Trait depression predicts negatively biased encoding and retrieval of ambivalent movie: interoceptive and lexical analysis

Jinwoo Lee (adem1997@snu.ac.kr)
Department of Psychology, Seoul National University
Seoul, South Korea

Kyungjin Oh (castella@snu.ac.kr)
Department of Psychology, Seoul National University
Seoul, South Korea

Junhyung Kim (jihndy.kim@samsung.com)
Department of Psychiatry, Samsung Kangbuk Hospital, Sungkyunkwan University School of Medicine
Seoul, South Korea

Jiook Cha (connectome@snu.ac.kr)
Department of Psychology, Seoul National University
Seoul, South Korea

Abstract

Ambivalent affect, the most naturalistic emotion in daily life, is largely modulated by individual traits. In particular, trait depression weights negative affect in the recall of ambivalent events. However, as trait-affect interplay is differently expressed across affective contexts and memory stages - encoding and retrieval, their relationship and mechanism should be specified along with these dimensions. We predicted that 1) altered interoception would mediate trait depression and negative encoding of affective context and 2) its negatively biased retrieval would be manifested in free languages. To test them, we combined movie-watching and free-recall paradigm, deep representational learning of EEG and ECG, and LLM-based sentiment analysis of recall text. We found that trait depression predicted ambivalent encoding of pleasant scenes via inter-subject similar interoceptive representation. Also, trait depression predicted negatively biased recall of ambivalent scenes. We discuss the potential of ambivalent processing as a depression-specific risk marker and its clinical implications.

Keywords: depression, ambivalence, negative bias, interoception, affective lexicon

Introduction

Ambivalent affect, the co-occurrence of positive and negative affect, pervades daily emotional life (Moeller et al., 2018) and captures psychological idiosyncrasies more effectively than purely univalent ones (Barford & Smillie, 2016). Accordingly, investigating how people process ambivalent affect offers insight into how psychological traits shape emotional experience. Trait depression, in particular, is related to negative bias for ambivalent stimuli (Watters & Williams, 2011), and such bias predicts future psychopathology (Rude et al., 2002).

However, since trait-affect interplay is differently manifested across affective contexts and memory stages (i.e., encoding versus retrieval), their relationship and its basis need to be specified along these dimensions. For example, mixed emotions in stressful events are related to resilience (Braniecka et al., 2014), but in non-stressful events, they are linked to lower well-being (Oh, 2025). Also, negative bias in trait depression is implemented by different mechanisms across memory stages. Such bias is expressed during encoding as prediction of 'future' negativity, but during retrieval as weight on 'past' one (Marsh et al., 2019).

To study how trait depression leads to negatively biased ambivalent affect across these dimensions, we combined a film-viewing and free-recall paradigm with emotionally rich movie stimulus. To probe mechanisms underlying stage-specific trait-affect interplay, we further implemented two methodological approaches. First, we examined interoception as a linking mechanism between trait depression and biased encoding of ambivalence, given its key role in affect encoding (Barrett & Simmons, 2015; Damasio & Carvalho, 2013; Seth, 2013). We identified each participant's interoceptive representation through deep joint representational learning between EEG and ECG during film watching. Based on findings linking negative encoding to reduced neural idiosyncrasy (Finn et al., 2018; Iyer et al., 2024), we hypothesized that trait depression would relate to inter-subject similar interoceptive representations during movie watching.

Second, we used LLM-based sentiment analysis that allows for more efficient evaluation of free-form language without human annotations (Hur et al., 2024) to test whether trait depression correlates to negatively biased language when recalling ambivalent scenes, given language's essential role of implementing affective bias in retrieval (Rude, Gortner, & Pennebaker, 2004).

Methods

Experiment

Forty-five participants' BDI-II and STAI-X scores were assessed and their 19-channel EEG and single-channel ECG in resting state were recorded. Their self-report of ongoing positive and negative affect were collected with EEG/ECG signals while they watched *One Small Step*, validated to induce diverse contexts (Vaccaro et al., 2024). Afterward, they freely described the scenes and their feelings.

Analysis

Self-report analysis We parsed four phases within a movie—neutral, positive, negative, and ambivalent—based on 18 independent raters' ratings (**Fig 1a**). Correlations between traits and average ratings of positive, negative, and ambivalent affect in each phase were tested. Ambivalent score was defined as the minimum value of positive and negative ratings.

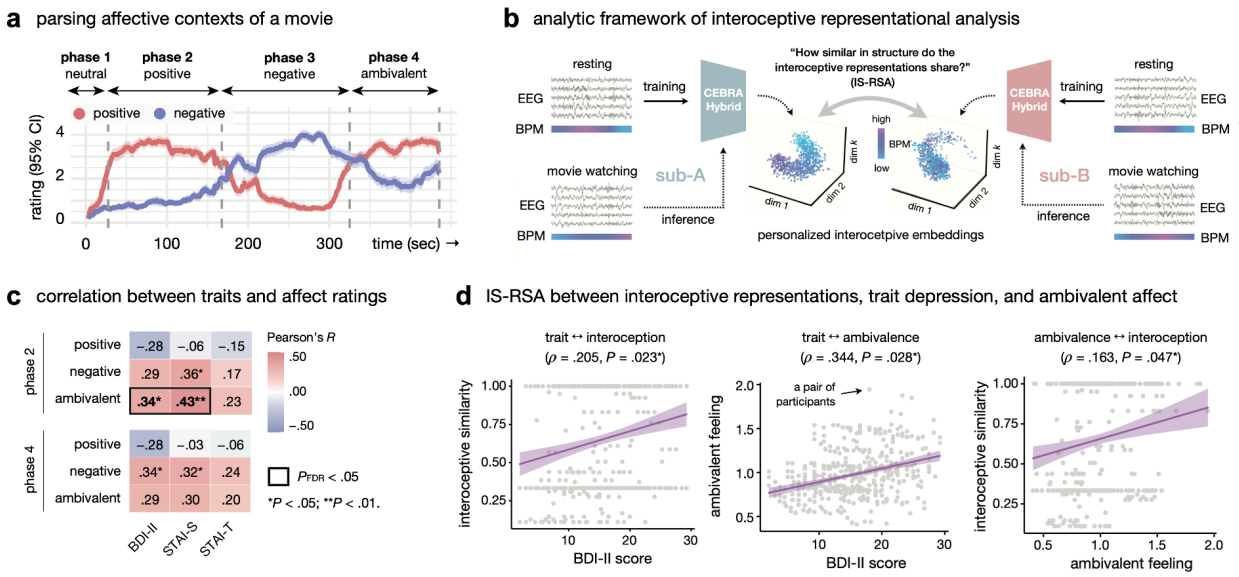

Figure 1: Analytic framework and key findings. a, parsing affective contexts of a movie. **b,** analytic framework of interoceptive representational analysis. **c,** correlation between traits and affect ratings. **d,** IS-RSA of similarity in interoceptive representations, average trait depression scores, and average ambivalent ratings of all possible participant pairs (trait and affect scores were from phase 2).

Interoceptive analysis To identify interoceptive representations, we used CEBRA (Schneider, Lee, & Mathis, 2023), a deep joint representational learning, which was trained separately for each subject (**Fig 1b**). Resting-state EEG and ECG were used for training to isolate interoceptive representation from exteroceptive processing, and film-viewing EEG was used for inference. δ, θ, α, β, and γ EEG bands and beats per minutes (BPM) from ECG served as input and auxiliary variables, respectively. Afterward, we applied IS-RSA to examine the association between similarity of interoceptive representation, trait score, and affect ratings.

Lexical analysis We asked ChatGPT-4o to assess positive, negative, and ambivalent sentiment of each participant's recall transcription for each phase on a 0-5 scale. We repeated this procedure five times, and the correlation between trait scores and average sentiment score for each phase was examined.

Results

Negatively biased encoding of ambivalence

Trait depression and state anxiety are significantly related to ratings of ambivalent affect in a pleasing phase (P_{FDR} < .05; **Fig 1c**).

Depression-driven interoceptive processing

Trait depression is significantly linked to inter-subject similarity of interoceptive representations and ratings of ambivalence in a pleasing phase (P < .05 from Mantel permutation test; **Fig 1d**). However, trait and state anxiety are not.

Verbally manifested biased recall

Trait depression significantly correlates to a more negative sentiment in the free recall of an ambivalent phase (R = .414, P_{FDR} < .05). No such association is observed for anxiety-related metrics.

Conclusion

In this study, we found that trait depression leads to negatively biased encoding and recall of ambivalent affect in a distinguishable way, which is driven by altered interoception and lingual expressions. Future studies need to test causality of body-trait-language interactions and its clinical implications.

References

Barford, K. A., & Smillie, L. D. (2016). Openness and other Big Five traits in relation to dispositional mixed emotions. *Personality and individual differences*, *102*, 118-122.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature reviews neuroscience*, *16*(7), 419-429.

Braniecka, A., Trzebińska, E., Dowgiert, A., & Wytykowska, A. (2014). Mixed emotions and coping: The benefits of secondary emotions. *PloS one*, *9*(8), e103940.

Damasio, A., & Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature reviews neuroscience*, *14*(2), 143-152.

Finn, E. S., Corlett, P. R., Chen, G., Bandettini, P. A., & Constable, R. T. (2018). Trait paranoia shapes inter-subject synchrony in brain activity during an ambiguous social narrative. *Nature communications*, *9*(1), 2043.

Hur, J. K., Heffner, J., Feng, G. W., Joormann, J., & Rutledge, R. B. (2024). Language sentiment predicts changes in depressive symptoms. *Proceedings of the National Academy of Sciences*, *121*(39), e2321321121.

Iyer, S., Collier, E., Broom, T. W., Finn, E. S., & Meyer, M. L. (2024). Individuals who see the good in the bad engage distinctive default network coordination during post-encoding rest. *Proceedings of the National Academy of Sciences*, *121*(1), e2306295121.

Marsh, L., Edginton, T., Conway, M. A., & Loveday, C. (2019). Positivity bias in past and future episodic thinking: Relationship with anxiety, depression, and retrieval-induced forgetting. *Quarterly Journal of Experimental Psychology*, *72*(3), 508-522.

Moeller, J., Ivcevic, Z., Brackett, M. A., & White, A. E. (2018). Mixed emotions: Network analyses of intra-individual co-occurrences within and across situations. *Emotion*, *18*(8), 1106.

Oh, V. (2025). Torn between valences? Associations between mixed emotions and well-being in stressful and nonstressful situations in a large-scale ecological momentary assessment study. *Emotion*.

Rude, S., Gortner, E. M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*(8), 1121-1133.

Rude, S. S., Wenzlaff, R. M., Gibbs, B., Vane, J., & Whitney, T. (2002). Negative processing biases predict subsequent depressive symptoms. *Cognition & Emotion*, *16*(3), 423-440.

Schneider, S., Lee, J. H., & Mathis, M. W. (2023). Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, *617*(7960), 360-368.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in cognitive sciences*, *17*(11), 565-573.

Vaccaro, A. G., Wu, H., Iyer, R., Shakthivel, S., Christie, N. C., Damasio, A., & Kaplan, J. (2024). Neural patterns associated with mixed valence feelings differ in consistency and predictability throughout the brain. *Cerebral Cortex*, *34*(4), bhae122.

Watters, A. J., & Williams, L. M. (2011). Negative biases and risk for depression; integrating self-report and emotion task markers. *Depression and anxiety*, *28*(8), 703-718.