JYX
JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Lou, Yixue; Lei, Yi; Astikainen, Piia; Peng, Weiwei; Otieno, Suzanne; Leppänen, Paavo H. T.

Title: Brain responses of dysphoric and control participants during a self-esteem implicit association test

Year: 2021

Version: Accepted version (Final draft)

Copyright: © 2021 Society for Psychophysiological Research

Rights: In Copyright

Rights url: http://rightsstatements.org/page/InC/1.0/?language=en

Please cite the original version:

Lou, Y., Lei, Y., Astikainen, P., Peng, W., Otieno, S., & Leppänen, P. H. T. (2021). Brain responses of dysphoric and control participants during a self-esteem implicit association test. Psychophysiology, 58(4), Article e13768. https://doi.org/10.1111/psyp.13768

Brain responses of dysphoric and control participants during a self-esteem implicit association test

Yixue Lou[1, 2, 3], Yi Lei[1, 3, †], Astikainen Piia[2], Weiwei Peng[3], Otieno Suzanne[2], Leppänen Paavo H. T.[2]

[1] Institute for Brain and Psychological Sciences, Sichuan Normal University, Chengdu 610066, China

[2] Department of Psychology, Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä 40014, Finland

[3] School of Psychology, Shenzhen University, Shenzhen 518060, China

[†] **Corresponding author:**

Yi Lei

Institute for Brain and Psychological Sciences, Sichuan Normal University,

Jing'an Road 5, Jinjiang District, Chengdu, P.R. China, 610066

E-mail: *leiyi821@vip.sina.com*

Tel: +86 18126260618

Fax: +86 755 26994020

Abstract

Previous studies have reported lowered implicit self-esteem at the behavioral level among depressed individuals. However, brain responses related to the lowered implicit self-esteem have not been investigated in people with depression. Here, event-related potentials were measured in 28 dysphoric participants (individuals with elevated amounts of depressive symptoms) and 30 control participants during performance of an Implicit Association Task (IAT) suggested to reflect implicit self-esteem. Despite equivalent behavioral performance, differences in brain responses were observed between the dysphoric and the control groups in late positive component (LPC) within 400-1000 ms post-stimulus latency. For the dysphoric group, self-negativity mapping stimuli (*me* with *negative* word pairing and *not-me* with *positive* word pairing) induced significantly larger LPC amplitude as compared to self-positivity mapping stimuli (*me* with *positive* pairing and *not-me* with *negative* pairing), whereas the control group showed the opposite pattern. These results suggest a more efficient categorization towards implicit self-is-negative association, possibly reflecting lower implicit self-esteem among the dysphoric participants, in comparison to the controls. These results demonstrate the need for further investigation into the functional significance of LPC modulation during IAT and determination of whether LPC can be used as a neural marker of depressive-related implicit self-esteem.

Keywords: depressive symptoms, dysphoria, implicit self-esteem, event-related potentials (ERPs), late positive component (LPC), implicit association test (IAT)

1 Introduction

1.1 research background

Self-esteem refers to a person's overall attitude towards himself or herself (Rosenberg, 1965), and is thought to play an important role in maintaining one's mental health and well-being. According to Beck's cognitive theory of depression (1967), people with depression typically have lowered self-esteem, as reflected in low self-evaluation and unfavorable self-attitude. As suggested by previous studies, part of self-esteem, the explicit self-esteem, can be accessed by introspective methods, such as self-reported questionnaires or tasks (e.g. Rosenberg Self-Esteem Scale; (Rosenberg, 1965)). In contrast, the unconscious and introspectively unidentified (or inaccurately identified) part of the self-esteem, the implicit self-esteem, can only be measured by implicit experimental paradigms (e.g. Implicit Association Task; (Greenwald & Farnham, 2000)). The behavioral performance and brain activities related to explicit self-esteem have been well-explored among individuals with depression (for a review, see (Lou, Lei, Mei, Leppänen, & Li, 2019). However, only few studies have used behavioral paradigms to investigate implicit self-esteem in depression (Leeuwis, Koot, Creemers, & van Lier, 2015; Randenborgh, Pawelzik, Quirin, & Kuhl, 2016; Roberts, Porter, & Vergara-Lopez, 2015; Smeijers et al., 2017).

The Implicit Association Test (IAT) (Greenwald, McGhee, & Schwartz, 1998), which is a categorization task in nature, is the most commonly used paradigm for measuring a person's implicit attitude (Bosson, Swann, & Pennebaker, 2000; Vianello & Bar-Anan, 2020). In investigations of the implicit self-esteem, participants are

typically asked to sort a series of stimulus words (e.g., pronouns and adjectives) into self-related (e.g., me, I, my) or other-related (e.g., him, his, they) categories, or into positive (e.g., bright, noble, honest) or negative (e.g., ugly, vile, guilty) categories, by pressing two response-keys (e.g., a left key and a right key). All the stimuli words are presented one-by-one in two independent blocks. During one of the blocks, participants respond to self-related and positive words using the same key (e.g., left key), and correspondingly to other-related and negative words with another key (e.g., right key). During another block, they respond to self-related and negative words similarly with the same key, and to other-related and positive words with another key. The premise of the IAT is that when participants use the same key to respond to well-associated categories that are congruent with their implicit self-attitude (e.g., self + positive), their performance should be better (e.g., faster and more accurate) than when less associated, incongruent categories (e.g. self + negative) utilize the same key (Greenwald & Farnham, 2000). For example, researchers found that self-positivity bias, which is a tendency for people to relate themselves more with positive rather than negative items, commonly exists among healthy individuals (Y. Chen et al., 2014; Mezulis, Abramson, Hyde, & Hankin, 2004; Pahl & Eiser, 2005; Watson, Dritschel, Obonsawin, & Jentzsch, 2007). Under the influence of this kind of bias, healthy participants often exhibit faster responses when self and positive attributes, as compared to self and negative attributes, are paired to the same key (Egenolf et al., 2013; Greenwald & Farnham, 2000). Thus, the self + positive mapping is usually labeled as the congruent condition in this design, and correspondingly the self +

negative mapping is labeled as the incongruent condition (Egenolf et al., 2013). The differences in performance between the congruent and the incongruent categorizations are therefore suggested to implicitly measure one's self-esteem, where a longer reaction time (RT) in the incongruent condition, relative to that in the congruent condition, indicates higher implicit self-esteem (Greenwald & Farnham, 2000).

Using the IAT, some previous studies have found that both depressed patients and non-depressed controls have longer RTs in the incongruent relative to the congruent condition, but this difference between RTs in incongruent vs. congruent condition was significantly smaller for currently depressed patients (Jabben et al., 2014; Risch et al., 2010), recurrently depressed patients (Risch et al., 2010), and remitted depressive patients (Risch et al., 2010), as compared to non-depressed controls. These results have been interpreted to indicate lowered implicit self-esteem among individuals with depressive symptoms. Franck, Raedt, Dereu, and Abbeele (2007) reported no significant RT difference between congruent and incongruent conditions, suggesting a lack of self-positivity bias, among patients with depressive disorder (for null results, see also (Kesting, Mehl, Rief, Lindenmeyer, & Lincoln, 2011; Lemmens et al., 2014). However, neural responses during IAT have not been investigated in relation to depressive symptoms. To address this, EEG-based event-related potentials (ERPs) were used to monitor the time course of brain activity for a preclinical depression group (labeled here as dysphoric, meaning participants with elevated amounts of depressive symptoms) and a control group when performing IAT as a measure of

implicit self-esteem.

1.2 Brain response measures during the IAT

The parietally-distributed late positive component (LPC; sometimes considered a sustained P3 response and thus be labeled as P3-like or P3b-like response), which is elicited at approximately 300 ms latency and continues to the end of the stimulus (Gable, Adams, & Proudfit, 2015), is especially interested in the current study because it was previously identified as significant in IAT studies. It is suggested that the LPC amplitude is highly modulated by the informative value of the stimuli (Polich, 2007; Verleger, Jaśkowski, & Wascher, 2005). Larger LPC amplitude has been observed in emotional vs. neutral stimuli (Yuan, Yang, Meng, Yu, & Li, 2008), in high-arousal vs. low-arousal stimuli (Rozenkrants & Polich, 2008), and in rare vs. frequent stimuli (Verleger & Śmigasiewicz, 2016). In categorization tasks, such as the IAT, the efficient categorization (e.g., the congruent pairings) tends to elicit enhanced LPC amplitude because it involves increased decision-related aspects of attentional allocation and stimulus evaluation (Kok, 2001; Polich, 2007; Verleger et al., 2005).

For example, Coates and Campbell (2010) recorded the ERPs during administration of the IAT and found that the congruent condition (good + musical instruments and bad + weapons in their study), relative to the incongruent condition (good + weapons and bad + musical instruments), elicited larger positive ERP amplitude at the parietal sites between 400 and 600 ms post-stimulus. In a study that measured individuals'

attitudes about gay versus straight,J. K. Williams and Themanson (2011) also reported a significantly larger posterior LPC amplitude (from 500 – 1000 ms post-stimulus) in the 'congruent' condition (gay + negative and straight + positive in their study) relative to that in the 'incongruent' condition (gay + positive and straight + negative) among individuals with gay-negative bias. L. Chen et al. (2018) investigated internet-addicted individuals and also reported a greater amplitude of the late positive potential (around 300 ms latency) at the occipital sites in the congruent (internet + positive and mammal + neutral in their study) condition, relative to that in the incongruent (internet + neutral and mammal + positive) condition. The larger parietal-distributed LPC amplitude has thus been associated with the stronger association of stimuli parings that are congruent with the implicit bias of individuals during the IAT (L. Chen et al., 2018; J. K. Williams & Themanson, 2011).

LPC responses are also considered an important marker of implicit self-bias during administration of the IAT (Egenolf et al., 2013; Yang & Zhang, 2009). For instance, larger LPC amplitudes were observed in the congruent condition (e.g., self + positive and other + negative) than those in the incongruent condition (e.g., self + negative and other + positive) among non-depressed healthy participants (Wu, Gu, Cai, & Zhang, 2016; Yang & Zhang, 2009). The enhanced LPC amplitudes here suggest that, for the healthy individuals, the "self with positive" association might be more congruent with their implicit self-attitude than the "self with negative" association (Fleischhauer, Strobel, Diers, & Enge, 2014; Wu et al., 2016; Xiao, Zheng, Wang, Cui, & Chen,

2015). These results have thus been taken to indicate that healthy participants usually have implicit self-positivity bias (Yang & Zhang, 2009).

Based on these findings, we expected significantly faster behavioral responses and larger posterior LPC amplitudes in the *self* + *positive* and *other* + *negative* condition, relative to the *self* + *negative* and *other* + *positive* condition, among non-dysphoric controls. Considering the lack of self-positivity bias and the consequently lowered implicit self-esteem among individuals with depressive symptoms (Franck et al., 2007; Jabben et al., 2014; Risch et al., 2010), we therefore expected a smaller (or even no) difference in behavioral and LPC amplitudes between these two conditions for participants in the dysphoric group. Since it is unclear which mapping (self + positive or self + negative) would be more congruent with the implicit self-attitude of dysphoric people, this study did not use the labels of "congruent*"* and "incongruent*"* conditions. Instead, we used the term "self-positivity" to refer to *self* + *positive* and *other* + *negative* pairings and "self-negativity" to refer to *self* + *negative* and *other* + *positive* pairings.

Additional early components, such as frontal N1, occipital P1, occipital N170, and P2, were also measured because they were observable in recordings during IAT administration in previous studies (frontal N1: (van Nunspeet, Ellemers, Derks, & Nieuwenhuis, 2014); occipital P1: (Fleischhauer et al., 2014); occipital N170: (Ibáñez et al., 2010); P2: (Grundy, Benarroch, Lebarr, & Shedden, 2015; Healy, Boran, &

Smeaton, 2015; Xiao et al., 2015). However, according to a recent research which examined the full range of mental processes that occur during the IAT, only a late brain response (with a rather typical P3-like topography commencing around 450 ms post-stimulus at the posterior topographic areas) was suggested to be associated with individual differences in implicit bias during the IAT (Schiller et al., 2016). We thus expected no significant group difference in the early components in this study. In addition, since this experiment employed highly identical stimuli (two-character simplified Chinese words; for a full description see 2.2.2 Experimental stimuli) that did not differ with regard to stimulus characteristics or overall emotional valence, the self-positivity and the self-negativity conditions should not vary in selectivity of attention or perceptual processing. Thus, it is predicted that the early components (reflecting more automatic attention and perceptual processing) would not differ between the two IAT conditions. The detailed description, analyses, results, and interpretations of these components are reported in **Supplementary data 1**.

2 Method

2.1 Participants

Five hundred and sixty-seven students from Shenzhen University volunteered to complete the Beck Depression Inventory - II (BDI-II), a well-validated instrument for the assessment of depressive symptoms in both psychiatric and normal populations for ages 13 years and above (Beck, Steer, & Brown, 1996). Individuals with scores distributed in the top five percent of the overall BDI-II score distribution were invited

to participate in the dysphoric group (N = 30), whereas individuals with scores distributed in the bottom five percent were selected to participate in the control group (N = 32). All participants were right-handed and had normal or corrected-to-normal vision. No participants reported previous or current physiological, neurological, or psychiatric disorders. The use of a pre-clinical sample (the dysphoric participants) was designed to avoid potential confounding factors related to the use of depression medication. According to previous studies, for instance, some psychopharmacological treatments (e.g., agomelatine) could affect the brain structures involved in self-related processing in depression (Delaveau et al., 2016). An additional advantage of examining a preclinical sample was that these participants were generally free from diagnostic comorbidities, which are more common in clinical samples.

Of the original 62 participants, data from two dysphoric individuals and two control participants were excluded due to excessive body movement during the collection of the EEG data, leaving a final sample of 58 participants (Dysphoric group = 28, Control group = 30; see Table 1). As there was no previous study investigating cognitive aspects of IAT processing in both control and dysphoric or depressed participants, sample size in the present study was estimated based on a standard medium effect size (f = 0.25) (Cohen, 1988). Power analysis, conducted with G*Power 3 (Faul, Erdfelder, Buchner, & Lang, 2009), showed a requirement of 27 participants in each group (dysphoric and control) with a statistical power of $(1 - \beta) = 0.95$ and a significance level of $\alpha = 0.05$. There was no significant difference for age (*t*

(56) = -1.05, p = .30) or gender (Pearson χ^2 (1) = .35, p = .55) between the dysphoric and the control groups. This study was approved by the local Review Board for Human Participant Research of Shenzhen University. Each subject signed an informed consent form before the experiment.

Please Insert **Table 1** Here

2.2 Experimental protocol

2.2.1 Implicit Association Test (IAT)

The task involved four categories of stimulus words: *me*-related pronouns (e.g. me, mine, us), *not-me*-related pronouns (e.g., other, his, they), *positive* adjectives (e.g., smart, brave, honest), and *negative* adjectives (e.g., fool, coward, dishonest). All stimulus words were presented in the center of the screen, one-by-one, and in written from. Participants were asked to complete a self-positivity block and a self-negativity block. Each block included three practice phases and one data-collection phase (as illustrated in **Figure 1a**). The first phase was a 10-trial practice phase for the *me* versus *not-me* categorization (five trials for each category). Participants were asked to sort the pronouns into me or not-me categories by pressing a left or a right key (e.g., *F* and *J* on the keyboard) with their left or right index fingers. Example of the instructions: Please press the *F* key when you see a *me*-related word (e.g., me, mine, us, …). Press the *J* key when you see a *not-me*-related word (e.g., other, his, they, …). The second phase was a 10-trial practice phase for *positive* versus *negative* categorization (five trials for each category). Participants were asked to sort the

adjectives into positive or negative categories by using the same keys. Example of the instructions: Please press the *F* key when you see a *positive* describing word (e.g., smart, brave, honest, …). Press the *J* key when you see a *negative* describing word (e.g., fool, coward, dishonest, …). The third phase included 20 practice trials (five trials for each category). Participants were asked to sort all the words, pronouns, and adjectives which they already practiced during the first and the second phases, to one of the existing four categories. The fourth phase was a 320-trial (80 trials for each category) data-collection phase during which the behavioral responses and EEG signals were recorded. During this phase, the requirement was the same as that of the third phase, but with more trials. Participants were given a break after every 80 trials to avoid fatigue.

The self-positivity and the self-negativity blocks differed in the key-assignment. During the self-positivity block, the *me* words and the *positive* words shared the same key, whereas the *not-me* words and the *negative* words shared another key. During the self-negativity block, however, the *me* words and the *negative* words shared the same key, while the *not-me* words and the *positive* words shared another. Example of the self-positivity instruction: Please press the *F* key when you see either *me* related words (e.g., me, mine, us, ….) OR *positive* describing words (e.g., smart, brave, honest, ….). Press the *J* key when you see either *not-me* related words (e.g., other, his, they, ….) OR *negative* describing words (e.g., fool, coward, dishonest, ….).

All words were presented in a completely random order. The assignment of the left and right keys and the order of the self-positivity and the self-negativity blocks were both counterbalanced across participants. Similar to the design of Wu et al. (2016), each stimulus trial began with a fixation cross with a random duration between 1000 and 2000 ms. A stimulus word was then presented for 1000 ms, during which time the participants were asked to respond to the word by pressing the *F* or *J* keys as quickly as possible. Next, a new fixation cross appeared to indicate the beginning of the next trial (see **Figure 1b**). Compared to behavioral IAT studies, two revisions were made in this protocol to facilitate EEG data collection, as suggested by a recent electrophysiological study (Wu et al., 2016). First, the labels that are usually presented in the upper left- and right-corners of the screen, which aim to remind the participants of the correct responses, were omitted to reduce additional eye movements during data recording. Second, to ensure a sufficient number of valid trials (with correct responses) for off-line ERP data analysis, participants were asked to practice until reaching a relatively high accuracy rate (85%) before beginning the data-collection session.

Please Insert **Figure 1** Here

2.2.2 Experimental Stimuli

One hundred and sixty adjectives (in Chinese, 80 positive and 80 negative) were used in the current study (see illustration list in **Supplementary data 2,** Table 1). Most of these words were selected from a pool of 562 personality-trait adjectives that was previously developed by Huang and Zhang (1992). The remaining attribute words

were selected from the *Chinese Affective Words System* (CAWS) that was established by Wang, Zhou, and Luo (2008). According to our pilot pre-experiment (N = 25), the positive and negative words were matched in terms of numbers of character strokes, meaningfulness, familiarity, and arousal ratings. The only difference between the two stimulus categories was in the dimension of desirability (see **Figure 2**). Due to the limited variations in the Chinese language, five words were used in the *me* category (self, me, I, mine, and us), five words were used in the *not-me* category (his, other ("他人" in Chinese), other ("别人" in Chinese), others, and they). During the data-collection phase, each of the *me* and the *not-me* word was presented 16 times, and the *positive* and the *negative* words were all presented without repetition. All stimuli included two Chinese characters and were presented on a gray background in black Song font, with a vertical visual angle of 0.45° and a horizontal visual angle of 0.9°.

Please Insert **Figure 2** here

2.2.3 Questionnaire

The Rosenberg Self-esteem Scale is a 10-item scale for the measurement of global feelings of self-worth or self-acceptance (Rosenberg, 1965) and was used to assess the explicit self-esteem level of participants. In this assessment, participants rate their agreement with each self-describing item by using a 4-point Likert scale (from 1 = totally agree to 4 = totally disagree). The full range of the RSES score is from 4 to 40, and higher scores indicate higher explicit self-esteem. We compared the RSES scores

between the dysphoric and the control groups to check if explicit self-esteem was, as consistently suggested in previous studies (Roberts et al., 2015; Smeijers et al., 2017; van Tuijl et al., 2016), significantly lower among individuals with elevated depressive symptoms, in comparison to controls without depressive symptoms.

2.3 EEG recording and preprocessing

EEG signals were recorded from 64 scalp sites using Ag-AgCl electrodes mounted on an elastic cap (Brain Products, Munich, Germany), with the online reference electrode on the FCz site and the ground electrode on the mid-line of frontal scalp area (AFz site). Electrooculograms (EOGs) were recorded with an electrode below the right eye. Both EEG and EOG signals were amplified using a 0.05–100-Hz band-pass filter and continuously sampled at 500 Hz. All inter-electrode impedances were maintained below 5 kΩ for on-line recording.

During offline pre-processing, EEG signals were re-referenced to the average signal at the mastoid electrodes (Luck, 2005) and a low-pass filter was applied (30 Hz; 24dB/octave). A semi-automatic ocular correction based on independent component analysis (ICA) was used to eliminate potential eye movement-related artifacts. The ERP waveforms were time-locked to the onset of the stimuli, and the time-window included a 200-ms pre-stimulus baseline and a post-stimulus duration of 1000 ms. Trials with EOG voltage that exceeded ±100 μV or ones that were contaminated with artifacts due to amplifier clipping of peak-to-peak deflection greater than ±100 μV

during the analyzed epochs were excluded from averaging. Trials with incorrect responses were also excluded. The ERPs for all the remaining trials within the self-positivity condition and the self-negativity condition were then separately averaged. The mean number of trials contributing to the average ERPs was 308 for the self-positivity condition and 298 for the self-negativity condition. There was no significant difference of the accepted trials between the dysphoric and the control groups (for the self-positivity condition: t (56) = .77, p = .44; for the self-negativity condition: t (56) = .14, p = .89). The grand mean ERPs for the self-positivity and the self-negativity conditions were then calculated by averaging the individual ERPs in each group.

2.4 Data analysis

2.4.1 Behavioral data analysis

Three indices were applied to evaluate implicit self-esteem: reaction time, accuracy, and *D*-score (an index of the IAT effect that is calculated form reaction time) (Greenwald & Banaji, 1995; Greenwald, Nosek, & Banaji, 2003). The means of reaction time and accuracy were separately calculated for the self-positivity and the self-negativity condition after excluding trials with incorrect responses. The analysis of variance (ANOVA) of both reaction time and accuracy were submitted with self-association (self-positivity and self-negativity) as the within-subject variable, while group (dysphoric and control) as the between-subject variable. For the *D*-score, we first calculated the difference of self-negativity means minus self-positivity means

in reaction time, and then divided that difference by the standard deviation for all reaction times in these two conditions (Greenwald et al., 2003). The one-sample *t* test was conducted separately for the dysphoric and the control groups to compare their *D* score to zero. The significantly higher *D* score (as compared to zero) indicates more positive self-attitude as compared to attitude towards others. The independent-sample *t* test was applied to investigate the difference in *D* scores between the dysphoric and the control groups to explore whether the dysphoric individuals exhibited less positive self-attitude than the controls. The means of the RESE scores were calculated separately for the dysphoric and control groups, and a *t* test of the mean scores was conducted to compare the difference of explicit self-esteem between those two groups.

2.4.2 EEG data analysis

As suggested by previous IAT studies, LPC responses usually occur over a long temporal course, where early occurring LPC (within approximately 300 – 400 ms post-stimulus latency) best reflects automatic attentional allocation (Grundy et al., 2015; Yang & Zhang, 2009) and late occurring LPC best reflect the efficacy of stimulus evaluation and categorization (Wu et al., 2016). Guided by previous literature (Grundy et al., 2015) and the visual inspection of the grand-averaged waveforms, mean amplitudes of the LPC were calculated separately for time windows of 300–400 ms, 400–600 ms, and 600–1000 ms after stimulus onset. Several electrodes were selected over central, parietal, and occipital sites (CP1, CPz, CP2, P1,

Pz, P2, PO3, POz, and PO4) based on visual inspection of the grand averaged topographies. The same electrode selection was also applied in a previous study that similarly employed Chinese word stimuli and a healthy Chinese sample (Wu et al., 2016).

For each time window, mean amplitude values were entered into a four-way ANOVA, with self-association (self-positivity and self-negativity), anterior-posterior (central-parietal, parietal, and parietal-occipital), and laterality (left, midline, and right) as the within-subject factors, while group (dysphoric and control) was used as the between-subject factor. The main purpose of these analyses was to examine the potential behavioral and brain response differences between the dysphoric group and the control group during performance of implicit self-positivity and self-negativity categorizations. We report the interaction effects including self-association and group in the main text, and all topographic effects of the ERP analysis are reported in **Supplementary data 3**.

Whenever a significant interaction was found, post-hoc analyses were conducted to test the main effect of group in the self-positivity and the self-negativity conditions, and also the main effect of self-association in the dysphoric and the control groups. For both behavioral and ERP analysis, a significance level of .05 was used and the degrees of freedom of the *F*-ratio were corrected for violations of spherical assumptions using the Greenhouse-Geisser method. The Bonferroni correction

method was used for both ANOVA results and post hoc comparisons to control for possible type I error due to multiple comparisons. Partial eta squared (η_p^2) values were calculated and reported to demonstrate the effect size of significant ANOVA results.

3 Results

3.1 Behavioral results

3.1.1 Implicit self-esteem

Reaction time

The analysis showed no other significant main or interaction effects related to reaction time except a significant main effect of self-association [F (1, 56) = 183.10, p < .001, η_p^2 = 0.77]. Participants responded faster in the self-positivity condition (M = 623.23 ms, SD = 43.80) relative to their response time in the self-negativity condition (M = 673.52 ms, SD = 43.70; Figure 3a-1).

Accuracy

The analysis revealed no other significant main or interaction effects related to accuracy other than a significant effect of self-association [F (1, 56) = 53.18, p < .001, η_p^2 = 0.49]. The accuracy was higher in the self-positivity condition (M = 0.96, SD = 0.02) than that in the self-negativity condition (M = 0.93, SD = 0.04; see Figure 3a-2).

D-score (the IAT effect for reaction time)

No significant group difference was observed for *D*-scores (t (56) = 0.13, p = 0.90, Cohen's d = 0.004; see Figure 3a-3). The *D*-scores were significantly higher than zero for both the dysphoric group (M = 0.45, SD = 0.28; t (27) = 8.60, p < .001; Cohen's d = 1.62) and the control group (M = 0.46, SD = 0.22; t (29) = 11.60, p < .001; Cohen's d = 2.12).

3.1.2 Explicit self-esteem

RSES

The t test of the RSES scores showed a significant group effect (t (56) = 2.90, p = 0.005; Cohen's d = 0.76). The dysphoric group (M = 24.18, SD = 4.27) exhibited lower RSES scores than the control group (M = 28.03, SD = 5.71; see Figure 3b).

Please Insert **Figure 3** Here

3.2 ERP results

LPC (300–400 ms)

The results showed neither a significant effect of self-association [F (1, 56) = 0.57, p = 1.00, η_p^2 = 0.01] nor group [F (1, 56) = 1.53, p = 0.66, η_p^2 = 0.03]. The self-association × group interaction was also non-significant [F (1, 56) = 4.36, p = 0.12, η_p^2 = 0.07].

LPC (400–600 ms)

There was a significant interaction effect of self-association and group [F (1, 56) =

18.58, $p < .001$, $\eta_p^2 = 0.25$]. The post-hoc analysis of this interaction showed a significant effect of self-association in both the control group [F (1, 29) = 14.06, $p < .001$, $\eta_p^2 = 0.33$] and the dysphoric group [F (1, 28) = 5.18, $p = 0.03$, $\eta_p^2 = 0.16$]. In the control group, the self-positivity condition (M = 6.04 uV, SD = 2.69) induced greater amplitudes than the self-negativity condition (M = 5.24 uV, SD = 2.81). In the dysphoric group, however, the self-negativity condition induced larger amplitudes (M = 6.63 uV, SD = 2.36) than the self-positivity condition (M = 6.23 uV, SD = 2.21). Moreover, a significant group difference was found in the self-negativity condition [F (1, 56) = 4.11, $p = .04$, $\eta_p^2 = 0.07$] but not in the self-positivity condition [F (1, 56) = 0.09, $p = 0.77$, $\eta_p^2 = 0.002$]. Compared to the control group (M = 5.24 uV, SD = 2.81), the dysphoric group (M = 6.63 uV, SD = 2.36) exhibited larger amplitudes in the self-negativity condition (see Figure 4).

LPC (600-1000 ms)

Similar to the ERPs in the previous time window, the results showed a significant interaction of self-association and group [F (1, 56) = 15.76, $p < .001$, $\eta_p^2 = 0.22$]. The post-hoc analysis of this interaction showed a significant main effect of self-association in the dysphoric group [F (1, 27) = 19.56, $p < .001$, $\eta_p^2 = 0.42$]. The self-negativity condition (M = 3.30 uV, SD = 2.25) induced larger amplitudes than the self-positivity condition (M = 2.30 uV, SD = 1.85). The main effect of self-association was not significant in the control group [F (1, 29) = 0.51, $p = 0.48$, $\eta_p^2 = 0.02$]. Moreover, the post-hoc analysis of this interaction also showed a significant group

effect in the self-negativity condition [F (1, 56) = 5.48, p = 0.02, η_p^2 = 0.09], but not in the self-positivity condition [F (1, 56) = 0.28, p = 0.60, η_p^2 = 0.01]. Relative to the control group (M = 1.91 uV, SD = 2.25), the dysphoric group (M = 3.30 uV, SD = 2.25) exhibited larger amplitude in the self-negativity condition (see Figure 4).

Please Insert **Figure 4** Here

3.3. Correlation analysis

According to these results, opposite LPC response patterns were observed for the dysphoric group and the control group within both 400 – 600 ms and 600 – 1000 ms time windows. However, there was no significant difference between these two groups for behavioral performance during the IAT procedure. Correlation analyses were then conducted to investigate whether the behavioral performance or the brain responses would be associated with the participants' depressive state. First, correlation analyses by Pearson's correlation coefficient analysis were conducted between the participants' depressive state (as measured by BDI-II scores) and their explicit self-esteem level and IAT performance (respectively reflected by RSES scores and *D*-scores). We also conducted correlation analyses between the behavioral indices (BDI-II scores, RSES scores and *D*-scores) and the ERP responses (means of LPC amplitudes in the self-positivity condition and the self-negativity condition, respectively) during the 400 – 600 ms and 600 – 1000 ms time windows. Other time windows were not analyzed because no significant effects were found there.

The results showed a significant negative correlation between the BDI-II scores and the RSES scores (r (58) = -0.32, p = 0.01), indicating that the higher the depressive score, the lower the explicit self-esteem. The correlation between the BDI-II scores and the *D*-scores was not significant (r (58) = -0.03, p = 0.82). There were significant positive correlations between the BDI-II scores and the LPC amplitude in the self-negativity conditions within both 400 – 600 ms (r (58) = 0.29, p = 0.03) and 600 – 100 ms (r (58) = 0.28, p = 0.03) time windows. **Table 2** provides the resultant correlation matrix. These results suggested that the participants' behavioral performance was not related to their depressive level. In addition, the participants' brain responses were related to their depressive level, but were not related to their behavioral performance during the IAT.

Please Insert **Table 2** Here

4 Discussion

The aim of this study was to explore the differences in electrophysiological brain event-related responses between the dysphoric group and the control group when they were performing IAT. As we expected, the significant group-related differences were observed during 400-1000 ms post-stimulus (LPC), but not for earlier time windows. Interestingly, the dysphoric participants and the controls exhibited opposite LPC responses to the self-positivity condition (in which *self* was associated with *negative* words while *others* were associated with *positive* words) and the self-negativity condition (in which *self* was associated with *negative* words while *others* were

associated with *positive* words) during the IAT.

For the control group, consistent with our hypothesis, larger LPC amplitude was observed in the self-positivity condition, relative to the self-negativity condition, within the 400 ms to 600 ms time window. This result is consistent with previous findings that, in healthy participants, the self-positivity condition usually elicited larger LPC amplitude as compared to the self-negativity condition (labeled as "congruent" versus "incongruent" conditions, or "compatible" versus "incompatible" conditions in these studies) ((Fleischhauer et al., 2014; Wu et al., 2016); for opposite response pattern, see also (Grundy et al., 2015)). In these studies, the increased LPC amplitude was interpreted as indicative of more voluntary attention and enhanced stimulus evaluation, thus reflecting more efficient categorization during the IAT (Fleischhauer et al., 2014; Wu et al., 2016; Xiao et al., 2015). Our result therefore suggests that the control participants might more efficiently categorize the self-positivity pairings compared to the self-negativity pairings. Together with the previous findings, we thus provide additional electrophysiological evidence on the implicit self-positivity bias among healthy individuals (Y. Chen et al., 2014; Egenolf et al., 2013; Wu et al., 2014).

For the dysphoric group, we predicted no significant difference in brain responses between the self-positivity and the self-negativity conditions. However, greater LPC amplitudes were observed in the self-negativity condition relative to the self-positivity

condition, from 400 ms to 1000 ms post-stimulus latency. Consistent with the interpretation for the control group, the observed LPC response patterns suggest that the dysphoric participants, as opposed to the controls, might continuously engage more voluntary attention and stimulus evaluation in the self-negativity condition, and thus be more efficient in self-negativity categorization than in self-positivity categorization. The result could thus be taken to indicate stronger association of self and negative attributes compared to self and positive attributes in people with depressive symptoms. This finding provides support for Beck's cognitive theory of depression (Beck, 1967). According to that theory, negative self-schema is a core symptom of depression (Beck, Rush, Shaw, & Emery, 1979; Clak, Beck, & Alford, 1999) and plays an important role in the development, maintenance, and relapse of depressive disorder (J. M. G. Williams, 1997). It is reasonable that individuals with elevated depressive symptoms might start to show negative self-schema, thus tending to associate themselves with negative attributes, and consequently, be more efficient in categorization of the self-is-negative pairing relative to the self-is-positive pairings. The result thus implies that facilitated self-negativity categorization probably contributes to lowered implicit self-esteem among dysphoric people.

There was no significant condition difference or group difference in the earliest LPC time window (within 300 – 400 ms latency). As previously mentioned, the earlier occurring LPC (sometimes labeled as P3a) is usually associated with automatic stimulus processing, such as automatically attentional capture to novelty or emotional

salient stimuli (Polich, 2007). The absence of a conditional difference here is thus consistent with the fact that the stimulus words we used in this study were identical Chinese words without apparent perceptual difference, so none of the word categories should have advantages for the capture of participants' automatic attention. In addition, ERP responses did not differ between the dysphoric group and the control group in the 300 – 400 ms time window. This result is consistent with the previous findings that individual differences during the IAT are mainly driven by the late mental processes that are related to cognitive control, rather than by early processes that are related to perceptual processing (Schiller et al., 2016).

The LPC results presented here suggest a lowered implicit self-esteem among the dysphoric participants compared to that of the controls. Unexpectedly, however, the dysphoric group and the control group did not show significant differences in IAT behavioral performance. Both groups exhibited faster and more accurate key-responses in the self-positivity condition, relative to the self-negativity condition. The behavioral results are thus not in line with the ERP results by showing an undifferentiated positive bias in implicit self-esteem between the dysphoric and control groups. However, for the following reasons, the behavioral indices used here might not be as sensitive as the ERP responses for the detection of group differences. First and most importantly, the IAT behavioral indices (e.g., reaction time and accuracy) might be affected by practice (Röhner, Schröder-Abé, & Schütz, 2011). In our study, we asked the participants to continue practicing until they reached a

relatively high accuracy (85%) before moving to the data-collection phase. This practice session was important because it helped the participants to fully understand the requirement of each experimental phase, and thus enabled us to have a sufficient number of valid trials (trials with correct response) for analysis of the ERP data. However, the additional practice might have also contributed to ceiling effects in both reaction time and accuracy (e.g., as suggested by the high mean accuracies and low standard deviations in both the self-positivity condition (M = 0.96, SD = 0.02) and the self-negativity condition (M = 0.93, SD = 0.04)), limiting detection of any potential group differences in behavioral performance. Second, as compared to the behavioral indices, the ERP responses have high temporal-resolution, so should be more informative in detecting individual differences, especially for differences that might pertain to only some specific phase of the ongoing processing (such as the observed late processing during the IAT).

We explored if behavioral IAT performance or ERP responses were correlated with the depressive levels of participants. The results showed no significant correlation between participants' behavioral IAT performance and amount of depressive symptoms. However, the LPC amplitudes in the self-negativity condition were positively correlated with participants' depressive symptoms, indicating that the facilitated self-negativity categorization was positively related to the increase of individuals' depressive symptoms. This result therefore supports our speculation that ERP responses, rather than behavioral performance, are more associated with the

depressive-related group differences in this study. The enhanced LPC amplitudes in the self-negativity association and its correlation to higher scores in depressive symptom scale (e.g., BDI-II) could therefore be interpreted as a neural index of lowered implicit self-esteem in the current study.

Taken together, these findings provide neural evidence for lowered implicit self-esteem in individuals with elevated depressive symptoms, probably driven by facilitated self-negativity association. To the best of our knowledge, this is the first investigation of brain activity related to implicit self-esteem in individuals with depressive symptoms. These findings extend our understanding of the relationship between implicit self-esteem and depression. However, we are cautious about drawing strong conclusions given the inconsistence between our ERP results and the behavioral results. Future studies could further test these findings by using a variety of implicit paradigms. In addition, the current study invited pre-clinical individuals, instead of clinical patients, because we wanted to eliminate potential effects of drug interference in clinical samples. Future studies could test the generality of these findings among clinically depressed samples and in different depression subgroups. Moreover, it is important to investigate whether improvement in implicit self-esteem, for instance due to a successful intervention, induces changes in ERP responses.

5 Conclusion

Employing the IAT paradigm in conjunction with EEG recordings, this study explored

the brain responses related to implicit self-esteem among individuals with elevated amounts of depressive symptoms (dysphoric participants). Interestingly, although the dysphoric and the control groups did not differ in behavioral performance, they showed opposite response patterns in brain activities during the IAT. The controls exhibited significantly larger LPC amplitudes, reflecting more efficient categorization, in the self-positivity condition, relative to the self-negativity condition, while the opposite pattern was observed for the dysphoric group. The results suggest facilitated categorization for self-negativity word pairings in dysphoric participants, implying that the self-is-negative association, as compared to the self-is-positive association, might be more congruent with their implicit self-attitude. These findings provide the first electrophysiological evidence for lowered implicit self-esteem among individuals with elevated amounts of depressive symptoms.

Acknowledgements

This work was supported by the National Natural Science Foundation of China [Grant NO. 31871130 and 31571153], the Innovative Team Program in Higher Education of Guangdong, China [Grant NO. 2015KCXTD009], the Major Program of Guangdong, China [Grant NO. 2016KZDXM009], and the Shenzhen Basic Research Scheme [Grant NO. JCYJ20150729104249783]

References

Beck, A. T. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. New York, USA: Hoeber Medical Division, Harper and Row.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression*. New York, USA: Guilford Press.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory - Second edition*. San Antonio, Texas, USA: Psychological Corporation.

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: the blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*(4), 631-643. https://doi.org/10.1037/0022-3514.79.4.631

Chen, L., Zhou, H., Gu, Y., Wang, S., Wang, J., Tian, L., . . . Zhou, Z. (2018). The Neural Correlates of Implicit Cognitive Bias Toward Internet-Related Cues in Internet Addiction: An ERP Study. *Frontiers in Psychiatry, 9*, 421-421. https://doi.org/10.3389/fpsyt.2018.00421

Chen, Y., Zhong, Y., Zhou, H., Zhang, S., Tan, Q., & Fan, W. (2014). Evidence for implicit self-positivity bias: an event-related brain potential study. *Experimental Brain Research, 232*(3), 985-994. https://doi.org/10.1007/s00221-013-3810-z

Clak, D. A., Beck, A. T., & Alford, B. A. (1999). *Scientific foundations of cognitive theory and therapy of depression.* New York, USA: Wiley.

Coates, M. A., & Campbell, K. B. (2010). Event-related potential measures of processing during an Implicit Association Test. *Neuroreport, 21*(16), 1029-1033. https://doi.org/10.1097/WNR.0b013e32833f5e7d

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, Michigan. New Jersey Lawrence Erlbaum Associates.

Delaveau, P., Jabourian, M., Lemogne, C., Allaïli, N., Choucha, W., Girault, N., . . . Fossati, P. (2016). Antidepressant short-term and long-term brain effects during self-referential processing in major depression. *Psychiatry Research: Neuroimaging, 247*, 17-24. https://doi.org/10.1016/j.pscychresns.2015.11.007

Egenolf, Y., Stein, M., Koenig, T., Holtforth, M. G., Dierks, T., & Caspar, F. (2013). Tracking the implicit self using event-related potentials. *Cognitive Affective and Behavioral Neuroscience, 13*(4), 885-899. https://doi.org/10.3758/s13415-013-0169-3

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Fleischhauer, M., Strobel, A., Diers, K., & Enge, S. (2014). Electrophysiological evidence for early perceptual facilitation and efficient categorization of self-related stimuli during an Implicit Association Test measuring neuroticism. *Psychophysiology, 51*(2), 142-151. https://doi.org/10.1111/psyp.12162

Franck, E., Raedt, R. D., Dereu, M., & Abbeele, D. V. d. (2007). Implicit and explicit self-esteem in currently depressed individuals with and without suicidal ideation. *Journal of Behavior Therapy and Experimental Psychiatry, 38*(1), 75-85. https://doi.org/10.1016/j.jbtep.2006.05.003

Gable, P. A., Adams, D. L., & Proudfit, G. H. (2015). Transient tasks and enduring emotions: the impacts of affective content, task relevance, and picture duration on the sustained late positive

potential. *Cognitive, Affective, and Behavioral Neuroscience, 15*(1), 45-54. https://doi.org/10.3758/s13415-014-0313-8

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4-27. https://doi.org/10.1037/0033-295x.102.1.4

Greenwald, A. G., & Farnham, D. S. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*(6), 1022-1038. https://doi.org/10.1037/0022-3514.79.6.1022

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring Individual Differences in Implicit Cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197-216. https://doi.org/10.1037/0022-3514.85.2.197

Grundy, J. G., Benarroch, M. F. F., Lebarr, A. N., & Shedden, J. M. (2015). Electrophysiological correlates of implicit valenced self-processing in high vs. low self-esteem individuals. *Social Neuroscience, 10*(1), 100-112. https://doi.org/10.1080/17470919.2014.965339

Healy, G. F., Boran, L., & Smeaton, A. F. (2015). Neural patterns of the Implicit Association Test. *Frontiers in Human Neuroscience, 9*(605). https://doi.org/10.3389/fnhum.2015.00605

Huang, X., & Zhang, S. (1992). Desirability, meaningfulness and familiarity ratings of 562 personality-trait adjectives. *Psychological Science (in Chinese), 5*, 17 - 22.

Ibáñez, A., Gleichgerrcht, E., Hurtado, E., González, R., Haye, A., & Manes, F. F. (2010). Early Neural Markers of Implicit Attitudes: N170 Modulated by Intergroup and Evaluative Contexts in IAT. *Frontiers in Human Neuroscience, 4*(188), 188. https://doi.org/10.3389/fnhum.2010.00188

Jabben, N., Jong, P. J. d., Kupka, R. W., Glashouwer, K. A., Nolen, W. A., & Penninx, B. W. J. H. (2014). Implicit and explicit self-associations in bipolar disorder : A comparison with healthy controls and unipolar depressive disorder. *Psychiatry Research, 215*(2), 329-334. https://doi.org/10.1016/j.psychres.2013.11.030

Kesting, M.-L., Mehl, S., Rief, W., Lindenmeyer, J., & Lincoln, T. M. (2011). When paranoia fails to enhance self-esteem: explicit and implicit self-esteem and its discrepancy in patients with persecutory delusions compared to depressed and healthy controls. *Psychiatry Research, 186*(2-3), 197-202. https://doi.org/10.1016/j.psychres.2010.08.036

Leeuwis, F. H., Koot, H. M., Creemers, D. H., & van Lier, P. A. (2015). Implicit and explicit self-esteem discrepancies, victimization and the development of late childhood internalizing problems. *Journal of Abnormal Child Psychology, 43*(5), 909-919. https://doi.org/10.1007/s10802-014-9959-5

Lemmens, L. H., Roefs, A., Arntz, A., van Teeseling, H. C., Peeters, F., & Huibers, M. J. (2014). The value of an implicit self-associative measure specific to core beliefs of depression. *Journal of Behavior Therapy and Experimental Psychiatry, 45*(1), 196-202. https://doi.org/10.1016/j.jbtep.2013.10.006

Lou, Y., Lei, Y., Mei, Y., Leppänen, P. H. T., & Li, H. (2019). Review of Abnormal Self-Knowledge in Major Depressive Disorder. *Front Psychiatry, 10*, 130. https://doi.org/10.3389/fpsyt.2019.00130

Luck, S. J. (2005). An introduction to event-related potentials and their neural origins. In Luck, S.J. (Ed.), *An introduction to the eventrelated potential technique* (Vol 1, pp 2-48). *Cambridge, Mass, London, UK*. MIT Press

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

Pahl, S., & Eiser, J. R. (2005). Valence, comparison focus and self-positivity biases: does it matter whether people judge positive or negative traits? *Experimental Psychology, 52*(4), 303-310. https://doi.org/10.1027/1618-3169.52.4.303

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10). https://doi.org/10.1016/j.clinph.2007.04.019

Röhner, J., Schröder-Abé, M., & Schütz, A. (2011). Exaggeration is harder than understatement, but practice makes perfect! Faking success in the IAT. *Experimental Psychology, 58*(6), 464-472. https://doi.org/10.1027/1618-3169/a000114

Randenborgh, A. v., Pawelzik, M., Quirin, M., & Kuhl, J. (2016). Bad roots to grow: deficient implicit self‐evaluations in chronic depression with an early onset. *Journal of Clinical Psychology, 72*(6), 580-590. https://doi.org/10.1002/jclp.22275

Risch, A. K., Buba, A., Birk, U., Morina, N., Steffens, M. C., & Stangier, U. (2010). Implicit self-esteem in recurrently depressed patients. *Journal of Behavior Therapy and Experimental Psychiatry, 41*(3), 199-206. https://doi.org/10.1016/j.jbtep.2010.01.003

Roberts, J. E., Porter, A., & Vergara-Lopez, C. (2015). Implicit and explicit self-esteem in previously and never depressed individuals: Baseline differences and reactivity to rumination. *Cognitive Therapy and Research, 40*(2), 164-172. https://doi.org/10.1007/s10608-015-9732-2

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Schiller, B., Gianotti, L. R. R., Baumgartner, T., Nash, K., Koenig, T., & Knoch, D. (2016). Clocking the social mind by identifying mental processes in the IAT with electrical neuroimaging. *Proceedings of the National Academy of Sciences, 113*(10), 2786. https://doi.org/10.1073/pnas.1515828113

Smeijers, D., Vrijsen, J. N., Oostrom, I. v., Isaac, L., Speckens, A., Becker, E. S., & Rinck, M. (2017). Implicit and explicit self-esteem in remitted depressed patients. *Journal of Behavior Therapy and Experimental Psychiatry, 54*, 301-306. https://doi.org/10.1016/j.jbtep.2016.10.006

van Nunspeet, F., Ellemers, N., Derks, B., & Nieuwenhuis, S. (2014). Moral concerns increase attention and response monitoring during IAT performance: ERP evidence. *Social Cognitive and Affective Neuroscience, 9*(2), 141-149. https://doi.org/10.1093/scan/nss118

van Tuijl, L. A., Glashouwer, K. A., Bockting, C. L., Tendeiro, J. N., Penninx, B. W., & de Jong, P. J. (2016). Implicit and explicit self-esteem in current, remitted, recovered, and comorbid depression and anxiety Disorders: The NESDA study. *PLoS One, 11*(11), e0166116. https://doi.org/10.1371/journal.pone.0166116

Vianello, M., & Bar-Anan, Y. (2020). Can the Implicit Association Test Measure Automatic Judgment? The Validation Continues. *Perspectives on Psychological Science*. https://doi.org/10.1177/1745691619897960

Wang, Y., Zhou, L., & Luo, Y. (2008). The Pilot Establishment and Evaluation of Chinese Affective Words System. *Chinese Mental Health Journal (in Chinese), 22*(8), 608-612.

Watson, L. A., Dritschel, B., Obonsawin, M. C., & Jentzsch, I. (2007). Seeing yourself in a positive light: brain correlates of the self-positivity bias. *Brain Research, 1152*(1), 106-110.

https://doi.org/10.1016/j.brainres.2007.03.049

Williams, J. K., & Themanson, J. R. (2011). Neural correlates of the implicit association test: evidence for semantic and emotional processing. *Social Cognitive and Affective Neuroscience, 6*(4). https://doi.org/10.1093/scan/nsq065

Williams, J. M. G. (1997). Depression. In D. M. Clark & C. G. Fairburn (Eds.), *Science and practice of cognitive behaviour therapy* (pp. 259–283). Oxford, UK: Oxford University Press.

Wu, L., Cai, H., Gu, R., Luo, Y. L. L., Zhang, J., Yang, J., . . . Ding, L. (2014). Neural manifestations of implicit self-esteem: an ERP study. *PLoS One, 9*(7), e101837. https://doi.org/10.1371/journal.pone.0101837

Wu, L., Gu, R., Cai, H., & Zhang, J. (2016). Electrophysiological evidence for executive control and efficient categorization involved in implicit self-evaluation. *Social Neuroscience, 11*(2), 153-163. https://doi.org/10.1080/17470919.2015.1044673

Xiao, F., Zheng, Z., Wang, Y., Cui, J., & Chen, Y. (2015). Conflict monitoring and stimulus categorization processes involved in the prosocial attitude implicit association test: Evidence from event-related potentials. *Social Neuroscience, 10*(4), 1-10. https://doi.org/10.1080/17470919.2014.1003598

Yang, J., & Zhang, Q. (2009). P300 as an index of implicit self-esteem. *Neurological Research.* https://doi.org/10.1179/174313209x431138

Tables

Table 1 Demographic description for participants in the dysohoric and control groups.

Description	Units	Dysphoric	Control
Participants	*N* (females)	28 (18)	30 (17)
Age	*M* ± *SD* (range) years	20.39±1.81 (18~24)	19.90±1.94 (18~24)
BDI-II	*M* ± *SD* (range) points	20.04±5.86 (14~42)	2.13±2.47 (0~11)

Note. *N* = number; *M* ± *SD* = means ± standard deviations; BDI-II = Beck Depression Inventory – II

Table 2 Correlations among the depressive state, explicit self-esteem level, and behavioral performance of participants during the IAT, and the LPC amplitudes under two conditions during 400 - 600 ms and 600 - 1000 ms time windows

Variables	BDI-II	RSES	*D*-score	LPC Amp. *Self-positivity*	LPC Amp. *Self-negativity*
BDI-II	–				
RSES	-0.32*	–			
D-score	-0.03	-0.24	–		
400–600 ms					
LPC Amp. *Self-positivity*	0.12	0.05	0.11	–	
LPC Amp. *Self-negativity*	0.29*	-0.07	0.14	0.89***	–
600–1000 ms					
LPC Amp. *Self-positivity*	0.03	-0.08	0.01	–	
LPC Amp. *Self-negativity*	0.28*	-0.14	0.19	0.86***	–

Note. IAT = Implict Association Task; BDI-II = Beck Depressive Inventory - II; RSES = Rosenberg Self-Esteem Scale; LPC = Late Positive Component; Amp. = Amplitude. *p < .05; ***p < .001

Figure Captions

Figure 1 (a) Illustration of the four phases during the self-positivity and the self-negativity blocks. The labeled black dots indicate the correct responses in each phase. The order of the two blocks and the assignment of the *F* key and *J* key were both counter-balanced between subjects. (b). Illustration of the IAT procedure during the data-collection phase. The stimulus words were presented one-by-one in written form according to a fully random order. Participants sorted the words to one of the four categories (*me*, *not-me*, *positive*, or *negative*) by pressing two keys (*F* or *J*). In the self-positivity block (see b (1)), the *me* words (e.g. me, mine, us, ….) and the *positive* words (e.g. smart, brave, honest, ….) shared the same key (e.g., *F*), whereas the *not-me* words (e.g., other, his, they, ….) and the *negative* words (e.g., fool, coward, dishonest, ….) shared a separate key (e.g., *J*). In the self-negativity block (see b (2)), the *me* words and the *negative* words shared the same key, whereas the *not-me* words and the *positive* words shared a separate key. All the stimulus words consisted of two simplified Chinese characters (for fully translated examples: see **Supplementary data 2** Table 1).

Figure 2 Means of the characters' stroke numbers, meaningfulness, familiarity, arousal, and desirability as rated in the pilot pre-experiment for positive and negative stimulus words. Error bars represent standard errors. $^{***}p < .001$

Figure 3 Indices of (a). Implicit self-esteem: (1). Means of reaction times (RTs), (2). Means of accuracy (ACC), and (3) Means of the *D*-score (calculated as the division of the difference of self-negativity RT means minus self-positivity RT means by the standard deviation of all RTs in the two conditions) in the IAT; and (b). Explicit self-esteem: Means of the Rosenberg Self-Esteem Scale (RSES) scores. Error bars represent standard errors, $^{***}p < .001$, $^{**}p < .01$

Figure 4 Illustration of LPC waveforms within 300–1000 ms (presented separately for time windows of 300–400 ms, 400–600 ms, and 600–1000 ms) in the dysphoric group (solid lines) and the control group (dashed lines) during the self-positivity (orange lines) and the self-negativity (blue lines) conditions. The Pz site and the corresponding scalp topographies of difference waves (self-positivity minus self-negativity conditions) are illustrated for both the two groups, separately. Orange shading indicates where the self-positivity condition showed greater waveforms than the self-negativity condition, and blue shading indicates where the self-negativity condition showed greater waveforms.