Received: 12 April 2021 | Accepted: 14 September 2021

DOI: 10.1002/ejsp.2809

RESEARCH ARTICLE

EASP WILEY

The effect of serial day on the measurement of positivity and emotional complexity in diary studies

Yoav Ganzach[1] | Ben Bulmash[2]

[1] Ariel University, Ariel, Israel

[2] Holon Institute of Technology, Holon, Israel

Correspondence
Yoav Ganzach, Tel Aviv, Business Administration, Tel Aviv, Israel.
Email: yoavgn@post.tau.ac.il

Abstract

We study the effects of a diary's serial day (the number of days from the beginning of the study) on participants' (n = 2022) reports about positive and negative affect (NA). We find that (1) the number of reported positive events and the number of reported negative events decrease with serial day; (2) positivity increases with serial day: Reported Positive Affect (PA) increases, and reported NA decreases; (3) emotional complexity—the tendency to differentiate between various types of emotions—decreases with serial day, both within and between affective dimensions. We attribute these effects to decrease in the effort exerted by participants in answering the diary questions, and suggest that these effects are consistent with the distinction between experienced and reported emotions and with a heuristic and biases perspective in which when effort decreases reported emotions regress to an easier-to-generate default response.

KEYWORDS
diary studies, emotions, positive and negative affect, regression to the mean

1 | INTRODUCTION

A constructionist view of emotions distinguishes between the experience of affect and the judgement of this affect. As Russell (2017, p. 111)) suggests 'Feeling bad is one thing, judging something to be bad is another'. In the current article we take a similar approach to analyse how key features of reports about emotions in a daily diary study vary as a function of the number of days from the beginning of the study (serial days), suggesting that these changes are due not to fluctuations in 'true', or experienced, emotions, but rather to systematic changes in reports about emotions. We attribute these changes to the processes by which reports about emotions are constructed, and in particular, to the fact that when effort decreases there is an increased tendency to rely on easy-to-use heuristics and readily available default responses (Danziger et al., 2011; Kahneman & Frederick, 2002; Shah & Oppenheimer, 2008).

Diary studies have become more and more popular (Bolger & Laurenceau, 2013; Gunthert & Wenze, 2012; Ohly et al., 2010), mainly due to the belief that the shorter the time gap between the experiences and their reports, the better the recall of these experiences and the more accurate the reports.[1] Indeed, many studies that have compared the accuracy of delayed reports to the accuracy of contemporary (i.e., diary based) reports show that the latter are more accurate than the former (Al Baghal et al., 2014; Brenner, 2017; Cannell et al., 1981; Townshend & Duka, 2002). These studies, however, focused on reports about activities or events rather than reports about emotions, and thus

[1] Diary studies involves asking participants to report on their thoughts, feelings, behaviours and/or environment on multiple occasions over time. They may involve methods such daily diaries in which reports are elicited once a day or Experience Sampling Methods in which reports are elicited more frequently. The distinction between the various diary methods is, however, not clear cut. As Horstmann (2021) argues, 'ESM and daily diary may be subsumed under the general method ecological momentary assessment (EMA), which also includes ambulatory assessment (AA). Whereas AA may employ other assessment methods apart from self-report (e.g., audio snippets, pictures, video recordings), ESM mostly refers to a repeated self-assessment of a person's current experiences. That being said, the use of terminology varies and is not always consistent across labs, disciplines or countries. For example, in ESM studies surveys can be taken every 3 h, while daily diary studies entail a survey once a day (mostly at the end of the day)'.

could rely on 'objective' information (e.g., church attendance records) as a criterion to assess the accuracy of the reports (e.g., reported church attendance).

Because of the difficulties in obtaining objective measures of emotions, most of the studies about the accuracy of reports about emotions took a different approach and focused primarily on comparing contemporary affective reports to retrospective reports. Early studies took for granted that diary-based affective reports are accurate, and used them as a criterion to evaluate the accuracy of delayed reports (Ptacek et al., 1994; Thomas & Diener, 1990).[2] But later research took a more cautious approach, since it became clear that contemporary and retrospective reports are fundamentally different (e.g., Ganzach & Yaor, 2019; Ready et al., 2007; Miron-Shatz et al., 2009; Neubauer et al., 2020), that they are associated with unique cognitive processes (Robinson & Clore, 2002a, 2002b), and that they represent two distinct constructs, retrospective affect and experienced affect, respectively (Kahneman et al., 1997).

In addition to shortening the time gap between experience and report, diary studies are characterised by prolonged tedious reporting. Such continuous reporting may adversely affect accuracy as a result of reactivity—the biasing effect of the sheer measurement (Barta et al., 2012).[3] For example, Lucas et al. (2019) found a low agreement between continuous measurement of affect and its measurement via the DRM (the Day Reconstruction Method, a non-continuous method), which they view as the result of reactivity in the continuous measurement. Sharpe and Gilbert (1998) found that test scores of instruments that measure negative mood states decrease with repeated administration of the tests. Knowles et al. (1996) found measurement-induced decrease in anxiety tests, and Merrilees et al. (2008) found that although self-reports indicated a temporal decrease in the perceptions of marital quality during a diary study, there were no apparent effects of time on couples' emotions regarding marital interactions.

Whereas all these studies showed that continuous reporting may lead to measurement biases in diary studies, the direction of these biases is not clear. Indeed, a main difficulty in studying the effect of continuous reporting on the accuracy of diary studies is that a gold standard against which reports can be compared is necessary. Such a standard is rarely available in diary studies of emotions, since people's 'true' emotions are not known. In the current study, we circumvent this problem by focusing on systematic changes during the reporting period in daily studies of emotions, studies in which subjects are asked to provide daily reports about their feelings. Since when the day of the week is controlled for, differences between daily experiences over a short period of time are a product of random daily circumstances in people's lives, any pattern of systematic changes in responses over time reflects a systematic reporting error. Thus, in the current study we examine the effect of time on changes in reports about positive and negative affect in a daily diary study, where time is operationalised as serial day, the position of the day in the sequence of the diary (to be distinguished from day of the week). Both experimental control (participants were randomly assigned to a starting weekday) and statistical control (weekday serves as a control variable in all the analyses), allow for this separation between the effects of serial day and weekday.

1.1 | Time and effort in diary studies

In our search for systematic changes in the accuracy of diary studies we rely on the literature documenting that the longer the questionnaire, the worse the response quality, a pattern that is attributed to a decrease in respondents' effort. This has been clearly documented in numerous studies showing that the quality of responses deteriorates with the length of the questionnaire (Iida et al., 2012; Deutskens et al., 2004; Heberlein & Baumgartner, 1978; Yammarino et al., 1991). Notably, such effects were also documented in diary studies. Reynolds et al. (2016) found that with time, agreement between mother and child diary reports about conflict weakens and the within-participant associations between child-reported negative mood and parent–child conflict strengthens; Van Berkel et al. (2019) found decrease in recall accuracy after 2 weeks of participation in a diary study; and most recently Eisele et al. (2020) found that compliance decreased with serial day (see table 6, p. 9). These results, however, were not replicated when careless responding and subjective burden were used as dependent variables.[4]

Following these previous studies, our basic assumption is that effort decreases with the duration of a diary study. This assumption, labelled the time-effort assumption, is our guide to studying accuracy in diary studies. We examine whether in a daily diary study serial day has a systematic effect on response that can be attributed to reduced effort. The fundamental idea behind this examination is that over time responses become more stereotypical, less a reflection of 'true' affect and more a reflection of an easy-to-generate default response. Note that this idea implicitly suggests that the 'gold standard' for response accuracy is the first day's responses, or at least early responses, and in this respect it is very different from the idea underlying Shrout et al. (2018), who argued for an 'elevation bias' in the early phase of a diary study (but see Arslan et al., 2021, for different results). In the general discussion section, we provide in depth comparison between Shrout et al.'s (2018) approach and ours.

In the following sub-sections, we describe three hypotheses derived from this general premise about the relationship between time and response stereotypicality.

[2] An exception is a study by Oishi & Sullivan (2006), who found that retrospective judgements of emotions better predict later relationship than daily ratings. However, this superior predictive validity of retrospective judgements is most likely due to the principle of *attitude specificity* (Heberlein & Black, 1976), which suggests that more accurate behavioural prediction is achieved when the predictor (e.g., global emotions) is measured on the same level of the behaviour (e.g., global behaviour).

[3] Reactivity is not limited to diary studies, and may occur whenever measurement results in changes in the people being measured. Although diary studies are most sensitive to reactivity, reactivity may occur also in studies that do not involve continuous reporting (e.g., French & Sutton, 2010).

[4] Eisele et al. (2020) also found that reports deteriorate with the length of the questionnaire, but not with the frequency of the questioning. These results also provide partial support for the idea that effort affects the quality of responses in diary studies.

1.2 | Time, effort and default responses in diary studies

We distinguish between two types of reports on affective experiences: reports on retrospective affect (e.g., how good/bad one felt during the day) and reports on short affective experiences, events that carry with them affective consequences (e.g., what were the good/bad experiences you had during the day). Although the formation of reports about retrospective affect is different from the formation of reports about short affective experiences, they both require effortful recall (Hertenstein & Campos, 2004) and attention (e.g., Hajcak & Olvet, 2008; Öhman et al., 2001). Due to fatigue, boredom and inattention, individuals who *repeatedly* respond to the very same questions may decrease their effort in recalling the information necessary to form these reports and rely more on default responses, which are less mentally demanding to generate.

In the current article we examine hypotheses regarding three types of default responses: (1) the default response associated with the *number* of recalled affective experiences, (2) the default response associated with the valence (i.e., positive-negative) of the recalled affect and (3) the default response associated with the complexity of reported affect. Thus, our hypotheses do not necessarily concern time-dependent changes in 'true' emotions, but rather time-dependent changes in *reported* emotions. In particular, there is no reason to believe that, in a daily diary study, serial day (as opposed to weekday) has a systematic effect on experienced emotions, so that any systematic pattern of changes in reported emotions is most likely due to the reporting process. As subjects dedicate less effort to accessing their daily emotions, they derive their reports not from their experienced emotions, but rather from a more simplified and less accurate representation of these emotions.

1.3 | The number of recalled experiences

We begin with an analysis of the effect of serial day on reports on the number of daily affective experiences. This task requires active and effortful search in memory for the relevant experiences: the more intensive the search, the larger the number of recalled—and reported—experiences. Thus, if serial day is negatively associated with the intensity of memory search, it should be negatively linked to the number of reported affective experiences (H1), both positive experiences (H1a) and negative experiences (H1b). We label these hypotheses the time-quantity hypotheses.

1.4 | The valence of recalled feelings

Our starting point in studying the relationship between serial day and the valence of the recalled feelings is the idea that, because it is more socially desirable to acknowledge positive than negative feelings—in fact, to view the word in positive terms—reporting about attitudes and feelings tends to be positively biased (Dodds et al., 2015; Mezulis et al., 2004; Oishi, 2002; Robinson et al., 1998; Scollon et al., 2009).

There is evidence that positivity bias increases with uncertainty, and that when the source of uncertainty is faded memory, this effect is moderated by effort dedicated to recall. First, there is strong evidence that when uncertainty about judgements increases, judgements tend to regress towards a positive default response (Denrell & Fang, 2010; Ganzach & Krantz, 1991). Among other things, this is reflected in memory-based judgements (as opposed to online judgements). In these judgements, as a result of faded memories, uncertainty about the precise features of the judged object results in reliance on a (positive) default response, which leads memory-based judgements to be more positive than online judgements (Ganzach & Mazursky, 1995; Skowronski, 2011; Walker et al., 2003). Furthermore, this effect is moderated by the effort dedicated to forming the judgements. When this effort decreases, the positivity bias increases (Mazursky & Ganzach, 1998). This effect is attributed to the influence of effort on the intensity of memory search: the lower the intensity of the search, the higher the uncertainty, and the stronger the reliance on the positive default response (ibid.). Thus, we suggest that when serial day increases and effort decreases, reports about feelings become more positive (H2). We label this hypothesis the time-positivity hypothesis.[5]

Our indicators for testing the time-positivity hypothesis are the reported daily frequencies of positive and negative feelings. Accurate reporting of these frequencies requires active search for, and evaluation of, positive and negative feelings that occurred during the day. This, too, requires investing effort in memory search. With time, as the effort dedicated to this search decreases, we expect that reports will be based less on actual feelings and more on socially desirable, positively biased, default values, increasing the reported frequency of positive feelings with serial day (H2a) and decreasing the reported frequency of negative feelings (H2b).

1.5 | Time, effort and emotional complexity

In addition to studying time-dependent changes in the number of recalled experiences and in the valence of retrospective affect we also study time-dependent changes in affective complexity. Our hypothesis here is that the default response is a low-complexity response, and therefore that emotional-complexity will show a time-dependent decrease—a negative relationship between time and complexity (H3). We label this hypothesis the time-complexity hypothesis. Note that this hypothesis is consistent with the more general notion of a time-stereotypicality relationship discussed above, since in our case complexity and stereotypicality are reversely linked.

[5] We note that there are also indications for a time-positivity effect in diary studies. Johnson et al. (2009) found that among psychiatric patients, reports about alcohol use decreased with time and reports about personal hygiene increased with time. They viewed these results as stemming from 'these behaviors' potential to be influenced by social desirability effects' (p. 51). The results of Ouweneel et al.'s (2012) study discussed below also show strong evidence for a time-positivity effect.

The time-complexity hypothesis is derived from the idea that the fewer the mental resources devoted to the reporting about one's emotions, the lower the emotional complexity revealed by these reports. This relationship has not yet been observed in diary studies, but has been observed in a number of studies in which mental resources were varied and emotions were measured. Thus, for example, Conway (2000) argues that 'participants with more attentional resources evidence … more complex representations of others' emotions' (p. 5), and Thompson et al. (2021) write that 'only when motivated to apply mental resources do people exhibit emotional complexity' (p. 8).

Relevant here is also the literature about the effect of stress on emotions, which suggests that lower levels of stress are associated with a higher tendency for emotional complexity (see for example Davis et al., 2004; Ong & Bergeman, 2004; Ong et al., 2006; Carstensen et al., 2000; but see also Scott et al., 2014). Thus, for example, in their summary of the literature on emotional complexity, Lindquist and Barrett (2008), using what they call dialecticism as an indicator for complexity, state that 'Greater dialecticism … is associated with greater resilience and lower stress'. This relationship between stress and complexity is consistent with our time-complexity hypothesis because less effort and more stress are analogous in that they both involve lack of availability of mental resources for sensing complex emotions. Thus, if diary studies involve time-dependent decrease in effort, the effect of time on emotional complexity should be similar to the effect of stress,

There is much debate about the measurement of emotional complexity (Grossmann et al., 2016). In the current article we rely on two simple measures of emotional complexity, one for between-dimensions complexity and another for within-dimension complexity. Between-dimension complexity is measured by the correlation between positive and negative affect (Larsen et al., 2017). A highly negative correlation between PA and NA is indicative of low differentiation between the two dimensions, suggesting that emotions are unidimensional, whereas a moderately negative or zero correlation is indicative of a high complexity, indicating that positive and negative emotions can coexist simultaneously.

Whereas between-dimensions complexity is associated with differentiation between the two affective dimensions, within-dimension complexity is associated with differentiation between emotions within each of the two dimensions as a result of discrimination between the various shades of feelings that characterised each of the two dimensions. It is measured by the within-dimension variance of the specific feelings of the dimension (Grühn et al., 2013).

In sum, our examination of the time-complexitly hypothesis involves both the examination of the between-dimensions correlation and the within-dimension variance. We hypothesise that the between-dimensions correlation strengthens (becomes more negative) with serial day (H3a) and that the within-dimension variance decreases (H3b).[6]

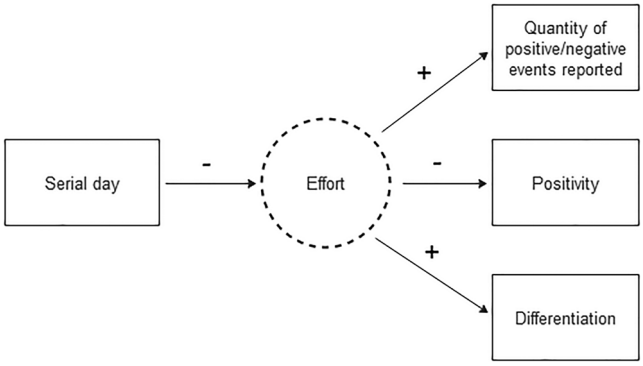

FIGURE 1 The study's theoretical model

1.6 | Summary

Figure 1 summarises our model. In this model serial day affects effort, which in turn leads to more stereotypical responses characterised by a decrease in the number of recalled affective experiences, an increase in the positivity of the reported feelings and a decrease in emotional complexity. In the figure effort is represented by a broken circle since it is assumed to mediate the effect of serial day on our dependent measures, but is not empirically measured in our data.

2 | METHOD

This study is based on data from MIDUS (Midlife in the United States), a national study of adult Americans. In order to investigate patterns of change in participants' affective reports, we analyse MIDUS' Project 2 (https://www.icpsr.umich.edu/icpsrweb/NACDA/studies/26841/), which is a daily stress study with 2022 respondents conducted in the years 2004–2006. Participants' age ranges from 33 to 84 with an average of 56.2. Females were 57.2%. The information was obtained using a daily telephone interview. Monetary compensation of $60 was given for participation in the diary survey + a base survey. Participants were randomly assigned to a starting weekday, but the starting weekday itself was not random—there were more interviews that started at the beginning of the week than in the weekends (see the Supplementary

[6] Two points are worth noting regarding H3b. First, in extreme cases, we may find a 'straight-lining' effect (see Cole, McCormick, & Gonyea, 2012; Schonla and Toepoel, 2015), where all feelings receive the very same response, indicating no differentiation between various feelings.Second, in evaluating H3b it is important to note that temporal changes in the within-dimension variance are influenced not only by temporal changes in emotional complexity but also by temporal changes in the random error variance. However, a temporal decrease in the within-dimension variance can be explained only by a decrease in emotional complexity, and not by a decrease in error variance. To see why, consider the relationship between subject i response at time t to item j that measures the affect of this subject at time t on one of our two dimensions (represented by D):$J_{jit} = D_{it} + E_{Djit} + E_{Rjit}$ (1)where J_{jit} is the judgement (response), of item j by subject i at time t; D_{it} is the subject i dimensional affect – his/her true affect at time t on this dimension; E_{Djit} represents the differentiation, the extent to which the affect associated with item j is different from the dimensional affect; E_{Rjit} is a random error associated with the response to item j at time t.Under our assumption that true affect does not change over time, this suggests that$VAR(J_{it}) = VAR(E_{Dit}) + VAR(E_{Rit})$ (2) Within our theoretical framework, the effect of time on $VAR(E_{Rit})$is the opposite of its effect on $VAR(E_{Dit})$. While time-induced decrease in effort leads to decrease in $VAR(E_{Dit})$, it leads to an increase in $VAR(E_{Rit})$, because effort is negatively related to measurement error (Kreuter, Müller, & Trappmann, 2010). Therefore, a decrease in $VAR(J_{it})$ can be attributed only to a decrease in $VAR(E_{Dit})$.

Material S1 for details).[7] Therefore, we controlled in our analysis for the day of the week effect.

2.1 | Procedure

Participants in the daily study completed a short telephone interview about their daily experiences. The first interviews lasted approximately 15–20 min, while the other interviews lasted 10–15 min. Our study uses the first 7 consecutive days, since day eight used a different set of questions. A total of 14,154 daily interviews were conducted for the purpose of our study.

2.2 | Measures

2.2.1 | Positive and negative daily affect

The MIDUS measures positive and negative affect based on items developed by Watson et al. (1988) for the PANAS scale. The PANAS scale contains 13 positive affect (PA) terms—good spirits, cheerful, extremely happy, calm and peaceful, satisfied, full of life, close to others, feeling like you belong, enthusiastic, attentive, proud, active and confident ($\alpha = .91$). The negative affect (NA) terms include restless or fidgety, nervous, worthless, so sad nothing cheers you up, everything is an effort, hopeless, lonely, afraid, jittery, irritable, ashamed, upset, angry, and frustrated, 14 items in total ($\alpha = .86$). The question asks participants to indicate how much time during the day they felt this way, ranging from none of the time (0) to all of the time (4).

2.2.2 | Number of negative events

This question presents subjects with a list of six negative events and asks them to indicate whether they have experienced such events during the day. The negative events are: an interpersonal conflict, a situation that could end in an argument but they decided to avoid, a problem at work, a problem at home, something bad happening to a close other, perceived discrimination, and any other stressful experiences not covered by the previous categories. The number of negative events is the sum of the negative events indicated by the subject. This scale, as well as the positive events scale below, was adopted from Almeida et al. (2002).

2.2.3 | Number of positive events

This question presents a list of four positive events and asks subjects to indicate whether they have experienced these events during the day. The positive events are: a positive interaction with someone, a positive event at work, a positive event at home, something good happening to a close other, and any other pleasant events not covered by the previous categories

2.2.4 | Serial day (time)

Throughout the year, each subject was expected to participate in an 8-day sequence of interviews. Notably, there is a certain decrease in the number of participants over time. At Day 1 all 2022 participants complied with the interview. The lowest number of participants was on Day 7, with 1802 responses in total (89.1% of the full sample). Since we were interested in the effect of time on the reported affective experiences, we used the interview day as a continuous variable. Importantly, there is no link between interview day and weekday, since each interview started at a different calendric day. Hence, in the analysis we were able to control for the effects of the calendric day, which has received attention in previous studies (Bolger et al., 2003; Sheldon et al., 1996). As explained earlier, since some modifications were made in the affective measurements at Day 8, only the first 7 first days were included in the analysis in order to ensure consistency. However, when Day 8 was included in the analyses, no meaningful differences were found.

2.3 | Analyses

For each of the dependent variables we conducted a multi-level analysis (using SAS PROC MIXED; the code of the analyses are available in the Supplementary Material S2) with random intercept and random slope for serial day in which each respondent had up to seven measures for each of the study variables (a log transformation was performed on the number of positive and negative events).[8] In all the analyses we controlled for the day of the week, which is known to have a substantial effect on daily feelings (Stone et al., 2012). In our initial analyses we included age, gender and income as covariates and found only a few significant weak effects. In particular, older individuals reported on average higher positive affect and lower negative affect, and also demonstrated lower variability in the negative affect responses. However, in order to decrease the amount of missing cases (about 10% did not answer the income question) and since they are outside the scope of our article, we did not include these covariates in the final analyses. In addition, in order to make sure that attrition did not influence our results, we also estimated models only for subjects that stayed in the study until the 7th day. The results of these two additional analyses were very similar to the results of the full sample, with no changes in the significant effects. The results of these analyses are reported in the Supplementary Material S3.

An autoregressive covariance structure was used for the data analysis. This structure does not assume variance homogeneity between

[7] The reason for that was that once notified about the time of the first interview participants could reschedule it.

[8] We use a logarithmic transformation here because this is the recommended method to model count data (Coxe et al., 2009). It captures the idea that changes from 1 to 2 are more meaningful than changes from 10 to 11.

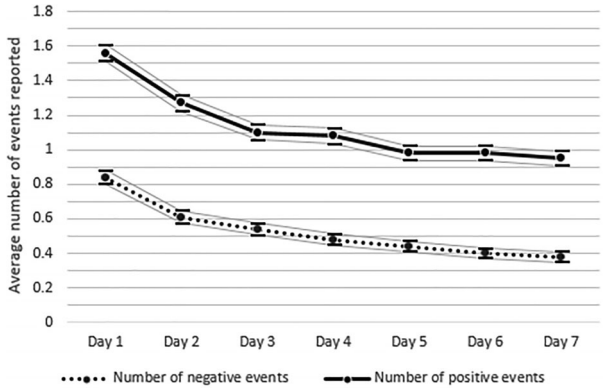

FIGURE 2 The number of reported positive and negative affective experiences by serial day. Confidence intervals are ±2SE above and below the mean

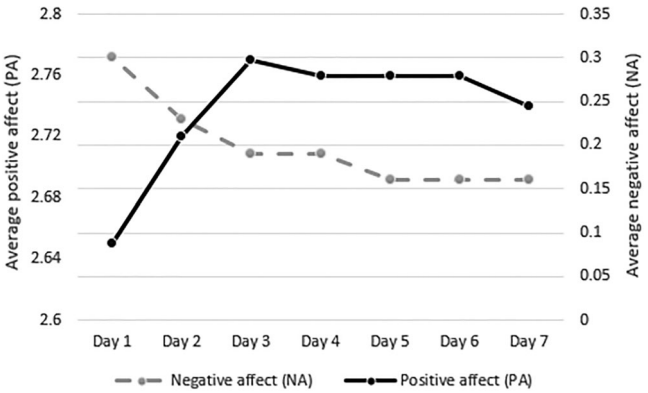

FIGURE 3 The average PA and NA by serial day. The Y-axis scale for PA is on the left side of the figure, and for NA it is on the right side of the figure. The standard deviations of average PA are 0.59, 0.60, 0.62, 0.61, 0.61, 0.61 and 0.61 for days 1–7, respectively. The standard deviations of average NA are 0.0089, 0.0078, 0.0069, 0.0070, 0.0064, 0.0062 and 0.0067 for days 1–7, respectively

separate measurements. This is important, since we argue here that quality of responses will vary between the days of the study. This covariance structure was also compared to other commonly used structures and demonstrated the best fit for our data (Akaike's Information Criterion, AIC).

3 | RESULTS

3.1 | The number of recalled experiences

Figure 2 plots the number of reported positive and negative experiences across participants for each serial day. It is clear from this figure that, consistent with H1, the number of reported experiences, both positive and negative, decreases with serial day, a decrease of about 40% for positive events and 55% for negative events. Indeed, when we regressed the logged number of reported experiences on serial day, controlling for weekday, the coefficient of serial day was significant for both positive ($b = -0.352$, $p < .001$) and negative experiences ($b = -0.376$, $p < .001$). Table 1 provides the details of the analysis.

Figure 2 also suggests that most of the decline in the number of reported affective experiences occurred in the early days of the study. Thus, when we added the squared term of serial day to the regression, we found that the coefficient of this term was significantly positive, $b = 0.067$ and 0.068, both p's $< .001$, for the positive and negative events, respectively, suggesting a curvilinear decline in reported events (these analyses are reported in the Supplementary Material S3). This result is consistent with the idea that most of the decline in effort occurs in the early stages of the study.

3.2 | Positivity

Figure 3 plots the average reported positive and negative daily affect across participants for each serial day. It is clear from this figure that, positive affect increases with time while negative affect decreases. In a regression analysis, the coefficient of serial day is significantly positive in predicting positive affect ($b = 0.007$, $p = .004$) and signif-

TABLE 1 The effect of serial day on the log number of positive and negative reported affective experiences

	Number of positive experiences reported		Number of negative experiences reported	
	B	SE	b	SE
Serial day	**−0.376*****	**0.017**	**−0.352*****	**0.018**
Monday	−0.391**	0.118	0.826***	0.135
Tuesday	−0.474***	0.121	1.049***	0.135
Wednesday	−0.511***	0.122	0.962***	0.137
Thursday	−0.571***	0.122	0.643***	0.136
Friday	−0.643***	0.123	0.492***	0.136
Saturday	−0.397**	0.120	0.381**	0.135
AIC	75882.294		73157.966	

Abbreviation: AIC, Akaike's Information Criterion.
$^{**}p < .01$, $^{***}p < .001$; weekdays are compared to Sunday.

TABLE 2 The effect of serial day on daily PA and NA

	Positive affect (PA) reported		Negative affect (NA) reported	
	B	SE	b	SE
Serial day	**0.007****	**0.002**	**−0.017*****	**0.001**
Monday	−0.059**	0.012	0.054***	0.007
Tuesday	−0.057***	0.013	0.043***	0.007
Wednesday	−0.055***	0.013	0.043***	0.007
Thursday	−0.050***	0.013	0.038***	0.007
Friday	−0.027*	0.013	0.029***	0.007
Saturday	0.000	0.012	0.007	0.007
AIC	18416.775		1707.084	

Abbreviation: AIC, Akaike's Information Criterion.
*$p < .05$, **$p < .01$, ***$p < .001$; Weekdays are compared to Sunday.

icantly negative in predicting negative affect ($b = -0.017$, $p < .001$) (Table 2 presents the details of this analysis). These findings are consistent with H2a and H2b, respectively, and together they support the time-positivity hypothesis (H2).

Two points are worth noting regarding these results. First, like the number of reported affective experiences, the larger part of the increase in positive and the decrease in negative affect occurs in the early phases of the study, and any change after Day 3 is minor to non-existent. In an additional analysis, we added the squared serial day term to the regression models. Consistent with the idea that the major effects of serial-day occur in the early phases of the study, the square term was significantly negative ($b = -0.007$, $p = .002$) for PA and significantly positive ($b = 0.005$, $p < .001$) for NA (the analyses are reported in the Supplementary Material S3).

Second, as evidenced in Figure 3, changes in affect are much more pronounced for negative affect (a decrease of about 50%) than for positive affect (an increase of about 3%). One possible explanation for this difference is that in evaluating their daily PA and NA, people attempt to actively recall the number of positive and negative events that occurred during the day. For the evaluation of daily PA, the influence of this active recall is in the opposite direction of the influence of positivity (the former decreases PA while the latter increases PA), whereas for daily NA this active recall is in the opposite direction of positivity (they both decrease NA). In the Supplementary Material S4, we present path models that estimate the joint influence of positivity and active recall of events on daily PA and NA.

3.3 | Emotional complexity

3.3.1 | Between-dimensions complexity

Figure 4 shows the relationship between PA and NA as a function of serial day by plotting the regression coefficients obtained from regressing PA on NA across subjects (controlling for weekday) for each of the 7 days of the study. It is clear from this figure that the negative relationship between PA and NA becomes stronger with serial day (an increase of about 41% in the slope), consistent with the idea that emotional complexity decreases and unidimensionality increases.

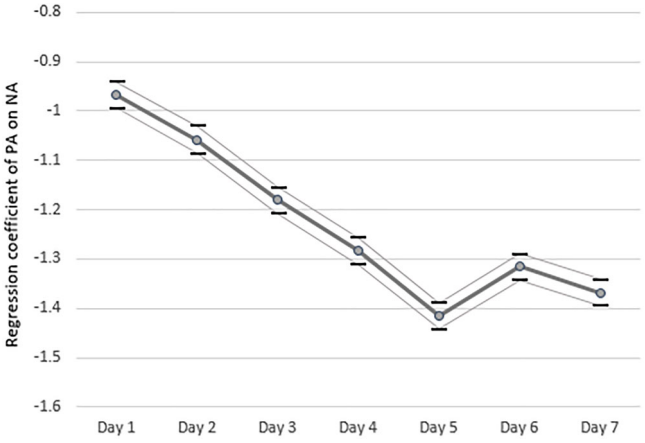

FIGURE 4 Daily regression coefficients of PA on NA. Confidence intervals are ±2SE above and below the mean

To test the hypothesis that the between-dimensions complexity decreases with serial day we performed a multi-level regression of PA on NA, the serial day and their interaction term, controlling for the weekday effect, where NA and serial day were both level-1 variables nested within subjects (see Table 3 for the details of this regression). As expected, the coefficient of NA was significantly negative ($b = -0.947$, $p < .001$) and the Serial Day × NA interaction term was significantly negative as well ($b = -0.042$, $p < .001$).[9]

3.3.2 | Within-dimension complexity

Figure 5 plots the average within-subject variance of PA and NA across subjects for each serial day. It is clear from this figure that both vari-

[9] In a recent study that was aimed at increasing emotional differentiation of individuals with major depressive disorder, Widdershoven et al. (2019) found that emotional differentiation did indeed increase. However, in this study subjects received treatment (e.g., through feedback provided by clinicians) that was aimed to increase this differentiation.

TABLE 3 The effect of serial day on the relationship between PA and NA

	Positive affect (PA)	
	b	SE
Serial day	**−0.006****	0.002
Negative affect (NA)	**−0.949*****	0.024
Serial day X Negative affect	**−0.043*****	0.007
Monday	−0.017	0.011
Tuesday	−0.026*	0.011
Wednesday	−0.017	0.011
Thursday	−0.011	0.011
Friday	−0.002	0.011
Saturday	0.009	0.011
AIC	15460.942	

Abbreviation: AIC, Akaike's Information Criterion.
*$p < .05$, **$p < .01$, ***$p < .001$; Weekdays are compared to Sunday.

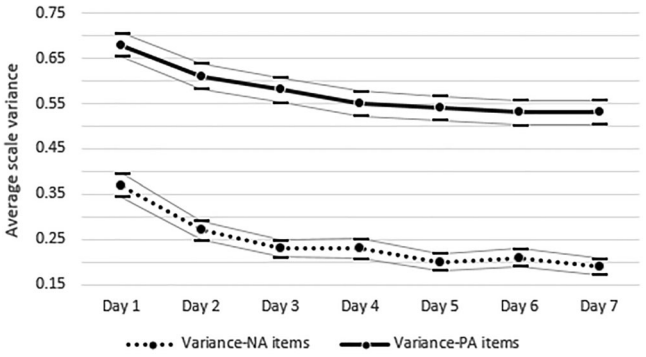

FIGURE 5 Average positive affect (PA) and negative affect (NA) scale variance as a function of serial day. Confidence intervals are ±2SE above and below the mean

ances decrease with serial day (about 25% decrease in the variance of PA and about 47% decrease in the variance of NA). In regression analyses, the coefficient for serial day is significantly negative both for PA scale variance ($b = -0.020$, $p < .001$) and for NA scale variance ($b = -0.018$, $p < .001$), fully supporting H3b. Table 4 presents the details of this analysis.

Like previous effects of serial day, most of the decrease in the variances of both PA and NA occurred in the early days of the diary study. Indeed, when we included the square term for serial day in the analyses, it was positive for both PA and NA variances (b's = 0.006 and 0.005, respectively, both p's < .001), suggesting diminished decrease in variability over time (see the Supplementary Material S3 for details).

4 | DISCUSSION

Most, if not all, research in emotions has not distinguished between 'true' emotions and reports about these emotions. True, researchers took into account the fact that reports about emotions are error ridden and that the reliability of these reports should be considered (Thomas & Diener, 1990). However, in the current paper we examine the validity, rather than the reliability, of reports about emotions by assessing time-dependent biases in these reports. Russell (2003, 2017) did raise questions regarding the validity of reported emotions, and called for considering the effects of psychological construction in analysing people's judgements of their affect. However, he did not offer an empirical method to study this validity, most likely because as long as 'true' affect (core affect in Russell's terminology) is unknown, it is not clear how the validity of reports about this affect can be examined. The current study approaches this issue indirectly. It does not attempt to assess the correspondence between true emotions and reported emotions, but it does try to identify biases in these reports and assess the direction of these biases. This assessment is based on the idea that, in a diary study, serial day (as opposed to weekday) does not have a systematic effect on experienced emotions, so that any systematic pattern of changes in reported emotions must be due to the reporting process.

Although our approach is somewhat similar to that of Russell (2003, 2017), it is worthwhile noting that while Russell suggested that 'Feeling bad is one thing, judging something to be bad is another', we suggest that 'Feeling bad is one thing, reporting a bad feeling is another'. The difference is that Russell (2003, 2017) distinguishes between the experience of contemporary affect on the one hand and the judgement of the affective qualities of the situation as pleasant or unpleasant on the other hand. We distinguish between the delayed evaluation of affect and the report about this affect. However, to the extent that one's own affect is the judged stimuli, our view coincides with Russell's.

Caution should be exercised in generalisations based on the current results. First, the current work examines the effect of time on delayed evaluation of extended (daily) affective experiences. The picture may be different when studying the effect of time on the evaluation of contemporary affective experiences, since the evaluation of contemporary affect may be less sensitive to constructionist biases: delayed affective reports involve retrieval of affective information from memory, whereas contemporary affective reports are primarily based on direct access to one's feelings. Second, our frequency measure of affect may be more sensitive to cognitive construction of emotions than other measures, particularly intensity measures of affect, since the former calls for cognitive construction, whereas the latter involves direct access of experienced affect. This argument is consistent with the finding that a frequency format is characterised by stronger bidimensionality (Watson, 1988) and with Russell's view that bidimensionality characterises cognitive evaluations of affect more than it characterises core affect.[10] Finally, the interviews we analysed involved direct interactions (via phone calls) between interviewers and interviewees. The effect of social desirability, and therefore

[10] On the other hand, there are reasons to believe that we should not expect considerable method-dependent differences in the effect of effort on reported affect; that is, that diary studies based on contemporary intensity measures of momentary experiences will yield similar results. First, even accessing current feelings is a cognitive process that requires effort (Bradley, 2014; Ellis, Thomas & Rodriguez, 1984); and second, by and large, the pattern of delayed affective reports tends to be similar to the pattern of contemporary affective reports and so does the pattern of reports elicited by frequency format and reports elicited by intensity format (Watson, 1988; Ganzach & Yaor, 2019).

TABLE 4 The effect of serial day on the within-dimension variabilities

	Positive affect (PA) scale variance		Negative affect (NA) scale variance	
	b	SE	b	SE
Serial day	**−0.020*****	**0.002**	**−0.018*****	**0.001**
Monday	0.049***	0.012	0.070***	0.007
Tuesday	0.026*	0.013	0.049***	0.007
Wednesday	0.022	0.013	0.055***	0.007
Thursday	0.026*	0.013	0.048***	0.007
Friday	0.018	0.012	0.032**	0.007
Saturday	0.007	0.012	0.014	0.007
AIC	16580.443		10987.148	

Abbreviation: AIC, Akaike's Information Criterion.
$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$; Weekdays are compared to Sunday.

the effect of time, in other modes of interviews (e.g., anonymous filling of questionnaires vs. face-to-face interviews) may be quite different.

It is interesting to compare here the current study to a recent study by De Vuyst et al. (2019). Although these authors were not interested in the effect of serial day on affective reports,[11] their data are still relevant to our work since they reveal no effect of time on the level of reported emotions (see Figure >1, p. 8). The difference between these results and ours may be due to the small sample size in the Vuyst et al. study (90 participants per condition) compared to the much larger sample size of about 2000 in our study. However, it is also possible that the difference has to do with the fact that De Vuyst et al. (2019) was an experience sampling study in which subjects were contacted quite a few times during each day and asked to report about their immediate momentary feelings. Reports about immediate feelings may be less prone to memory biases than retrospective reports about daily feelings such as the ones analysed in the current work.[12] Thus, there is a need for more work to examine whether the effects observed in the current study generalise to other types of experience sampling methods.

We turn now to a comparison between our study and Shrout et al. (2018), who also found a clear effect of time on emotions. First, despite the similarities between the results of Shrout et al. (2018) and ours with regard to negative affect (they also found a temporal decrease in NA),[13] there are also important differences. First, unlike our study, most of Shrout et al.'s (2018) data were collected from subjects awaiting important examinations; thus the effect of serial day in their data was confounded with the effect of proximity to the exam.[14] Second, our results concerning decrease in emotional complexity are unrelated to the initial elevation effect suggested by Shrout et al. (2018), but are consistent with the idea that decreased emotional complexity is a default response.

Finally, a most important difference between our results and Shrout et al. (2018) is that they observed a decreased PA with time (which they explain by an initial elevation bias), whereas we observed an increase in PA over time (which we explain by a late social desirability bias). Thus, the pattern of temporal changes in PA can be viewed as a critical test of our approach versus Shrout et al.'s approach. In this respect it is interesting to refer the reader to an earlier article, Ouweneel et al. (2012), who present results of a diary study about the within-person relationship between *positive* emotions and work engagement and report that they unexpectedly found a strong effect of serial day which they could not explain, in which participants' positive feelings about their work became more positive over time: 'in our analyses we only controlled for the linear effect of time. This is not common in diary studies in the field of organizational psychology. However, results show that it is advisable to include time in multilevel analyses' (p. 1146). Clearly, these results are consistent with the pattern of increased positivity over time that we find in our data, but not with the pattern of decreased positivity found by Shrout et al.

4.1 | Limitations, alternative explanations and future research

A significant limitation of the current study is that our mediator variable−−effort—is not measured. While the data presented in the article suggest clear and strong relationships between our independent variable (time, or serial day) and our dependent variables (number of recalled events and intensity and complexity of emotions), there is no empirical evidence in our data that supports the idea that effort (or related constructs) mediates these relationships. Obviously, stronger support for our model could be provided if effort were to be measured, or even manipulated. The preferred measurement of effort would be

[11] The study showed that continuously reporting on either positive emotions, negative emotions or non-emotional internal state has no impact, either on the global level of emotional experience or on the level of depressive symptoms.
[12] In this respect our discussion of De Vuyst et al (2019) is relevant also to that of Eisele et al. (2020) discussed in Section 1. The two studies are similar in that they were both low-powered experience sampling studies with frequent daily contacts.
[13] This effect of time on negative affect appears to be robust—it was also found in Knowles et al. (1996) and Sharpe & Gilbert (1998), which were reviewed in Section 1.
[14] The initial elevation and later decline may involve a therapeutic effect, which was not convincingly ruled out by Shrout et al. (2018)

unobstructed measures such as questionnaire filling time, although such measures may by themselves be problematic (see Van Berkel et al., 2019). An even better support for the role of effort on the effect of time on reporting could be provided by experimental manipulation of effort or related constructs, for example by examining the effect of compensation on the relationship between serial day and affective reports.

As our literature review indicates, decreased effort is but one type of reactivity to continuous reporting, and alternative explanations for the changes in affective reports cannot be completely ruled out. For example, continuous reporting may increase reflections on one's feeling, which in turn leads to increase [decrease] in positive [negative] affect. It may be even possible that by reporting about negative events, people realise what are the circumstances that lead to such events and learn to avoid them. However, we believe that the idea that over time responses become more stereotypical, reflecting an easy-to-generate default response, provides the most comprehensive account for the data. For example, the positive effect of reflecting on one's feeling may explain the increase in positive affect and decrease in negative affect, but cannot explain the decrease in reporting about both positive and negative events; or learning to avoid negative events may explain the decrease in reports about negative events, but is inconsistent with the decrease in reports about positive affect.

The general notion of time–stereotypicality relationship is also consistent with the decrease in emotional complexity, both between dimensions and within dimensions. We suggest that this decreased complexity is associated with declining effort leading to simplified affective reports. Note, however, that it is not clear what are the heuristics that underlie this simplified reporting. One possible explanation is that when effort decreases, affective reports are derived more from dispositions (e.g., trait PA/NA) and less from actual affective experiences (e.g., the number and strength of positive/negative events). This proposition is an interesting subject for future research.

Finally, it is important to keep in mind that, by and large, reactivity is influenced by a number of processes that simultaneously shape the effect of continuous reporting, of which decreased effort is just one. Some of these processes may improve the accuracy of reporting, some of them may impair it, and some of them may actually change the phenomena which we measure. Future research could examine how different diary survey designs are associated with different systematic changes in affective reporting and which mechanisms underlie these changes.

5 | CONCLUSIONS

In the article we demonstrate three effects of serial day on affect: decrease in the remembered number of affective experiences, increase in positivity and decrease in emotional complexity. All these effects are associated with subjects' responses becoming more and more stereotypical, increasingly based on a default response. It is clear, however, that although the psychological processes that lead to these default responses are similar in that they are all associated with decreased effort, they involve different default responses and different mechanisms by which reduced effort leads to reliance on these default responses.

ETHICS STATEMENT

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with human participants or animals performed by any of the authors.

OPEN PRACTICE STATEMENT

Data are available at https://www.icpsr.umich.edu/web/NACDA/studies/4652

CONFLICTS OF INTEREST

The authors declare no conflict of interests.

ORCID

Yoav Ganzach https://orcid.org/0000-0002-9821-9785

REFERENCES

Al Baghal, T., Belli, R. F., Phillips, A. L., & Ruther, N. (2014). What are you doing now? Activity-level responses and recall failures in the American time use survey. *Journal of Survey Statistics and Methodology*, *2*(4), 519–537. https://doi.org/10.1093/jssam/smu020

Almeida, D. M., Wethington, E., & Kessler, R. C. (2002). The daily inventory of stressful events: An interview-based approach for measuring daily stressors. *Assessment*, *9*(1), 41–55.

Arslan, R. C., Reitz, A. K., Driebe, J. C., Gerlach, T. M., & Penke, L. (2021). Routinely randomize potential sources of measurement reactivity to estimate and adjust for biases in subjective reports. *Psychological Methods*, *26*(2), 175–185. https://doi.org/10.1037/met0000294

Barta, W. D., Tennen, H., & Litt, M. D. (2012). Measurement reactivity in diary research. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 108–123). Guilford Press.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Bolger, N., & Laurenceau, J. P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. Guilford Press.

Bradley, M. M. (2014). Emotional memory: A dimensional analysis. In van Goozen, S. H. M., Van de Poll, N. E., Sergeant, S. A., van Goozen, S. H. M. & Sergeant, J. A. (Eds), *Emotions: Essays on emotion theory* (pp. 111–148). Psychology Press.

Brenner, P. S. (2017). Narratives of response error from cognitive interviews of survey questions about normative behavior. *Sociological Methods & Research*, *46*(3), 540–564.

Cannell, C. F., Miller, P. V., & Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology*, *12*, 389–437. https://doi.org/10.2307/270748

Carstensen, L. L., Pasupathi, M., Mayr, U., & Nesselroade, J. R. (2000). Emotional experience in everyday life across the adult life span. *Journal of Personality and Social Psychology*, *79*(4), 644–655. https://doi.org/10.1037/0022-3514.79.4.644

Cole, J. S., McCormick, A. C. & Gonyea, R. M. (2012). Respondent use of straight-lining as a response strategy in education survey research: Prevalence and implications. In Annual Meeting of the American Educational Research Association (AERA).

Conway, M. (2000). Individual differences in attentional resources and social cognition: Elaboration and complexity in representations of others and self. *Generative mental processes and cognitive resources* (pp. 5–38). Dordrecht: Springer.

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, *91*(2), 121–136. https://doi.org/10.1080/00223890802634175

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Reply to Weinshall-Margel and Shapard: Extraneous factors in judicial decisions persist. *Proceedings of the National Academy of Sciences*, *108*(42), E834–E834. https://doi.org/10.1073/pnas.1112190108

Davis, M. C., Zautra, A. J., & Smith, B. W. (2004). Chronic pain, stress, and the dynamics of affective differentiation. *Journal of Personality*, *72*, 133–1159. https://doi.org/10.1111/j.1467-6494.2004.00293.x

De Vuyst, H. J., Dejonckheere, E., Van der Gucht, K., & Kuppens, P. (2019). Does repeatedly reporting positive or negative emotions in daily life have an impact on the level of emotional experiences and depressive symptoms over time? *PLoS ONE*, *14*(6), e0219121. https://doi.org/10.1371/journal.pone.0219121

Denrell, J., & Fang, C. (2010). Predicting the next big thing: Success as a signal of poor judgment. *Management Science*, *56*(10), 1653–1667. https://doi.org/10.1287/mnsc.1100.1220

Deutskens, E., De Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing letters*, *15*(1), 21–36. https://doi.org/10.1023/B:MARK.0000021968.86465.00

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., Mitchell, L., Harris, K. D., Kloumann, I. M., Bagrow, J. P., Megerdoomian, K., McMahon, M. T., Tivnan, B. F., & Danforth, C. M. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, *112*(8), 2389–2394. https://doi.org/10.1073/pnas.1411678112

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2020). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment*, *27*, 1–16.

Ellis, H. C., Thomas, R. L., & Rodriguez, I. A. (1984). Emotional mood states and memory: Elaborative encoding, semantics processing, and cognitive effort. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 470–482.

French, D. P., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *British Journal of Health Psychology*, *15*(3), 453–468. https://doi.org/10.1348/135910710X492341

Ganzach, Y., & Krantz, D. H. (1991). The psychology of moderate prediction: II. Leniency and uncertainty. *Organizational Behavior and Human Decision Processes*, *48*(2), 169–192. https://doi.org/10.1016/0749-5978(91)90011-H

Ganzach, Y., & Mazursky, D. (1995). Time dependent biases in consumer multi-attribute judgment. *Journal of Economic Psychology*, *16*(2), 331–349. https://doi.org/10.1016/0167-4870(95)00005-9

Ganzach, Y., & Yaor, E. (2019). The retrospective evaluation of positive and negative affect. *Personality and Social Psychology Bulletin*, *45*(1), 93–104 https://doi.org/10.1177/0146167218780695

Grossmann, I., Huynh, A. C., & Ellsworth, P. C. (2016). Emotional complexity: Clarifying definitions and cultural correlates. *Journal of Personality and Social Psychology*, *111*(6), 895–914. https://doi.org/10.1037/pspp0000084

Grühn, D., Lumley, M. A., Diehl, M., & Labouvie-Vief, G. (2013). Time-based indicators of emotional complexity: Interrelations and correlates. *Emotion*, *13*(2), 226–237. https://doi.org/10.1037/a0030363

Gunthert, K. C., & Wenze, S. J. (2012). Daily diary methods. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 144–159). Guilford Press.

Hajcak, G., & Olvet, D. M. (2008). The persistence of attention to emotion: Brain potentials during and after picture presentation. *Emotion (Washington, D.C.)*, *8*(2), 250. https://doi.org/10.1037/1528-3542.8.2.250

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, *43*, 447–462. https://doi.org/10.2307/2094771

Heberlein, T. A., & Black, J. S. (1976). Attitudinal specificity and the prediction of behavior in a field setting. *Journal of Personality and Social Psychology*, *33*(4), 474–479. https://doi.org/10.1037/0022-3514.33.4.474

Hertenstein, M. J., & Campos, J. J. (2004). The retention effects of an adult's emotional displays on infant behavior. *Child Development*, *75*(2), 595–613. https://doi.org/10.1111/j.1467-8624.2004.00695.x

Horstmann, K. T. (2021). Experience sampling and daily diary studies: Basic concepts, designs, and challenges. *The handbook of personality dynamics and processes* (pp. 791–814). Academic Press.

Iida, M., Shrout, P. E., Laurenceau, J.-P., & Bolger, N. (2012). Using diary methods in psychological research. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 277–305). American Psychological Association.

Johnson, E. I., Grondin, O., Barrault, M., Faytout, M., Helbig, S., Husky, M., Granholm, E. L., Loh, C., Nadeau, L., Wittchen, H. U., & Swendsen, J. (2009). Computerized ambulatory monitoring in psychiatry: A multi-site collaborative study of acceptability, compliance, and reactivity. *International Journal of Methods in Psychiatric Research*, *18*(1), 48–57. https://doi.org/10.1002/mpr.276

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In Gilovich, T., Griffin, D., & Kahneman, D. (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge University Press.

Kahneman, D., Wakker, P. P., & Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*, *112*(2), 375–406. https://doi.org/10.1162/003355397555235

Knowles, E. S., Coker, M. C., Scott, R. A., Cook, D. A., & Neville, J. W. (1996). Measurement-induced improvement in anxiety: Mean shifts with repeated assessment. *Journal of Personality and Social Psychology*, *71*(2), 352–367. https://doi.org/10.1037/0022-3514.71.2.352

Kreuter, F., Müller, G., & Trappmann, M. (2010). Nonresponse and measurement error in employment research: Making use of administrative data. *Public Opinion Quarterly*, *74*(5), 880–906. https://doi.org/10.1093/poq/nfq060

Larsen, J. T., Hershfield, H., Stastny, B. J., & Hester, N. (2017). On the relationship between positive and negative affect: Their correlation and their co-occurrence. *Emotion*, *17*(2),323–326. https://doi.org/10.1037/emo0000231

Lindquist, K. A., & Barrett, L. F. (2008). Emotional complexity. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (pp. 513–530). Guilford Press.

Lucas, R. E., Wallsworth, C., Anusic, I., & Donnellan, M. B. (2019). A direct comparison of the Day Reconstruction Method and the Experience Sampling Method. *Journal of Personality and Social Psychology*, *120*(3), 819–335. https://doi.org/10.1037/pspp0000289

Mazursky, D., & Ganzach, Y. (1998). Does involvement moderate time-dependent biases in consumer multiattribute judgment? *Journal of Business Research*, *41*(2), 95–103. https://doi.org/10.1016/S0148-2963(97)00016-7

Merrilees, C. E., Goeke-Morey, M., & Cummings, E. M. (2008). Do event-contingent diaries about marital conflict change marital interactions?

Behaviour Research and Therapy, *46*(2), 253–262. https://doi.org/10.1016/j.brat.2007.11.007

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, *130*(5), 711–747. https://doi.org/10.1037/0033-2909.130.5.711

Miron-Shatz, T., Stone, A., & Kahneman, D. (2009). Memories of yesterday's emotions: Does the valence of experience affect the memory-experience gap? *Emotion*, *9*(6), 885–896. https://doi.org/10.1037/a0017823

Neubauer, A. B., Scott, S. B., Sliwinski, M. J., & Smyth, J. M. (2020). How was your day? Convergence of aggregated momentary and retrospective end-of-day affect ratings across the adult life span. *Journal of Personality and Social Psychology*, *119*(1), 185–201. https://doi.org/10.1037/pspp0000248

Ohly, S., Sonnentag, S., Niessen, C., & Zapf, D. (2010). Diary studies in organizational research. *Journal of Personnel Psychology*, *9*(2), 79–93. https://doi.org/10.1027/1866-5888/a000009

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*(3), 381–396. https://doi.org/10.1037/0022-3514.80.3.381

Oishi, S. (2002). Experiencing and remembering of well-being: A cross-cultural analysis. *Personality and Social Psychology Bulletin*, *28*, 1398–1406. https://doi.org/10.1177/014616702236871

Oishi, S., & Sullivan, H. W. (2006). The predictive value of daily vs. retrospective well-being judgments in relationship stability. *Journal of Experimental Social Psychology*, *42*(4), 460–470. https://doi.org/10.1016/j.jesp.2005.07.001

Ong, A. D., & Bergeman, C. S. (2004). The complexity of emotions in later life. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *59*(3), P117–P122. https://doi.org/10.1093/geronb/59.3.P117

Ong, A. D., Bergeman, C. S., Bisconti, T. L., & Wallace, K. A. (2006). Psychological resilience, positive emotions, and successful adaptation to stress in later life. *Journal of Personality and Social Psychology*, *91*(4), 730–749. https://doi.org/10.1037/0022-3514.91.4.730

Ouweneel, E., Le Blanc, P. M., Schaufeli, W. B., & van Wijhe, C. I. (2012). Good morning, good day: A diary study on positive emotions, hope, and work engagement. *Human relations*, *65*(9), 1129–1154. https://doi.org/10.1177/0018726711429382

Ptacek, J. T., Smith, R. E., Espe, K., & Raffety, B. (1994). Limited correspondence between *daily coping reports and retrospective coping recall. Psychological Assessment*, *6*(1), 41–49. https://doi.org/10.1037/1040-3590.6.1.41

Ready, R. E., Weinberger, M. I., & Jones, K. M. (2007). How happy have you felt lately? Two diary studies of emotion recall in older and younger adults. *Cognition and Emotion*, *21*(4), 728–757. https://doi.org/10.1080/02699930600948269

Reynolds, B. M., Robles, T. F., & Repetti, R. L. (2016). Measurement reactivity and fatigue effects in daily diary research with families. *Developmental Psychology*, *52*(3), 442–454. https://doi.org/10.1037/dev0000081

Robinson, M. D., Johnson, J. T., & Shields, S. A. (1998). The gender heuristic and the database: Factors affecting the perception of gender-related differences in the experience and display of emotions. *Basic and Applied Social Psychology*, *20*, 206–219. https://doi.org/10.1207/s15324834basp2003_3

Robinson, M. D., & Clore, G. L. (2002a). Belief and feeling: evidence for an accessibility model of emotional self-report. *Psychological bulletin*, *128*(6), 934–951.

Robinson, M. D., & Clore, G. L. (2002b). Episodic and semantic knowledge in emotional self-report: evidence for two judgment processes. *Journal of personality and social psychology*, *83*(1), 198–2012.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145

Russell, J. A. (2017). Mixed emotions viewed from the psychological constructionist perspective. *Emotion Review*, *9*(2), 111–117. https://doi.org/10.1177/1754073916639658

Schonlau, M., & Toepoel, V. (2015). Straightlining in Web survey panels over time. *Survey Research Methods*, *9*(2), 125–137.

Scollon, C. N., Howard, A. H., Caldwell, A. E., & Ito, S. (2009). The role of ideal affect in the experience and memory of emotions. *Journal of Happiness Studies*, *10*(3), 257–269. https://doi.org/10.1007/s10902-007-9079-9

Scott, S. B., Sliwinski, M. J., Mogle, J. A., & Almeida, D. M. (2014). Age, stress, and emotional complexity: Results from two studies of daily experiences. *Psychology and Aging*, *29*(3), 577–587. https://doi.org/10.1037/a0037282

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: An effort-reduction framework. *Psychological Bulletin*, *134*(2), 207–222. https://doi.org/10.1037/0033-2909.134.2.207

Sharpe, J. P., & Gilbert, D. G. (1998). Effects of repeated administration of the Beck Depression Inventory and other measures of negative mood states. *Personality and Individual Differences*, *24*(4), 457–463. https://doi.org/10.1016/S0191-8869(97)00193-1

Sheldon, K. M., Ryan, R., & Reis, H. T. (1996). What makes for a good day? Competence and autonomy in the day and in the person. *Personality and Social Psychology Bulletin*, *22*(12), 1270–1279. https://doi.org/10.1177/01461672962212007

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports. *Proceedings of the National Academy of Sciences*, *115*(1), E15–E23. https://doi.org/10.1073/pnas.1712277115

Skowronski, J. J. (2011). The positivity bias and the fading affect bias in autobiographical memory: A self-motives perspective. In M. D. Alicke & C. Sedikides (Eds.), *Handbook of self-enhancement and self-protection* (pp. 211–231). Guilford Press.

Stone, A. A., Schneider, S., & Harter, J. K. (2012). Day-of-week mood patterns in the United States: On the existence of 'Blue Monday', 'Thank God it's Friday' and weekend effects. *The Journal of Positive Psychology*, *7*(4), 306–314. https://doi.org/10.1080/17439760.2012.691980

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology*, *59*(2), 291–297. https://doi.org/10.1037/0022-3514.59.2.291

Thompson, R. J., Springstein, T., & Boden, M. (2021). Gaining clarity about emotion differentiation. *Social and Personality Psychology Compass*, *15*(3), e12584. https://doi.org/10.1111/spc3.12584

Townshend, J. M., & Duka, T. (2002). Patterns of alcohol drinking in a population of young social drinkers: A comparison of questionnaire and diary measures. *Alcohol and Alcoholism*, *37*(2), 187–192. https://doi.org/10.1093/alcalc/37.2.187

Van Berkel, N., Goncalves, J., Koval, P., Hosio, S., Dingler, T., Ferreira, D. & Kostakos, V. (May). (2019) Context-informed scheduling and analysis: Improving accuracy of mobile self-reports. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–12).

Walker, W. R., Skowronski, J. J., & Thompson, C. P. (2003). Life is pleasant—and memory helps to keep it that way!. *Review of General Psychology*, *7*(2), 203–210. https://doi.org/10.1037/1089-2680.7.2.203

Watson, D. (1988). The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect. *Journal of Personality and Social Psychology*, *55*(1), 128–141. https://doi.org/10.1037/0022-3514.55.1.128

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. https://doi.org/10.1037/0022-3514.54.6.1063

Widdershoven, R. L., Wichers, M., Kuppens, P., Hartmann, J. A., Menne-Lothmann, C., Simons, C. J., & Bastiaansen, J. A. (2019). Effect of self-monitoring through experience sampling on emotion differentiation in depression. *Journal of Affective Disorders*, *244*, 71–77. https://doi.org/10.1016/j.jad.2018.10.092

Yammarino, F. J., Skinner, S. J., & Childers, T. L. (1991). Understanding mail survey response behavior: A meta-analysis. *Public Opinion Quarterly*, *55*(4), 613–639. https://doi.org/10.1086/269284 https://doi.org/10.1086/269284

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Ganzach, Y, Bulmash, B. (2021). The effect of serial day on the measurement of positivity and emotional complexity in diary studies. *European Journal of Social Psychology*. *51*, 1213–1225. https://doi.org/10.1002/ejsp.2809