



Exploring Affect Recall Bias and the Impact of Mild Depressive Symptoms: An Ecological Momentary Study

Desirée Colombo¹(✉), Carlos Suso-Ribera¹,
Javier Fernandez-Álvarez², Isabel Fernandez Felipe¹,
Pietro Cipresso^{2,3}, Azucena Garcia Palacios¹, Giuseppe Riva^{2,3},
and Cristina Botella¹

¹ Department of Basic Psychology, Clinic and Psychobiology,
Universitat Jaume I, 12071 Castellón, Spain
{dcolombo, susor, azucena, botella}@uji.es,
al260660@alumail.uji.es

² Department of Psychology, Università Cattolica del Sacro Cuore,
20100 Milan, Italy
{javier.fernandezkirszman, pietro.cipresso,
giuseppe.riva}@unicatt.it

³ Applied Technology for Neuro-Psychology Lab,
IRCCS Istituto Auxologico Italiano, 20149 Milan, Italy
{p.cipresso, g.riva}@auxologico.it

Abstract. Traditional clinical and research assessments rely on retrospective questionnaires, that ask individuals to retrospectively summarize how they felt during the last period. Nevertheless, people are not accurate at recalling past experiences without altering the content, especially when they are required to report their affect. In this study, we adopted a smartphone-based ecological momentary assessment (EMA) to collect daily assessments of positive (PA) and negative (NA) affect throughout two weeks in a sample of healthy students ($n = 47$). Results showed that both PA and NA are subject to the recall bias; more specifically, people tended to overestimate both affects during the retrospective assessment. This bias was influenced by the presence of mild depressive symptoms as measured by the Beck Depression Inventory (BDI), which led participants to a greater overestimation of NA and higher underestimation of PA. While NA bias was more context-dependent, PA bias showed more stability across time.

Keywords: Ecological momentary assessment · Momentary affect · Recall bias

1 Introduction

Most of traditional assessment techniques both in research and clinical practice are retrospective, i.e. people are asked to summarize their affective experience or symptoms throughout the previous weeks [1]. However, people are not always able to recall past experiences without altering their content and, especially in depressed individuals,

systematic biases can be observed, such as increased elaboration of negative information or greater recall of negative rather than positive stimuli [2]. This recall bias has been detected also in affect recall, pointing out a general tendency to retrospectively exaggerate both positive (PA) and negative (NA) affect [3–5]. Interestingly, clinically depressed individuals show greater inaccuracy for NA [6], which would be explained by different factors such as personal beliefs, memory salience, cognitive styles or past affective experiences [7, 8]. Nevertheless, no study has investigated the role of mild as opposed to moderate/severe depressive symptoms on affect recall bias, so the symptom severity level at which recall bias emerges is unclear.

Ecological momentary assessment (EMA) is an alternative approach to laboratory experiments to collect repeated daily self-reports [9] and/or objective data [10–13] by means of paper-and-pencil diaries or mobile devices, which can be performed in naturalistic settings and close-in-time to the real experience [14, 15]. Not surprisingly, an increasing number of researchers are adopting this approach to explore affect dynamics in daily life [16, 17]. Instead of using retrospective questionnaires, indeed, EMA allows to capture affect fluctuations with higher precision and accuracy and thus to delete the aforementioned recall bias of traditional retrospective assessments.

Here, we explored affect recall bias in a healthy population by comparing two-week EMA ratings of PA and NA collected with a mobile smartphone against affect retrospectively recalled using a paper-and-pencil approach. The aims of the study were to (1) investigate affect recall bias in healthy individuals, (2) explore the direction of such bias (i.e. retrospective affect overestimation and/or underestimation), and (3) deepen into the role that depressive symptoms may play in this phenomenon.

2 Methods

2.1 Participants

Participants were 48 students whose age ranged from 18 to 36 years ($M = 22.26$; $SD = 4.12$). The sample was mostly composed of women (71%). Recruitment was conducted via online advertisements at the Jaume I University (Castellon, Spain).

Participants were first contacted by telephone by one of the researchers and were provided with a web link to fulfil the baseline questionnaires. Subsequently, a face-to-face laboratory meeting was scheduled. Data from one of the participants was excluded because responses markedly deviated from other observations in the sample and within-individual inconsistencies were observed, which made us think of careless or random response style. Therefore, the final sample was composed of 47 participants.

2.2 Measures

At the beginning and at the end of the study, participants completed the Patient Health Questionnaire – 9 (PHQ-9) [18]. The PHQ-9 is a self-report tool for the assessment of depressive symptoms, based on DSM-IV depression diagnostic criteria. This

questionnaire is composed of 9 items which refer to symptoms experienced during the past two weeks. At the end of the study, participants were also asked to fulfil the Positive and Negative Affect Schedule (PANAS) [19], a self-report measure of positive (10 items) and negative (10 items) affect. Specifically, participants were asked to rate PA and NA experienced during the past two weeks (i.e., to retrospectively report their affect throughout the duration of the study). The descriptive statistics of all these questionnaires are reported in Table 1.

Table 1. Questionnaires descriptive statistics.

Measure	Min-Max	Mean	St. Dev.
PHQ-9 (pre)	0–18	4.49	4.49
PHQ-9 (post)	0–18	3.53	3.53
PANAS-PA	16–44	29.87	6.72
PANAS-NA	0–41	18.64	6.88

2.3 Procedure

After the completion of baseline questionnaires, participants were invited to attend the laboratory in order to sign the informed consent. Participants were provided with an identification number to download and access “EMA Móvil”, an Android mobile application created by our team to administer ecological assessments. This application can be easily monitored and programmed from a web platform, where items, type of answer, number of prompts and sampling method can be chosen. No programming skills are therefore required.

Over the following 14 days, the application prompted three daily assessments at random times within three-time intervals (9:30–14:00; 14:00–18:30; and 18:30–23:00). During each evaluation, participants were asked to complete single-items of momentary affect (PA: “*To what extent are you experiencing positive emotions in this moment?*”; NA: “*To what extent are you experiencing negative emotions in this moment?*”). Participants were just asked to enter the notification and to complete the questionnaire. To prevent backfilling, participants were given one hour to answer the current assessment. If they did not respond in time, the evaluation was marked as missing. At the end of the study, participants were asked to return to the laboratory to complete post-assessment questionnaires and receive a monetary compensation of 10 euros.

2.4 Data Analysis

EMA affect values were obtained by calculating the mean of PA and NA item scores across the study (42 possible assessments for each participant), while PANAS affect values were calculated by dividing the total PANAS score (positive and negative separately) administered at the end of the study by the total number of positive/negative

affect items in the PANAS. The range scores for both forms of assessment (two weeks of daily, single-item assessments with an app and a single retrospective evaluation using the full-length scale at the end of the study) were the same (1 = lowest affect to 5 = highest affect).

To test the construct validity of the EMA affect items, we carried out a correlation between PA scores (Pearson correlation) and NA scores (Spearman correlation) obtained via EMA and the PANAS. We subsequently compared daily and retrospective PA means (paired-samples T-test) and NA means (Wilcoxon Signed Ranks Test) to test participants' ability to estimate PA and NA retrospectively.

To explore affect recall bias and distinguish between retrospective overestimation and underestimation of affect, we calculated delta scores between PA and NA measured via the PANAS and EMA. Positive delta scores would reflect affect overestimation during the retrospective assessment, while negative values would reveal retrospective underestimation of affect. We compared PA and NA delta scores (Wilcoxon Signed Ranks Test) and conducted four correlations to investigate the association between PA (Pearson correlation) and NA delta scores (Spearman correlation) and depressive symptoms measured by means of the PHQ-9 at the beginning and at the end of the study (i.e., to explore whether depressive symptoms were associated with the ability to estimate affect).

In the analyses, non-parametric tests were used when the assumptions for the use of parametric tests (i.e., normality of scores) were not met. Parametric tests were used elsewhere.

3 Results

Results showed a significant correlation between daily and retrospective PA measures and daily and retrospective NA measures, as showed in Table 2.

Table 2. Correlations between NA and PA measured via EMA or PANAS. Bivariate associations with NA (PANAS) were calculated with Spearman correlations. The remaining are calculated using Pearson correlations. * $p < .05$, ** $p < .01$, *** $p < .001$.

	Mean (SD)	Bivariate associations		
		NA (EMA)	PA (PANAS)	NA (PANAS)
PA (EMA)	2.75 (0.64)	.08	.36*	.043
NA (EMA)	1.47 (0.33)		-.32*	.48***
PA (PANAS)	2.99 (0.71)			-.22
NA (PANAS)	1.89 (0.74)			

The comparison of PA measured via EMA and with the PANAS evidenced significant mean difference in scores ($t = -2.25, p = .03, 95\% \text{ IC} = -0.453, 0.025$). Furthermore, the comparison of NA between assessment methods also resulted in statistically significant differences in rank scores ($Z = -4.11, p < .001$). Specifically, both PA and NA indicated higher scores when recalled retrospectively with the PANAS.

To further explore the observed recall bias and distinguish between retrospective affect overestimation or underestimation, we calculated delta scores between PA and NA measured by means of the PANAS and EMA. As shown in Fig. 1, a higher variability in the distribution of deltas was observed for PA. However, the analysis of differences in PA and NA delta scores did not result in a statistically significant difference in rank scores ($Z = -.810, p < .418$).

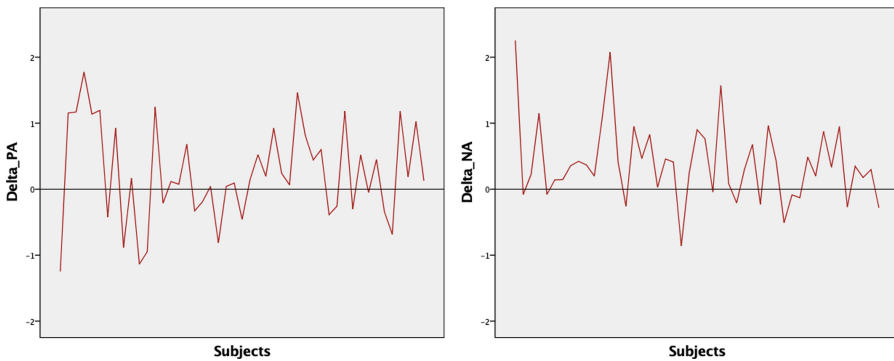


Fig. 1. Distribution of NA and PA delta scores across participants.

We finally investigated the role of depressive symptoms in affect recall bias. PA delta values negatively correlated with PHQ-9 scores measured both at baseline and the end of the study. NA delta scores positively correlated with PHQ-9 measured at the end of the study (Table 3 and Fig. 2).

Table 3. Correlations between delta scores and pre and post-PHQ-9. Bivariate associations with Delta NA are calculated using Spearman correlations. The remaining are Pearson correlations. * $p < .05$, ** $p < .01$, *** $p < .001$.

	Mean (SD)	PHQ-9 (pre)	PHQ-9 (post)
Delta PA	2.75 (0.64)	-.530***	-.653***
Delta NA	1.47 (0.33)	.077	.300*

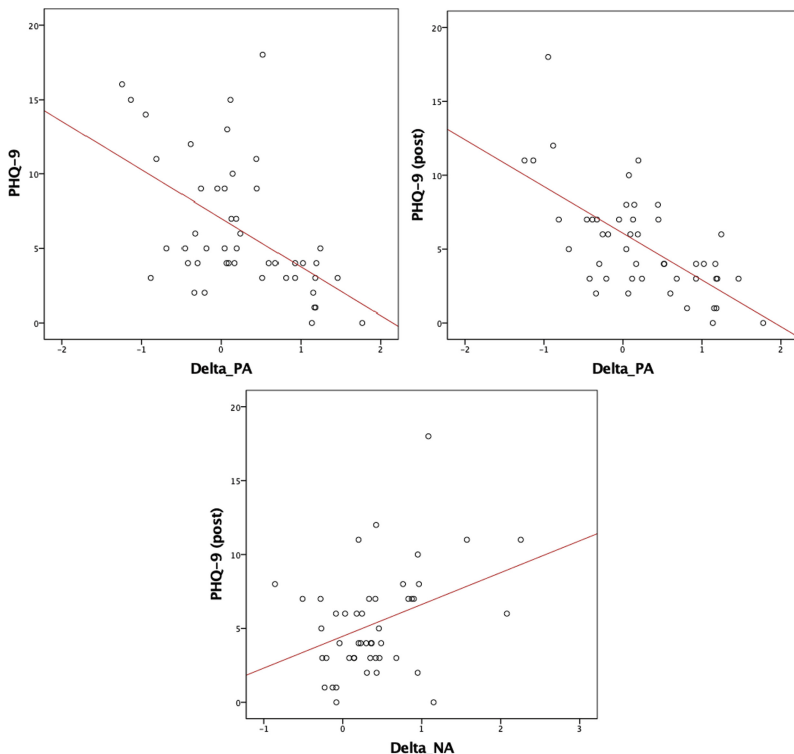


Fig. 2. Correlation between delta scores and pre/post PHQ-9.

4 Discussion

The aim of this study was to investigate affect recall bias in a healthy population by comparing two-week EMA affect assessments against the PANAS administered via paper-and-pencil at the end of the study. To the best of our knowledge, this is the first investigation exploring the role of mild depressive symptomatology on affect recall bias in a healthy population.

Daily EMA measures and retrospective PANAS scores showed a significant correlation, suggesting the construct validity of our single items to assess daily PA and NA. Importantly, one of the main challenges when designing EMA protocols is adherence [20], that is, the percentage of completed assessments obtained from each participant. Our results suggest that the use of single items to assess PA and NA as opposed to long questionnaires is feasible and conceptually valid, which makes it an adequate solution to be adopted in EMA protocols.

According to previous literature [3–5], people tend to retrospectively exaggerate both PA and NA. Here, we replicated this result, as participants showed a general retrospective overestimation of both affects. We were also interested in exploring mild depressive symptomatology as a potential variable affecting affect recall. In their study, Ben-Zeev and colleagues found that clinical depression leads to the retrospective intensifications of

both PA and NA, with greater inaccuracy for NA recall [6]. Here, we showed that the presence of mild depressive symptoms in healthy individuals also influences affect recall. According to our results, individuals with higher PHQ-9 scores show a greater overestimation of NA and a greater underestimation of PA. By contrast, participants with low or no depressive symptoms are more likely to overestimate PA and underestimate NA during the retrospective assessment. This is in line with the hypothesis of illusion of control that non-depressed individuals have shown to be positively biased and to benefit from positive illusions, that in turn would foster well-being [21].

Interestingly, PA recall bias correlated negatively with depressive symptoms assessed both at the baseline (i.e. assessment of depressive symptoms during the two weeks prior to the beginning of the study) and at the end of the study (i.e. assessment of depressive symptoms throughout the two-weeks of the EMA study), while NA recall bias was only positively associated with post-PHQ-9. In other words, we may hypothesize that the tendency to over- or underestimate PA may be considered as a trait as opposed to a state, and would therefore show a greater stability across time, regardless of daily events. On the other hand, our results suggest that the tendency to over- or underestimate NA would, on the contrary, be more context-dependent, and would be determined by momentary experiences of emotions and by the occurrence of specific events in daily life.

There are limitations in the present investigation, including the reduced sample size and the correlational nature of the study, which affect the generalizability and causal inferences that can be drawn from this study. Due to the small sample size, it is also not possible to address the hierarchy of data that are nested within participants. Once that more data are collected, hierarchical mixed nonlinear models, or similar, can be considered. A final important aspect that needs to consider revolves around the content validity. While it may be true that in our preliminary examination a proper correlation of PANAS and the single item exists, it is necessary to contrast this finding in larger populations in order to guarantee content validity. It may be the case that such a complex construct like affect may not be accurately grasped by means of a single item. Beyond these issues, future studies should also consider the impact that other variables have on affect recall, such as the presence of anxiety symptoms or high levels of stress, as well as focus on the development of standardized ad hoc items to be used in mobile devices for the daily assessment of affect.

Nevertheless, we believe that this study sheds new light into the importance and utility of EMA in the study of affect, as well as the need to study the influence of recall bias for a wider range of depressive severity scores, including milder cases as conducted in the present investigation. These findings are important for clinical purposes, as they indicate that recall bias can occur even when depressive symptoms are mild, especially for PA. Accordingly, the evaluation of affect should be preferably performed ecologically and repeatedly using EMA.

References

1. Shiffman, S., Stone, A.A., Hufford, M.R.: Ecological momentary assessment. *Annu. Rev. Clin. Psychol.* **4**(1), 1–32 (2008)

2. Gotlib, I.H., Joormann, J.: Cognition and depression: current status and future directions. *Annu. Rev. Clin. Psychol.* **6**, 285–312 (2010)
3. Wirtz, D., Kruger, J., Scollon, C.N., Diener, E.: *Psychol. Sci.* **14**(5), 520–524 (2003)
4. Thomas, D.L., Diener, E.: Memory accuracy in the recall of emotions. *J. Pers. Soc. Psychol.* **59**(2), 291–297 (1990)
5. Kardum, I., Tićac Daskijević, K.: Absolute and relative accuracy in the retrospective estimate of positive and negative mood. *Eur. J. Psychol. Assess.* **17**(1), 69–77 (2001)
6. Ben-Zeev, D., Young, M.A., Madsen, J.W.: Retrospective recall of affect in clinically depressed individuals and controls. *Cogn. Emot.* **23**(5), 1021–1040 (2009)
7. Fredrickson, B.L.: Extracting meaning from past affective experiences: the importance of peaks, ends, and specific emotions. *Cogn. Emot.* **14**(4), 577–606 (2000)
8. Levine, L.J., Prohaska, V., Burgess, S.L., Rice, J.A., Lauhere, T.M.: Remembering past emotions: the role of current appraisals. *Cogn. Emot.* **15**(4), 393–417 (2001)
9. Suso-Ribera, C., Castilla, D., Zaragozá, I., Ribera-Canudas, M.V., Botella, C., García-Palacios, A.: Validity, reliability, feasibility, and usefulness of pain monitor, a multidimensional smartphone app for daily monitoring of adults with heterogeneous chronic pain. *Clin. J. Pain* **34**(10), 900–908 (2018)
10. Mohr, D.C., Zhang, M., Schueller, S.M.: Personal sensing: understanding mental health using ubiquitous sensors and machine learning. *Annu. Rev. Clin. Psychol.* **13**(1), 23–47 (2017)
11. Gaggioli, A., et al.: A mobile data collection platform for mental health research. *Pers. Ubiquit. Comput.* **17**(2), 241–251 (2013)
12. Gaggioli, A., et al.: Positive technology: a free mobile platform for the self-management of psychological stress. *Annu. Rev. CyberTherapy Telemed.* **199**, 25–29 (2014)
13. Cipresso, P., et al.: Is your phone so smart to affect your state? An exploratory study based on psychophysiological measures. *Neurocomputing* **84**, 23–30 (2012)
14. Stone, A.A., Shiffman, S., Atienza, A.A., Nebeling, A.: Historical roots and rationale of ecological momentary assessment (EMA). In: *The Science of Real-Time Data Capture: Self-Reports in Health Research*, pp. 3–10 (2007)
15. Csikszentmihalyi, M., Larson, R.: Validity and reliability of the experience-sampling method. *J. Nerv. Ment. Dis.* **175**(9), 526–536 (1987)
16. Ebner-Priemer, U.W., Trull, T.J.: Ecological momentary assessment of mood disorders and mood dysregulation. *Psychol. Assess.* **21**(4), 463–475 (2009)
17. Bylsma, L.M., Rottenberg, J.: Uncovering the dynamics of emotion regulation and dysfunction in daily life with ecological momentary assessment. In: Nyklíček, I., Vingerhoets, A., Zeelenberg, M. (eds.) *Emotion Regulation and Well-Being*. Springer, New York (2011). https://doi.org/10.1007/978-1-4419-6953-8_14
18. Kroenke, K., Spitzer, R.L., Williams, J.B.W.: The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**(9), 606–613 (2001)
19. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* **54**(6), 1063–1070 (1988)
20. Colombo, D., Cipresso, P., Fernández Alvarez, J., García Palacios, A., Riva, G., Botella, C.: An overview of factors associated with adherence and dropout to ecological momentary assessments in depression. *Annu. Rev. CyberTherapy Telemed.* (2018)
21. Taylor, S.E., Brown, J.D.: Positive illusions and well-being revisited: separating fact from fiction. *Psychol. Bull.* **116**(1), 21–27 (1994)