Furthermore, new methods suggest other ways to partial out the influence of the mean on the *SD* (Mestdagh et al., 2018). In sum, additional methodological research is needed to determine how ER variability and flexibility is best captured.

We used four rich ESM data sets to address the question of how ER variability is associated with NA in daily life. This meant that we were able to replicate our findings across different studies conducted by different labs. Data from these studies have previously been used to explain NA dynamics with other constructs, such as mindfulness (Blanke et al., 2018), daily events (i.e., whether small hassles / stressors have occurred; Blanke et al., 2018; Brose et al., 2017; Koval et al., 2015), and affective memory updating (Pe et al., 2013). In other studies, these constructs that have already been examined in our datasets were related to emotion regulation strategies (e.g., Brockman, Ciarrochi, Parker & Kashdan, 2017; Brose, Schmiedek, Lövdén, & Lindenberger, 2011). The previous studies conducted with our data suggest that these processes are important for emotion in daily life in our samples, and it may be that ER variability is also implicated in how these constructs predict NA. However, since our four studies originally targeted different research questions and thus do not contain the same variables, it was not possible for us to test whether ER variability reliably predicts NA above and beyond these different constructs, or interacts with these other constructs in predicting NA. Here, we think that it is important to note that more research is needed to replicate our findings on adaptiveness. In particular, targeted confirmatory research is necessary to investigate the potential role of ER variability in relationships between other variables and NA. Such future research should also adhere to current best practices in the field such as preregistration and a priori sample size determination, which was not possible in the present work.

Conclusion

In this research, we investigated whether two types of ER variability were associated with reduced NA, an indicator of adaptive regulation, in daily life. Using data from four ESM studies, the findings provide support for the adaptive value of between-strategy variability, an indicator capturing ER variability at one point in time. Between-strategy variability was associated with reduced NA both at the average level and at the level of within-person dynamics. Thus, this study provides support for modern theoretical perspectives, which emphasize flexible choice rather than universal efficacy of ER strategies. This study also points towards the necessity of advancing current thinking about how best to conceptualize and capture emotion regulation variability in daily life.

References

- Aldao, A., & Nolen-Hoeksema, S. (2010). Specificity of cognitive emotion regulation strategies: A transdiagnostic examination. *Behaviour Research and Therapy*, *48*, 974–983. doi: 10.1016/j.brat.2010.06.002
- Aldao, A., & Nolen-Hoeksema, S. (2012). The influence of context on the implementation of adaptive emotion regulation strategies. *Behaviour Research and Therapy*, *50*, 493–501. doi: 10.1016/j.brat.2012.04.004
- Aldao, A., & Nolen-Hoeksema, S. (2013). One versus many: Capturing the use of multiple emotion regulation strategies in response to an emotion-eliciting stimulus. *Cognition & Emotion*, *27*, 753–760. doi: 10.1080/02699931.2012.739998
- Aldao, A., Sheppes, G., & Gross, J. J. (2015). Emotion regulation flexibility. *Cognitive Therapy and Research*, *39*, 263–278. doi: 10.1007/s10608-014-9662-4
- Appleton A. A. & , Kubzansky L. D. (2014). *Emotion regulation and cardiovascular disease risk*. In J. Gross (ed.) *Handbook of Emotion Regulation* (2nd edition, pp. 596-612). New York, Guilford Press.
- Barrett, D. J., & Barrett, L. F. (2000). The experience sampling program (ESP) [Software]. Retrieved from <http://www.experience-sampling.org/>
- Bastian, B., Koval, P., Erbas, Y., Houben, M., Pe, M. L., & Kuppens, P. (2015). Sad and alone: Social expectancies for experiencing negative emotions are linked to feelings of loneliness. *Social Psychological and Personality Science*, *6*, 496-503. doi: 10.1177/1948550614568682
- Benson, L., Ram, N., Almeida, D., Zautra, A., & Ong, A. D. (2018). Fusing biodiversity metrics into investigations of daily life and healthy aging: Illustrations and recommendations with emodiversity. *Journal of Gerontology: Psychological Sciences*, *73*, 74–86. doi: 10.1093/geronb/gbx025

- Birk, J. L., & Bonanno, G. A. (2016). When to throw the switch: The adaptiveness of modifying emotion regulation strategies based on affective and physiological feedback. *Emotion, 16*, 657–670. doi: 10.1037/emo0000157
- Blanke, E.S. & Brose, A. (2017). Mindfulness in Daily Life: a Multidimensional Approach *Mindfulness, 8*, 737–750. doi:10.1007/s12671-016-0651-4
- Blanke, E.S., Riediger, M., & Brose, A. (2018). Pathways to happiness are multidirectional: Associations between state mindfulness and everyday affective experience. *Emotion, 18*, 202–211. doi: 10.1037/emo0000323.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616. doi: 10.1146/annurev.psych.54.101601.145030
- Bonanno, G. A., & Burton, C. L. (2013). Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on Psychological Science, 8*, 591–612. doi: 10.1177/1745691613504116
- Bonanno, G. A., Papa, A., Lalande, K., Westphal, M., & Coifman, K. (2004). The importance of being flexible: The ability to both enhance and suppress emotional expression predicts long-term adjustment. *Psychological Science, 15*, 482–487. doi: 10.1111/j.0956-7976.2004.00705.x
- Bonanno, G. A., Pat-Horenczyk, R., & Noll, J. (2011). Coping flexibility and trauma: The Perceived Ability to Cope With Trauma (PACT) scale. *Psychological Trauma: Theory, Research, Practice, and Policy, 3*, 117–129. doi: 10.1037/a0020921
- Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis (version 2)*. Englewood, NJ: Biostat, 104.
- Brans, K., Koval, P., Verduyn, P., Lim, Y. L., & Kuppens, P. (2013). The regulation of negative and positive affect in daily life. *Emotion, 13*, 926–939. doi: 10.1037/a0032400

- Brockman, R., Ciarrochi, J., Parker, P., & Kashdan, T. (2017). Emotion regulation strategies in daily life: mindfulness, cognitive reappraisal and emotion suppression. *Cognitive Behaviour Therapy, 46*, 91–113. doi: 10.1080/16506073.2016.1218926
- Brose, A., Scheibe, S., & Schmiedek, F. (2013). Life contexts make a difference: Emotional stability in younger and older adults. *Psychology and Aging, 28*, 148–159. doi:10.1037/a0030047
- Brose, A., Schmiedek, F., Lövdén, M., & Lindenberger, U. (2012). Daily variability in working memory is coupled with negative affect: the role of attention and motivation. *Emotion, 12*, 605–617. doi: 10.1037/a0024436
- Brose, A., Wichers, M., & Kuppens, P. (2017). Daily stressful experiences precede but do not succeed depressive symptoms: Results from a longitudinal experience sampling study. *Journal of Social and Clinical Psychology, 36*, 196–220. doi: 10.1521/jscp.2017.36.3.196
- Cheng, C. (2001). Assessing coping flexibility in real-life and laboratory settings: A multimethod approach. *Journal of Personality and Social Psychology, 80*, 814–833. doi: 10.1037/0022-3514.80.5.814
- Cheng, C., Kogan, A., & Chio, J. H. M. (2012). The effectiveness of a new, coping flexibility intervention as compared with a cognitive-behavioural intervention in managing work stress. *Work & Stress, 26*, 272–288. doi: 10.1080/02678373.2012.710369
- Cheng, C., Lau, H.-P. B., & Chan, M.-P. S. (2014). Coping flexibility and psychological adjustment to stressful life changes: A meta-analytic review. *Psychological Bulletin, 140*, 1582–1607. doi: 10.1037/a0037913

- Dejonckheere, E., Mestdagh, M., Houben, M., Erbas, Y., Pe, M., Koval, P., ..., & Kuppens P. (2018). The bipolarity of affect and depressive symptoms. *Journal of Personality and Social Psychology, 114*, 323-341. doi: 10.1037/pspp0000186
- Dixon-Gordon, K. L., Aldao, A., & De Los Reyes, A. (2015). Emotion regulation in context: Examining the spontaneous use of strategies across emotional intensity and type of emotion. *Personality and Individual Differences, 86*, 271–276. doi: 10.1016/j.paid.2015.06.011
- Eldesouky, L., & English, T. (2018). Another year older, another year wiser? Emotion regulation strategy selection and flexibility across adulthood. *Psychology and Aging, 33*, 572–585. doi: 10.1037/pag0000251
- Erbas, Y., Ceulemans, E., Kalokerinos, E., Houben, M., Koval, P., Pe, M., & Kuppens, P. (in press). Why I don't always know what I'm feeling: The role of stress in within-person fluctuations in emotion differentiation. *Journal of Personality and Social Psychology*.
- Erbas, Y., Ceulemans, E., Koval, P., & Kuppens, P. (2015). The role of valence focus and appraisal overlap in emotion differentiation. *Emotion, 15*, 373–382. doi:10.1037/emo0000039
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*, 72-91. doi: 10.1037/a0032138
- Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology, 2*, 271–299. doi: 10.1037/1089-2680.2.3.271
- Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology, 39*, 281–291. doi: 10.1017.S0048577201393198

- Haines, S. J., Gleeson, J., Kuppens, P., Hollenstein, T., Ciarrochi, J., Labuschagne, I., ... & Koval, P. (2016). The wisdom to know the difference: Strategy-situation fit in emotion regulation in daily life is associated with well-being. *Psychological Science*. Advance online publication. doi: 0956797616669086.
- Hautzinger, M., & Bailer, M. (1993). *ADS—Allgemeine Depressionsskala*. [General Depression Scale]. Göttingen, Germany: Beltz Test GmbH.
- Houben, M., Van den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*, 901–930. doi:10.1037/a0038822
- Hollenstein, T., Lichtwarck-Aschoff, A., & Potworowski, G. (2013). A model of socioemotional flexibility at three time scales. *Emotion Review*, *5*, 397–405. doi: 10.1177/1754073913484181
- Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult: criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Archives of Internal Medicine*, *159*, 1701–1819. doi: 1701-1704. doi:10.1001/archinte.159.15.1701
- Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods*, *13*, 354–375. doi: 10.1037/a0014173
- Kalokerinos, E. K., Résibois, M., Verduyn, P., & Kuppens, P. (2016). The temporal deployment of emotion regulation strategies during negative emotional episodes. *Emotion*. Advance online publication. doi: 10.1037/emo0000248
- Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*, 865–878. doi: 10.1016/j.cpr.2010.03.001

- Kato, T. (2012). Development of the Coping Flexibility Scale: Evidence for the coping flexibility hypothesis. *Journal of Counseling Psychology, 59*, 262–273. doi: 10.1037/a0027770
- Koval, P., Brose, A., Pe, M. L., Houben, M., Erbas, Y., Champagne, D., & Kuppens, P. (2015). Emotional inertia and external events: The roles of exposure, reactivity, and recovery. *Emotion, 15*, 625–636. doi: 10.1037/emo0000059
- Koval, P., Ogrinz, B., Kuppens, P., Van den Bergh, O., Tuerlinckx, F., & Sütterlin, S. (2013). Affective instability in daily life is predicted by resting heart rate variability. *PLoS ONE, 8*, e81536. doi:10.1371/journal.pone.0081536.
- Koval, P., Pe, M. L., Meers, K., & Kuppens, P. (2013). Affect dynamics in relation to depressive symptoms: Variable, unstable or inert? *Emotion, 13*, 1132–1141. doi: 10.1037/a0033579
- Mestdagh, M., Pe, M. L., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*. Advance online publication. doi: 10.1037/met0000153.
- Pe, M. L., Brose, A., Gotlib, I. H., & Kuppens, P. (2016). Affective updating ability and stressful events interact to prospectively predict increases in depressive symptoms over time. *Emotion, 16*, 73–82. doi:10.1037/emo0000097.
- Pe, M. L., Koval, P., & Kuppens, P. (2012). Executive well-being: Updating of positive stimuli in working memory is associated with subjective well-being. *Cognition, 126*, 335–340. doi : 10.1016/j.cognition.2012.10.002.
- Pe, M. L., Koval, P., Houben, M., Erbas, Y., Champagne, D., & Kuppens, P. (2015). Updating in working memory predicts greater emotion reactivity to and facilitated

- recovery from negative emotion-eliciting stimuli. *Frontiers in Psychology*, 6(372). doi: 10.3389/fpsyg.2015.00372.
- Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385–401. doi: 10.1177/014662167700100306
- Ram, N., & Gerstorf, D. (2009). Time-structured and net intraindividual variability: Tools for examining the development of dynamic characteristics and processes. *Psychology and Aging*, 24, 778–791. doi:10.1037/a0017915
- Rauers, A., Blanke, E., & Riediger, M. (2013). Everyday empathic accuracy in younger and older couples: Do you need to see your partner to know his or her feelings? *Psychological Science*, 24, 2210–2217. doi: 10.1177/0956797613490747
- Richter, D., & Schupp, J. (2015). The SOEP Innovation Sample (SOEP IS). *Schmollers Jahrbuch*, 135, 389–399. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie*, 46, 344–364.
- Riediger, M., Schmiedek, F., Wagner, G. G., & Lindenberger, U. (2009). Seeking pleasure and seeking pain: Differences in prohedonic and contra-hedonic motivation from adolescence to old age. *Psychological Science*, 20, 1529–1535. doi: 10.1111/j.1467-9280.2009.02473.x
- Röcke, C. & Brose, A. (2013). Intraindividual variability and stability of affect and well-being: Short-term and long-term change and stabilization processes. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry*, 26, 185–199. doi:10.1024/1662-9647/a000094
- Röcke, C., Li, S. C., & Smith, J. (2009). Intraindividual variability in positive and negative affect over 45 days: Do older adults fluctuate less than young adults?. *Psychology and Aging*, 24, 863–878. doi: 0.1037/a0016276

Russell, J. A. (2003). Core affect and the psychological construction of emotion.

Psychological Review, *110*, 145–172. doi: 10.1037//0033-295X.110.1.145

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York, NY: Russell Sage Foundation.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Troy, A. S., Shallcross, A. J., & Mauss, I. B. (2013). A person-by-situation approach to emotion regulation: Cognitive reappraisal can either help or hurt, depending on the context. *Psychological Science*, *24*, 2505–2514. doi: 10.1177/0956797613496434

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. doi: 10.1037/0022-3514.54.6.1063

Footnotes

¹ Coping usually refers to the downregulation of NA or stress whereas ER also considers processes such as the maintenance or upregulation of positive affect (Gross, 1998). In the following, we do not differentiate between literature on ER and coping.

² We had no hypotheses regarding associations between ER variability and positive affect (PA). Omega scores for within- and between-person reliability for PA were as follows: Study 1: .77/.93; Study 2: .71/.91; Study 3: .74/.91; Study 4: .80/.90). However, we also computed a meta-analytic correlation based on semi-partial correlations between PA and variability controlling for mean ER endorsement and depressive symptoms. This analysis did not reveal an effect for between-strategy variability ($r=.014$, $p=.753$, 95% CI [$-.071$; $.098$]), or within-strategy variability ($r=.029$, $p=.502$, 95% CI [$-.056$; $.114$]).

³ We also computed a meta-analytic correlation based on semi-partial correlations between depression and variability controlling for mean ER endorsement. This analysis did not reveal an effect for between-strategy variability ($r=.020$, $p=.650$, 95% CI [$-.065$; $.104$]), but a small positive association with within-strategy variability ($r=.112$, $p<.01$, 95% CI [$.028$; $.195$]).

⁴ Due to a mistake in the programming of the task, a 5-point scaling instead of the original scaling was used.

⁵ We also tested the effect of within-strategy variability on NA for each single strategy separately for each study. The results are in Table 1 of the supplementary material. We also conducted multiple regression analyses, entering mean ER endorsement, the CES-D score, as well as the within-strategy *SDs* for all single strategies (all predictors were grand-mean centered). In Studies 1 and 3, none of the within-strategy *SDs* significantly predicted NA above and beyond the other predictors. In Study 2, within-strategy variability for distraction was significantly related to lower NA above and beyond the (non-significant) effect of the other strategies. In Study 4, within-strategy variability for distraction was significantly

associated with lower NA, and within-strategy variability for rumination was significantly associated with higher NA. These results are in Table 2 of the supplementary material.

Table 1.
Overview of the Measures

| | Study 1 (<i>N</i> = 70, Germany, 9 consecutive days, 6 beeps per day) | Study 2 (<i>N</i> = 95, Belgium, 7 consecutive days, 10 beeps per day) | Study 3 (<i>N</i> = 200, Belgium, 7 consecutive days, 10 beeps per day) | Study 4 (<i>N</i> = 179, Germany, 3 x 4 days, 6 beeps per day) |
|----------------------------|--|---|---|---|
| ER strategies (ESM) | | | | |
| | <ul style="list-style-type: none"> • Rumination on thoughts • Rumination on feelings • Distraction from thoughts • Distraction from feelings • Reflection on thoughts • Reflection on feelings | <ul style="list-style-type: none"> • Rumination • Distraction • Reflection • Other perspective/ reappraisal • Expressive suppression • Social sharing | <ul style="list-style-type: none"> • Rumination about the past • Rumination about the future • Distraction • Other perspective/ reappraisal • Expressive suppression • Social sharing | <ul style="list-style-type: none"> • Rumination • Distraction • Reflection • Positive reappraisal • Acceptance |
| <i>Answering scale</i> | 7-point scale from 0 (<i>does not apply at all</i>) to 6 (<i>applies strongly</i>) | Slider scale from 0 (<i>not at all</i>) to 100 (<i>very much</i>) | Slider scale from 0 (<i>not at all</i>) to 100 (<i>very much</i>) | 7-point scale from 0 (<i>does not apply at all</i>) to 6 (<i>applies strongly</i>) |
| <i>Reference frame</i> | Since waking up/ since the last beep | Since the last beep | Since the last beep | Since waking up/ since the last beep |
| NA (ESM) items | | | | |
| | <ul style="list-style-type: none"> • Nervous • Downhearted • Distressed | <ul style="list-style-type: none"> • Angry • Sad • Anxious • Depressed | <ul style="list-style-type: none"> • Angry • Sad • Anxious • Depressed | <ul style="list-style-type: none"> • Angry • Nervous • Downhearted • Upset • Jittery • Distressed |
| <i>Answering scale</i> | 7-point scale from 0 (<i>does not apply at all</i>) to 6 (<i>applies strongly</i>) | Slider scale from 0 (<i>not at all</i>) to 100 (<i>very much</i>) | Slider scale from 0 (<i>not at all</i>) to 100 (<i>very much</i>) | 7-point scale from 0 (<i>does not apply at all</i>) to 6 (<i>applies strongly</i>) |
| <i>Reference frame</i> | Since waking up/ since the last beep | Current (at the moment of the beep) | Current (at the moment of the beep) | Current (at the moment of the beep) |
| Depressive symptoms | | | | |
| | 20-item CES-D | 20-item CES-D | 20-item CES-D | 10-item CES-D |
| <i>Answering scale</i> | 5-point scale from 0 (never) to 4 (always) ⁴ | 4-point scale from 0 (<i>rarely or none of the time, less than 1 day</i>) to 3 (<i>most or all of the time, 5–7 days</i>) | 4-point scale from 0 (<i>rarely or none of the time, less than 1 day</i>) to 3 (<i>most or all of the time, 5–7 days</i>) | 4-point scale from 0 (<i>rarely or none of the time, less than 1 day</i>) to 3 (<i>most or all of the time, 5–7 days</i>) |

Notes. CES-D = Center for Epidemiologic Studies Depression Scale (Radloff, 1977); ER = emotion regulation, *SD* = standard deviation, NA =

negative affect, ESM = experience-sampling methodology.

Table 2.

Descriptive Information: Mean Values (Standard Deviations).

| | Study 1 (<i>N</i> = 70) | Study 2 (<i>N</i> = 95) | Study 3 (<i>N</i> = 200) | Study 4 (<i>N</i> = 179) |
|---------------------------------|-----------------------------|-----------------------------|------------------------------|------------------------------|
| Mean ER endorsement | 1.87 (0.70) | 23.92 (10.62) | 21.26 (11.32) | 2.42 (0.93) |
| <i>SD</i> within ER strategies | 1.23 (0.33) | 17.78 (5.53) | 17.04 (5.83) | 1.35 (0.42) |
| <i>SD</i> between ER strategies | 1.34 (0.40) | 17.47 (7.60) | 17.06 (8.47) | 1.44 (0.61) |
| NA (ESM) | 1.40 (0.90) | 15.65 (10.75) | 14.31 (8.44) | 1.05 (0.81) |
| Depressive symptoms | 1.44 (0.61) ⁴ | 0.73 (0.48) | 0.63 (0.39) | 0.85 (0.51) |

Notes. ER = emotion regulation, *SD* = standard deviation, NA = negative affect, ESM = experience-sampling methodology.

Table 3.

Correlations Between Mean ER Strategy Endorsement, Within-strategy Variability, and Between-strategy Variability

| | Mean SD within ER strategies | | | | Mean SD between ER strategies | | | |
|-------------------------------------|-------------------------------------|-----------------------------|------------------------------|------------------------------|--------------------------------------|-----------------------------|------------------------------|------------------------------|
| | Study 1 (<i>N</i> = 70) | Study 2 (<i>N</i> = 95) | Study 3 (<i>N</i> = 200) | Study 4 (<i>N</i> = 179) | Study 1 (<i>N</i> = 70) | Study 2 (<i>N</i> = 95) | Study 3 (<i>N</i> = 200) | Study 4 (<i>N</i> = 179) |
| Mean ER endorsement | .471** | .681** | .669** | -.038 | .322** | .679** | .742** | .437** |
| Mean SD within ER strategies | - | - | - | - | .553** | .758** | .702** | .424** |

Notes. ER = emotion regulation, *SD* = standard deviation. ** $p < .01$, * $p < .05$.

Table 4.
Stepwise Multiple Regression Analysis: Predicting Aggregated NA (ESM) with Within-strategy Variability

| | Estimates (<i>SE</i>) [95% CI] | | | |
|--------------------------------|----------------------------------|---------------------------------|---------------------------------|-------------------------------|
| | Study 1 (<i>N</i> = 70) | Study 2 (<i>N</i> = 95) | Study 3 (<i>N</i> = 200) | Study 4 (<i>N</i> = 179) |
| <i>Step 1</i> | | | | |
| Intercept | 1.40** (0.09) [1.21–1.59] | 15.65** (0.87) [13.93–17.37] | 14.31** (0.41) [13.50–15.12] | 1.05** (0.06) [0.93–1.16] |
| Mean ER endorsement | 0.66** (0.13) [0.39–0.92] | 0.63** (0.08) [0.47–0.79] | 0.54** (0.04) [0.47–0.62] | 0.15* (0.06) [0.02–0.28] |
| Adjusted <i>R</i> ² | .25 | .38 | .53 | .02 |
| <i>Step 2</i> | | | | |
| Intercept | 1.40** (0.09) [1.21–1.59] | 15.65** (0.86) [13.94–17.37] | 14.31** (0.41) [13.51–15.12] | 1.05** (0.06) [0.93–1.16] |
| Mean ER endorsement | 0.72** (0.15) [0.41–1.02] | 0.74** (0.11) [0.51–0.96] | 0.59** (0.05) [0.50–0.69] | 0.14* (0.06) [0.02–0.27] |
| Within-strategy <i>SD</i> | –0.28 (0.32) [–0.92–0.37] | –0.30 (0.21) [–0.72–0.13] | –0.15 (0.09) [–0.34–0.04] | –0.22 (0.14) [–0.50–0.06] |
| Adjusted <i>R</i> ² | .25 | .39 | .53 | .03 |
| <i>Step 3</i> | | | | |
| Intercept | 1.40** (0.08) [1.24–1.56] | 15.65** (0.69) [14.28–17.03] | 14.31** (0.39) [13.55–15.08] | 1.05** (0.05) [0.94–1.15] |
| Mean ER endorsement | 0.56** (0.13) [0.29–0.82] | 0.65** (0.09) [0.47–0.83] | 0.56** (0.05) [0.47–0.66] | 0.16** (0.06) [0.04–0.27] |
| Depressive Symptoms | 0.70** (0.14) [0.43–0.98] | 11.13** (1.54) [8.07–14.18] | 5.03** (1.07) [2.92–7.14] | 0.69** (0.11) [0.48–0.90] |
| Within-strategy <i>SD</i> | –0.39 (0.28) [–0.94–0.16] | –0.51** (0.17) [–0.85–0.16] | –0.21* (0.09) [–0.39–0.03] | –0.28* (0.13) [–0.53–0.02] |
| Adjusted <i>R</i> ² | .45 | .61 | .58 | .21 |

Notes. Unstandardized regression estimates. NA = negative affect, ER = emotion regulation, *SD* = standard deviation, *SE* = standard error.

** *p* < .01, * *p* < .05.

Table 5.
Stepwise Multiple Regression Analysis: Predicting Aggregated NA (ESM) with Between-strategy Variability

| | Estimates (<i>SE</i>) [95% CI] | | | |
|--------------------------------|----------------------------------|---------------------------------|---------------------------------|---------------------------------|
| | Study 1 (<i>N</i> = 70) | Study 2 (<i>N</i> = 95) | Study 3 (<i>N</i> = 200) | Study 4 (<i>N</i> = 179) |
| <i>Step 1</i> | | | | |
| Intercept | 1.40** (0.09) [1.21–1.59] | 15.65** (0.87) [13.93–17.37] | 14.31** (0.41) [13.50–15.12] | 1.05** (0.06) [0.93–1.16] |
| Mean ER endorsement | 0.66** (0.13) [0.39–0.92] | 0.63** (0.08) [0.47–0.79] | 0.54** (0.04) [0.47–0.62] | 0.15* (0.06) [0.02–0.28] |
| Adjusted <i>R</i> ² | .25 | .38 | .53 | .02 |
| <i>Step 2</i> | | | | |
| Intercept | 1.40** (0.08) [1.25–1.55] | 15.65** (0.87) [13.92–17.38] | 14.31** (0.38) [13.56–15.07] | 1.05** (0.05) [0.94–1.15] |
| Mean ER endorsement | 0.87** (0.12) [0.64–1.11] | 0.59** (0.11) [0.37–0.82] | 0.74** (0.05) [0.64–0.84] | 0.35** (0.06) [0.23–0.48] |
| Between-strategy <i>SD</i> | −1.18** (0.20) [−1.58–−0.77] | 0.08 (0.16) [−0.24–0.39] | −0.36** (0.07) [−0.50–−0.23] | −0.71** (0.10) [−0.90–−0.52] |
| Adjusted <i>R</i> ² | .49 | .38 | .59 | .25 |
| <i>Step 3</i> | | | | |
| Intercept | 1.40** (0.07) [1.27–1.54] | 15.65** (0.70) [14.26–17.04] | 14.31** (0.36) [13.60–15.02] | 1.05** (0.05) [0.95–1.14] |
| Mean ER endorsement | 0.70** (0.11) [0.48–0.91] | 0.62** (0.09) [0.44–0.80] | 0.71** (0.05) [0.61–0.80] | 0.33** (0.06) [0.22–0.45] |
| Depressive Symptoms | 0.55** (0.12) [0.31–0.78] | 12.17** (1.69) [8.80–15.53] | 5.32** (0.99) [3.38–7.26] | 0.55** (0.10) [0.35–0.74] |
| Between-strategy <i>SD</i> | −0.99** (0.18) [−1.36–−0.63] | −0.35* (0.14) [−0.62–−0.07] | −0.40** (0.06) [−0.53–−0.27] | −0.61** (0.09) [−0.79–−0.43] |
| Adjusted <i>R</i> ² | .61 | .60 | .64 | .36 |

Notes. Unstandardized regression estimates. NA = negative affect, ER = emotion regulation,

SD = standard deviation, *SE* = standard error.

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

Table 6.
*Multilevel Modeling: Predicting Negative Affect with Between-Strategy Variability
 (Controlled for Mean ER Strategy Endorsement)*

| | Estimates (<i>SE</i>) [95% CI] | | | |
|--------------------------------|----------------------------------|----------------------------------|---------------------------------|--------------------------------|
| | Study 1 (<i>N</i> = 70) | Study 2 (<i>N</i> = 95) | Study 3 (<i>N</i> = 200) | Study 4 (<i>N</i> = 179) |
| <i>Fixed effects</i> | | | | |
| Intercept | 1.40 (0.11)** [1.19–1.62] | 15.64 (1.10)** [13.45–17.83] | 14.26 (0.60)** [13.09–15.43] | 1.04 (0.06)** [0.92–1.16] |
| Between- strategy <i>SD</i> | −0.09 (0.04)* [−0.18—0.01] | −0.07 (0.03)* [−0.13 —<−0.01] | −0.08 (0.02)** [−0.12—0.04] | −0.15 (0.02)** [−0.20—0.10] |
| Mean ER endorsement | 0.40 (0.03)** [0.34–0.46] | 0.37 (0.03)** [0.32–0.42] | 0.33 (0.02)** [0.29–0.38] | 0.11 (0.02)** [0.07–0.15] |
| <i>Random effects</i> | | | | |
| Intercept | 0.80 (0.14) ¹ | 112.84 (16.86) ¹ | 68.67 (7.11) ¹ | 0.64 (0.07) ¹ |
| Between- strategy <i>SD</i> | 0.08 (0.02) ¹ | 0.06 (0.01) ¹ | 0.06 (0.01) ¹ | 0.07 (0.01) ¹ |
| Mean ER endorsement | 0.03 (0.01) ¹ | 0.04 (0.01) ¹ | 0.07 (0.01) ¹ | 0.06 (0.01) ¹ |
| Residual | 0.66 (0.02) | 100.76 (2.19) | 91.51 (1.32) | 0.62 (0.01) |
| Pseudo <i>R</i> ² | .22 | .20 | .19 | .14 |

Notes. Unstandardized multilevel estimates. ER = emotion regulation, *SD* = standard deviation, *SE* = standard error.

¹Random effect exceeds the .01 critical χ^2 value obtained by the deviance statistic (see Singer & Willett, 2003), indicating that the random effects should not be restricted to zero.

Autoregressive error-structure and covariances between intercept and slopes were estimated, but are not displayed.

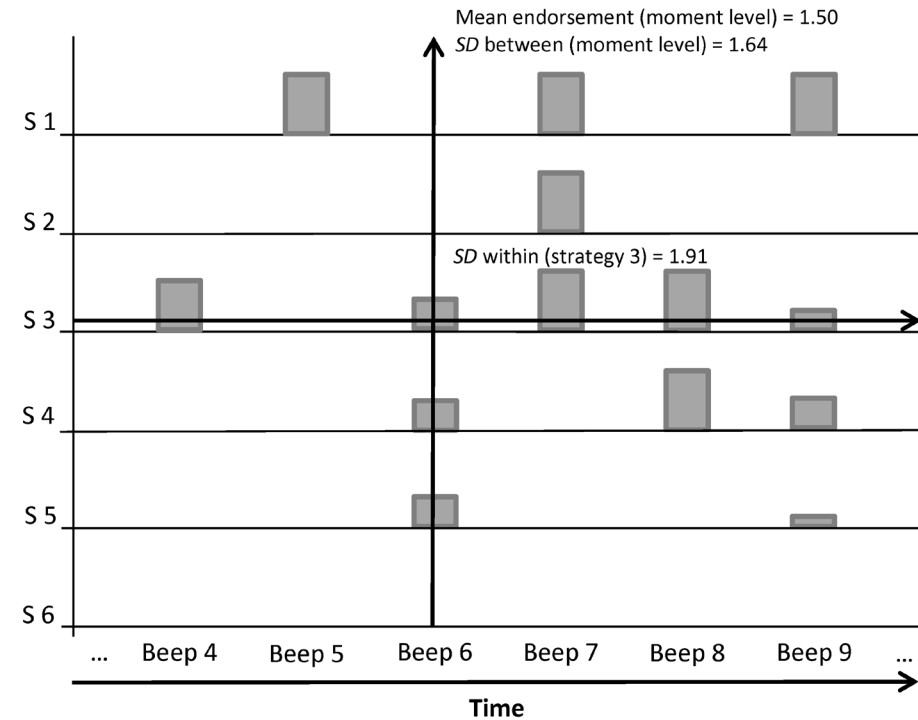
** $p < .01$, * $p < .05$.

Figure 1.
Illustration of Within-Strategy and Between-Strategy SD.

Panel A

| Beep No | S 1 | S 2 | S 3 | S 4 | S 5 | S 6 | <i>SD</i> Between | <i>M</i> |
|---------------------|-----|-----|-----|-----|-----|-----|----------------------|-------------|
| 1 | 4 | 4 | 4 | 3 | 3 | 3 | 0.55 | 3.5 |
| 2 | 1 | 1 | 1 | 2 | 2 | 2 | 0.55 | 1.5 |
| 3 | 6 | 1 | 3 | 1 | 1 | 1 | 2.04 | 2.17 |
| 4 | 0 | 0 | 5 | 0 | 0 | 0 | 2.04 | 0.83 |
| 5 | 6 | 0 | 0 | 0 | 0 | 0 | 2.45 | 1 |
| 6 | 0 | 0 | 3 | 3 | 3 | 0 | 1.64 | 1.5 |
| 7 | 6 | 6 | 6 | 0 | 0 | 0 | 3.29 | 3 |
| 8 | 0 | 0 | 6 | 6 | 0 | 0 | 3.1 | 2 |
| 9 | 6 | 0 | 2 | 3 | 1 | 0 | 2.28 | 2 |
| 10 | 0 | 0 | 4 | 4 | 0 | 0 | 2.07 | 1.33 |
| 11 | 6 | 1 | 3 | 2 | 3 | 0 | 2.07 | 2.5 |
| | | | | | | | 2.01 | 1.94 |
| <i>SD</i> Within | 2.9 | 2 | 1.9 | 1.9 | 1.3 | 1 | 1.85 | |

Panel B



Notes. *SD* Within = within-strategy variability, *SD* Between = between-strategy variability. S = ER strategy.

Panel A: Fictitious data from one person that rated six ER strategies on a scale from 0–6, displaying different means and *SD*s. Bold numbers denote person-level indicators.

Panel B: Illustration of Beeps 4 to 9 from Panel A and the corresponding *SD*s.