JYU DISSERTATIONS 700

Yixue Lou

Depression and Self-Knowledge

Behavioral and Brain Responses of Reflected Self-Evaluation and Implicit Self-Esteem in Sub-Clinical Depression

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF EDUCATION AND PSYCHOLOGY

JYU DISSERTATIONS 700

Yixue Lou

Depression and Self-Knowledge

Behavioral and Brain Responses of Reflected Self-Evaluation and Implicit Self-Esteem in Sub-Clinical Depression

Esitetään Jyväskylän yliopiston kasvatustieteiden ja psykologian tiedekunnan suostumuksella julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa S212 lokakuun 31. päivänä 2023 kello 12.

Academic dissertation to be publicly discussed, by permission of the Faculty of Education and Psychology of the University of Jyväskylä, in building Seminariun, Old Festival Hall S212, on October 31, 2023, at 12 o’clock.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2023

Editors
Noona Kiuru
Department of Psychology, University of Jyväskylä
Päivi Vuorio
Open Science Centre, University of Jyväskylä

ISBN 978-951-39-9763-2 (PDF)
URN:ISBN:978-951-39-9763-2
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-9763-2

ABSTRACT

Lou, Yixue
Depression and self-knowledge: Behavioral and brain responses of reflected self-evaluation and implicit self-esteem in sub-clinical depression
Jyväskylä: University of Jyväskylä, 2023, 76 p. + original articles
(JYU Dissertations
ISSN 2489-9003; 700)
ISBN 978-951-39-9763-2 (PDF)

Depression often involves negative self-knowledge, with individuals viewing themselves in a distorted, negative way. The neurobiological mechanism underlying this phenomenon is not yet fully understood. This dissertation consists of three individual studies that investigate behavioral and brain responses of negative self-knowledge in depression. **Study I** reviewed existing studies and found that, for depressed individuals, negative self-knowledge has been associated with enhanced responses at late positive components and altered activity in cortical midline structures during negative biased direct self-evaluation, where individuals evaluate the self through their own perspective. **Study II** investigated brain activity of reflected self-evaluation, where individuals evaluate the self through another person's perspective, in participants with enhanced depressive symptoms (labeled as a dysphoric group). The functional magnetic resonance images were scanned while the participants were evaluating themselves according to others' opinions. Compared to a control group, the dysphoric group exhibited negative bias in behavioral ratings and altered brain activity in the bilateral tempo-parietal junction during the reflected self-evaluation. **Study III** investigated brain responses of depression-related low self-esteem, which is considered a consequence of the negative self-evaluation, in both a dysphoric group and a control group. The electroencephalogram was recorded during an implicit association task measuring implicit self-esteem. The results showed that, contrary to the control group, the dysphoric group exhibited an enhanced late positive brain response when the self was unconsciously associated with negative personality traits, compared to when the self was associated with positive personality traits, within the time window of 400–1,000 ms post-stimulus latency. The results suggest a facilitated self-is-negative association, reflecting low implicit self-esteem, in sub-clinical depression. Overall, this dissertation extends our understanding of Beck's cognitive theory of depression by providing behavioral and neuroimaging evidence for the negative reflected self-evaluation and the low implicit self-esteem related to depression. It also suggests that the self-negativity bias does not occur only in clinical depression, but also in sub-clinical populations with enhanced depressive symptoms.

Keywords: depression, dysphoria, self-negativity bias, reflected self-evaluation, implicit self-esteem, functional magnetic resonance imaging, event-related potentials

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Lou, Yixue
Masennus ja itsetunto: käyttäytymis- ja aivovasteet itsearviointiin ja implisiittiseen itsetuntoon subkliinisessä masennuksessa.
Jyväskylä: Jyväskylän yliopisto, 2023, 76 s. + alkuperäiset artikkelit
(JYU Dissertations
ISSN 2489-9003; 700)
ISBN 978-951-39-9763-2 (PDF)

Negatiivisesti korostunut itsetuntemus on yksi masennuksen keskeisistä oireista. Tällaisen negatiivisesti korostuneen minäkuvan vaikutuksesta masennuksesta kärsivillä henkilöillä on taipumus nähdä itsensä vääristyneellä negatiivisella tavalla. Tämän ilmiön taustalla olevaa neurobiologista mekanismia ei kuitenkaan vielä tunneta kovin hyvin. **Tutkimuksessa I** kävin läpi olemassa olevia tutkimuksia ja huomasin, että masentuneilla henkilöillä negatiivinen itsetuntemus on yhdistetty vahvistuneisiin reaktioihin myöhäisten positiivisten komponenttien osalta ja muuttuneeseen toimintaan aivokuoren keskilinjan rakenteissa negatiivisesti vääristyneen suoran itsearvioinnin aikana, jossa yksilöt arvioivat itseään omasta näkökulmastaan. **Tutkimuksessa II** aivotoimintaa tutkittiin ”heijastuneen” itsearvioinnin aikana, jolloin yksilöt arvioivat itseään ulkopuolisen henkilön näkökulmasta. Toiminnalliset magneettikuvat skannattiin, kun osallistujat arvioivat itseään muiden mielipiteiden mukaan. Verrattuna kontrolliryhmään, dysforinen ryhmä (henkilöitä joilla oli lisääntyneitä masennusoireita ilman varsinaista diagnoosia) arveli muiden liittävän heihin enemmän negatiivisia piirteitä. Dysforisella ryhmällä oli myös kontrolliryhmään nähden poikkeavaa aivotoimintaa tempo-parietaalisessa liitoskohdassa muiden mielipiteitä heijastavan itsearvioinnin aikana. **Tutkimuksessa III** tutkittiin sekä kontrolli- että dysforisessa ryhmässä EEG:hen (elektroenkefalogrammi) aivojen herätevasteita, jotka liittyvät alhaiseen itsetunnon, jota puolestaan pidetään negatiivisen itsearvioinnin seurauksena. Herätevasteita mitattiin implisiittistä itsetuntoa mittaavan implisiittisen assosiaatiotehtävän aikana. Tulokset osoittivat, että toisin kuin kontrolliryhmässä, dysforisessa ryhmässä esiintyi suurentunut positiivinen aivovaste myöhäisellä aikaikkunalla (400-1000 ms). Tämä suurentunut positiivisuus näkyi silloin, kun itseen yhdistettiin tiedostamatta negatiivisia persoonallisuuden piirteitä verrattuna siihen, kun itseen yhdistettiin positiivisia persoonallisuuden piirteitä. Nämä tulokset viittaavat siihen, että subkliiniseen masennukseen (dysforiaan) liittyy kasvanut negatiivinen assosiaatio itseen, joka heijastaa matalaa implisiittistä itsetuntoa. Kaiken kaikkiaan tämä väitöskirja laajentaa ymmärrystämme Beckin kognitiivisesta masennusteoriasta tarjoamalla käyttäytymis- ja neurokuvantamisnäyttöä masennukseen liittyvästä negatiivisesta heijastuneesta itsearvioinnista ja alhaisesta implisiittisestä itsetunnosta. Tulokseni viittaavat myös siihen, että itseen liitettyä negatiivista vinoumaa ei esiinny ainoastaan kliinisessä masennuksessa vaan myös subkliinisissä väestöryhmissä, joilla on lisääntyneitä masennusoireita.

Avainsanat: masennus, dysforia, korostunut negatiivinen minäkuva, toisten mielipiteitä heijastava itsearviointi, implisiittinen itsetunto, toiminnallinen magneettikuvaus, aivojen herätevaste

Author Yixue Lou
Department of Psychology
University of Jyväskylä
Email: yixue.y.lou@jyu.fi
ORCID: 0000-0002-0261-8344

Supervisors Professor Paavo H. T. Leppänen
Department of Psychology
University of Jyväskylä

Associate Professor Piia S. Astikainen
Department of Psychology
University of Jyväskylä

Senior Researcher Susannah C.S.A. Leppänen
Department of Psychology
University of Jyväskylä

Professor Hong Li
Department of Psychology
Sichuan Normal University

Professor Yi Lei
Department of Psychology
Sichuan Normal University

Reviewers Senior Research Fellow Mikko Peltola
Faculty of Social Sciences
University of Tampere

Professor Georgia Panayiotou
Department of Psychology
University of Cyprus

Opponent Senior Research Fellow Mikko Peltola
Faculty of Social Sciences
University of Tampere

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Paavo H.T. Leppänen, for his unwavering patience, invaluable guidance, unflagging support, and insightful feedback throughout my doctoral journey. I would also like express my heartfelt thanks to my co-supervisors, Associate Professor Piia S. Astikainen and Dr. Susannah C.S.A. Otieno-Leppänen, for their continuous care, assistance, and support, not only for my academic pursuits but also for the personal challenges I encountered during my time in Finland. Additionally, I am immensely thankful to my supervisors in China, Professor Hong Li and Professor Yi Lei, for their invaluable suggestions and steadfast support.

I also wish to express my appreciation to my opponent, Senior Research Fellow Mikko Peltola, and the external reviewer, Professor Georgia Panayiotou, for generously dedicating their time and effort to review my dissertation. Their kind comments and valuable suggestions have significantly contributed to the enhancement of my work.

I would like to extend my sincere gratitude to the Faculty of Education and Psychology at the University of Jyväskylä, and to the School of Psychology at Shenzhen University, for providing me with a stimulating and supportive academic environment that not only encouraged but also inspired me to continually push my boundaries and achieve my goals in all aspects of my work. I would also like to acknowledge and thank all my colleagues there for the enjoyable discussions and camaraderie we shared over the past years.

Big thank you to the scientific editor, Professor Noona Kiuru, as well as the editors and proofreaders who meticulously worked on this dissertation. Their careful attention to detail was instrumental in the successful publication of this work. I wish to also extend my appreciation to Tiina Volanen for her invaluable assistance in organizing and coordinating all aspects related to my graduation. I must also express my gratitude to all my co-authors and the participants in the studies included in my dissertation. Their contributions were indispensable to the success of the studies.

Lastly, I am profoundly grateful to my family and friends for their unwavering love and encouragement throughout my life. Special thanks are reserved for my husband, Hao Zhang, and my little pumpkin, Lotta, for their unwavering support and trust. Their understanding and belief in me sustained me through the long hours of research, and I cherish the sweet memories we created together.

Thank you all for being a part of my journey.

Espoo 12.09.2023
Yixue Lou

LIST OF ORIGINAL PUBLICATIONS

I. Lou, Y., Lei, Y., Mei, Y., Leppänen, P., & Li, H. (2019). Review of abnormal self-knowledge in major depressive disorder. *Frontiers in Psychiatry*, 10, 130. https://doi.org/10.3389/fpsyt.2019.00130

II. Lou, Y., Chen, S., Hu, G., Astikainen, P., Otieno-Leppänen, S., Lei, Y., & Leppänen, P. (2023). Altered temporoparietal junction activity during reflected self-evaluation in sub-clinical depression. *Cerebral Cortex*, bhad060. Advance online publication. https://doi.org/10.1093/cercor/bhad060

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

Taking into account the instructions given and the comments made by the co-authors, the author of this thesis contributed to the original publications as follows: the author designed the experiments, collected the data, conducted the analyses, wrote the manuscripts, and contributed to the review process.

FIGURES

TABLES

CONTENTS

1 INTRODUCTION

1.1 Depression and self-knowledge

“Know thyself” is an ancient Greek aphorism inscribed in the forecourt of the Temple of Apollo at Delphi. From a psychologist’s point of view, knowing the self is not only an aphorism but also one of the “most puzzling puzzles” to solve in psychological research (James, 1890). According to the definition by the American Psychological Association (*APA*, n.d.), the “self” comprises all of the characteristic attributes of a person, consciously and unconsciously, mentally and physically. According to the National Institute of Mental Health (*NIMH*, n.d.-b), the perception and understanding of the self includes being aware of the self, getting access to knowledge about the self, and/or making judgments about the self. And it can include current cognitive or emotional internal states of the self, traits of the self, and/or abilities of the self, either in isolation or in relationships with others (*NIMH*, n.d.-b). The perception and understanding of the self is organized into two sub-constructs: self-agency and self-knowledge. Self-agency refers to the ability to recognize oneself as the agent of one’s actions and thoughts, and the self-knowledge refers to the ability to make judgments about one’s current cognitive or emotional internal states, traits, and/or abilities (*NIMH*, n.d.-b). In the present study, I have mainly focused on investigating the self-knowledge in depression.

Depression, which is also known as major depressive disorder or clinical depression, is a mood disorder that causes symptoms that affect how people feel, think, and handle their daily activities (*NIMH*, n.d.-a). Typical depressive symptoms include neurovegetative dysfunction (appetite or sleep disturbances), cognitive dissonance (inappropriate guilt, feelings of worthlessness), aberrant psychomotor activities (agitation or retardation), and elevated suicidal ideation (Chesney et al., 2014; Fava et al., 2000; Klonsky et al., 2016). One of the core symptoms related to self in depression is the abnormally negative biased self-knowledge, which causes depressed individuals to think of themselves in a negative way (Beck, 1967). Based on the increasing literature on this topic, a generally accurate and positive biased self-knowledge, which is called self-positivity bias, is adaptive in maintaining one's mental health and well-being (Chen et al., 2014; Mezulis et al., 2004; Showers, 1992). Thus, for most people, information about the self is reasonably connected with positive feelings (Mezulis et al., 2004; Watson et al., 2007). However, according to Beck's cognitive theory of depression (Beck, 1967; Beck et al., 1979), a negative biased self, which is also called negative self-schema, is observed in depression. Thus, for those depressives, such as patients with clinical depression and sub-clinically dysphoric individuals, the self is abnormally associated with negative feelings (Bradley & Mathews, 1983; Swann, Wenzlaff, Krull, et al., 1992). Such negative biased self-knowledge can be reflected through overly negative self-evaluation and in inappropriate self-blame and low self-esteem (Beck et al., 1979; Derry & Kuiper, 1981; Lemogne et al., 2009; Northoff, 2007; Swann, Wenzlaff, & Tafarodi, 1992; Thew et al., 2017)

1.1.1 Negative direct and reflected self-evaluation in depression

According to classic social psychological theories, a person's self-knowledge is usually formed in two ways: the direct personal experience of meaningful life events; and the perception of how one is "seen" by others, developed through interaction with or feedback from others (Cooley, 1902; Mead, 1934; Srivastava, 2012). Therefore, researchers use the term "direct self-evaluation/direct self-appraisal" to refer to the self-knowledge that is directly abstract from one's previous experience, whereas "reflected self-evaluation/reflected self-appraisal" is used to refer to the self-knowledge that is drawn upon one's beliefs about how other people "see" them (see examples at (Ochsner et al., 2005; Pfeifer et al., 2017; Pfeifer et al., 2009; Van der Cruijsen et al., 2019; Yue & Huang, 2012). Specifically, in studies of direct self-knowledge, participants need to consider the question, "How do you think of yourself?", whereas in reflected self-knowledge studies they consider the question, "How do you think someone else will think of you?"

A commonly used paradigm for investigating direct self-evaluation is the Self-Referential Encoding Task (SRET) (Derry & Kuiper, 1981; Kuiper & Derry, 1982). In the SRET, participants are asked to evaluate whether a series of positive and negative personality trait words (such as bleak, dismal, loyal, or organized) can be used to describe themselves, and the self-evaluation is usually followed by an unexpected free-recall/recognition task that requires the participants to recall/recognize as many of the presented words as possible (Derry & Kuiper, 1981; Kuiper & Derry, 1982). The basis of the paradigm is that, during the self-referential processing, words that are consistent with the participants' self-knowledge would be processed more elaborately than those that are inconsistent with their self-knowledge, so the consistent words can be remembered more easily (Derry & Kuiper, 1981; Kuiper & Derry, 1982). By using the SRET, previous studies have found that depressed individuals endorsed more negative and less positive words as self-descriptive than non-depressed controls did (Dainer-best et al., 2017; Ji et al., 2017; Li et al., 2017; Shestyuk & Deldin, 2010). The depressed individuals also recalled/recognized more of the negative words that they had endorsed than the positive words that they had endorsed (Dainer-best et al., 2017; Shestyuk & Deldin, 2010), suggesting a negative bias in the direct self-evaluation in depression.

However, reflected self-evaluation has rarely been studied in the depressed population. Based on the literature, at the beginning of one's life, in particular during childhood and adolescence, individuals form a sense of the self first by learning and inferring how others perceive them (Srivastava, 2012). They compare themselves to others, look for a place within a certain community, and try to learn about the self by imagining other people's impression of them (Srivastava, 2012). Over time, those perceptions are gradually internalized into one's self-knowledge and become so-called perceived/reflected self-knowledge (Srivastava, 2012). The development of the reflected self-knowledge requires the theory of mind (ToM) (D'Argembeau et al., 2007; Pfeifer et al., 2009), which allows people to reason others' mental states and to predict people's behavior based on the reasoning (Saxe & Kanwisher, 2003). For instance, in order to evaluate how others see us, we have to put ourselves into other's "shoes", estimating whether they think of us positively or negatively. Based on previous studies, most people overestimate how positively they are seen in the "eyes" of others because of the influence of self-positivity bias (Carlson & Barranti, 2016; Carlson & Kenny, 2012). However, in studies examining the role of depression in reflected self-evaluation, researchers have found that depression negatively biased how people believe they are seen by others (Moritz & Roberts, 2018). For example, participants who reported having more depressive symptomatology tended to believe that their partners, especially those new acquaintances, viewed them as less conscientious, less agreeable, and less emotionally stable, independent of their partners' actual views of them (Moritz & Roberts, 2018). The finding suggests a negative biased reflected self-evaluation in depression.

1.1.2 Low explicit and implicit self-esteem in depression

Terminologically, self-esteem can be seen as a result of self-evaluation (Bailey, 2003; Knyazev et al., 2021). To date, the term "esteem" was originally used in markets where a certain object required a fair monetary evaluation in order to be traded. In this context, objects that had been evaluated as having more worth and value were considered to have high esteem; on the contrary, objects that had been evaluated as having less worth and value were considered to have low esteem (Bailey, 2003). The term was gradually used to describe humans and the term "self-esteem" was then used to refer to a person's overall evaluation of his or her own worth and value (Bailey, 2003). For instance, if a person evaluates that his/her positives outweigh his/her negatives, that person will establish an increased self-confidence; thus, the person's view of the self is that of having high self-esteem and, if the reverse, low self-esteem (Bailey, 2003). Reasonably high self-esteem is considered to play a role in mental health, whereas low self-esteem and feelings of worthlessness or lack of value are common depressive symptoms (*APA*, n.d.).

According to the literature, people can be conscious or unconscious of self-esteem; therefore, researchers use the term "explicit self-esteem" to refer to the self-esteem that can be assessed through conscious introspection, whereas they use the term "implicit self-esteem" to refer to the self-esteem that cannot be reflected consciously (see examples at (Smeijers et al., 2017; van Tuijl et al., 2016; Vázquez et al., 2008; Wu et al., 2014)). Since explicit self-esteem can be reflected through active introspection, it is usually measured by using self-reported questionnaires, such as the Rosenberg Self-Esteem Scale (RSES) (Rosenberg, 1965). By using these questionnaires, the depressive disorder was consistently correlated with low explicit self-esteem in previous studies (Orth & Robins, 2013; Orth et al., 2008; Orth et al., 2009; Orth et al., 2014; Steiger et al., 2014). And since implicit self-esteem is not as consciously accessible as explicit self-esteem, it cannot be measured using self-report methods (Greenwald & Banaji, 1995). Thus, unconscious methods, such as the Implicit Association Task (IAT) (Greenwald & Banaji, 1995), were adopted to approach the implicit self-esteem (for example, (Greenwald & Farnham, 2000)).

The IAT is a categorization task in nature, measuring a person's implicit attitude to an object by unconsciously associating the object with positive or negative items (Greenwald & Farnham, 2000). For instance, during the IAT that measures implicit self-esteem, participants are usually presented with four types of stimulus words: two types of pronouns that refer to either *self* or *not-self* (for example, either me, I, …, or him, he, ….), and two types of adjectives that indicate either *positive* or *negative* attributes (for example, either smart, kind, …, or ugly, stupid….). All of the stimulus words are presented one by one in two independent blocks. During one of the blocks, participants are asked to press one key (for example, the *F* key on the keyboard) when they see either *self* or *positive* words, and to press another key (for example, the *J* key) when they see either *not-self* or *negative* words. During another block, the responses to *positive* and *negative* words are switched; that is, the participants press one key (for example, *F*) when they see either *self* or *negative* words, and another key (for instance, *J*) when they see either *not-self* or *positive* words. The premise of the IAT is that the participants' performance in the task should be better (for example, faster and more accurately) during the block where they use the same key to sort the stimulus pairings that are congruent with their implicit self-attitude, in comparison to during the block where they use the same key to sort the stimulus parings that are incongruent with their implicit self-attitude (Greenwald & Farnham, 2000). For example, people with a positive self-attitude responded significantly faster when *self* and *positive* words shared the same key, in comparison to when *self* and *negative* words shared the same key, suggesting high implicit self-esteem (Egenolf et al., 2013; Greenwald & Farnham, 2000). However, such a difference in the reaction times between these two blocks was significantly smaller for currently depressed patients (Jabben et al., 2014; Risch et al., 2010), for recurrently depressed patients (Risch et al., 2010), and for remitted depressive patients (Risch et al., 2010) than it was for those non-depressed controls. These results suggest that implicit self-esteem is lower (that is, less positive) in depressives than in non-depressed individuals. Some researchers have reported no significant responding difference between these two blocks among patients with clinical depression, which suggests a lack of self-positivity bias in the population (Franck et al., 2007).

1.2 Using neuroimaging to study self-knowledge in depression

The non-invasive neuroimaging techniques, for instance functional magnetic resonance imaging (fMRI) and event-related potential (ERP), have been widely employed to investigate the brain activities that underlie the processing of self-knowledge. The fMRI reflects brain activity by measuring changes in blood flow of the brain (Glover, 2011). To be specific, a brain region's demand for oxygen increases when the region is handling an ongoing cognitive activity. As a result, blood flow increases in this region so that oxygen can be delivered to neurons by red blood cells, and such Blood Oxygen Level Dependent (BOLD) contrast can be detected by MRI, which therefore is able to provide good spatial resolution (Glover, 2011). Based on this working principle, fMRI studies investigate brain activity by comparing BOLD signals between two or more different mental states; for example, between resting state and task-based stimulated state (Glover, 2011). As another popular imaging method, the ERP reflects brain activity by measuring electrical activity in response to cognitive events (Luck, 2005; Sur & Sinha, 2009). To be specific, the brain produces a specific pattern of electrical activity when a person is engaged in a cognitive event (for example, orientating attention to an object, detecting a novel item from a series of similar items, and so on), and such an electrical activity can be dynamically recorded at the milliseconds level by using electroencephalogram (EEG). The recorded electroencephalographic signals can be time locked and, in further, be related to task-based cognitive events, reflecting brain activities underlying such events (Luck, 2005; Sur & Sinha, 2009).

1.2.1 Brain activity underlying direct and reflected self-evaluation

By employing the fMRI in conjunction with the SRET paradigm, researchers have observed that the processing of direct self-evaluation usually engages the activity in the cortical midline structures (CMS), which includes the medial prefrontal cortex (mPFC), the anterior cingulate cortex (ACC), the posterior cingulate cortex (PCC), and the precuneus (Northoff, 2013; Northoff et al., 2006). Compared to non-depressed healthy controls, clinically depressed patients and sub-clinical individuals with a high risk of depression exhibited altered activity and abnormal connectivity within the CMS during the direct self-evaluation (Auerbach et al., 2015; Bradley et al., 2016; Davey et al., 2017; Ji et al., 2017; Li et al., 2017; Liu et al., 2020). Specifically, the clinically depressed patients, in comparison to those healthy controls, showed hyperactivity in the mPFC and the rostral ACC when they were evaluating the self by using negative words; and such hyperactivity can be positively correlated with the patients' depressive symptom severity (Yoshimura et al., 2010).

Regarding the reflected self-evaluation, previous studies have rarely investigated brain activity in depression, but we can still find some references from studies on healthy population. To study the reflected self-evaluation, researchers have adopted a reflected self-evaluation task that is similar to the SRET, but with additional conditions that require the participants to evaluate the self through others' perspectives (for example, (Pfeifer et al., 2017; Pfeifer et al., 2009; Van der Cruijsen et al., 2019)). By using the fMRI in conjunction with the reflected self-evaluation task, previous studies have observed that the processing of reflected self-evaluation usually engages the activity in the temporal parietal junction (TPJ) (Pfeifer et al., 2017; Pfeifer et al., 2009; Van der Cruijsen et al., 2019), which is roughly characterized as an area at the border between the temporal and parietal lobes surrounding the ends of the Sylvian fissure (Schurz et al., 2014). The observation is reasonable since the reflected self-evaluation largely relies on ToM, and the TPJ has been found to play a central role in ToM (Mahy et al., 2014; Schurz et al., 2014). For instance, the engagement of the TPJ activity was observed in a variety of ToM tasks among healthy individuals (for example, (Denny et al., 2012; Gweon et al., 2012; Mahy et al., 2014; Richardson & Saxe, 2020; Saxe, 2010; Saxe & Kanwisher, 2003; Saxe et al., 2006; Saxe & Powell, 2006; Schurz et al., 2014)). Lesion of the TPJ was selectively associated with deficits in ToM performance among clinical patients who had brain injury in the TPJ area (Apperly et al., 2004; Samson et al., 2004). Despite a lack of depression-related studies, we can expect an altered TPJ activity during the reflected self-evaluation among this population.

1.2.2 Electroencephalographic responses underlying IAT measuring implicit self-esteem

By using the ERP in conjunction with the IAT paradigm, previous studies have investigated the electroencephalographic responses underlying the implicit association in the IAT (for example, (Fleischhauer et al., 2014; Grundy et al., 2015; Healy et al., 2015; van Nunspeet et al., 2014; Wu et al., 2014; Wu et al., 2016; Xiao et al., 2015)). Based on those studies, several ERP components were observed during the IAT administration. For example, the frontal N1, peaking around 100 ms post-stimulus latency with a negative polarity, has been observed in previous IAT studies as an indicator of selective attention in relation to visual stimuli (van Nunspeet, Ellemers, Derks, & Nieuwenhuis, 2014); the occipital P1, peaking around 100 ms post-stimulus latency with a positive polarity, has been associated with automatic visual processing during IAT administration (Fleischhauer et al., 2014); the occipital N170, a negative deflection occurring approximately 200 ms after stimulus onset, has been found responding to emotional and contextual stimuli in the IAT context (Ibáñez et al., 2010); the P2 component, a positive deflection occurring approximately 200 ms post-stimulus over frontal and parieto-occipital areas, has been related to emotional arousal in previous IAT studies (Grundy et al., 2015; Healy et al., 2015; Xiao et al., 2015); and at last, the late positive component (LPC), which is also labeled as a P3-like or a P3b-like response in some studies, has been consistently reported in previous IAT studies (Egenolf et al., 2013; Fleischhauer et al., 2014; Yang & Zhang, 2009). These observations suggest that the performance of the IAT involves both early perceptual processing and late cognitive processing. However, according to a study that examined the full range of mental processes occurring in the IAT, although the early responses (such as N1, P1, and P2) were visible in the IAT context, only one late response (starting around 450 ms post-stimulus latency with a positive polarity at the posterior topographic areas) can be associated with individual differences in implicit bias (Schiller et al., 2016). The finding suggests a significant role of the LPC (P3-like) response in detecting individual differences in implicit attitude by using the IAT.

Based on previous studies, the LPC response is highly modulated by the informative value of experimental stimulus (Polich, 2007). For instance, during the IAT administration, amplitude of the LPC may be suppressed if the participants have doubts about their decision, as the informative value of the stimulus is not fully extracted (Coates & Campbell, 2010). Thus, the LPC amplitude should be elicited more significantly by easy tasks (for example, the IAT block that is congruent with one's implicit attitude), rather than by difficult tasks (for example, the IAT block that is incongruent with one's implicit attitude) (Coates & Campbell, 2010). Consistent with this idea, larger LPC amplitudes were reported in IAT blocks that were congruent with one's implicit self-attitudes, in comparison with IAT blocks that were incongruent with such attitudes (Saulnier et al., 2021; Wu et al., 2016; Yang & Zhang, 2009). Specifically, in IAT studies that investigated implicit self-esteem among non-depressed healthy individuals, larger LPC amplitudes were observed during the block where *self* and *positive* words shared a same key, in comparison to another block where *self* and *negative* words shared another key (Wu et al., 2016; Yang & Zhang, 2009). These results suggest that, for healthy individuals, the self and positive association is more congruent with their attitude than the self and negative association is, supporting a self-positivity bias and high implicit self-esteem in this population (Wu et al., 2016; Yang & Zhang, 2009). Although the existing ERP studies have not yet examined the brain responses during the IAT measuring implicit self-esteem among depressed individuals, we can expect these individuals to show enhanced LPC amplitude during the block when *self* and *negative* words share a same key, compared to the block when *self* and *positive* words share another key, because of the influence of negative self-schema.

1.3 Purpose and hypothesis of the research

Despite growing research on the topic of depression-related self-knowledge, some questions still need to be explored. Thus, my dissertation had two main purposes. The first was to review the previous publications and propose research topics that the previous studies had not investigated (**Study I**), and the second was to accomplish these proposals using empirical methods (**Study II** and **Study III**). Three individual studies were designed to achieve the purpose.

Study I aimed to review previous publications investigating abnormal self-knowledge in depression. Both behavioral and neuroimaging studies were involved. In the review, I organized the studying methods (for example, questionnaires and experimental paradigms) that had been commonly used for measuring self-knowledge, including self-evaluation and self-esteem. I then summarized previous findings that had been reported by using those methods. Based on these findings, I illustrated several behavioral and brain-responding abnormalities during the processing of self-knowledge that can be related to depression. Finally, I proposed possible directions that might be worth considering in future studies.

Study II aimed to examine one of the research topics that was proposed in **Study I**; namely, using the fMRI technique to investigate brain activity underlying the reflected self-evaluation in depression. In this study, individuals who had enhanced depressive symptoms (labeled as dysphoric group in my dissertation) and individuals who had few depressive symptoms (labeled as control group) were invited to participate in my experiment. The participants were asked to evaluate the extent to which a series of positive and negative personality trait adjectives describe them, according to either their own opinion (direct self-evaluation condition) or others' opinions (reflected self-evaluation condition). Considering the self-negativity bias in depression (Derry & Kuiper, 1981; Fava et al., 2000), I hypothesized that the dysphoric group, in comparison to the control group, would rate more negative adjectives and fewer positive adjectives as self-describing, regardless of whose perspective they needed to take. The fMRI technique was used to scan the participants' brain activity during the self-evaluations, and the difference in the brain activity would be compared between the dysphoric group and the control group. Regions of interest (ROIs) were mainly located within the bilateral TPJ areas because of their role in reflected self-evaluation (Pfeifer et al., 2017; Pfeifer et al., 2009; Van der Cruijsen et al., 2019). I hypothesized that the control group, in line with the previous findings, would exhibit enhanced TPJ activity during the reflected versus the direct self-evaluation (Pfeifer et al., 2009; Van der Cruijsen et al., 2019). For the dysphoric group, I hypothesized an altered TPJ activity during the reflected self-evaluation because of impaired ToM ability in clinical depression (Bora & Berk, 2016; Nestor et al., 2022) and mild/sub-clinically depressed populations (Erle et al., 2019; Lee et al., 2005; Manstead et al., 2013).

Study III aimed to examine another research topic that was proposed in **Study I**; namely, using the ERP to investigate electroencephalographic responses underlying the low implicit self-esteem in depression. Similar to **Study II**, a dysphoric group and a control group of individuals were invited to participate in my experiment. The IAT paradigm was used to measure the participants' implicit self-esteem. The EEG signals were recorded during the IAT administration. I was primarily interested in the LPC response within 400–1,000 ms post-stimulus latency because of its central role in reflecting individual differences in implicit bias during the IAT (Chen et al., 2018; Coates & Campbell, 2010; Schiller et al., 2016; Williams & Themanson, 2011). I hypothesized that the control group, in line with previous studies, would show faster reaction and enhanced LPC amplitudes during the block where *self* and *positive* words share the same responding key (labeled as self-positivity block in my study), in comparison to the block where the *self* and *negative* words share another key (labeled as self-negativity block), because the *self-is-positive* association is congruent with their implicit self-esteem (Wu et al., 2016; Yang & Zhang, 2009). Conversely, considering the self-negativity bias in depression (Beck, 1967; Beck et al., 1979), I hypothesized that the *self-is-negative* association should be congruent with the implicit self-esteem in the dysphoric group. Therefore, I further hypothesized that there would be faster reaction and enhanced LPC amplitudes during the self-negativity block than during the self-positivity block in this group.

Figure 1 summarized the research background and the hypothesis for each of the individual studies that were included in my dissertation.

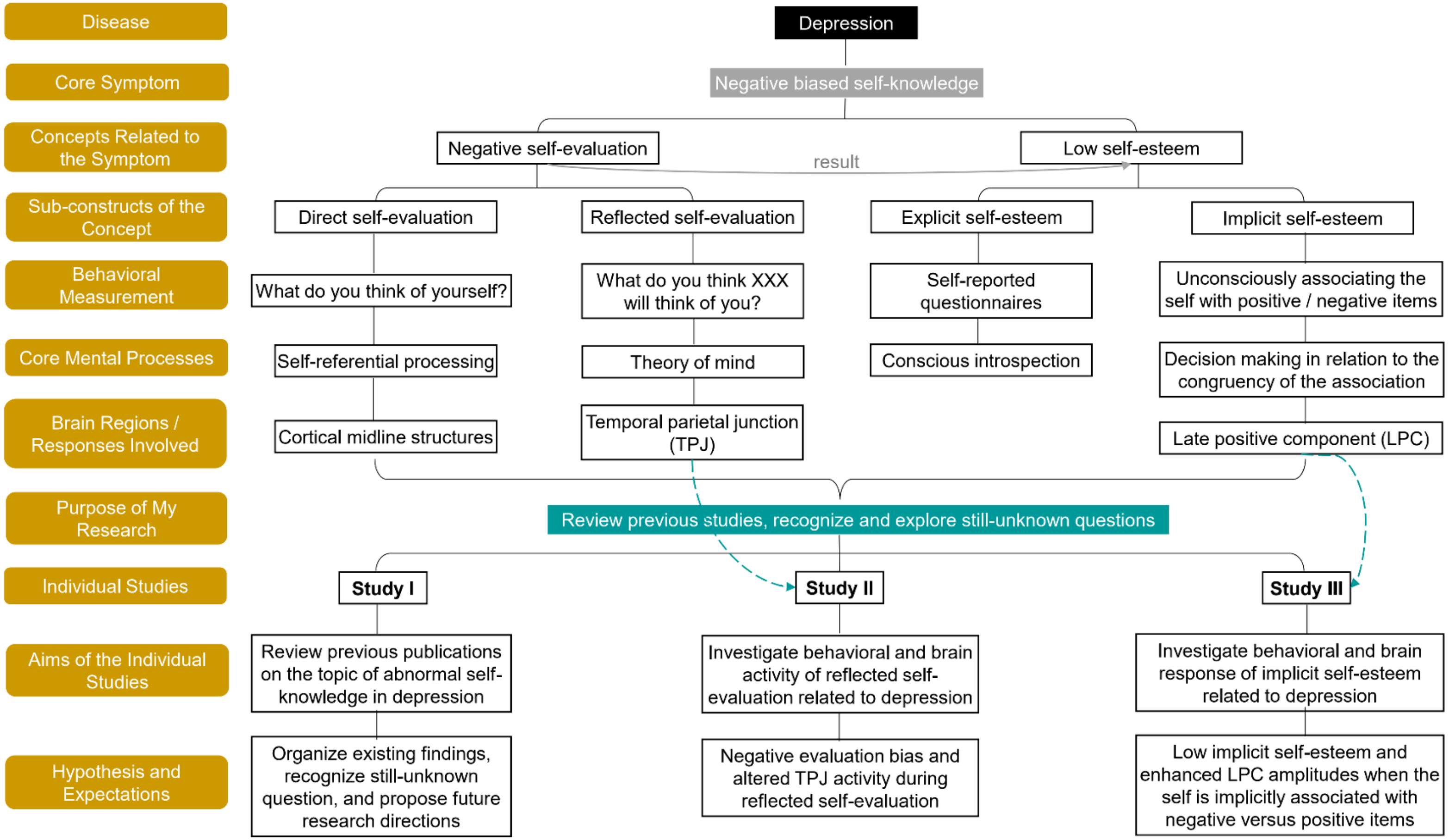

FIGURE 1 Summary of the research background and research purpose.

2 METHODS

2.1 Participants

Human participants were involved in **Study II** and **Study III**. In **Study II**, the sample size was determined based on previous neuroimaging studies that investigated the brain function of reflected self-evaluation among normal population (Pfeifer et al., 2017; Shiota et al., 2017). Before the formal recruitment stage, 275 university students who lived in Beijing City volunteered to fill out the Beck Depression Inventory – Second edition (BDI-II) (Beck et al., 1996) online. The BDI-II, which is a 21-item instrument with a total score that ranges from 0 to 64, is used for assessing depressive symptoms in both psychiatric and normal populations aged 13 and over (Beck et al., 1996). The cutoffs for the BDI-II were (a) 0 to 13 = minimum depression, (b) 14 to 19 = mild depression, (c) 20 to 28 = moderate depression, and (d) 29 to 63 = severe depression (Beck et al., 1996). Among all the volunteers, students with enhanced depressive symptoms (namely those who scored 14 or above on the BDI-II, labeled as dysphoric group in my study, N = 30) and students with few depressive symptoms (those who scored 13 or below on the BDI-II, and never got clinical depression before the study, labeled as control group, N = 30) were invited to participate in the study. Of the 60 original participants, data from four dysphoric participants and two control participants were excluded due to excessive head movement (≥ 2mm) during the fMRI scanning, leaving a final sample of 54 participants (dysphoric group = 26, control group = 28). Age (t (52) = 0.98, p = 0.33) and gender ratio (Pearson $\chi 2$ (1) = 0.30, p = 0.59) did not differ between the two groups.

In **Study III,** similar to **Study II**, the sample consisted of a group of dysphoric participants and a group of control participants. Since the study involved only two factors in the data analysis stage – one within-subject factor (self-association: self-positivity and self-negativity) and one between-subject factor (group: dysphoric and control) – I was able to determine the sample sizes by conducting power analysis with G*Power 3.1.9.2 (Faul et al., 2009). The input parameters of the power analysis included: Effect size f = 0.25, α err prob = 0.05, power (1-β err prob) = 0.95, number of groups = 2, number of measurements = 2, corr among rep measures = 0.5, non-sphericity correction ε = 1. The results suggested that a total sample of 54 participants should be able to achieve a medium effect size (f = 0.25; (Cohen, 1988)). Therefore, I planned to recruit 30 participants for each of the two groups. Before the recruitment, 567 students from Shenzhen University volunteered to fill out the BDI-II online. Since the sample pool was sufficient, I was able to invite students with scores distributed in the top 5 percent of the overall BDI-II score as the dysphoric group (N = 30; BDI-II ≥ 14), whereas students with scores distributed in the bottom 5 percent formed the control group (N = 32; BDI-II < 14). Of the 62 original participants, data from two dysphoric participants and two control participants were excluded due to excessive head movement during the EEG recording, leaving a final sample of 58 participants (Dysphoric group = 28, Control group = 30). Age (t (56) = −1.05, p = 0.30) and gender ratio (Pearson χ2 (1) = 0.35, p = 0.55) did not differ between the two groups.

All of the participants in **Studies II** and **III** were native Chinese speakers, right-handed, and had normal or corrected-to-normal vision. No participants reported previous or current physiological, neurological, or psychiatric disorders. Sub-clinical, rather than clinical, samples (that is, the dysphoric participants) were used to avoid confounding factors related to potential depression treatments. For example, some psychopharmacological treatments (such as agomelatine) were found to affect brain structures involved in self-related processing in depression (Delaveau et al., 2016). Another advantage of examining sub-clinical samples was that these participants were generally free of diagnostic comorbidities, which are more common in clinical samples. Similar sub-clinical samples have been used in previous studies of depressive self-knowledge (for example, (Shiota et al., 2017)). **Table I** presents the demographic information of the participants in **Studies II** and **III**.

TABLE 1 Demographic description of participants in Studies II and III

Study	Description	Units	Dysphoric	Control
Study II	N participants	N (females)	26 (18)	28 (11)
	Age	*M* ± *SD* (range) y/o	21.12 ± 2.86 (18 ~ 28)	21.86 ± 2.69 (18 ~ 28)
	BDI-II	*M* ± *SD* (range) points	19.85 ± 4.95 (14 ~ 32)	2.11 ± 2.17 (0 ~ 8)
Study III	N participants	N (females)	28 (18)	30 (17)
	Age	*M* ± *SD* (range) y/o	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 = numbers; *M* = means; *SD* = standard deviations; BDI-II = Beck Depression Inventory – II; y/o = years old.

2.2 Research ethics

Study II and **Study III** were both approved by the local Review Board for Human Participant Research of Shenzhen University. Before each study began, the participants were informed, both in writing and verbally, about the study's purpose, procedure, and data management methods. The participants were informed that they could withdraw from the study at any time without any negative consequences. Researchers requested the right to use the participants' anonymized data for only scientific research purposes and for publication of results. All participants volunteered to participate in the experiments and signed an informed consent form before the study began. The participants' brain responding data (that is, fMRI data in **Study II** and EEG data in **Study III**) and behavioral data (**Study II** and **Study III**) were stored separately in folders labeled with the participant's ID in an encrypted computer that is owned by the researchers' laboratory at Shenzhen University. Only researchers involved in these studies have access to these files. All written informed consent forms were preserved in a locked cabinet in a secure office space within Shenzhen University. Requests to access the data outside the research group can be granted only for scientific purposes and with anonymized data.

2.3 Materials and stimuli

Study I involved a search of previous publications. I targeted articles published between January 1960 and August 2018. Two databases – Web of Science (WOS) and PubMed – were used for the search. Search parameters were: searching fields = all field, and searching terms = "depression AND self-evaluation" OR "depression AND self-esteem" OR "depression AND self-reference". Searching filters were: article language = English, and species = humans. For the WOS database, an additional filter was: research domains = psychology or psychiatry. The first round of the search returned articles including empirical studies, systematic reviews, and meta-analyses. I manually screened the articles by examining the articles' titles, then their abstracts, and finally their full texts. To be included, an article had to focus on behavioral or neurological changes in certain aspect(s) of self-knowledge, such as self-evaluation and/or self-esteem, in depression. Moreover, the study had to involve at least one of the following types of population: (1) clinical patients who were diagnosed as currently having depression; (2) remitted patients who previously had a diagnosis of depression; or (3) sub-clinical populations who were currently under a depressive state. Gender and age were not specifically restricted because those two factors were not critical to the study. Exclusion criteria included conference abstracts that were not published in a scientific journal, and publications whose full text was written in a language other than English. The results consisted of 50 studies. **Figure 2** presents the search procedure.

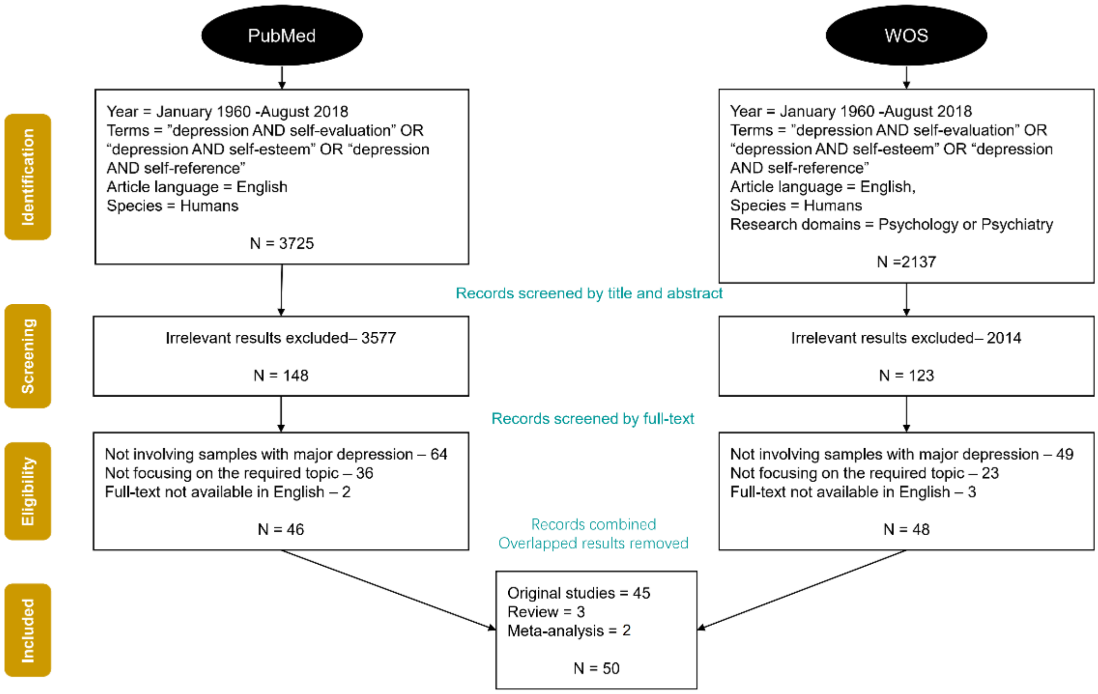

FIGURE 2 Flow of the literature search procedure in Study I. WOS = Web of Science; N = number of articles that remained after each exclusion.

Studies II and **III** involved stimulus words that describe positive or negative personality traits. The original stimulus pool consisted of 300 Chinese adjectives, selected from either the 562 personality trait adjectives pool (Huang & Zhang, 1992) or the Chinese Affective Words System (CAWS) (Wang et al., 2008). Before the formal experiments, I invited 25 volunteers to evaluate each of the adjectives' desirability, comprehensibility, familiarity, and arousal level. The volunteers were all university students who were interested in the material evaluation task, and none of them were invited to participate in the formal experiment. According to their evaluation, 40 adjectives (20 positive and 20 negative) were selected for **Study II** and 120 additional adjectives (160 in total, 80 positive and 80 negative) were selected for **Study III**. For both the two studies, the positive and the negative adjectives did not differ from each other in terms of the average ratings of comprehensibility, familiarity, and arousal level. The only rating difference concerned the desirability, with the positive adjectives being rated as more desirable than the negative adjectives. In addition, all the adjectives consisted of two Chinese characters, and the average stroke numbers of the two characters did not differ between the positive and the negative categories.

Study III involved extra stimulus words describing *self* or *not-self*. Due to the limited variations in Chinese language, five pronouns were used as *self*-describing words (labeled as *me*-related words in my study, consisted of self, me, I, mine, and us), and five other pronouns were used as the *not-self*-describing words (labeled as *not-me*-related words, consisted of his, other ("他人" in Chinese), other ("别人" in Chinese), others, and they). As with those personality trait adjectives, all of the pronouns consisted of two Chinese characters. During **Study III**, each of the pronouns was repeatedly presented 16 times to match the amount of the adjectives (N = 160). **Table 2** presented comparisons between the positive and the negative stimulus adjectives in **Study II** and **Study III**.

TABLE 2 Statistical data of stimuli used in Studies II and III

Study	Dimension	Positive	Negative	*t*(*df*)	*p*
		N / *M* ± *SD*	N / *M* ± *SD*		
Study II	N adjectives	20	20	–	–
	desirability	7.76 ± 0.21	2.75 ± 0.31	-59.61 (38)	< 0.001
	comprehensibility	4.22 ± 0.19	4.12 ± 0.24	- 1.48 (38)	0.15
	familiarity	3.97 ± 0.33	3.89 ± 0.36	-0.71 (38)	0.48
	arousal	4.95 ± 0.79	5.12 ± 0.92	0.63 (38)	0.53
	strokes	16.55 ± 4.32	18.60 ± 4.38	1.49 (38)	0.15
Study III	N adjectives	80	80	–	–
	desirability	7.68 ± 0.22	2.56 ± 0.45	91.95 (158)	< 0.001
	comprehensibility	4.17 ± 0.24	4.13 ± 0.23	1.17 (158)	0.24
	familiarity	3.96 ±0.32	3.91 ±0.32	0.91 (158)	0.36
	arousal	5.10 ± 0.83	4.99 ± 0.81	0.88 (158)	0.38
	strokes	17.59 ± 4.50	18.20 ± 4.43	-0.87 (158)	0.39

Note. N = numbers; *M* = means; *SD* = standard deviations.

2.4 Behavioral measurement procedure

2.4.1 Study II: Reflected self-evaluation task

Study II used a reflected self-evaluation task to examine how the participants would think of themselves through reflecting others' opinions about the self. Here, the "others" were restricted to two of the participant's acquaintances, one of whom must be the participant's close other (such as a best friend) and another must be an unclose other (such as an unfamiliar classmate). The acquaintances must be the same gender and of a similar age (± 2 years) as the participant was. Before the formal task, the participants were required to provide the names of the two acquaintances. The closeness between the participants and the two acquaintances was estimated separately using the Inclusion of Other in the Self (IOS) scale, a one-item graphic measure that depicts the closeness between the respondent and other person, with the least closed graph being coded as one point and the most closed graph being coded as seven points (Aron et al., 1992). According to the IOS scores, the close others (mean ± standard deviation = 4.59 ± 1.45) were rated as closer to the participants than the unclose others (1.81 ± 0.73) were, regardless of the participants' depressive symptoms (t (53) = 16.09, $p < .000$, Cohen's d = 2.21). During the task, the participants were asked to evaluate the extent to which the selected 40 personality trait adjectives (20 positive and 20 negative) described themselves, according to their own opinion (Self to Self condition, S2S), the close other's opinion (Close-other to Self, C2S), and the unclose other's opinion (Unclose-other to Self, U2S). We also asked participants to evaluate how these adjectives described the close other (Self to Close-other, S2C) and the unclose other (Self to Unclose-other, S2U). The baseline condition was to evaluate the desirability of each adjective. All the evaluations were made by using four-point Likert scales. The instructions and rating scales for each condition are illustrated in **Figure 3(a)**.

The procedure was block-designed, with one block including only one of the six evaluation conditions illustrated above. Since the positive and the negative adjectives were presented in separate blocks, the six evaluation conditions then formed 12 blocks in total. To avoid fatigue, the data were collected in four scanning runs, so that the participants were able to take a break between each run, but one scanning run should still include all 12 blocks. Thus, the 20 positive and the 20 negative adjectives were assigned in equal amounts across the four runs, meaning that each run included five positive and five negative adjectives. The five adjectives with the same valence were used repeatedly in each evaluation block. For instance, during one run, five positive adjectives (for example, smart, optimistic, confident, brave, and kind) would be repeatedly used in six blocks that each measured one of the six evaluation conditions (that are, the S2S, C2S, U2S, S2C, S2U, and baseline condition), and five negative adjectives (for example, stupid, diffident, pessimistic, cowardly, and indifferent) would be repeatedly used in other six blocks that measured the same evaluation conditions. Therefore, each block included five trials, with each trial presenting only one adjective. The order of the 12 blocks was completely random and the order of the four scanning runs was counterbalanced by Latin-square design across participants.

The time setting of the procedure was as follows: Prior to the onset of each block, an introduction appeared at the center of the screen for six seconds, indicating whose perspective the participants should take to evaluate whom. Five adjectives were then presented, one by one, in five continuous trials. Each trial lasted for five seconds, with the initial two seconds presenting an adjective and the last three seconds presenting a four-point Likert scale. Participants were asked to make their evaluation during the presentation of the Likert Scale. Once a block finished, a fixation cross that lasted for either six, eight, or 10 seconds, appeared at the center of the screen to indicate the beginning of the next block. The next block initiated when another instruction appeared on the screen. All of the text (that is, the instructions, the stimulus adjectives, the rating scales, and the fixation cross) was presented in black font at the center of the screen with a gray background. The procedure is illustrated in **Figure 3(b)**.

(a)

Condition	Instruction	Rating scale
S2S	Please evaluate, in your opinion, how appropriate each of the following words is to describe you?	1 - not at all 2 - a little 3 - very well 4 - completely
C2S	Please evaluate, according to 'friend's name', how appropriate each of the following words is to describe you?	
U2S	Please evaluate, according to 'classmate's name', how appropriate each of the following words is to describe you?	
S2C	Please evaluate, in your opinion, how appropriate each of the following words is to describe 'friend's name'?	
S2U	Please evaluate, in your opinion, how appropriate each of the following words is to describe 'classmate's name'?	
Baseline	Please evaluate, in general, how does each of the following words describe a desirable personality-trait?	1 - very bad 2 - bad 3 - good 4 - very good

Note. S2S = self to self; C2S = close other (e.g., one of the participant's best friend) to self; U2S = unclose other (e.g., one of the participant's unfamiliar classmate) to self; S2C = self to close other (e.g., the best friend); S2U = self to unclose other (e.g., the classmate); Baseline = evaluating the desirability of each stimulus word

(b)

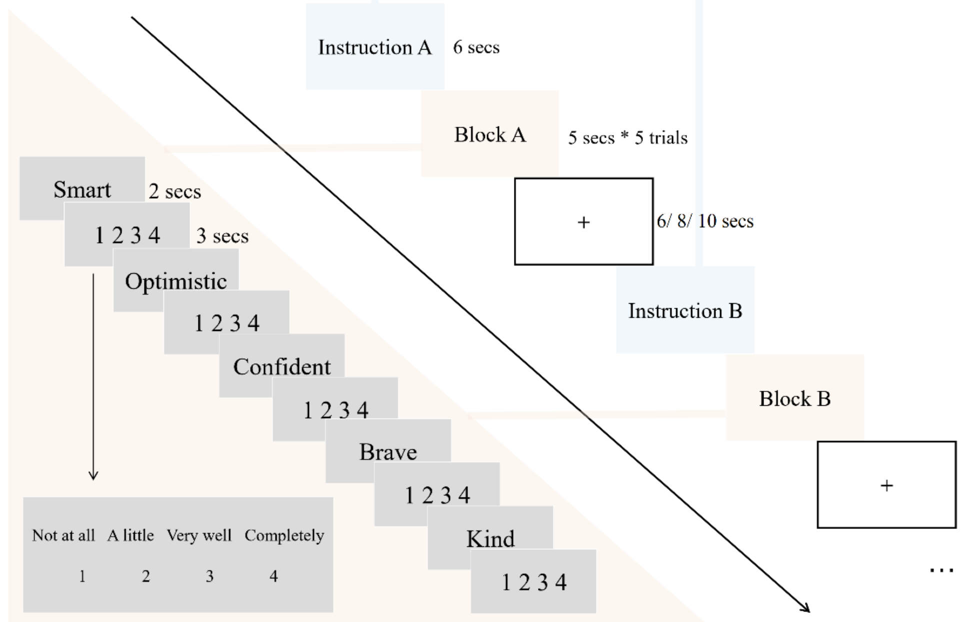

FIGURE 3 Instructions and experimental procedure of the reflected self-evaluation task in Study II. The upper panel **(a)** presents instructions and rating scales for each of the six evaluation conditions; the lower panel **(b)** illustrates one scanning run of the experimental procedure. Each run consisted of 12 blocks because the positive and the negative stimulus adjectives were presented in separate blocks.

2.4.2 Study III: The IAT measuring implicit self-esteem

Study III adopted the IAT that measures implicit self-esteem (Greenwald & Farnham, 2000). During the IAT, participants were asked to use two responding keys (the *F* and *J* keys on the keyboard) to sort four categories of stimulus words:

me-related pronouns, *not-me*-related pronouns, *positive* adjectives, and *negative* adjectives. All the stimulus words were repeatedly presented in two separate blocks: a self-positivity block and a self-negativity block. Each block consisted of three practice phases and one data-collection phase. The first phase was a 10-trial practice phase in which the participants learned to sort the *me* (five trials) and the *not-me* (five trials) categories by using two responding keys. The second phase was also a 10-trial practice phase in which the participants learned to sort the *positive* (five trials) and the *negative* (five trials) categories by using the same responding keys. The third phase was a 20-trial practice phase in which the words in the former two phases were mixed, and the participants learnt to sort all the four categories by using the responding keys that they just used. The fourth phase was a 320-trial data-collection phase that had the same requirement as the third practice phase, except with more trials (80 trials for each category). The difference between the self-positivity and the self-negativity blocks was that, in the self-positivity block, the participants used a same key for sorting the stimulus words in the *me* category and the *positive* category, whereas they used another key for sorting the stimulus words in the *not-me* category and the *negative* category. In the self-negativity block, the responding keys to the *positive* and the *negative* categories switched. Specifically, participants used the same key for sorting the stimulus words in the *me* category and the *negative* category, while they used another key for sorting the stimulus words in the *not-me* category and the *positive* category. In such a procedure, the self was implicitly associated with positive personality traits in the self-positivity block and was implicitly associated with negative personality traits in the self-negativity block. The behavioral responses and the EEG signals were recorded only during the fourth phase. The requirements in the two blocks are illustrated in **Figure 4(a)**.

The time setting of the procedure was as follows: Each block included 320 trials and the participants were able to take a break after every 80 trials to avoid fatigue. Each trial started with a fixation cross, presenting at the center of the screen, with a random duration between 1,000 and 2,000 ms. The cross was followed by a stimulus word with a fixed duration of 1,000 ms. The participants were asked to give a response (by pressing the *F* key or the *J* key) as quickly as they could during the presentation of each word. After that, a new fixation cross appeared on the screen to indicate the beginning of the next trial. All the stimulus words were presented, one by one, in a black font on a gray background at the center of the screen with a vertical visual angle of 0.45° and a horizontal visual angle of 0.9°. The order of the words was completely random in both the self-positivity and the self-negativity blocks, and the order of the two blocks was counterbalanced across participants. The assignment of the *F* key and the *J* key was also counterbalanced.

Compared to classic IAT in behavioral studies, two revisions were made, as suggested by a previous ERP study (Wu et al., 2016), to ensure the quality of the brain data. First, in the classic IAT there are usually labels displayed on the upper left- and upper right-corners of the screen, reminding the participants of the correct response during the block. For instance, during the self-positivity block, the labels of "*me*" and "*positive*" would be always displayed on the upper-left corner of the screen, while the labels of "*not-me*" and "*negative*" would be displayed on the upper-right corner of the screen. In ERP studies, however, the labels were omitted to reduce excessive eye movements during EEG signal recording. Second, to ensure a sufficient number of valid trials (namely trials with correct responses) for offline ERP data analysis, participants were asked to practice until they reached a relatively high accuracy rate (85 percent) before beginning the data collection phase. The data-collection phase is presented in **Figure 4(b)**.

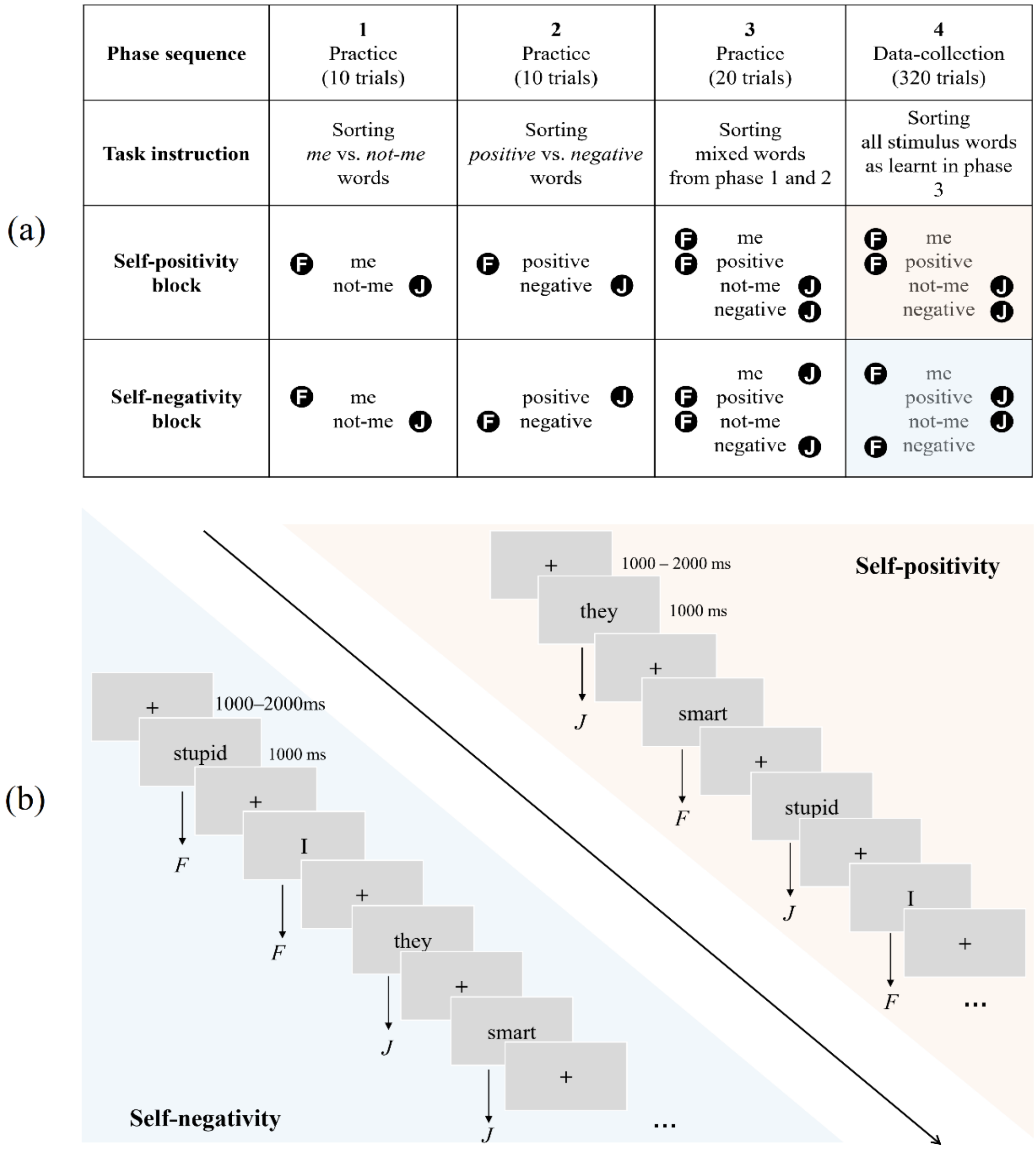

FIGURE 4 Experimental requirements and procedures of the IAT measurement in Study III. The upper panel (a) shows the requirements of each experimental phases for the self-positivity and the self-negativity blocks, with the labeled black dots indicating the correct responses for each phase; the lower panel (b) illustrates the procedure of the data-collection phase (that is, the fourth phase) in each of the two blocks.

2.5 Brain data acquisition and preprocessing

In **Study II**, the fMRI data were acquired using a 3.0 Tesla MRI scanner (Magnetom Prisma, Siemens, Erlangen, Germany). The imaging parameters were as follows: Task-based fMRI data were acquired using an EPI sequence with TR = 2000 ms, TE = 30 ms, flip angle = 90°, FOV = 224×224 mm, matrix size = 64×64, slice thickness = 2 mm, voxel size = 2×2×2 mm^3, and slice number = 64. T1-weighted structural images were acquired using a sagittal 3D-magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence: TR = 2530 ms, TE = 2.98 ms, inversion time = 1100 ms, FA = 7°, FOV = 256 × 256 mm2, slice thickness = 1 mm, voxel size = 0.5×0.5×1 mm3, and sagittal planes = 192. Statistical Parametric Mapping 12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12) was used for the fMRI data analysis with regular preprocessing steps of realignment, volume registration, spatial normalization (resampled into 2 mm isotropic voxels), and spatial smoothing with a Gaussian kernel of 6 mm full width at half maximum. Head movement estimates derived from the realignment step were included as nuisance regressors in subsequent general linear modeling (GLM) to diminish the impact of movement-related effects.

In **Study III**, the 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 midline of the 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 interelectrode impedances were maintained below 5 kΩ for on-line recording. During offline preprocessing, EEG signals were re-referenced to the average signal at the mastoid electrodes (Luck, 2005) and a low-pass filter was applied (30 Hz; 24 dB/octave). A semiautomatic 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 1,000 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, leaving an average amount of 308 valid trials contributing to the mean ERPs in the self-positivity block, and 298 valid trials in the self-negativity block, at the individual level. The grand mean ERPs for the self-positivity and the self-negativity blocks were then calculated by averaging the individual ERPs in each group, respectively.

2.6 Statistical analyses

For both the behavioral and the brain imaging data, a significance level of .05 was accepted 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.

2.6.1 Study II: Evaluation ratings and fMRI data analysis

For the behavioral data, means of the participants' evaluation ratings in each experimental condition were entered into a three-way mixed ANOVA with valence (positive vs. negative) and evaluation (S2S, S2C, S2U, C2S, U2S) as the within-subjects variables, while group (dysphoric vs. control) as the between-subjects variable.

For the fMRI data, the two self-to-other evaluation conditions (namely the S2C and the S2U conditions) were not included because the purpose of analyzing the imaging data was to investigate brain activity differences between direct self-evaluation (that is, the S2S condition) versus reflected self-evaluation (that is, the C2S and U2S conditions) among dysphoric participants. In the first-level individual analysis, six contrasts were created to subtract the brain activity of each evaluation condition from the baseline conditions (i.e., S2S(pos) > Baseline(pos), C2S(pos) > Baseline(pos), U2S(pos) > Baseline(pos), S2S(neg) > Baseline(neg), C2S(neg) > Baseline(neg), U2S(neg) > Baseline(neg)) for each participant. Two ROIs – the left TPJ [MNI coordinates: x = −53, y = −59, z = 20] and the right TPJ [x = 56, y = −56, z = 18] – each consisted of 8 mm spheres, were defined based on previous empirical studies and meta-analyses in which personality trait stimulus words were used to investigate brain responses of theory of mind (Schurz et al., 2014; Van der Cruijsen et al., 2019). An additional ROI within the mPFC area [MNI coordinates: x = 3, y = 51, z = −7] consisting of an 8 mm sphere was defined and analyzed for exploratory purposes because the activity of the mPFC has been reported in previous studies investigating the self-related processing (for examples, (Jankowski et al., 2014; Lemogne et al., 2012; Li et al., 2017; Liu et al., 2020; Nejad et al., 2013; Pfeifer et al., 2017; Shiota et al., 2017; Van der Cruijsen et al., 2019)).

For each ROI, the mean percentage signal changes (PSCs) of each participant were extracted using the Marsbar toolbox (Brett et al., 2002) and were exported to the SPSS for further analysis. In the second-level group analysis, the mean PSCs were submitted into three-way mixed ANOVAs with valence (positive vs. negative) and evaluation (S2S, C2S, U2S) as the within-subjects variables, while group (dysphoric vs. control) as the between-subjects variable. Pearson's correlation coefficient was calculated with the whole participant sample to investigate whether the participants' brain activity in the bilateral TPJs can be associated with their depressive severities (reflected by the BDI-II scores).

2.6.2 Study III: IAT indices and ERP data analysis

For the behavioral data, three IAT indices – namely reaction time, accuracy, and *D*-score (an index of the IAT effect that is calculated from reaction time) (Greenwald & Banaji, 1995; Greenwald et al., 2003) – were applied to reflect the participants' implicit self-esteem. For all the three indices, trials with incorrect responses were not included in the data analysis. For the reaction time and the accuracy, means were calculated separately for the self-positivity and the self-negativity conditions. The means of the two indices were then submitted into two-way ANOVAs with self-association (self-positivity and self-negativity) as the within-subjects variable, while group (dysphoric and control) as the between-subjects variable. For the *D*-score, I first used the means of reaction time in the self-negativity condition to subtract the means of reaction time in the self-positivity condition, then divided their 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 in order to compare their *D*-score to zero. A higher *D*-score indicates higher implicit self-esteem (Greenwald & Farnham, 2000; Greenwald et al., 2003). The independent sample *t*-test was applied to investigate the difference of *D*-scores between the dysphoric and the control groups. I also used the RESE (Rosenberg, 1965) to measure the participants' explicit self-esteem. The means of the RESE scores were calculated separately for the dysphoric and control groups, and an independent *t*-test of the mean RESE scores was conducted to compare the difference in explicit self-esteem between those two groups.

For the ERP data, the LPC was measured from 300 ms post-stimulus latency to end of the stimulus (namely, 1,000 ms post-stimulus latency). According to previous studies, 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 (approximately after 400 ms post-stimulus latency) best reflects the efficacy of stimulus evaluation and categorization (Wu et al., 2016). Guided by the literature (Grundy et al., 2015) and by 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 600–1,000 ms post-stimulus latency. Nine 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 can also be found in a previous IAT study that, similar to my study, involved Chinese stimulus words and Chinese participants (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-subjects variables, while group (dysphoric and control) as the between-subjects variable. By including the laterality and the anterior-posterior location as separate factors instead of calculating an average electrophysiological response across all these channels, I was able to compare my results to the results of previous studies that used the same analysis method (for example, (Grundy et al., 2015; Wu et al., 2016)).

Pearson's correlation coefficient was calculated with the whole participant sample to investigate whether the participants' explicit and implicit self-esteem (reflected by the RSES scores and *D*-scores, respectively) can be associated with their depressive severities (reflected by the BDI-II scores). The correlation analyses were also conducted to investigate whether the ERP responses (reflected by means of LPC amplitudes in the self-positivity condition and the self-negativity condition) can be associated with the participants' depressive severities or their behavioral performance during the IAT (reflected by the BDI-II scores and the *D*-scores, respectively).

3 SUMMARY OF RESULTS

3.1 Study I: Literature review of abnormal self-knowledge in depression

Study I reviewed scientific publications in the field of depression-related self-knowledge. After screening, 50 publications remained. Based on these publications, I summarized two types of research measures, the explicit measures and the implicit measures, that had been commonly used to investigate abnormal self-knowledge in depression. For explicit measures, in addition to the SRET paradigm and the RSES questionnaire that I introduced in the Introduction section, researchers also used the self-worth sub-scale of the World Assumption Scale (Janoffbulman, 1989) and the self-acceptance sub-scale of the Scales of Psychological Well-Being (Ryff & Keyes, 1995) to measure participants' explicit attitudes towards themselves. The BDI-II also contains specific items, such as the self-blame measuring items, that assess individuals' explicit self-attitude (Beck et al., 1996). For implicit measures, in addition to the IAT paradigm that I introduced in the Introduction section, the Name-Letter Test (NLT; (Gu et al., 2014; Koole & Pelham, 2003; Nuttin, 1985), which reflects participants' implicit self-esteem by comparing their preference of their name's initial letter to other letters, was also commonly used. The Go/No-Go Association Task (GNAT; (Brian A. Nosek & Banaji, 2001), which unconsciously associates the self with positive or negative personality trait, is another measure that assesses a person's implicit self-esteem.

Here, I summarized studies that adopted the empirical paradigms (for instance, the SRET, the IAT, the NLT, or the GNAT) to investigate at least one aspect of self-evaluation (that is, direct or reflected self-evaluation) or self-esteem (that is, explicit or implicit self-esteem) that can be related to depression (for instance, by involving clinical or sub-clinical depressed participants). I primarily listed work published after 2000 for two reasons: First, most studies on self-evaluation in depression have already been summarized in previous reviews (for example, in (Northoff, 2007; Wisco, 2009)). Second, the empirical paradigms related to implicit self-esteem in depression were introduced only after 2000 (for example, (Brian A. Nosek & Banaji, 2001; Greenwald & Farnham, 2000; Koole et al., 2001)). Thus, twenty-seven publications were listed in **Table 3**.

TABLE 3 Overview of articles that were mainly discussed in Study I

Study[a]	Population Description	Sample Size	Depression Assessment	Self-Knowledge	Measure	Neuro-imaging	Index	Result[b] (Depressed vs. Controls)	
								Behavioral	Neurological
Auerbach et al. (2015)	Depressed female adolescents, Healthy female controls	22 28	BDI-II	Direct self-evaluation	SRET	ERP	Endorsement, Recall rate, Recognition rate, P1, P2, Early and late LPP	Endorsed ↑ negative and ↓ positive words as self-descriptive, Recalled and recognized ↓ positive words	↑ P1 amplitudes for negative words ↑ early and late LPP amplitudes for negative vs. positive words
Bradley et al. (2016)	Adolescents with MDD, Healthy controls	20 15	DSM-IV, CDRS-R	Direct self-evaluation	SRET	fMRI	Endorsement, RT, amPFC, dmPFC, vmPFC, PCC/precuneus	↑ negative self-perceptions	Both recruited the CMS in self-reflection, but depressed individuals recruited the PCC/precuneus more for positive self-judgments
Dainer-best et al. (2017)	Adults with MDD, Healthy controls	21 23	CESD, MINI, DSM-V	Direct self-evaluation	SRET	ERP	Endorsement, P1, P2, Early and late LPP	Endorsed ↑ negative and ↓ positive words as self-descriptive	↑ early and late LPP amplitude difference (negative vs. positive words)
Frank et al. (2007)	Depressed patients with suicidal ideation, Depressed patients without suicidal ideation, Matched controls	14 15 15	MINI, DSM-IV	Explicit self-esteem, Implicit self-esteem	RSES, IAT	–	RSES score IAT effect[c]	Patients without suicidal ideation: ↓ explicit and ↓ implicit self-esteem Patients with suicidal ideation: ↓ explicit but ↑ implicit self-esteem	–
Frank et al. (2008)	Currently depressed individuals, Formerly depressed	29 35	MINI, DSM-IV, HRSD, BDI-II	Explicit self-esteem, Implicit self-esteem	RSES, IAT	–	RSES score IAT effect[c]	Formerly depressed: normal explicit but ↑ implicit self-esteem Currently depressed:	–

Study[a]	Population Description	Sample Size	Depression Assessment	Self-Knowledge	Measure	Neuro-imaging	Index	Result[b] (Depressed vs. Controls)	
								Behavioral	Neurological
	individuals, Never depressed controls	38						↓ explicit but normal implicit self-esteem	
Gemar et al. (2001)	Currently depressed patients, Formerly depressed patients, Never depressed controls	32 23 27	SCID, DSM-IV, BDI-II, HRSD	Implicit self-esteem	IAT	–	IAT effect[c]	Currently depressed: ↓ positive self-bias than the formerly depressed individuals	–
Grimm et al. (2009)	Depressed subjects with an acute MDD episode, Healthy controls	25 25	DSM-IV, HDRS, BDI	Direct self-evaluation	SRET-Like	fMRI	Endorsement, dmPFC, supragenual ACC, precuneus, VS, bilateral DMT	↑ self-relatedness of negative emotional stimuli	↓ signal intensities in the dmPFC, supragenual ACC, precuneus, VS, and the DMT
Roberts et al. (2015)	Individuals with a history of depression, Individuals with no history of depressive episodes	28 33	PHQ-9	Explicit self-esteem, Implicit self-esteem	RSES, IAT	–	RSES score IAT effect[c]	↓ explicit but ↑ implicit self-esteem	–
Jabben et al. (2014)	Patients with unipolar depressive disorder, Non-clinical controls	1236 387	CIDI	Implicit self-esteem	IAT	–	IAT effect[c]	↑ depressive self-associations	–
Ji et al. (2017)	Participants scoring high in depression scores, Participants scoring low in depression score	23 22	DASS-21	Direct self-evaluation	SRET	–	RT	↓ attention to positive information that had been processed in a self-referential manner	–
Kesting et al. (2011)	Patients diagnosed with a	21	CIDI, ICD-10	Explicit self-esteem,	RSES, IAT	–	RSES score IAT effect[c]	↓ explicit but normal implicit self-esteem	–

Study[a]	Population Description	Sample Size	Depression Assessment	Self-Knowledge	Measure	Neuro-imaging	Index	Result[b] (Depressed vs. Controls)	
								Behavioral	Neurological
	depressive disorder, Healthy controls	59		Implicit self-esteem					
Kiang et al. (2017)	Outpatients with nonpsychotic MDD, Healthy controls	16 16	MINI, DSM-IV, HAMD17	Direct self-evaluation	SRET	ERP	Endorsement, N400	Endorsed ↑ negative and ↓ positive words as self-descriptive	↓ N400 amplitudes for negative words
Lemmens et al. (2014)	Depressed patients, Healthy controls	87 30	SCID, DSM-IV	Explicit self-esteem, Implicit self-esteem	SLCS-R, IAT	–	SLCS-R score IAT effect[c]	↓ explicit but normal implicit self-esteem	–
Lemogne et al. (2009)	Patients with a major depression, Healthy subjects	15 15	MINI, DSM-IV, MADRS, BDI	Direct self-evaluation	SRET	fMRI	Endorsement, MFG	Endorsed ↑ negative and ↓ positive words as self-descriptive	Activated extended regions (e.g., dorsal MFG and dlPFC) in self-reflection
Li et al. (2017)	Unipolar depressive patients, Healthy controls	19 21	SCID, DSM-IV, BDI	Direct self-evaluation	SRET	fMRI	Endorsement, RT, Subregions of the mPFC	Endorsed ↑ negative vs. positive words as self-descriptive, showed slower response to these negative words	↑ activation of the cmPFC and ↓ activation of the dmPFC in self-reflection
Poulsen et al. (2009)	Depressed participants, Non-depressed controls	39 97	DSM-IV, BDI	Direct self-evaluation	SRET	ERP	Endorsement, N1, P1, P2, MFN	↑ depression was associated with ↓ endorsement for positive words	N1 and P2-MFN were attenuated or absent to the endorsed negative words
Raedt et al. (2006)	study i Inpatients with depressive symptoms, Healthy control study ii Inpatients with depressive	study i 15 15 study ii 16 16	DSM-IV, BDI, HRSD	Implicit self-esteem	study i IAT study ii NLT	–	study i RT study ii Standardized NLT rating	study i Normal implicit self-esteem study ii Normal implicit self-esteem	–

Study[a]	Population Description	Sample Size	Depression Assessment	Self-Knowledge	Measure	Neuro-imaging	Index	Result[b] (Depressed vs. Controls)	
								Behavioral	Neurological
	symptoms, Healthy person								
Randen-borgh et al. (2016)	Chronically de-pressed patients with an early on-set, Chronically de-pressed patients with a late onset, Episodic depres-sion	17 13 29	SCID, LIFE-Inter-view	Explicit self-esteem, Implicit self-esteem	RSES, NLT	—	RSES score NLT rating	Chronic patients with early onset: ↓ explicit self-esteem than epi-sodic patients, ↓ im-plicit self-esteem than episodic patients and chronic patients with late onset	—
Risch et al. (2010)	First-onset cur-rently depressed patients, Recurrently de-pressed patients, Remitted de-pressed patients, Never depressed controls	24 28 33 34	IDC, DSM-IV	Implicit self-esteem	IAT	—	IAT effect[c]	First onset currently depressed: ↓ implicit self-esteem Recurrently de-pressed: ↓ implicit self-esteem Remitted depressed: normal implicit self-esteem	—
Romero et al. (2016)	MDD group, Never-depressed controls	38 40	SCID, DSM-IV	Explicit self-esteem, Implicit self-esteem, Direct self-evaluation	RSES, GNAT, SRET	—	RSES score GNAT index[d] Endorsement, Recall rate	↓ explicit and ↓ im-plicit self-esteem, en-dorsed ↑ negative and ↓ positive words as self-descriptive, re-called ↑ depressed and ↓ positive words	—
Shestyuk et al. (2010)	Patients with cur-rent MDD, Patients with re-mitted MDD, Healthy compari-son subjects	17 18 17	SCID, DSM-IV, BDI	Direct self-evaluation	SRET	ERP	Endorsement, recall rate, P2, LPC	Currently depressed: endorsed ↑ negative vs. positive words as self-descriptive, re-called ↓ positive words Remitted depressed:	Currently depressed: showed ↑ P2 and LPC amplitudes to negative vs. positive words Remitted depressed: showed ↑ P2

Study[a]	Population Description	Sample Size	Depression Assessment	Self-Knowledge	Measure	Neuro-imaging	Index	Result[b] (Depressed vs. Controls)	
								Behavioral	Neurological
								endorsed ↑ positive vs. negative words as self-descriptive, recalled ↑ positive vs. negative words	amplitudes to positive vs. negative words
Shiota et al. (2017)	Students with subthreshold depression: Intervention group, No-treatment controls	29 30	Japanese version of the SCID, CIDI, BDI-II	Reflected self-evaluation	Reflected self-evaluation task	fMRI	Δ (post- vs. pre-treatment) Depressive symptoms, Δ brain activation in mPFC, Δ reaction times	Before BA: longer RT to other-to-self-negative words After BA: improvement in depressive symptoms, longer RT to other-to-self-positive words	After BA: ↑ activation in the dmPFC to the other-to-self-positive words, such an enhanced dmPFC activity was positively correlated with the improvement of depressive symptoms
Smeijers et al. (2017)	Remitted depressed patients, Never depressed controls	75 75	SCID-I, MINI, DSM-IV	Explicit self-esteem, Implicit self-esteem	RSES, IAT, NLT	–	RSES score, IAT effect[c] & NLT ratings	↓ explicit but normal implicit self-esteem	–
Van Tuijl et al. (2016)	Current MDD; Remitted MDD; Recovered MDD; Comparison group	60 41 136 382	Recruitment criteria was published elsewhere	Explicit self-esteem, Implicit self-esteem	RSES, IAT	–	RSES score IAT effect[c]	Depressed groups showed ↓ explicit but normal implicit self-esteem	–
Watson et al. (2008)	Dysphoric group; Non-dysphoric group	17 10	BDI-II	Direct self-evaluation	SRET	–	Endorsement	No endorsement differences across positive, negative, and neutral words	–
Yoshimura et al. (2010)	Patients with unipolar major depression; Healthy controls	13 13	SCID, DSM-IV, BDI-II	Direct self-evaluation	SRET	fMRI	mPFC, rostral ACC	–	Hyperactivity in the mPFC and the rostral ACC during the self-referential processing of negative words

Study[a]	Population Description	Sample Size	Depression Assessment	Self-Knowledge	Measure	Neuro-imaging	Index	Result[b] (Depressed vs. Controls)	
								Behavioral	Neurological
Yoshimura et al. (2014)	Depressive patients; Healthy controls	23 15	SCID, DSM-IV	Direct self-evaluation	SRET	fMRI	mPFC, ventral ACC	After CBT: improvement in depressive symptoms	Before CBT: hyperactivity in the mPFC during self-referential processing of negative words After CBT: activity in the mPFC and the ventral ACC during self-referential processing was ↑ for positive words and ↓ for negative words

Note:
[a] Only studies that were published after 2000 and that adopted empirical paradigms, such as the SRET, the IAT, the NLT, or the GNAT (see the meaning of the abbreviations as below) are listed in the table.
[b] Only the main results that are related to the current topic, namely the depression-related self-knowledge, are reported. I summarized here the results in the depressed groups in comparison to the controls. "↑" refers to greater/enhanced; "↓" refers to smaller/decreased.
[c] The IAT effect is an index for measuring implicit self-esteem; it is usually computed from reaction time or accuracy as suggested by Greenwald and colleagues (1995, 2003). The higher the IAT effect, the higher the implicit self-esteem.
[d] Reaction time in the self-negativity block less reaction time in the self-positivity block. Positive scores indicate positive implicit self-esteem.
Abbreviations:
Column 2 (population description): MDD = major depressive disorder
Column 4 (depression assessment): BDI = Beck Depression Inventory; BDI-II = Beck Depression Inventory - Second Edition; MINI = mini international neuropsychiatric interview; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition; DSM-V = Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; MADRS = Montgomery–Asberg Depression Rating Scale; HDRS = 21-item Hamilton Depression Rating Scale; SCID = clinical interview for DSM-IV; CDRS-R = Children's Depression Rating Scale – Revised; CESD = Center for Epidemiologic Studies – Depression Scale; DASS-21 = 21-item Depression, Anxiety, and Stress Scale; HAMD17 = 17-item Hamilton Depression Rating Scale; CIDI = composite international diagnostic interview; MOD = Mood Disorder; HRSD = Hamilton Rating Scale for Depression; ICD-10 = The Tenth Revision of the International Statistical Classification of Diseases and Related Health Problems Checklists; PHQ-9 = The Patient Health Questionnaire-9; LIFE-Interview = longitudinal interval follow-up evaluation interview

Column 6 (measure): SRET = self-referential encoding task; RSES = Rosenberg Self-Esteem Scale; SLCS-R = Self-Liking and Self-Competence Scale Revised; IAT = implicit association task; NLT = name letter task; GNAT = go/no-go association task.
Column 7 (neuroimaging): ERP = event-related potentials; fMRI = functional magnetic resonance imaging.
Columns 8, 9, 10 (index and result): RT = reaction time; CBT = cognitive behavioral therapy; BA = behavioral activation; MFN = medial frontal negativity; LPP = late positive potentials; LPC = late positive components; MFG = medial frontal gyrus; mPFC = medial prefrontal cortex; amPFC = anterior medial prefrontal cortex; cmPFC = central medial prefrontal cortex; dmPFC = dorsal medial prefrontal cortex; vmPFC = ventral medial prefrontal cortex; dlPFC = dorsolateral prefrontal cortex; ACC = anterior cingulate cortex; PCC = posterior cingulate cortex; DMT = dorsomedial thalamus; VS = ventral striatum; CMS = cortical midline structures

As these previous studies have shown, one of the depression-related abnormality that had been consistently observed was the enhanced endorsement for negative words and the decreased endorsement for positive words during direct self-evaluation (see examples at ((Auerbach et al., 2015; Dainer-best et al., 2017; Kiang et al., 2017; Li et al., 2017; Poulsen et al., 2009; Romero et al., 2016; Shestyuk & Deldin, 2010)). In addition, the memory for the endorsed negative words was found to be better than the memory for the endorsed positive words in depression (Auerbach et al., 2015; Romero et al., 2016; Shestyuk & Deldin, 2010). The results were interpreted as suggesting an enhanced negative self-perception in depression, in comparison to those non-depressed healthy individuals (Bradley et al., 2016). Such a negative self-perception was found to mainly occur at a later effortful information processing stage (for example, reflected by the LPC/P3b amplitudes) (Auerbach et al., 2015; Dainer-best et al., 2017; Shestyuk & Deldin, 2010). Moreover, during the processing of the negative self-related information, the depressed individuals, in comparison to the healthy controls, exhibited altered brain activity in the mPFC and its subregions, such as the dorsal mPFC and the central mPFC (Grimm et al., 2009; Li et al., 2017; Yoshimura et al., 2010). Furthermore, the depressed individuals uniquely recruited an extended medial prefrontal network, such as the dorsal part of the medial frontal gyrus and the dorsolateral PFC, during the processing of the self-related information, suggesting the involvement of greater cognitive controls in this population (Lemogne et al., 2009). In the treatment of depression, interventions such as cognitive behavioral therapy were found to not only significantly improve the depressive symptoms, but also change the abnormal brain activities such as the altered mPFC activity during the direct self-evaluation (Yoshimura et al., 2014).

Interestingly, unlike the consistent evidence for a negative direct self-evaluation, more heterogeneous results had been reported in relation to the self-esteem in depression. For instance, although some studies observed both low explicit and low implicit self-esteem in depressed patients (Romero et al., 2016), others observed discrepancy between explicit and implicit self-esteem (for example, low explicit but normal implicit self-esteem (Franck et al., 2008; Kesting et al., 2011; Lemmens et al., 2014; van Tuijl et al., 2016), or even low explicit but high implicit self-esteem (Roberts et al., 2015)) in depressives. After reviewing these studies, I suggest that the discrepant self-esteem may occur only in certain types of depression. For example, chronic patients with early-onset depression seem to have had congruently low explicit and low implicit self-esteem, whereas the discrepancy was found in episodic patients and chronic patients with late-onset depression (Randenborgh et al., 2016). In another example, the discrepancy was observed in currently depressed patients with suicidal ideation, but not in patients without such an ideation (Franck et al., 2007). The same discrepancy was also found in remitted individuals with a history of depression (Risch et al., 2010; Roberts et al., 2015; Smeijers et al., 2017), but not in patients with first onset of current depression, nor patients with recurrent depression (Risch et al., 2010). However, due to the heterogeneous samples across these studies, and also due to the challenges of measuring the unconscious aspect of self-knowledge, no convincing explanation has been given so far for the inconsistent findings regarding implicit self-esteem in depression.

Based on the review, I proposed two research topics that future studies could consider investigating. The first is the reflected self-evaluation in depression. Although the direct self-evaluation has been well studied in depression, the behavioral and brain activity of the reflected self-evaluation in this population remains unclear. To the best of my knowledge, only one study has attempted to explore the brain activity of the reflected self-evaluation in sub-threshold depression (Shiota et al., 2017). However, the main purpose of that study was to test the effect of behavioral activation intervention on depression at the pre- and post-treatment stages, therefore the experiment did not involve a non-depressed control group (Shiota et al., 2017). Future studies could consider investigating the behavioral and brain activity difference of the reflected self-evaluation between depressed and non-depressed individuals.

The second research topic is the brain response of implicit self-esteem in depression. Although many studies have explored depression-related implicit self-esteem by using behavioral measures (for example, (Roberts et al., 2015; Smeijers et al., 2017; Tuijl et al., 2014; van Tuijl et al., 2016)), none has reached the neural responses underlying the processing of these measures in depression. Besides, the argument about whether explicit and implicit self-esteem are congruent in depressives is not yet settled, so it is meaningful to test the explicit and the implicit self-esteem in a single study with more heterogeneous samples (for example, sub-clinical depressed individuals without complications nor clinical intervention) by using neuroimaging approaches, which can probably reveal additional information underlying the processing of the behavioral tasks measuring the implicit self-esteem.

3.2 Study II: Behavioral and brain activity of reflected self-evaluation in sub-clinical depression

Study II investigated behavioral and brain activity differences between the dysphoric and the control groups during the processing of reflected self-evaluation. At the behavioral level, analysis of the evaluation ratings showed a significant valence × evaluation × group interaction [F (4, 208) = 6.37, p < 0.001, η_p^2 = 0.11]. The post hoc analysis of this interaction showed significant evaluation × group interactions in both the positive valence [F (4, 208) = 4.49, p = 0.006, η_p^2 = 0.08] and the negative valence [F (4, 208) = 6.57, p < 0.001, η_p^2 = 0.11]. For the positive valence, there were significant group differences in the S2S (t (52) = 3.13, p = 0.003, Cohen's d = 0.85), the C2S (t (52) = 2.33, p = 0.023, Cohen's d = 0.63), and the U2S (t (52) = 2.05, p = 0.046, Cohen's d = 0.54) evaluations. According to these group differences, the dysphoric participants rated themselves less positively than the control participants did, regardless of whose perspective the participants took. Moreover, the group difference was not significant in the S2C

(t (52) = 0.88, p = 0.384) nor in the S2U (t (52) = -.76, p = 0.450) evaluations, suggesting that the dysphoric participants rated the others as positively as the control group did. For the negative valence, the group difference was significant in all the conditions (S2S (t (52) = -7.94, $p < 0.001$, Cohen's d = 2.13); S2C (t (52) = -4.51, $p < 0.001$, Cohen's d = 1.23); S2U (t (52) = -3.34, p = 0.002, Cohen's d = 0.88); C2S (t (52) = -7.45, $p < 0.001$, Cohen's d = 1.99); U2S (t (52) = -5.00, $p < 0.001$, Cohen's d = 1.34)). The dysphoric participants rated everyone more negatively than the control group did, regardless of whose perspective they took. **Figure 5** presents the means of behavioral ratings for each of the evaluations in the two groups.

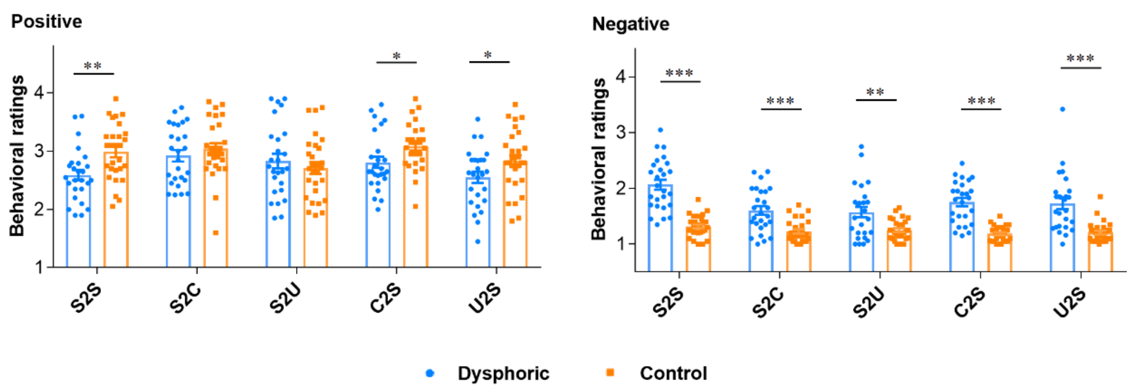

FIGURE 5 Scatter plots of behavioral ratings in Study II. Dots represent each of the evaluation ratings for the dysphoric group and the control group. S2S = self to self; S2C = self to close other (for example, the best friend); S2U = self to unclose other (for example, the unfamiliar classmate); C2S = close other to self; U2S = unclose other to self. Bars and error bars respectively represent means and standard errors. *p < .05, **p < .01, ***p < .001

On the other hand, with the positive valence, there were significant evaluation effects in both the dysphoric group [F (4, 100) = 4.48, p = 0.009, η_p^2 = 0.15] and the control group [F (4, 108) = 7.22, $p < 0.001$, η_p^2 = 0.21]. For the dysphoric participants, the S2S evaluation was significantly less positive than the S2C evaluation, while there was no difference between the S2S and the S2U evaluation. For the control participants, however, the S2S evaluation was significantly more positive than the S2U evaluation, while there was no difference between the S2S and the S2C evaluation. With the negative valence, the condition effect was also significant in both the dysphoric group [F (4, 100) = 10.55, $p < 0.001$, η_p^2 = 0.30] and the control group [F (4, 108) = 3.09, p = 0.028, η_p^2 = 0.10]. For the dysphoric participants, the S2S evaluation was significantly more negative than all of the four other evaluations, and for the control participants, the S2S evaluation showed no difference compared to the S2U evaluation, but it was also more negative than other evaluations. **Table 4** provides all the statistical data, including the means and standard deviations of each evaluation, as well as comparisons between each two evaluations.

TABLE 4 Statistical data of behavioral ratings in Study II

Group	Valence	Condition	*M*(*SD*)	*t* (df)				
				S2S	S2C	S2U	C2S	U2S
Dysphoric	Positive	S2S	2.59(.46)	—				
		S2C	2.92(.50)	-3.28(25)**	—			
		S2U	2.83(.63)	-1.76(25)	0.77(25)	—		
		C2S	2.81(.49)	-3.82(25)**	1.10(25)	0.19(25)	—	
		U2S	2.55(.50)	0.54(25)	3.02(25)**	2.22(25)*	3.32(25)**	—
	Negative	S2S	2.07(.46)	—				
		S2C	1.60(.39)	6.24(25)***	—			
		S2U	1.57(.48)	4.43(25)***	0.33(25)	—		
		C2S	1.75(.38)	6.43(25)***	-2.77(25)**	-1.96(25)	—	
		U2S	1.73(.52)	3.71(25)**	-1.56(25)	-1.33(25)	0.31(25)	—
Control	Positive	S2S	2.99(.48)	—				
		S2C	3.04(.50)	-0.59(27)	—			
		S2U	2.71(.53)	2.88(27)**	3.38(27)**	—		
		C2S	3.09(.40)	-1.90(27)	-0.72(27)	-3.98(27)***	—	
		U2S	2.83(.53)	2.30(27)*	2.24(27)*	-1.33(27)	4.27(27)***	—
	Negative	S2S	1.32(.19)	—				
		S2C	1.22(.20)	2.15(27)*	—			
		S2U	1.25(.19)	1.45(27)	-0.47(27)	—		
		C2S	1.18(.14)	4.01(27)***	1.09(27)	1.64(27)	—	
		U2S	1.21(.18)	3.04(27)**	0.37(27)	1.13(27)	-0.61(27)	—

Note. S2S = self to self; S2C = self to close other (for example, the best friends); S2U = self to unclose other (for example, the unfamiliar classmate); C2S = close other to self; U2S = unclose other to self; *M* = means; *SD* = standard deviations. *p < .05, **p < .01, ***p < .001. The table was modified from Lou, et al. (2023)

At the brain level, for the left TPJ, there was a significant evaluation × group interaction [F (2, 104) = 7.05, p = 0.001, η_p^2 = 0.12], although the valence × evaluation × group interaction was not significant [F (2, 104) = 1.78, p = 0.175]. The post hoc analysis of the evaluation × group interaction showed a significant evaluation effect in the control group [F (2, 54) = 21.73, $p < 0.001$, η_p^2 = 0.45], but not in the dysphoric group [F (2, 50) = 0.003, p = 0.997]. For the control participants, brain activity during the C2S evaluation (means ± standard deviation = 0.14 ± 0.11 %; t (27) = -4.81, $p < 0.001$, Cohen's d = 0.97) and during the U2S evaluation (0.17 ± 0.14 %; t (27) = -5.80, $p < 0.001$, Cohen's d = 1.08) were significantly greater than it was during the S2S evaluation (0.04 ± 0.11 %), while the activation difference between the C2S and the U2S evaluations was not significant (t (27) = -1.60, p = 0.24). For the right TPJ, there was also a significant evaluation × group interaction [F (2, 104) = 3.61, p = 0.032, η_p^2 = 0.07], although the valence × evaluation × group interaction was not significant [F (2, 104) = 0.09, p = 0.910]. The post hoc analysis of the evaluation × group interaction showed a significant evaluation effect, again, in the control group [F (2, 54) = 6.99, p = 0.003, η_p^2 = 0.21], but not in the dysphoric group [F (2, 50) = 0.060, p = 0.938]. For the control participants, brain activity during the U2S evaluation (0.13 ± 0.10 %) was significantly greater than it was during the S2S evaluation (0.03 ± 0.15 %; t (27) = -4.03, $p < 0.001$, Cohen's d = 0.80). The activation difference between the C2S evaluation (0.09 ± 0.11 %) and the S2S evaluation was marginal (t (27) = -2.13, p = 0.084), while the difference between the C2S and the U2S evaluations was not significant (t (27) = -1.36, p = 0.370). **Figure 6** illustrates the fMRI results.

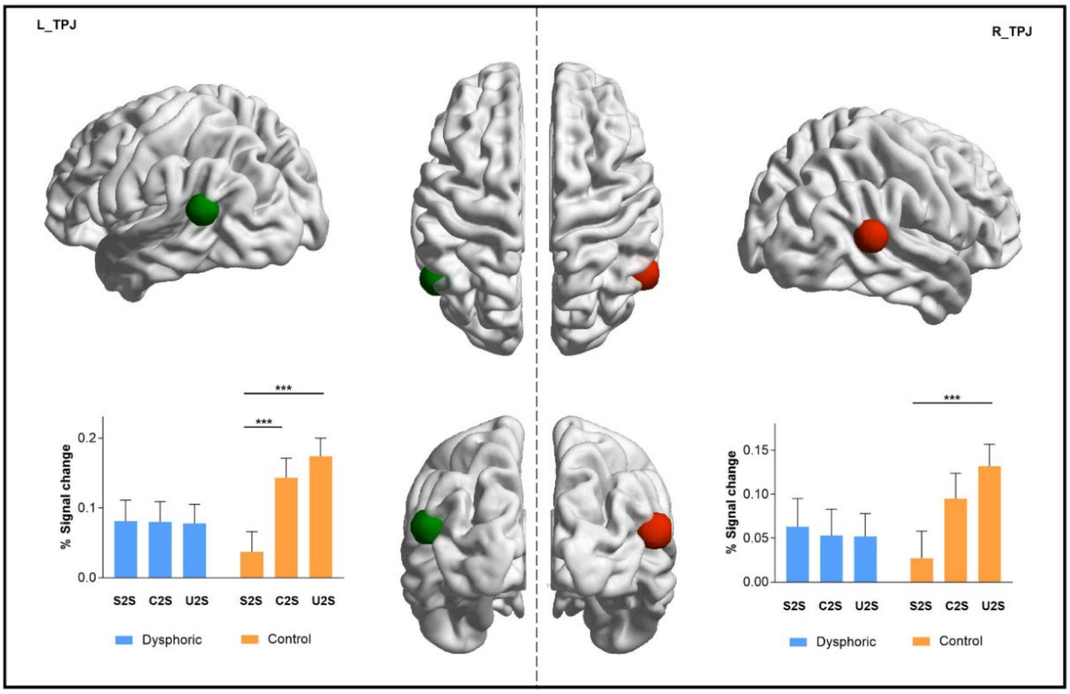

FIGURE 6 Temporal parietal junction (TPJ) activity in Study II. Mean percentage signal changes of each evaluation relative to baseline in the left TPJ (L_TPJ; illustrated by the green ball on the left panel) and the right TPJ (R_TPJ; illustrated by the red ball on the right panel). S2S = self to self; C2S = close other (for example, the best friend) to self; U2S = unclose other (for example, the unfamiliar classmate) to self. Error bars represent standard errors, ***p < .001.

Pearson's correlation coefficient analysis showed a significant negative correlation between the participants' BDI-II scores and brain activity during the C2S and the U2S evaluations, in both the left and right TPJs. The resulting correlation matrix is presented in **Table 5**. These results suggest that the more depressive symptoms the participant had, the smaller the bilateral TPJs activation was when taking others' perspectives to evaluate the self.

TABLE 5 Correlations between depressive state and TPJ activity

		BDI-II	**TPJ (L / R)**		
			S2S	C2S	U2S
L_TPJ	S2S	0.09	1		
	C2S	-0.32*	0.53**	1	
	U2S	-0.34*	0.40**	0.52**	1
R_TPJ	S2S	0.09	1		
	C2S	-0.29*	0.40**	1	
	U2S	-0.33*	0.41**	0.45**	1

Note: BDI-II = Beck Depression Inventory - II; L_TPJ = left temporal parietal junction; R_TPJ = right temporal parietal junction; S2S = self to self; C2S = close other (for example, the best friends) to self; U2S = unclose other (for example, the unfamiliar classmate) to self.
*p < .05, **p <.01

For the mPFC, results showed no significant interaction among valence, condition, and group [F (2, 104) = 0.15, p = .857], nor between valence and group [F (1, 52) = 0.52, p = .472]. The main effect of group [F (1, 52) = 0.65, p = .422] was not significant either. Although the interaction between condition and group was significant [F (2, 104) = 3.40, p = .037, η_p^2 = .06], the simple effects of condition were not significant in neither the control group [F (2, 54) = 1.65, p = .202] nor the dysphoric group [F (2, 50) = 2.27, p = .115] after breaking down the two-way interaction. When investigating the group effects in each of the three conditions, the simple effects of group were not significant for the S2S condition (t (52) = −0.73, p = .469), the C2S condition (t (52) = 1.71, p = .093), nor the U2S condition (t (52) = 0.96, p = .336).

In addition to studying the ROIs, a whole-brain analysis was exploratorily conducted to examine possible statistically significant group effects at the whole-brain level. This analysis was performed using the general lineal model (GLM), and the multiple comparison was corrected by a level of significance of $p < 0.05$ (few-small volume corrected). Both the group × valence interaction and the group × evaluation interaction were entered in a factorial design module during the second-level analysis, and the main effect of group was also examined. However, the results showed that no region had passed the significant threshold, neither for the interactions nor the main effect.

3.3 Study III: Behavioral and brain responses of IAT measuring implicit self-esteem in sub-clinical depression

Study III investigated differences in behavioral and brain responses between the dysphoric and the control groups during the IAT that measures implicit self-esteem. At the behavioral level, data analysis of reaction time showed no other significant main nor interaction effects, except for a significant main effect of self-association [F (1, 56) = 183.10, $p < 0.001$, $\eta_p^2 = 0.77$]. Participants responded faster for the self-positivity association (623.23 ± 43.80 ms) than they did for the self-negativity association (673.52 ± 43.70 ms). The analysis of accuracy also revealed no other significant main nor interaction effects, except for a significant main effect of self-association [F (1, 56) = 53.18, $p < 0.001$, $\eta_p^2 = 0.49$]. The accuracy was higher in the self-positivity association (0.96 ± 0.02) than it was in the self-negativity association (0.93 ± 0.04). The analysis of *D*-score (the IAT effect for reaction time) showed no significant group difference (t (56) = 0.13, p = 0.90, Cohen's d = 0.004), and the *D*-scores were significantly higher than zero for both the dysphoric group (0.45 ± 0.28 points; t (27) = 8.60, $p < 0.001$; Cohen's d = 1.62) and the control group (0.46 ± 0.22 points; t (29) = 11.60, $p < 0.001$; Cohen's d = 2.12). These results seem to suggest positive implicit self-esteem in both the dysphoric and the control groups. However, the *t*-test of the RSES scores showed a significant group effect (t (56) = 2.90, p = 0.005; Cohen's d = 0.76), with the dysphoric group (24.18 ± 4.27 points) exhibiting lower RSES scores than the control group did (28.03 ± 5.71 points), suggesting low explicit self-esteem in the dysphoric group compared to the control group. All these behavioral results are visualized in **Figure 7**.

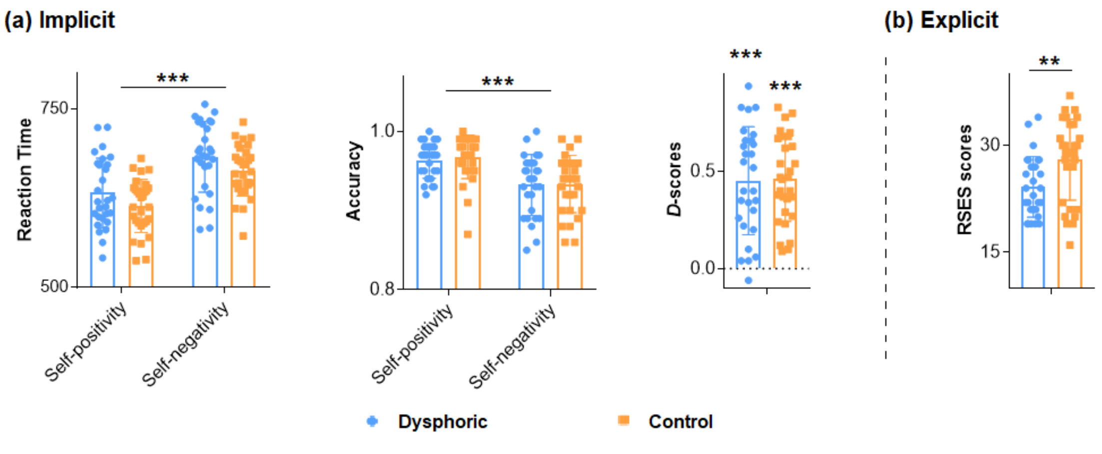

FIGURE 7 Scatter plots of behavioral indices in Study III. The left panel (a) shows indices measuring implicit self-esteem: reaction time (left), accuracy (middle), and *D*-score (right) in the Implicit Association Task. The right panel (b) shows an index measuring explicit self-esteem: Rosenberg Self-Esteem Scale (RSES) scores. Bars and error bars represent means and standard errors, respectively. **$p < .01$, ***$p < .001$

At the brain level, the analysis of the LPC amplitudes within 300–400 ms post-stimulus latency did not show 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 not significant [F (1, 56) = 4.36, p = 0.12, η_p^2 = 0.07]. The analysis of the LPC amplitudes within 400–600 ms post-stimulus latency showed a significant self-association × group interaction [F (1, 56) = 18.58, p < 0.001, η_p^2 = 0.25]. The post hoc analysis of this interaction showed a significant self-association effect in both the control group [F (1, 29) = 14.06, p < 0.001, η_p^2 = 0.33] and the dysphoric group [F (1, 28) = 5.18, p = 0.03, η_p^2 = 0.16]. In the control group, the self-positivity association (6.04 ± 2.69 uV) induced greater amplitudes than the self-negativity association did (5.24 ± 2.81 uV). In the dysphoric group, however, the self-negativity association (6.63 ± 2.36 uV) induced larger amplitudes than the self-positivity association did (6.23 ± 2.21 uV). Moreover, a significant group difference was found in the self-negativity association [F (1, 56) = 4.11, p = 0.04, η_p^2 = 0.07], but not in the self-positivity association [F (1, 56) = 0.09, p = 0.77, η_p^2 = 0.002]. Compared to the control group (5.24 ± 2.81 uV), the dysphoric group (6.63 ± 2.36 uV) exhibited larger amplitudes for the self-negativity association. The analysis of the LPC amplitudes within 600–1000 ms post-stimulus latency again showed a significant self-association × group interaction [F (1, 56) = 15.76, p < 0.001, η_p^2 = 0.22]. The post hoc analysis of this interaction showed a significant self-association effect in the dysphoric group [F (1, 27) = 19.56, p < 0.001, η_p^2 = 0.42], but not in the control group [F (1, 29) = 0.51, p = 0.48, η_p^2 = 0.02]. For the dysphoric group, the self-negativity association (3.30 ± 2.25 uV) induced larger amplitudes than the self-positivity association did (2.30 ± 1.85 uV). Moreover, the post hoc analysis of self-association × group interaction also showed a significant group effect in the self-negativity association [F (1, 56) = 5.48, p = 0.02, η_p^2 = 0.09], but not in the self-positivity association [F (1, 56) = 0.28, p = 0.60, η_p^2 = 0.01]. Relative to the control group (1.91 ± 2.25 uV), the dysphoric group (3.30 ± 2.25 uV) exhibited larger amplitude in the self-negativity association. **Figure 8** presents the neurological results.

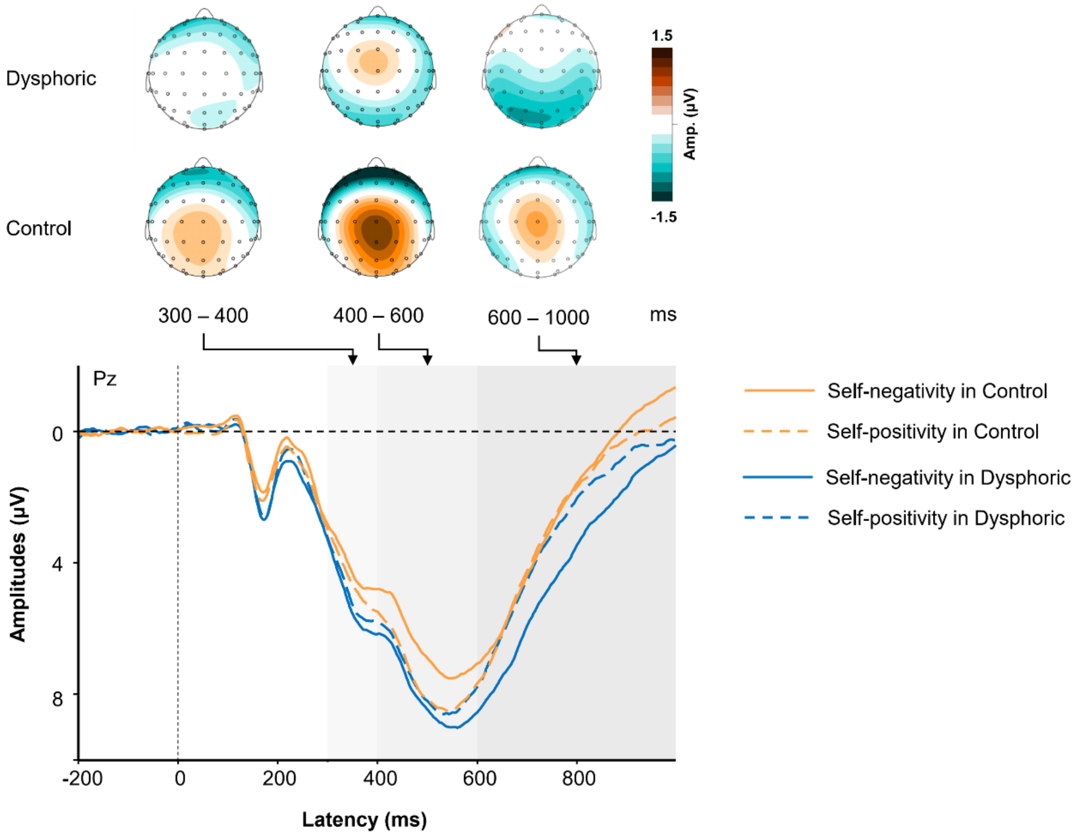

FIGURE 8 Late positive component (LPC) responses in Study III. The waveforms are presented separately for time windows of 300–400 ms, 400–600 ms, and 600–1,000 ms in the dysphoric group (blue lines) and the control group (orange lines) during the self-positivity (dashed lines) and the self-negativity (solid lines) associations. The Pz site and the corresponding scalp topographies of difference waves (self-positivity minus self-negativity associations) are illustrated for both groups separately. Orange shading indicates where the self-positivity association showed greater waveforms than the self-negativity association, and blue shading indicates where the self-negativity association showed greater waveforms.

Pearson's correlation coefficient analyses were conducted in the time windows of 400–600 ms post-stimulus latency and 600–1,000 ms post-stimulus latency. The time window of 300–400 ms post-stimulus latency was 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 ($r\ (58) = -.32$, $p = 0.01$), 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 ($r\ (58) = -.03$, $p = 0.82$). 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 600–1000 ms ($r\ (58) = 0.28$, $p = 0.03$) time windows. **Table 6** provides the resulting correlation matrix. These results suggest that the participants' behavioral performance during the IAT 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.

TABLE 6 Correlations between behavioral and LPC responses

			BDI-II	RESE	*D*-Scores	LPC Amp. (400 – 600 ms / 600 – 1000 ms)	
						Self-positivity	Self-negativity
BDI-II			1				
RSES			-0.32*	1			
D-Scores			-0.03	-0.24	1		
LPC Amp.	400-600 ms	Self-positivity	0.12	0.05	0.11	1	
		Self-negativity	0.29*	-0.07	0.14	0.89***	1
	600-1000 ms	Self-positivity	0.03	-0.08	0.01	1	
		Self-negativity	0.28*	-0.14	0.19	0.86***	1

Note: LPC = late positive component; Amp. = amplitudes; BDI-II = Beck Depressive Inventory-II; IAT = Implicit Association Task; RSES = Rosenberg Self-Esteem Scale. *p < .05; ***p < .001. The table was modified from Lou, et al. (2021)

4 DISCUSSION

The main purpose of this dissertation was to investigate the pattern of self-knowledge, including negative biased self-evaluation and abnormal self-esteem, in depression. I started the dissertation by reviewing psychological/psychiatric research that studied self-knowledge in depression (**Study I**). Based on the review, I proposed two research topics that can be considered for future studies: One pertains to the behavioral and brain activity of reflected self-evaluation in depression, another concerns the brain responses during the IAT measuring implicit self-esteem in depression. I then designed two empirical studies (**Study II** and **Study III**) to examine these two topics by using neuroimaging techniques, such as the fMRI and the ERP. In the two studies, I observed depression-related behavioral responses and brain activity during the reflected self-evaluation, and during the implicit self-association.

4.1 Negative direct self-evaluation and abnormal self-esteem in depression

In **Study I**, I reviewed the accumulating publications in the field of self-knowledge in depression and found that the depressive self-knowledge can manifest as negative self-evaluation and abnormal self-esteem.

Specifically, one's self-evaluation consists of two aspects, the direct self-evaluation and the reflected self-evaluation. For the direct self-evaluation in depression, findings in the previous studies were in line with Beck's cognitive theory, which suggested a negative self-schema in depression (Beck, 1967). At the behavioral level, such negative self-schema can be observed through the negative biased self-evaluation. The depressed individuals endorsed more negative and less positive words as self-descriptive (Auerbach et al., 2015; Bradley et al., 2016; Dainer-best et al., 2017) and recalled/recognized more of the endorsed negative words than the endorsed positive words (Auerbach et al., 2015; Romero et al., 2016; Shestyuk & Deldin, 2010). At the neurological level, researchers have observed enhanced LPC amplitudes for the negative self-endorsement compared to the positive self-endorsement among the depressives, whereas opposite LPC responses have been observed for the non-depressed controls (Auerbach et al., 2015; Dainer-best et al., 2017; Risch et al., 2010). Although some researchers have also observed the ERP difference between the depressives and the controls in early components (for example, P1 and P2; see (Auerbach et al., 2015; Poulsen et al., 2009; Shestyuk & Deldin, 2010)), the results were not always congruent across studies, as other researchers have reported no group difference in these early components (Dainer-best et al., 2017). Therefore, the results suggest that the difference in direct self-evaluation between depressed and non-depressed individuals probably occurs at a late cognitive control stage, rather than early perceptual stages (Dainer-best et al., 2017). Findings from the fMRI studies seem to support the idea that the depression-related group difference in direct self-evaluation can be found at the late cognitive control stage (Lemogne et al., 2009). For instance, compared to the non-depressed controls, the depressives exhibited altered brain activity in the mPFC (Grimm et al., 2009; Li et al., 2017; Yoshimura et al., 2010), which is considered part of the prefrontal cortex that is central for cognitive control (Menon & D'Esposito, 2022). In addition, during the processing of direct self-evaluation, the depressives uniquely recruited the extended medial prefrontal network that the non-depressed people did not engage, suggesting enhanced cognitive control in the depressed population (Lemogne et al., 2009). Taken together, these findings indicate a negative direct self-evaluation, probably driven by a late mental stage that is related to cognitive control, in depression.

Regarding the reflected self-evaluation in depression, to date, only one study (Shiota et al., 2017) has explored this topic. However, the main purpose of that study was to test the effect of psychological intervention on the improvement of depressive symptoms and on the changes of brain responses underlying the reflected self-evaluation (Shiota et al., 2017). It is still unknown how depression itself may affect an individual's perception of others' opinions about themselves. I therefore proposed that future studies could consider investigating the behavioral and brain-responding difference between a depressed group and a non-depressed group during the reflected self-evaluation.

As a result of self-evaluation, self-esteem has been extensively studied in previous research related to depression, by utilizing behavioral methods. For the explicit self-esteem, which can be consciously perceived, previous studies have provided consistent evidence for low explicit self-esteem among individuals with clinical or sub-clinical depression, in comparison to non-depressed controls (Franck et al., 2007; Franck et al., 2008; Jabben et al., 2014; Kesting et al., 2011; Lemogne et al., 2009; Roberts et al., 2015; Romero et al., 2016; Smeijers et al., 2017; van Tuijl et al., 2016). For the implicit self-esteem, which cannot be consciously perceived, previous studies have yielded conflicting results. For instance, some researchers observed low implicit self-esteem in depression (Risch et al., 2010; Romero et al., 2016), while some observed normal implicit self in depression (Kesting et al., 2011; Lemogne et al., 2009; Smeijers et al., 2017; van Tuijl et al., 2016), and others even observed high implicit self-esteem in depression (Franck et al., 2007; Gemar et al., 2001; Roberts et al., 2015), when compared to those non-depressed controls. Nevertheless, no convincing conclusion can be drawn so far, as these studies differed from each other in various aspects, such as the criteria of including/excluding the participants, the diagnosis for clinical depression, the complications that accompany depression, and the medical or psychiatric intervention targeting depression. Further research is still required in order to explore the reasons for these conflicts and to clarify the pattern of implicit self-esteem in depression. I therefore propose that future studies could consider continuing the investigation of depression-related implicit self-esteem using neuroimaging methods, which may provide additional insights not observable through behavioral indices.

4.2 Altered temporal parietal junction activity during reflected self-evaluation in sub-clinical depression

In **Study II**, I accomplished the first proposal that had been suggested in Study I. That was, investigating the depression-related behavioral responses and brain activity during reflected self-evaluation. The fMRI technique was applied because of its high spatial resolution. In this study, I observed a positive bias for the control group, while a negative bias for the dysphoric group, during the reflected self-evaluation. Moreover, increased activities in bilateral TPJs were observed during the reflected versus the direct self-evaluation for the control group, but not for the dysphoric group. These results are discussed in detail as below.

For the control group, the participants rated themselves more positively than others, especially the unclose others, suggesting a self-positivity bias in this group (Hampton & Varnum, 2018; Mezulis et al., 2004; Watson et al., 2007). For the dysphoric group, such an adaptive self-positivity bias is missing, with the dysphoric participants rating themselves less positively than others, especially the close others. In addition, the dysphoric participants rated themselves more negatively and less positively than the control participants did, regardless of whether it was through their own opinion or the opinions of others. The finding in the self-to-self evaluation is in line with previous studies, which reported that the depressed individuals, compared to those non-depressed controls, tended to endorse more negative and less positive words as self-descriptive when they evaluated the self directly through their own opinion (Dainer-best et al., 2017; Kiang et al., 2017; Shestyuk & Deldin, 2010). In the present study, I add evidence that such a negative biased behavioral pattern also occurs in the sub-clinical population, and can be observed not only during the direct self-to-self evaluation, but also during the reflected other-to-self evaluation.

The neuroimaging result from the control group is consistent with the results from previous studies by showing a greater TPJ activity during the reflected self-evaluation versus the direct self-evaluation. For instance, in a previous study that involved a similar sample of young adults and used a similar paradigm of a SRET-like reflected self-evaluation task, researchers observed a greater TPJ activity during the reflected self-evaluation versus the direct self-evaluation (Pfeifer et al., 2009). In another study that involved a similar sample of Chinese young adults, researchers also observed an engagement of the TPJ when the participants were evaluating themselves through their peers' points of view (Pfeifer et al., 2017). The engagement of TPJ in reflected self-evaluation among young adults was again reported in the work of Jankowski and colleagues (Jankowski et al., 2014). Here, my neuroimaging result from the control group provides one more piece of evidence to support the idea that the TPJ plays an important role in reflecting on others' thoughts toward the self (Jankowski et al., 2014; Pfeifer et al., 2017; Pfeifer et al., 2009).

Importantly, the neuroimaging result from the dysphoric group demonstrates a lack of TPJ involvement, indicated by similarly low activity during the reflected and the direct self-evaluation, in such reflection. In addition, the TPJ activity during the reflected self-evaluation was found to be negatively correlated with the amount of the participants' depressive symptoms. The more depressive symptoms the participants had, the less the TPJ engagement there was when the participants evaluated themselves through others' perspectives. According to previous studies, the lack of TPJ engagement here might be explained by the impaired ToM ability in sub-clinical/mild depressed population (Erle et al., 2019; Lee et al., 2005; Manstead et al., 2013). The first evidence of impaired ToM in sub-clinical depression came from Lee and colleagues' work (Lee et al., 2005). In that study, the researchers sought to investigate the ToM ability in clinical depression; however, they included not only clinically depressed outpatients but also a sub-clinical population that did not have clinical diagnoses of depression (Lee et al., 2005). Results from the study showed a significantly impaired ToM ability in the depressed group, which consisted of both clinical and sub-clinical participants, compared to a non-depressed healthy control group (Lee et al., 2005). Importantly, in Lee's study, the severely depressed participants (most of whom were clinically depressed patients) and the moderately depressed participants (most of whom were sub-clinical individuals) did not differ from each other in terms of the ToM performance (Lee et al., 2005). Other researchers then suggested that Lee's work had indicated a possibility that the ToM ability is impaired in both the clinically and the sub-clinically depressed people (Manstead et al., 2013). To test this assumption, Manstead et al. (2013) used the same ToM task that Lee had used to measure the ToM ability in a group of sub-clinical dysphoric participants; and as the researchers expected, they observed worse ToM performance in the sub-clinical group, in comparison to a non-depressed control group. Later, by using a convenient sample that consisted of sub-clinical individuals with high depressive symptoms, Erle and colleagues (2019) found that the sub-clinically depressed group underperformed the non-depressed control group in particular on ToM tasks, while their performance on other cognitive tasks was equally good as the control group. These previous studies therefore suggest an impaired ToM ability in sub-clinical depression.

Additionally, there was no significant group effect in the mPFC, suggesting that this area was not engaged here in detecting individual differences during the reflected self-evaluation task. According to previous studies, the mPFC is more of a domain-general area that plays a relatively general role in a variety of psychological processes, such as default mode activity (Qin & Northoff, 2011), emotion processing (Etkin et al., 2011), fear conditioning and extinction (Giustino & Maren, 2015), self-control (Tang et al., 2016), memory and decision making (Euston et al., 2012), and so on. Although the mPFC engagement was observed in some of the previous studies investigating the direct and the reflected self-evaluation (for examples, (Li et al., 2017; Liu et al., 2020; Pfeifer et al., 2017; Shiota et al., 2017; Van der Cruijsen et al., 2019)), its activation might actually be mediated by some meta-cognitive processes that are recruited by both the two types of self-evaluations, rather than the self-referential processing (for example, during direct self-evaluation) or the perspective taking (for example, during reflected self-evaluation) per se (Otti et al., 2015). Thus, other previous research reported no mPFC engagement in certain ToM tasks, suggesting that mPFC engagement is not always necessary for ToM (Otti et al., 2015). In the present study, it is reasonable to have found no mPFC activity difference between the dysphoric group and the control group because there was no evidence of dysfunction at a general level among the dysphoric participants.

4.3 Enhanced LPC response during negative versus positive implicit self-association in sub-clinical depression

In **Study III**, I accomplished the second proposal that had been suggested in Study I. That was, using the ERP technique to investigate brain responses underlying the processing of depression-related implicit self-association. Significant group differences were observed in the LPC during the time window of 400–1,000 ms post-stimulus latency. During the IAT that measures implicit self-esteem, the control group showed enhanced LPC amplitudes for the self-positivity block (in which the *me* pronouns were implicitly associated with *positive* adjectives, while the *not-me* pronouns were associated with *negative* adjectives) versus the self-negativity block (in which the *me* pronouns were implicitly associated with *negative* words, while the *not-me* pronouns were associated with *positive* words). Nevertheless, the dysphoric group showed an opposite responding pattern to the control group did.

For the control group, the result is consistent with previous findings, which indicated larger LPC amplitudes for self-positivity condition versus self-negativity condition (labeled as "congruent" versus "incongruent", or "compatible" versus "incompatible" conditions in these studies) among healthy participants (Fleischhauer et al., 2014; Wu et al., 2016). According to these studies, the enhanced LPC amplitude during the IAT occurs because the participants usually engage more voluntary attention and effortful stimulus evaluation in the condition that is congruent with their implicit attitude (Fleischhauer et al., 2014; Wu et al., 2014; Xiao et al., 2015). As a result, they perform more efficiently in the congruent versus the incongruent condition (Fleischhauer et al., 2014; Wu et al., 2016; Xiao et al., 2015). However, for the condition that is incongruent with the participants' implicit attitude, the LPC amplitude might be suppressed because a stimulus' informative value cannot be fully extracted if the participants have doubts regarding their responses (Coates & Campbell, 2010). In line with this interpretation, the enhanced LPC amplitudes I observed for the self-positivity versus self-negativity block suggests that the self-is-positive association is more congruent, than the self-is-negative association is, with the implicit self-esteem in the control group. This funding further supports the presence of a self-positivity bias in the healthy population (Chen et al., 2014; Egenolf et al., 2013). For the dysphoric group, by contrast, larger LPC amplitudes were observed for the self-negativity versus self-positivity block from 400 ms to 1,000 ms post-stimulus latency. This result indicates sustained engagement of voluntary attention and effortful stimulus evaluation, resulting in more efficient performance, during the self-negativity versus self-positivity association. The finding therefore suggests low implicit self-esteem in sub-clinical depression.

Noticeably, there was no significant brain responding difference between the self-positivity block and the self-negativity block during the time window of 300-400 ms post-stimuli latency, where early perceptual processing occurs (Grundy et al., 2015; Yang & Zhang, 2009). The result might be explained by the fact that the stimulus words used in this study were identical Chinese words without apparent perceptual differences. Specifically, the early LPC (or P3-like) responses are usually associated with automatic stimulus processing, such as automatic attention capture to novelty or emotionally salient stimuli (Dainer-best et al., 2017; Polich, 2007). In the present study, however, all the stimulus words consist of two Chinese characters, and these characters across the four categories were visual similar, as indicated by their similar average stroke numbers. In addition, the positive and the negative stimulus words did not significantly differ from each other in terms of the average ratings of comprehensibility, familiarity, and arousal level. Taken together, none of the stimulus categories should have an advantage in capturing the participants' automatic attention. Furthermore, no significant group differences were observed during the same time window. This result aligns with previous findings indicating that individual differences in the IAT are primarily driven by late mental processes, such as cognitive control, rather than early perceptual processing (Schiller et al., 2016). The absence of group differences in early mental stages also supports prior research showing that, in the processing of self-related information, individuals with depression differ from non-depressed controls at a later cognitive stage rather than during early perceptual stages (Dainer-best et al., 2017). Overall, the ERP results in my study suggest a stronger implicit self-association with negative personality traits than with positive personality traits, indicating low implicit self-esteem in sub-clinical depression. Furthermore, these results support that the depressive implicit self-esteem, as measured by the IAT, is likely influenced by enhanced cognitive control under the self-is-negative association, rather than by early perception of the negative self-related stimuli.

However, an unexpected finding was that the dysphoric group and the control group did not show significant differences at the behavioral level during the IAT performance. Both groups exhibited faster and more accurate responses to the self-positivity condition compared to the self-negativity condition. The behavioral results seem to suggest an undifferentiated positive implicit self-esteem in both groups, so the results are not congruent with the ERP findings. However, I found that the behavioral indices (the reaction time and the accuracy) in my study might not be as sensitive as the ERP responses were in detecting group differences. First, according to a previous study, the IAT performance, such as reaction time and accuracy, can be largely affected by practice, as the practice may diminish individual difference by making fake success (Röhner et al., 2011). However, as an ERP study, sufficient practice was critical for my experiment because it helped the participants to fully understand the requirement of the measurement, and therefore ensured enough valid trials (that is, trials with correct response) for ERP data analysis. Inevitably, the practice might have improved the participants' behavioral performance during the IAT, making it impossible to detect potential group differences. This speculation can be partly proved by the high mean accuracies and low standard deviations in both the self-positivity condition (0.96 ± 0.02) and the self-negativity condition (0.93 ± 0.04). Second, compared with the behavioral index, ERP response may be more informative because of its advantage in unfolding mental states at the temporal dimension, and it should therefore be able to detect individual differences pertaining to certain phases of an ongoing mental processing (such as the observed late cognitive processing during the IAT). Finally, the results of the correlation analysis further confirmed my speculation. On one hand, it was the ERP responses (the LPC amplitudes), rather than the behavioral IAT performance (the *D*-scores), that correlated with the participants' depressive symptoms (their BDI-II scores) in the current study. On the other hand, the LPC amplitudes in the self-negativity block positively correlated with the participants' BDI-II scores, suggesting that the enhanced engagement of voluntary attention under the self-negativity association is associated with the higher level of depressive symptoms.

4.4 General discussion

Through the three individual studies, a negative biased self-knowledge is found in individuals with enhanced depressive symptoms. Specifically, by reviewing previous publications, **Study I** suggests a negative bias in direct self-evaluation and in explicit self-esteem among clinically depressed individuals. By conducting empirical investigations, **Study II** suggests that such a negative bias also exists in reflected self-evaluation and among sub-clinically depressed individuals. **Study III** then suggests that, as a consequence of negative self-evaluation, the sub-clinically depressed individuals may show low self-esteem – reflected through both conscious self-report and unconscious self-association – compared to the non-depressed individuals.

The neuroimaging results are the highlights of my dissertation. In **Study I**, I found that, as suggested by previous fMRI and ERP studies (for example, (Dainer-best et al., 2017; Lemogne et al., 2009)), the difference in direct self-evaluation between depressed and non-depressed individuals mainly occurs at a late mental stage related to cognitive control. In **Study II**, my fMRI findings suggest that the difference in reflected self-evaluation between sub-clinically depressed and non-depressed individuals also occurs at a late mental stage. Specifically, at the stage involves the utilization of one's Theory of Mind (ToM) ability, reflected by TPJ engagement. In **Study III**, my ERP findings suggest that the group difference in implicit self-esteem, measured by employing the IAT paradigm, can also be identified at a late mental stage. This stage involves voluntary attention and delicate stimulus evaluation, reflected by LPC responses. Overall, these findings support Beck's cognitive theory (Beck, 1965) that depression can influence individuals understanding of themselves. More importantly, these influences are probably driven by altered mental processes, such as enhanced cognitive control during direct negative self-evaluation, worsened perspective-taking during reflected self-evaluation, and stronger voluntary engagement during self-negativity association.

However, caution should be exercised when adopting the interpretations and generalizing the findings to a wider community. For instance, considering how little we know about the function of the TPJ, I cannot rule out the possibility that the TPJ activity in **Study II** is specific to social cognition. It is known that the TPJ is also activated in other tasks; for example, the right TPJ is found to be activated in spatial attention tasks (Dugué et al., 2018; Käsbauer et al., 2020; Krall et al., 2015; Krall et al., 2016). In addition, although the lack of TPJ engagement during the reflected self-evaluation is interpreted as impaired ToM ability in sub-clinical depression, future studies are needed to directly investigate the ToM abilities in the same population before a firm conclusion can be drawn. Besides, I tend to not draw strong conclusions in **Study III** considering the incongruency between my ERP results and my behavioral results. And the lack of a strong conclusion means that it is still unclear why existing studies that I have summarized in **Study I** had observed conflicting results by employing the IAT paradigm. Noticeably, although the IAT is currently the most used paradigm as a measure of implicit attitude, its validity has been criticized in multiple studies (Bading & Stahl, 2020; Schimmack, 2021a, 2021b; Vianello & Bar-Anan, 2021). Thus, future research could consider investigating the depression-related implicit self-esteem and testing the stability of my findings by using other implicit paradigms, such as the GNAT and so on. Moreover, although the determination of the sample size in **Study II** can be referred to previous research using similar methods, it was relatively limited and may carry the risk of perpetuating a tradition of underpowered studies, therefore the replicability of my current findings should be examined in bigger samples in future studies. Last, since all the participants in **Studies II and III** are Chinese, it is unknown whether the pattern of the depressive self-knowledge that I observed here can be generalized to individuals from other cultures. For instance, the collectivist culture of Eastern Asia, relative to individualism in Western culture, encourages people to think of/compare themselves in a wider social network. Therefore, the individuals' self-knowledge in eastern Asian countries can be largely affected by their social relationships and others' opinions about themselves (Chiao et al., 2009; Pfeifer et al., 2017; Yue & Huang, 2012). Thus, the pattern of depression-related self-knowledge that I observed in my studies, particularly in the reflected self-evaluation study (**Study II**), may not be fully applicable for Western participants. Future studies could consider investigating the effect of culture difference on depressive self-knowledge.

Except for the topics that I investigated in my dissertation, I also found several interesting directions that can be further studied. First and most importantly, it is meaningful to consider how to apply the existing findings to clinical practices. For example, can we use the findings to develop biomarkers that can identify the onset of clinical depression and track its development? To test this application, a large number of studies and meta-analyses are needed to, first, organize the bio-responding patterns that can identify different subgroups of depression from a healthy population; and second, compare these patterns with other mental disorders such as social anxiety and so on. Researchers could also consider how to use the neuroimaging methods to test the validity of psychological interventions that aim to improve the negative self-knowledge in depression. For example, previous studies have found that negative self-esteem can be improved by utilizing positive self-images (Hulme et al., 2012) and mindfulness (Keng et al., 2016). Particularly for treating depression, cognitive behavioral therapy and behavioral activation are found to be able to improve the patients' depressive symptoms, and change the altered brain responses underlying the negative self-knowledge (Shiota et al., 2017; Yoshimura et al., 2014). However, the existing studies are limited, more research is needed to test the effect of other interventions that have been used in clinical treatment. Overall, in the exploration of self-knowledge in depression, there are still many topics that are worth investigating, and hopefully my dissertation can serve as a steppingstone for other researchers.

YHTEENVETO (SUMMARY)

Masennus ja itsetunto: käyttäytymis- ja aivovasteet itsearviointiin ja implisiittiseen itsetuntoon subkliinisessä masennuksessa

Masennus on mielialahäiriö, joka vaikuttaa siihen, miten ihmiset ajattelevat itsestään. Masennuksesta kärsivät yksilöt taipuvat näkemään itsensä vääristyneellä, negatiivisella tavalla, riippumatta siitä, onko tällainen näkemys linjassa todellisuuden kanssa (Beck, 1967; Bradley & Mathews, 1983; Derry & Kuiper, 1981). Tämä väitöskirja tutkii, miten masennus vaikuttaa yksilön käsitykseen itsestä ja tähän itsetuntemukseen liittyvään aivojen toimintaan.

Tutkimus I oli kirjallisuuskatsaus, joka tiivisti aiempia käyttäytymiseen ja aivokuvantamiseen liittyviä tutkimuksia negatiivisesta itsearvioinnista ja poikkeavasta itsetunnosta masennuksessa. Katsauksessa havaittujen tutkimusaukkojen perusteella ehdotin kahta aihetta tulevia empiirisiä tutkimuksia varten. Ensimmäisessä tutkimuksessa käsiteltiin masennukseen liittyviä käyttäytymis- ja aivojen vasteita, jotka liittyvät reflektoivaan itsearviointiin eli itsearviointiin toisten näkökulmasta. Toinen koski aivojen vasteita, jotka liittyvät masennukseen kytkeytyvään implisiittiseen itsetuntoon. Näitä kahta teemaa tutkittiin erikseen **Tutkimuksessa II** ja **Tutkimuksessa III**.

Katsausartikkeli **tutkimuksessa I** havaitsin, että aiemmat tutkimukset fokusoivat negatiiviseen suoraan itsearviointiin eli itsearviointiin omasta näkökulmasta sekä kliinisillä että subkliinisillä masennuksesta kärsivillä henkilöillä (esimerkiksi, Auerbach et al., 2015; Dainer-best et al., 2017; Li et al., 2017; Liu et al., 2020). Poikkeava aivojen toiminta aivokuoren keskiosien rakenteissa liittyi negatiiviseen suoran itsearvioinnin vääristymiseen (Bradley et al., 2016; Grimm et al., 2009; Lemogne et al., 2009; Li et al., 2017). Aiemmat tutkimukset ovat johdonmukaisesti myös havainneet negatiivisen itsearvioinnin liittyvänmatalampaan eksplisiittiseen itsetuntoonniin kliinisillä kuin subkliinisillä masennuksesta kärsivillä henkilöillä verrattuna niihin, jotka eivät ole koskaan kärsineet masennuksesta (esimerkiksi, Franck et al., 2007; Franck et al., 2008; Kesting et al., 2011; Roberts et al., 2015). Näistä nousi kaksi näkökohtaa empiirisiin tutkimuksiin. Ensinnäkin aiemmat tutkimukset ovat harvoin tutkineet masennuksen vaikutusta reflektiiviseen itsearviointiin. Oli siten mielekästä verrata käyttäytymis- ja aivovasteiden eroja masennusta sairastavien ja masennusta sairastamattomien henkilöiden välillä reflektiivisen itsearvioinnin suorittamisen aikana. Toiseksi aiemmat löydökset liittyen masennukseen liittyvään implisiittiseen itsetuntoon ovat ristiriitaisia; tutkijat ovat havainneet, käyttäen käyttäytymiseen liittyviä kokeellisia paradigmoja, masentuneilla alhaisempaa, normaalia tai jopa korkeampaa implisiittistä itsetuntoa verrattuna niihin, jotka eivät ole koskaan kärsineet masennuksesta, (esimerkiksi, Randenborgh et al., 2016; Risch et al., 2010; Romero et al., 2016; Smeijers et al., 2017; van Tuijl et al., 2016). Koska näistä ristiriitaisista tuloksista ei voitu vetää yksiselitteisiä johtopäätöksiä, masennukseen liittyvää implisiittistä itsetuntoa tulisi tutkia jatkossakin käyttäen monipuolisia menetelmiä, kuten neurokuvantamistekniikoita.

Tutkimuksessa II käytin funktionaalista magneettikuvausta (functional magnetic resonance imaging, fMRI) tutkiakseni, miten masennus vaikuttaa käyttäytymis- ja aivojen vasteisiin reflektiivisen itsearvioinnin aikana. Kutsuin tutkimukseen ryhmän subkliinisen masennuksen oireita kokevia henkilöitä ja ryhmän henkilöitä, jotka eivät olleet koskaan kärsineet masennuksesta (joita tutkimuksessani nimitettiin vastaavasti dysforiseksi ryhmäksi ja kontrolliryhmäksi). Behavioraalisten ja fMRI-mittausten aikana osallistujia pyydettiin arvioimaan itseään ei ainoastaan omien mielipiteidensä perusteella vaan myös muiden näkökulmasta (esimerkiksi, parhaan ystävän tai tuntemattoman luokkatoverin näkökulmasta) käyttäen joukkoa positiivisia tai negatiivisia persoonallisuuspiirre-sanoja. Nämä tulokset paljastivat, että dysforinen ryhmä, verrattuna kontrolliryhmään, osoitti korkeampia negatiivisia arvioita ja matalampia positiivisia arvioita sekä suorassa että reflektiivisessä itsearvioinnissa, riippumatta siitä, minkä näkökulman he valitsivat. Nämä tulokset tukevat aiempien tutkimusten havaintoja negatiivisesta vääristymästä masentuneilla yksilöillä silloin, kun he arvioivat itseään suoraan. Tulokset tarjoavat näyttöä masennuksesta kärsivien negatiivisesta vääristymästä myös reflektiivisessä, eli muiden mielipiteitä heijastavassa itsearvioinnissa. Lisäksi, kuten olin odottanut, aivojen temporo-parietaalisen alueen (temporoparietal junction, TPJ) aktivoituminen oli poikkeavaa dysforisessa ryhmässä. Tässä ryhmässä ei havaittu eroa, toisin kuin kontrolliryhmässä, aktivaatiossa reflektiivisen itsearvioinnin aikana verrattuna suoraan itsearviointiin. Tämä havainto kontrolliryhmässä tukee aiempia havaintoja, jotka osoittavat, että TPJ:llä on rooli arvioitaessa itseä muiden mielipiteiden näkökulmasta (Pfeifer et al., 2017; Pfeifer et al., 2009; Van der Cruijsen et al., 2019). Mielenkiintoista oli, että vastaavaa TPJ-aktivaatioeroa ei havaittu dysforisessa ryhmässä, mikä viittaa toisten näkökulman ottamisen puutteeseen. Tämä todennäköisesti subkliinisen masennuksen heikentyneellä mielenteorian kyvyllä reflektiivisen itsearvioinnin aikana.

Tutkimuksessa III käytin implisiittistä assosiaatiotestiä (IAT) tapahtumasidonnaisten aivovasteiden (brain event-related potentials, ERPs) mittaamisenaikana tutkiakseni, miten aivojen vasteet liittyvät dysforisten implisiittiseen itsetuntoon. Kuten aiemmassa osa-tutkimuksessa, tutkimuksiin kutsuttiin sekä dysforinen ryhmä että kontrolliryhmä, jossa tutkittavat eivät olleet koskaan kokeneet masennusta. EEG-mittausten aikana aikana osallistujia pyydettiin vastaamaan itseään koskeviinsekä positiivisiin että negatiivisiin persoonallisuuden piirteisiin. ERP-tulokset osoittivat, että dysforisen ja kontrolliryhmän välillä oli eroa myöhäisessä aivovasteessa (late positive component, LPC). Kuten olin odottanut, kontrolliryhmä osoitti suurempia aivovasteita itseä koskeviin myönteisiin piirteisiin kuin kielteisiin piirteisiin, viitaten kykyyn yhdistää itsensä positiivisiin persoonallisuuden piirteisiin. Tulos tukee aiempia havaintoja, jotka viittaavat korkeaan implisiittiseen itsetuntoon henkilöillä, jotka eivät ole kärsineet masennuksesta (Chen et al., 2014; Egenolf et al., 2013; Wu et al., 2016). Sen sijaan dysforisessa ryhmässä LPC-vasteet olivat suurempia itseä koskeviin negatiivisiin kuin positiivisiin persoonallisuuden piirteisiin, mikä osoittaa, että itseen oli yhdistetty korostetusti negatiivisia persoonallisuuden piirteitä. Tulos viittaa siten alhaiseen implisiittiseen itsetuntoon dysforisissa henkilöissä. Nämä löydökset ovat ensimmäisiä neurokuvantamistuloksia masennukseen liittyvästä alhaisesta implisiittisestä itsetunnosta.

Kolmen yksittäisen tutkimuksen tulokset tulkittiin Beckin kognitiivisen masennusteorian (Beck, 1967) viitekehyksessä. Teorian mukaan yksi masennuksen keskeisistä oireista on negatiivinen kognitiivinen skeema, joka saa masennuksesta kärsivät yksilöt näkemään maailman korostetusti negatiivisena (Beck, 1967). Koska skeema itsestä (itse-skeema) on osa yksilön kognitiivista skeemaa, on ajateltu, että masennuksesta kärsivillä yksilöillä on itsestä negatiivinen vääristymä (Beck, 1967). Negatiivisen itse-skeeman käsite on saanut tukea aikaisemmista tutkimuksista, jotka ovat osoittaneet, että masennuksessa ilmenee esimerkiksi liiallista itsekritiikkiä, ylikorostunutta itsensä syyttämistä ja syyllisyyden tunteita (esimerkiksi, Alexander et al., 1999; Thew et al., 2017). Tutkimukseni osoittaa empiirisesti sekä käyttäytymis- että neurokuvantamismenetelmin, että negatiivinen itse-skeema ilmenee paitsi suorassa itsearvioinnissa myös reflektiivisessä itsearvioinnissa. Lisäksi väitöskirjani osoittaa, että negatiivinen itse-skeema voi heijastua sekä eksplisiittiseen että implisiittiseen itsetuntoon. Lisäksi tällainen negatiivinen itse-skeema voi ilmetä, ei vain kliinisessä masennuksessa, vaan myös subkliinisessä masennuksessa, dysforiassa. Tulevat tutkimukset voivat käyttää tuloksiani kehittäessään hoitoja, joiden tavoitteena on korjata itseen liittyvää harhakäsitystä masennuksessa.

REFERENCES

Alexander, B., Brewin, C. R., Vearnals, S., Wolff, G., & Leff, J. (1999). An investigation of shame and guilt in a depressed sample. *The British Journal of Medical Psychology*, *72*(3), 323–338. https://doi.org/10.1348/000711299160031

APA. (n.d.-a). APA Dictionary of Psychology. https://dictionary.apa.org/self

APA. (n.d.-b). APA Dictionary of Psychology. https://dictionary.apa.org/self-esteem

Apperly, I. A., Samson, D., Chiavarino, C., & Humphreys, G. W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *Journal of Cognitive Neuroscince*, *16*(10), 1773–1784. https://doi.org/10.1162/0898929042947928

Auerbach, R. P., Stanton, C. H., Proudfit, G. H., & Pizzagalli, D. A. (2015). Self-referential processing in depressed adolescents: A high-density event-related potential study. *Journal of Abnormal Psychology*, *124*(2), 233–245. https://doi.org/10.1037/abn0000023

Bading, K., & Stahl, C. (2020). Why a standard IAT effect cannot provide evidence for association formation: the role of similarity construction. *34*(1), 128-143. https://doi.org/10.1080/02699931.2019.1604322

Bailey, J. A., 2nd. (2003). The foundation of self-esteem. *Journal of the National Medical Association*, *95*(5), 388–393.

Beck, A. T. (1967). *Depression: Clinical, Experimental, and Theoretical Aspects*. Hoeber Medical Division, Harper & Row. https://books.google.fi/books?id=L3NHAAAAMAAJ

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

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory – Second edition*. Psychological Corporation.

Bora, E., & Berk, M. (2016). Theory of mind in major depressive disorder: A meta-analysis. *Journal of Affective Disorders*, *191*, 49–55. https://doi.org/10.1016/j.jad.2015.11.023

Bradley, B., & Mathews, A. (1983). Negative self-schemata in clinical depression. *British Journal of Clinical Psychology*, *22 (Pt 3)*, 173–181. https://doi.org/10.1111/j.2044-8260.1983.tb00598.x

Bradley, K. A., Colcombe, S., Henderson, S. E., Alonso, C. M., Milham, M. P., & Gabbay, V. (2016). Neural correlates of self-perceptions in adolescents with major depressive disorder. *Developmental Cognitive Neuroscience*, *19*, 87–97. https://doi.org/10.1016/j.dcn.2016.02.007

Brian A. Nosek, & Banaji, M. R. (2001). The Go/No-Go Association Task. *Social Cognition*, *19*(6), 625–666. https://doi.org/10.1521/soco.19.6.625.20886

Carlson, E. N., & Barranti, M. (2016). Metaperceptions: Do people know how others perceive them? In *The social psychology of perceiving others*

accurately. (pp. 165–182). Cambridge University Press. https://doi.org/10.1017/CBO9781316181959.008
Carlson, E. N., & Kenny, D. A. (2012). Meta-accuracy: Do we know how others see us? In *Handbook of self-knowledge.* (pp. 242–257). The Guilford Press.
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. 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
Chesney, E., Goodwin, G. M., & Fazel, S. (2014). Risks of all-cause and suicide mortality in mental disorders: a meta-review. *World Psychiatry Official Journal of the World Psychiatric Association*, *13*(2), 153. https://doi.org/10.1002/wps.20128
Chiao, J. Y., Harada, T., Komeda, H., Li, Z., Mano, Y., Saito, D., . . . Iidaka, T. (2009). Neural basis of individualistic and collectivistic views of self. *Human Brain Mapping*, *30*(9), 2813–2820. https://doi.org/10.1002/hbm.20707
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
Cooley, C. H. (1902). *Human Nature and the Social Order*. Scribner's.
D'Argembeau, A., Ruby, P., Collette, F., Degueldre, C., Balteau, E., Luxen, A., . . . Salmon, E. (2007). Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *Journal of Cognitive Neuroscience*, *19*(6), 935–944. https://doi.org/10.1162/jocn.2007.19.6.935
Dainer-best, J., Trujillo, L. T., Schnyer, D. M., & Beevers, C. G. (2017). Sustained engagement of attention is associated with increased negative self-referent processing in major depressive disorder. *Biological Psychology*, *129*, 231–241. https://doi.org/10.1016/j.biopsycho.2017.09.005
Davey, C. G., Breakspear, M., Pujol, J., & Harrison, B. J. (2017). A Brain Model of Disturbed Self-Appraisal in Depression. *American Journal of Psychiatry*, *174*(9), 895–903. https://doi.org/10.1176/appi.ajp.2017.16080883
Denny, B. T., Kober, H., Wager, T. D., & Ochsner, K. N. (2012). A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *24*(8), 1742–1752. https://doi.org/10.1162/jocn_a_00233
Derry, P. A., & Kuiper, N. A. (1981). Schematic processing and self-reference in clinical depression. *Journal of Abnormal Psychology*, *90*(4), 286. https://doi.org/10.1037/0021-843x.90.4.286

Dugué, L., Merriam, E. P., Heeger, D. J., & Carrasco, M. (2018). Specific Visual Subregions of TPJ Mediate Reorienting of Spatial Attention. *Cereb Cortex*, *28*(7), 2375–2390. https://doi.org/10.1093/cercor/bhx140

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

Erle, T. M., Barth, N., & Topolinski, S. (2019). Egocentrism in sub-clinical depression. *Cognition and Emotion*, *33*(6), 1239–1248. https://doi.org/10.1080/02699931.2018.1552120

Etkin, A., Egner, T., & Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Cognitive Sciences*, *15*(2), 85–93. https://doi.org/10.1016/j.tics.2010.11.004

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, *76*(6), 1057–1070. https://doi.org/10.1016/j.neuron.2012.12.002

Fava, Maurizio, Kendler, & Kenneth, S. (2000). Major Depressive Disorder. *Neuron*, *28*(2), 335. https://doi.org/10.1016/S0896-6273(00)00112-4

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

Franck, E., Raedt, R. D., & Houwer, J. D. (2008). Activation of latent self-schemas as a cognitive vulnerability factor for depression: The potential role of implicit self-esteem. *Cognition and Emotion*, *22*(8), 1588–1599. https://doi.org/10.1080/02699930801921271

Gemar, M. C., Segal, Z. V., Sagrati, S., & Kennedy, S. J. (2001). Mood-induced changes on the Implicit Association Test in recovered depressed patients. *Journal of Abnormal Psychology*, *110*(2), 282. https://doi.org/10.1037/0021-843x.110.2.282

Giustino, T. F., & Maren, S. (2015). The Role of the Medial Prefrontal Cortex in the Conditioning and Extinction of Fear. *Frontiers in Behavioral Neuroscience*, *9*, 298. https://doi.org/10.3389/fnbeh.2015.00298

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery Clinics of North America*, *22*(2), 133–139, vii. https://doi.org/10.1016/j.nec.2010.11.001

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., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality & Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197
Grimm, S., Ernst, J., Boesiger, P., Schuepbach, D., Hell, D., Boeker, H., & Northoff, G. (2009). Increased self-focus in major depressive disorder is related to neural abnormalities in subcortical-cortical midline structures. *Human Brain Mapping*, *30*(8), 2617–2627. https://doi.org/10.1002/hbm.20693
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
Gu, N., Yang, H., & Wang, F. (2014). Name Letter Effect: An Index of Implicit Self-Esteem. *Advances in Psychology (In Chinese)*, *4*, 284–293. https://doi.org/10.12677/ap.2014.42041
Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development*, *83*(6), 1853–1868. https://doi.org/10.1111/j.1467-8624.2012.01829.x
Hampton, R. S., & Varnum, M. E. W. (2018). Do cultures vary in self-enhancement? ERP, behavioral, and self-report evidence. *Social Neuroscience*, *13*(5), 566–578. https://doi.org/10.1080/17470919.2017.1361471
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.
Hulme, N., Hirsch, C., & Stopa, L. (2012). Images of the self and self-esteem: Do positive self-images improve self-esteem in social anxiety? *Cognitive Behaviour Therapy*, *41*(2), 163. https://doi.org/10.1080/16506073.2012.664557
Jabben, N., de Jong, P. J., Kupka, R. W., Glashouwer, K. A., Nolen, W. A., & Penninx, B. W. (2014). Implicit and explicit self-associations in bipolar disorder: a comparison with healthy controls and unipolar depressive disorder. *Psychiatry Research: Neuroimaging*, *215*(2), 329–334. https://doi.org/10.1016/j.psychres.2013.11.030
James, W. (1890). *The Principles of Psychology*. Henry Holt and Co. https://doi.org/10.2307/1423317

Jankowski, K. F., Moore, W. E., Merchant, J. S., Kahn, L. E., & Pfeifer, J. H. (2014). But do you think I'm cool? Developmental differences in striatal recruitment during direct and reflected social self-evaluations. *Developmental Cognitive Neuroscience*, *8*, 40–54. https://doi.org/10.1016/j.dcn.2014.01.003

Janoffbulman, R. (1989). Assumptive worlds and the stress of traumatic events: Applications of the schema construct. *Social Cognition*, *7*(2), 113–136.

Ji, J. L., Grafton, B., & MacLeod, C. (2017). Referential focus moderates depression-linked attentional avoidance of positive information. *Behaviour Research and Therapy*, *93*, 47–54. https://doi.org/10.1016/j.brat.2017.03.004

Käsbauer, A. S., Mengotti, P., Fink, G. R., & Vossel, S. (2020). Resting-state Functional Connectivity of the Right Temporoparietal Junction Relates to Belief Updating and Reorienting during Spatial Attention. *Journal of Cognitive Neuroscience*, *32*(6), 1130–1141. https://doi.org/10.1162/jocn_a_01543

Keng, S. L., Seah, S. T. H., Tong, E. M. W., & Smoski, M. (2016). Effects of brief mindful acceptance induction on implicit dysfunctional attitudes and concordance between implicit and explicit dysfunctional attitudes. *Behaviour Research and Therapy*, *83*, 1–10. https://doi.org/10.1016/j.brat.2016.05.004

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

Kiang, M., Farzan, F., Blumberger, D. M., Kutas, M., McKinnon, M. C., Kansal, V., . . . Daskalakis, Z. J. (2017). Abnormal self-schema in semantic memory in major depressive disorder: Evidence from event-related brain potentials. *Biological Psychology*, *126*, 41–47. https://doi.org/10.1016/j.biopsycho.2017.04.003

Klonsky, E. D., May, A. M., & Saffer, B. Y. (2016). Suicide, suicide attempts, and suicidal ideation. *Annual Review of Clinical Psychology*, *12*(1), 307. https://doi.org/10.1146/annurev-clinpsy-021815-093204

Knyazev, G. G., Savostyanov, A. N., Bocharov, A. V., & Rudych, P. D. (2021). How Self-Appraisal is Mediated by the Brain. *Frontiers in Human Neuroscience*, *15*, 700046. https://doi.org/10.3389/fnhum.2021.700046

Koole, S. L., Dijksterhuis, A., & Van, K. A. (2001). What's in a name: implicit self-esteem and the automatic self. *Journal of Personality & Social Psychology*, 80(4), 669. https://doi.org/10.1037/0022-3514.80.4.669

Koole, S. L., & Pelham, B. W. (2003). On the nature of implicit self-esteem: The case of the Name Letter Effect. In S. Spencer, S. Fein, & M. P. Zanna (Eds.), *Motivated social perception: The Ontario Symposium* (pp. 93–116). NJ: Lawrence Erlbaum.

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., . . . Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, *220*(2), 587–604. https://doi.org/10.1007/s00429-014-0803-z

Krall, S. C., Volz, L. J., Oberwelland, E., Grefkes, C., Fink, G. R., & Konrad, K. (2016). The right temporoparietal junction in attention and social interaction: A transcranial magnetic stimulation study. *Human Brain Mapping*, *37*(2), 796–807. https://doi.org/10.1002/hbm.23068

Kuiper, N. A., & Derry, P. A. (1982). Depressed and nondepressed content self-reference in mild depressives. *Journal of Personality*, *50*(1), 67–80. https://doi.org/10.1111/j.1467-6494.1982.tb00746.x

Lee, L., Harkness, K. L., Sabbagh, M. A., & Jacobson, J. A. (2005). Mental state decoding abilities in clinical depression. *Journal of Affective Disorders*, *86*(2), 247–258. https://doi.org/10.1016/j.jad.2005.02.007

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 & Experimental Psychiatry*, *45*(1), 196–202. https://doi.org/10.1016/j.jbtep.2013.10.006

Lemogne, C., le Bastard, G., Mayberg, H., Volle, E., Bergouignan, L., Lehericy, S., . . . Fossati, P. (2009). In search of the depressive self: extended medial prefrontal network during self-referential processing in major depression. *Social Cognitive and Affective Neuroscience*, *4*(3), 305–312. https://doi.org/10.1093/scan/nsp008

Li, Y., Li, M., Wei, D., Kong, X., Du, X., Hou, X., . . . Qiu, J. (2017). Self-referential processing in unipolar depression: Distinct roles of subregions of the medial prefrontal cortex. *Psychiatry Research: Neuroimaging*, *263*, 8–14. https://doi.org/10.1016/j.pscychresns.2017.02.008

Liu, P., Vandermeer, M. R. J., Joanisse, M. F., Barch, D. M., Dozois, D. J. A., & Hayden, E. P. (2020). Neural Activity During Self-referential Processing in Children at Risk for Depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(4), 429–437. https://doi.org/10.1016/j.bpsc.2019.12.012

Luck, S. J. (2005). An introduction to event-related potentials and their neural origins. *An introduction to the eventrelated potential technique*. 107. (Cambridge, MA: MIT)

Mahy, C. E., Moses, L. J., & Pfeifer, J. H. (2014). How and where: theory-of-mind in the brain. *Developmental Cognitive Neuroscience*, *9*, 68–81. https://doi.org/10.1016/j.dcn.2014.01.002

Manstead, A. S. R., Dosmukhambetova, D., Shearn, J., & Clifton, A. (2013). The influence of dysphoria and depression on mental state decoding. *Journal of Social and Clinical Psychology*, *32*(1), 116–133. https://doi.org/10.1521/jscp.2013.32.1.116

Mead, G. H. (1934). *Mind, Self and Society*. University of Chicago Press.

Menon, V., & D'Esposito, M. (2022). The role of PFC networks in cognitive control and executive function. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology*, *47*(1), 90–103. https://doi.org/10.1038/s41386-021-01152-w

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

Moritz, D., & Roberts, J. E. (2018). Self-Other Agreement and Metaperception Accuracy Across the Big Five: Examining the Roles of Depression and Self-Esteem. *Journal of Personality*, *86*(2), 296–307. https://doi.org/10.1111/jopy.12313

Nestor, B. A., Sutherland, S., & Garber, J. (2022). Theory of mind performance in depression: A meta-analysis. *Journal of Affective Disorders*, *303*, 233–244. https://doi.org/10.1016/j.jad.2022.02.028

NIMH. (n.d.-a). https://www.nimh.nih.gov/health/topics/depression

NIMH. (n.d.-b). https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc/constructs/perception-and-understanding-of-self

Northoff, G. (2007). Psychopathology and pathophysiology of the self in depression – neuropsychiatric hypothesis. *Journal of Affective Disorders*, *104*(1–3), 1–14. https://doi.org/10.1016/j.jad.2007.02.012

Northoff, G. (2013). Brain and self – a neurophilosophical account. *Child and Adolescent Psychiatry and Mental Health*, *7*(1), 28. https://doi.org/10.1186/1753-2000-7-28

Northoff, G., Heinzel, A., De, G. M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain – a meta-analysis of imaging studies on the self. *Neuroimage*, *31*(1), 440–457. https://doi.org/10.1016/j.neuroimage.2005.12.002

Nuttin, J. M. (1985). Narcissism beyond Gestalt and awareness: The name letter effect. *European Journal of Social Psychology*, *15*, 353–361. https://doi.org/10.1002/ejsp.2420150309

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D., Kihsltrom, J. F., & D'Esposito, M. (2005). The neural correlates of direct and reflected self-knowledge. *Neuroimage*, *28*(4), 797–814. https://doi.org/10.1016/j.neuroimage.2005.06.069

Orth, U., & Robins, R. W. (2013). Understanding the Link between low self-esteem and depression. *Current Directions in Psychological Science*, *22*(6), 455–460. https://doi.org/10.1177/0963721413492

Orth, U., Robins, R. W., & Roberts, B. W. (2008). Low self-esteem prospectively predicts depression in adolescence and young adulthood. *Journal of Personality and Social Psychology*, *95*(3), 695–708. https://doi.org/10.1037/0022-3514.95.3.695

Orth, U., Robins, R. W., Trzesniewski, K. H., Maes, J., & Schmitt, M. (2009). Low self-esteem is a risk factor for depressive symptoms from young

adulthood to old age. *Journal of Abnormal Psychology*, *118*(3), 472–478. https://doi.org/10.1037/a0015922
Orth, U., Robins, R. W., Widaman, K. F., & Conger, R. D. (2014). Is low self-esteem a risk factor for depression? Findings from a longitudinal study of Mexican-origin youth. *Developmental Psychology*, *50*(2), 622–633. https://doi.org/10.1037/a0033817
Otti, A., Wohlschlaeger, A. M., & Noll-Hussong, M. (2015). Is the Medial Prefrontal Cortex Necessary for Theory of Mind? *PLoS One*, *10*(8), e0135912. https://doi.org/10.1371/journal.pone.0135912
Pfeifer, J. H., Mahy, C. E. V., Merchant, J. S., Chen, C., Masten, C. L., Fuligni, A. J., . . . Chen, C. (2017). Neural systems for reflected and direct self-appraisals in Chinese young adults: Exploring the role of the temporal-parietal junction. *Cultural Diversity and Ethnic Minority Psychology*, *23*(1), 45–58. https://doi.org/10.1037/cdp0000122
Pfeifer, J. H., Masten, C. L., Borofsky, L. A., Dapretto, M., Fuligni, A. J., & Lieberman, M. D. (2009). Neural Correlates of Direct and Reflected Self-Appraisals in Adolescents and Adults: When Social Perspective-Taking Informs Self-Perception. *Child Development*, *80*(4), 1016–1038. https://doi.org/10.1111/j.1467-8624.2009.01314.x
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
Poulsen, C., Luu, P., Crane, S. M., Quiring, J., & Tucker, D. M. (2009). Frontolimbic activity and cognitive bias in major depression. *Journal of Abnormal Psychology*, *118*(3), 494–506. https://doi.org/10.1037/a0015920
Qin, P., & Northoff, G. (2011). How is our self related to midline regions and the default-mode network? *Neuroimage*, *57*(3), 1221–1233. https://doi.org/10.1016/j.neuroimage.2011.05.028
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
Richardson, H., & Saxe, R. (2020). Chapter 21 – Early signatures of and developmental change in brain regions for theory of mind. In J. Rubenstein, P. Rakic, B. Chen, & K. Y. Kwan (Eds.), *Neural Circuit and Cognitive Development (Second Edition)* (pp. 467–484). Academic Press. https://doi.org/10.1016/B978-0-12-814411-4.00021-4
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

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

Romero, N., Sanchez, A., Vázquez, C., & Valiente, C. (2016). Explicit self-esteem mediates the relationship between implicit self-esteem and memory biases in major depression. *Psychiatry Research*, *242*, 336–344. https://doi.org/10.1016/j.psychres.2016.06.003

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

Ryff, C. D., & Keyes, C. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719–727. https://doi.org/10.1037/0022-3514.69.4.719

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, *7*(5), 499–500. https://doi.org/10.1038/nn1223

Saulnier, K. G., Huet, A., Judah, M. R., & Allan, N. P. (2021). Anxiety Sensitivity and Arousal Symptom Implicit Association Task Performance: An Event-Related Potential Study of Cognitive Processing of Anxiety-Relevant Stimuli. *Journal of Affective Disorders*, *280*(Pt B), 7–15. https://doi.org/10.1016/j.jad.2020.11.067

Saxe, R. (2010). The right temporo-parietal junction: a specific brain region for thinking about thoughts. In A. Leslie & T. German (Eds.), *Handbook of Theory of Mind* (pp. 1–35). Taylor and Francis.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage*, *19*(4), 1835–1842. https://doi.org/10.1016/s1053-8119(03)00230-1

Saxe, R., Moran, J. M., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience*, *1*(3), 229–234. https://doi.org/10.1093/scan/nsl034

Saxe, R., & Powell, L. J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science*, *17*(8), 692–699. https://doi.org/10.1111/j.1467-9280.2006.01768.x

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

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Shestyuk, A. Y., & Deldin, P. J. (2010). Automatic and strategic representation of the self in major depression: trait and state abnormalities. *American*

Journal of Psychiatry, *167*(5), 536–544. https://doi.org/10.1176/appi.ajp.2009.06091444
Shiota, S., Okamoto, Y., Okada, G., Takagaki, K., Takamura, M., Mori, A., . . . Yamawaki, S. (2017). Effects of behavioural activation on the neural basis of other perspective self-referential processing in subthreshold depression: a functional magnetic resonance imaging study. *Psychological Medicine*, *47*(5), 877–888. https://doi.org/10.1017/S0033291716002956
Showers, C. (1992). Compartmentalization of positive and negative self-knowledge: keeping bad apples out of the bunch. *Journal of Personality and Social Psychology*, *62*(6), 1036–1049. https://doi.org/10.1037/0022-3514.62.6.1036
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
Srivastava, S. (2012). Other people as a source of self-knowledge. In *Handbook of self-knowledge.* (pp. 90–104). The Guilford Press.
Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and decreasing self-esteem during adolescence predict adult depression two decades later. *Journal of Personality and Social Psychology*, *106*(2), 325–338. https://doi.org/10.1037/a0035133
Sur, S., & Sinha, V. K. (2009). Event-related potential: An overview. *Ind Psychiatry J*, *18*(1), 70–73. https://doi.org/10.4103/0972-6748.57865
Swann, W. B., Wenzlaff, R. M., Krull, D. S., & Pelham, B. W. (1992). Allure of negative feedback: self-verification strivings among depressed persons. *Journal of Abnormal Psychology*, *101*(2), 293. https://doi.org/10.1037/0021-843x.101.2.293
Swann, W. B., Wenzlaff, R. M., & Tafarodi, R. W. (1992). Depression and the search for negative evaluations: More evidence of the role of self-verification strivings. *Journal of Abnormal Psychology*, *101*(2), 314–317. https://doi.org/10.1037/0021-843x.101.2.314
Tang, Y. Y., Tang, R., & Posner, M. I. (2016). Mindfulness meditation improves emotion regulation and reduces drug abuse. *Drug and Alcohol Dependence*, *163 Suppl 1*, S13–18. https://doi.org/10.1016/j.drugalcdep.2015.11.041
Thew, G. R., Gregory, J. D., Roberts, K., & Rimes, K. A. (2017). The phenomenology of self-critical thinking in people with depression, eating disorders, and in healthy individuals. *Psychology and psychotherapy*, *90*(4), 751–769. https://doi.org/10.1111/papt.12137
Tuijl, L. A. v., Jong, P. J. d., Sportel, B. E., Hullu, E. d., & Nauta, M. H. (2014). Implicit and explicit self-esteem and their reciprocal relationship with symptoms of depression and social anxiety: A longitudinal study in adolescents. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(1), 113–121. https://doi.org/10.1016/j.jbtep.2013.09.007

Van der Cruijsen, R., Peters, S., Zoetendaal, K. P. M., Pfeifer, J. H., & Crone, E. A. (2019). Direct and reflected self-concept show increasing similarity across adolescence: A functional neuroimaging study. *Neuropsychologia*, *129*, 407–417. https://doi.org/10.1016/j.neuropsychologia.2019.05.001

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

Vázquez, C., Diez-Alegría, C., Hernández-Lloreda, M. J., & Moreno, M. N. (2008). Implicit and explicit self-schema in active deluded, remitted deluded, and depressed patients. *Journal of Behavior Therapy and Experimental Psychiatry*, *39*(4), 587–599. https://doi.org/10.1016/j.jbtep.2008.01.006

Vianello, M., & Bar-Anan, Y. (2021). Can the Implicit Association Test Measure Automatic Judgment? The Validation Continues. *Perspect Psychol Sci*, *16*(2), 415-421. 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

Wisco, B. E. (2009). Depressive cognition: self-reference and depth of processing. *Clin Psychol Rev*, *29*(4), 382-392. https://doi.org/10.1016/j.cpr.2009.03.003

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

Yoshimura, S., Okamoto, Y., Onoda, K., Matsunaga, M., Okada, G., Kunisato, Y., . . . Yamawaki, S. (2014). Cognitive behavioral therapy for depression changes medial prefrontal and ventral anterior cingulate cortex activity associated with self-referential processing. *Social Cognitive and Affective Neuroscience*, *9*(4), 487–493. https://doi.org/10.1093/scan/nst009

Yoshimura, S., Okamoto, Y., Onoda, K., Matsunaga, M., Ueda, K., Suzuki, S., & Shigetoyamawaki. (2010). Rostral anterior cingulate cortex activity mediates the relationship between the depressive symptoms and the medial prefrontal cortex activity. *Journal of Affective Disorders*, *122*(1–2), 76–85. https://doi.org/10.1016/j.jad.2009.06.017

Yue, C., & Huang, X. (2012). A Review of the Properties of Reflected Self-appraisal. *Journal of Psychological Science*, *35*(3).

ORIGINAL PAPERS

I

REVIEW OF ABNORMAL SELF-KNOWLEDGE IN MAJOR DEPRESSIVE DISORDER

by

Yixue Lou, Yi Lei, Ying Mei, Paavo H.T. Leppänen, & Hong Li, 2019

Frontiers in psychiatry, 10, 130

Available online: https://doi.org/10.3389/fpsyt.2019.00130 frontiers in Psychiatry

REVIEW
published: 28 March 2019
doi: 10.3389/fpsyt.2019.00130

Review of Abnormal Self-Knowledge in Major Depressive Disorder

Yixue Lou[1,2,3], Yi Lei[1,3], Ying Mei[1,3], Paavo H. T. Leppänen[2] and Hong Li[1,3]*

[1] College of Psychology and Sociology, Shenzhen University, Shenzhen, China, [2] Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland, [3] Center for Neurogenetics, Shenzhen Institute of Neuroscience, Shenzhen, China

Background: Major depressive disorder (MDD) is an affective disorder that is harmful to both physical and mental health. Abnormal self-knowledge, which refers to abnormal judgments about oneself, is a core symptom of depression. However, little research has summarized how and why patients with MDD differ from healthy individuals in terms of self-knowledge.

Objective: To gain a better understanding of MDD, we reviewed previous studies that focused on the behavioral and neurological changes of self-knowledge in this illness.

Main Findings: On the behavioral level, depressed individuals exhibited negative self-knowledge in an explicit way, while more heterogeneous patterns were reported in implicit results. On the neurological level, depressed individuals, as compared with non-depressed controls, showed abnormal self-referential processing in both early perception and higher cognitive processing phases during the Self-Referential Encoding Task. Furthermore, fMRI studies have reported aberrant activity in the medial prefrontal cortex area for negative self-related items in depression. These results revealed several behavioral features and brain mechanisms underlying abnormal self-knowledge in depression.

Future Studies: The neural mechanism of implicit self-knowledge in MDD remains unclear. Future research should examine the importance of others' attitudes on the self-concept of individuals with MDD, and whether abnormal self-views may be modified through cognitive or pharmacological approaches. In addition, differences in abnormal self-knowledge due to genetic variation between depressed and non-depressed populations remain unconfirmed. Importantly, it remains unknown whether abnormal self-knowledge could be used as a specific marker to distinguish healthy individuals from those with MDD.

Conclusion: This review extends our understanding of the relationship between self-knowledge and depression by indicating several abnormalities among individuals with MDD and those who are at risk for this illness.

Keywords: major depressive disorder, self-knowledge, abnormality, behavioral abnormality, neurological abnormality

OPEN ACCESS

Edited by:
Fushun Wang,
Nanjing University of Chinese Medicine, China

Reviewed by:
Chiara Fabbri,
King's College London, United Kingdom
Juan Yang,
Southwest University, China

Correspondence:
Yi Lei
yutian@szu.edu.cn

Specialty section:
This article was submitted to Mood and Anxiety Disorders, a section of the journal Frontiers in Psychiatry

Received: *30 May 2018*
Accepted: *22 February 2019*
Published: *28 March 2019*

Citation:
Lou YX, Lei Y, Mei Y, Leppänen PHT and Li H (2019) Review of Abnormal Self-Knowledge in Major Depressive Disorder. Front. Psychiatry 10:130. doi: 10.3389/fpsyt.2019.00130

INTRODUCTION

Major depressive disorder (MDD) is a complicated affective disease characterized by abnormal clinical symptoms, including neurovegetative dysfunction (appetite or sleep disturbances), cognitive dissonance (inappropriate guilt, feelings of worthlessness), aberrant psychomotor activities (agitation or retardation) (1), and elevated suicide risk (2, 3). According to the World Health Organization, there are approximately 350 million people suffering from depression worldwide (4). In a recent survey, the proportion of years lived with disability (YLDs) caused by MDD was 4.2%, approximately 34.1 million of the total YLDs (5). Thus, MDD is thought to be a major global cause of disease burden and human suffering (5–7).

Abnormal perception and understanding of the self is a core symptom of MDD (1). This includes abnormal processes and/or representations involved in being aware of the self, abnormal knowledge about the self, and/or abnormal judgments about the self (National Institute of Mental Health; NIMH). As a sub-construct of perception and understanding of the self, self-knowledge, which refers to the ability to make judgments about one's current cognitive or emotional internal states, traits, and/or abilities (NIMH), is also impaired in individuals with MDD (8–11). For instance, individuals with MDD, unlike non-depressed healthy individuals, often exhibit negative self-evaluation, inappropriate self-blame, and excessive self-criticism (8, 12).

Although researchers have increasingly begun exploring abnormal self-knowledge in depression, few have compared existing findings in a single study. To enable a better understanding of how and why patients with MDD differ from healthy individuals in terms of self-knowledge, the current review focused on previous studies that examined behavioral patterns and brain mechanisms underlying abnormal self-knowledge in depression. Both explicit and implicit self-knowledge, which reflect conscious and unconscious self-views respectively, were discussed. Various abnormalities such as abnormal brain responses and aberrant neural circuits were illustrated. Furthermore, the present review pointed out some possible directions for future clinical studies (see **Figure 1**).

LITERATURE

Literature Review

A search of previous studies published between January 1960 and August 2018 was conducted using the databases Web of Science and PubMed. Self-knowledge is defined as a construct that includes self-evaluation, self-esteem, and self-reference. Thus, the search terms were designed as follows: "depression AND self-evaluation," OR "depression AND self-esteem," OR "depression AND self-reference." Search filters were set for publications written in English. Empirical research and reviews that examined the role of self-evaluation, self-attitude, self-view, self-reference, and/or self-esteem in MDD were found.

Eligibility Criteria

We screened for inclusion based on titles and abstracts, and again using full text. To be included, previous studies had to focus on behavioral and neurological changes of self-knowledge in MDD. All publications had to be reported on clinical populations currently or previously diagnosed with MDD, or populations who were currently in a depressive episode, regardless of gender and age. Conference abstracts were excluded if they were not published in a scientific journal. Publications were also excluded if they were published in a language other than English (see **Supplementary Figure 1**).

PARADIGMS

The majority of the research conformed to one of two methods. Specifically, these were explicit and implicit research paradigms.

Explicit Paradigms

Explicit methods are used to assess individuals' self-attitudes by using self-reported measures such as direct self-evaluation. The most commonly used explicit methods are the Self-Referential Encoding Task (SRET) (13) and self-reported questionnaires (14, 15).

Self-Referential Encoding Task, SRET

The self-referential encoding task (SRET) was designed to examine one's self-attitude (13, 16). Theoretically, individuals are more sensitive to information that is encoded as strongly related to oneself (17). Thus, self-related stimuli commonly display better recall and recognition performance, when compared to other-related stimuli (18). In the SRET, researchers present participants with positive and negative personality trait words, and ask them to decide whether each trait describes themselves (self-related condition), a familiar other (other-related condition) (19–22), or a socially desirable trait (semantic encoding condition; see **Figure 2**) (10, 23). After the judgment, the participants were asked to recall or recognize all the trait words that had been presented to them.

Individuals with positive self-attitudes, such as feelings of self-value, commonly endorse more positive traits relative to negative traits as self-describing, and show better recall and recognition rates of these words (18). Conversely, negative self-attitudes, such as feelings of worthlessness in individuals with MDD, often lead to more endorsement of negative traits and, in turn, better memory performance of these words (18, 24).

Other Explicit Approaches

Direct self-report questionnaires are often used in studies of depressive self-knowledge. For instance, researchers have used the Rosenberg Self-Esteem Scale (RSES) to measure explicit self-esteem in depression (15). In addition, the self-worth subscale of the World Assumption Scale (25) and the self-acceptance subscale of the Scales of Psychological Well-Being (26) are used to assess explicit self-attitude in depression. Moreover, the Beck Depression Inventory (BDI), which is commonly used to measure the depressive state, also

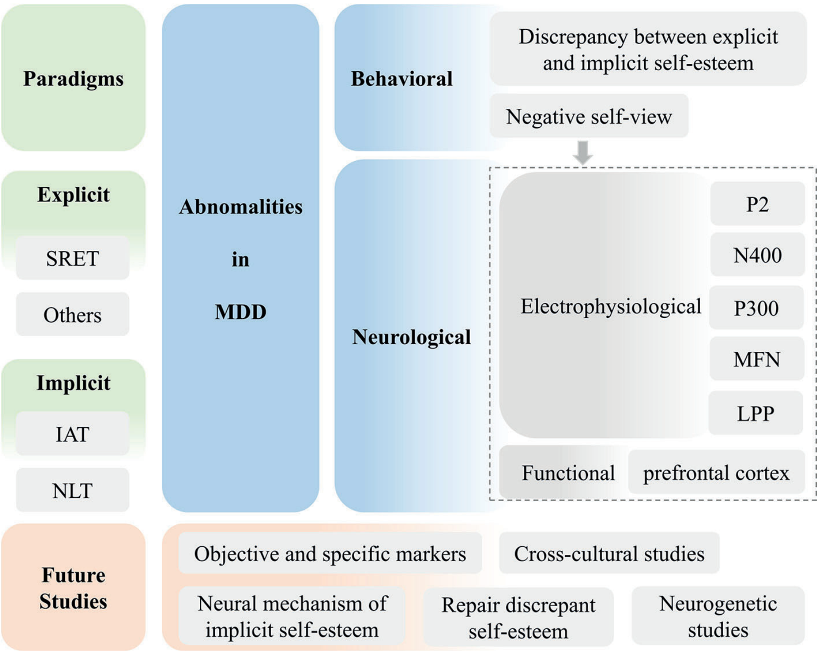

FIGURE 1 | Framework of the current review.

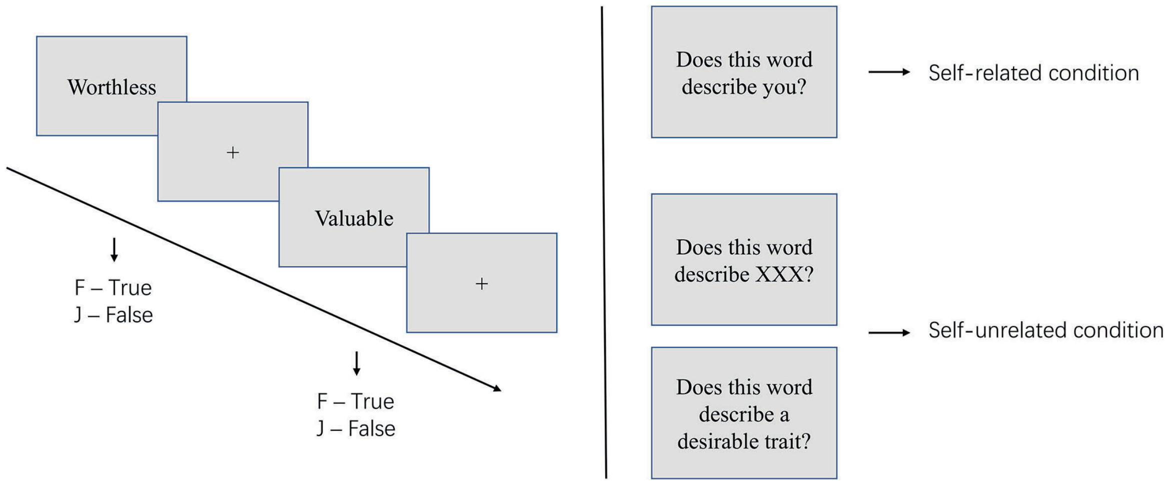

FIGURE 2 | Illustration of the self-referential encoding task (SRET).

contains self-evaluation factors, such as the self-blame factor, in its items (14).

Implicit Paradigms

The efficacy of explicit methods is debated by some researchers for the following reasons. First, according to cognitive theory, the self-concept involves automatic processes that occur without reflection and/or logical reasoning accessible within the conscious mind (27). Second, direct self-appraisal might be affected by social desirability and cultural differences (28, 29). In brief, explicit methods may not accurately reflect a person's real attitude about him/herself (30, 31). Thus, implicit paradigms were introduced into self-knowledge studies (32–34). The most commonly used implicit paradigms are the Implicit

TABLE 1 | Illustration of the self-evaluation Implicit Association Task (sIAT).

Task	Categorization	Stimulus	Key-press
Compatible	Self-related/unrelated words	Self	F
		Other	J
	Personality trait words	Valuable	F
		Worthless	J
Incompatible	Self-related/unrelated words	Self	F
		Other	J
	Personality trait words	Valuable	J
		Worthless	F

Association Test (IAT) (35, 36) and the Name-Letter Test (NLT) (37, 38).

Implicit Association Task, IAT

The self-evaluation IAT (sIAT) is a paradigm that has been commonly used to examine implicit self-attitudes of depression (39, 40). In the sIAT, it is assumed that information that is compatible with one's implicit attitude would be better processed as compared to that which is incompatible (36). Thus, participants are asked to complete two types of categorization (compatible and incompatible) by using a two key-press system. In the compatible condition, self-related stimulus words (e.g., one's own name or date of birth) shared the same key with "valuable" personality trait words (e.g., competent), while self-unrelated words (e.g., other's name or non-meaningful date) shared another key with "worthless" personality trait words (e.g., unsuccessful). The incompatible condition was reverse coded (34, 41, 42) (**Table 1**).

Differences in reaction times (RTs) and accuracy (ACC) between compatible and incompatible conditions were analyzed. Typically, the condition that is congruent with one's implicit self-attitude should show better performance when compared to the incongruent one. For instance, individuals with positive self-bias should demonstrate a faster and more accurate response in the compatible condition, relative to the incompatible condition (43, 44), while the negative self-attitude found in depression should lead to an opposite pattern.

Name-Letter Test, NLT

The name-letter test (NLT) has also been used in previous studies to measure implicit self-attitudes (38, 45–47). In the NLT, researchers presented participants with the 26 letters of the alphabet one-by-one, and asked them to judge the attractiveness or likability of each letter, relying on their first, intuitive reaction (48). According to the name letter effect, one's initial is thought to be highly associated with the self (49, 50). Thus, under the influence of positive self-bias, non-depressed individuals should show a rational preference toward their initials relative to other letters, even though they are generally unaware of this effect (38). However, an opposite pattern may be true for individuals with MDD (33).

The name letter effect has been shown to be a cross-cultural phenomenon, since it has also been reported in Thai, Japanese, and Korean studies (51–53). Thus, the NLT qualifies as an indirect assessment of self-attitude in depression (33).

MAIN FINDINGS

By using the aforementioned paradigms, researchers have found abnormal behavioral patterns and brain responses in individuals with MDD, when compared to non-depressed, healthy controls. Evaluation of the quality of included studies was listed in **Supplementary Table 1**.

Behavioral Abnormalities

Behavioral abnormalities include explicit/conscious and implicit/unconscious behaviors that have been observed in depression.

Explicit: Negative Self-View

At the explicit level, previous behavioral research has revealed a negative self-view in depression, as compared with a non-depressed healthy population. For instance, healthy individuals typically exhibit positive attitudes about themselves (54–57). For instance, they often attribute themselves with more positive, rather than negative, personality traits (54, 58), so that their self-esteem may be protected (18, 59). However, individuals with depression typically demonstrate an abnormally negative self-view (1, 60, 61).

For instance, under the influence of negative self-knowledge, individuals with MDD show less positive self-bias, less self-confidence, and lower self-esteem (62–65), as well as excessive self-criticism, negative self-evaluation, inappropriate self-blame, and shame (8, 12, 66–68). This negative self-representation has been associated with greater self-reported depression (69, 70), poor and slower recovery from a major depressive episode (71, 72), and higher probability of suicide attempt (73, 74). In addition, individuals with higher self-esteem may exhibit sudden improvements in depressive symptomatology even without treatment (75), while lower self-esteem is thought to be a prospective risk factor for depressive symptoms from young adulthood to old age (76–78).

In the SRET, individuals with depression, relative to healthy controls, endorsed more negative trait words as self-described, and showed faster response, better recall performance, and increased recognition rate for these words (9, 23, 79, 80). In a longitudinal study Derry and Kuiper (13), found that such negative self-bias might be a specific symptom in currently depressed patients, since the recall rate of negative self-related words decreased after recovery from the illness.

Implicit: Discrepancy Between Explicit and Implicit Self-Esteem

Although a large number of studies have indicated a lower self-attitude in MDD, relative to healthy individuals, at an explicit level (8–10, 20), more heterogeneous patterns have been reported in implicit studies (34, 36, 41, 42, 81).

For instance, when using the IAT and/or NLT to measure implicit self-esteem (ISE) and RSES to assess explicit self-esteem (ESE), some researchers have observed both low ESE and ISE in currently depressed individuals (39, 40, 42) and chronically depressed individuals with early onset (33), relative to never depressed healthy controls. However, more researchers have reported a normal ISE combined with a lower ESE in individuals with current depression (41, 42, 82–85), previous depression (41), remitted depression (11, 39, 86), and chronic depression with late onset (33), when compared to non-depressed individuals. Some researchers have even observed higher ISE and lower ESE in current depression (83, 85, 87) and previous depression (34, 82).

The discrepancy between explicit and implicit self-esteem, especially the combination of low ESE and high ISE, is thought to be associated with internalizing problems such as affective disorders (88–92). For major depression, such a discrepancy seems to be more severe in depressed individuals with suicidal ideation relative to those without such ideation (42). Moreover, depressed patients with congruent self-esteem, compared to those with incongruent self-esteem, exhibited better recovery from the illness throughout antidepressant treatment (93).

Neurological Abnormalities

Neurological abnormalities include abnormal electrophysiological responses and aberrant functional neural activities. These abnormalities were all detected using the SRET.

Abnormal Electrophysiological Response

To explore the brain mechanism of negative self-knowledge in depression, electroencephalography (EEG) technology was used in conjunction with the SRET. By collecting the event-related potentials (ERPs) during the SRET, researchers attempted to identify the key ERP components that are involved in negative self-referent processing in MDD.

For instance, Shestyuk and Deldin (62) observed increased P2 component, which was quantified as a positive peak in the 200- to 300-ms time window poststimulus, in individuals with depression while processing negative, relative to positive, self-referential items. The opposite, however, was true for the non-depressed healthy controls. A recent study reported decreased N400 amplitude, which was measured as mean voltage of the ERP average between 350–500 ms, in individuals with depression, as compared with healthy controls, in negative self-referent processing (9). Regarding the latter component, Poulsen et al. (94) found an attenuated or absent MFN response between 260 ms and 480 ms in depression, relative to non-depressed controls, when specifically endorsing negative trait descriptors. However, in a recent study, depressed individuals were found to exhibit enhanced MFN for both positive and negative endorsement (95). Consistently, an attenuated P300 response from 300- to 600- ms was observed in both of these two studies (94, 95). Concerning the more delayed late positive potential (LPP), larger LPP amplitudes were detected following negative vs. positive endorsement in depressed adults (62, 96), depressed adolescents (8), and young girls who were vulnerable to depression (97), when compared to healthy controls.

In these studies, the P2 component is thought to be related to automatic semantic processes (98). Thus, an increased P2 reflects a stronger automatic attentional capture and orientation in patients with depression under the negative, relative to positive, self-related condition (62). The N400 component was interpreted to be influenced by semantic memories about the self, and could be reduced by greater association of the stimuli with a preceding self-related context (99, 100). Therefore, this result indexed a congruent pattern between negative semantic memories and the self-concept in individuals with depression (9). In addition, the MFN is thought to be associated with early cognitive evaluation during self-referential processing (95). The altered MFN response may reflect abnormal self-evaluation among clinically depressed individuals. The greater P300, which is evoked by a saliency effect of self-referential information and positive affect (101), was attenuated in depression. One possible interpretation is that it was possibly associated with a chronically negative self-view in this population (95). Last, an increased LPP amplitude, which is associated with effortful encoding (102), indicates that individuals with depression engage more cognitive effort in processing self-related negative, relative to positive, items (62).

In all, in the time domain, abnormal self-knowledge in depression could be reflected in early phases of self-related processing, such as automatic attention and orientation toward negative self-descriptive items (62). Retrieval of negative memories about the self could also be involved (9). For later phases of self-referential processing, an attenuated bonding between positive affect and the self may be associated with negative self-view in depression (95). Furthermore, depressed individuals seem to engage more cognitive effort in negative, instead of positive, self-reference (62).

Abnormal Functional Neural Activities

The high spatial resolution of functional MRI technology makes it possible for researchers to determine abnormal brain activities in depression during the SRET. Several fMRI studies, thus, have suggested that the prefrontal cortex and its sub-regions might be abnormal in individuals with MDD (103). The prefrontal cortex is thought to play an important role in self-referential processing (104). In particular, dysfunction within the medial prefrontal cortex (mPFC) and in the circuits that connect the mPFC to other cortical and limbic structures is responsible for the cognitive dissonance found in depression (103).

For instance, the cortical midline structures (CMS), such as the mPFC, are critical for self-referential processing in healthy individuals (17, 105), adult patients (106–108), and adolescent patients with MDD (109). However, aberrant activity in the mPFC was reported in depression when compared to healthy controls (17, 23, 106). Additionally, Yoshimura et al. (108) found that individuals with depression, relative to healthy controls, exhibited hyperactivity in the mPFC and the rostral anterior cingulate cortex (rostral ACC) during self-referential processing of negative personality traits; such activity was shown to be associated with depressive symptoms (108).

Furthermore, abnormal activities of other sub-regions of the prefrontal cortex were also observed during the processing of self-related negative stimuli in depression (10, 23). For instance, by using the SRET, researchers found significantly higher activation of the central mPFC and significantly lower activation of the dorsal mPFC in depression, relative to healthy controls, during the self-referential condition (10). Local connectivity of the dorsal mPFC was also reduced during self-reflection in depressed adolescents (109). The activity of the dorsolateral prefrontal cortex (dlPFC) was also involved in self-referential processing in depression, but was absent in healthy controls (23). In addition, a meta-analysis revealed hyperactivation in the ventromedial prefrontal cortex (vmPFC) within major depression during resting state, which was discussed as a neural reflection of self-referential processing (110).

Therefore, aberrant activity of the prefrontal cortex and its sub-regions could index the abnormal brain activity that is a hallmark of depression, specifically during the processing of self-referential stimuli. In particular, hyperactivity in the mPFC during negative self-referential processing could possibly even be associated with the severity of depressive symptoms.

DISCUSSION

According to previous studies, abnormal self-knowledge, which is commonly found in MDD, is mainly reflected in abnormal behaviors and abnormal neurological responses during self-evaluation, self-esteem, and/or self-referential processing.

At the behavioral level, abnormal self-knowledge could be indexed by a negative explicit self-view (13, 80) and discrepant self-esteem, which involves relatively higher implicit self-esteem and lower explicit self-esteem (11, 33, 34, 111). Furthermore, a greater discrepancy between implicit and explicit self-esteem is related to more severe MDD, or a higher possibility of being affected by the illness (42, 111).

At the neurological level, several abnormalities have been found during abnormal self-referential processing, by using electrophysiological technology (8, 9, 62) and fMRI technology (10, 108, 112, 113). For instance, for abnormal electrophysiological processing, enhanced P2 and LPP and decreased N400 amplitudes were all detected in depression, relative to non-depressed controls, in the SRET. For aberrant brain activities, higher activation of the central mPFC, lower activation of the dorsal mPFC (10), and aberrant activity of the dlPFC (23) during self-referential processing can also distinguish MDD, as well as indicate the severity of symptoms.

By using the indexes above, researchers and clinicians could distinguish patients with MDD and non-depressed individuals more objectively and effectively. However, caution should be exercised for several reasons. First, some of the studies involved limited samples and poor replications. For instance, abnormalities in P2 and LPP amplitude in MDD were reported in a study with 17 patients with current depression, 17 patients with remitted depression, and 18 controls, and abnormalities of N400s were reported in a study including 16 patients with MDD and 16 controls. Considering this issue, larger samples are needed to confirm changes of electrophysiological response during depressive self-referential processing.

Second, abnormal self-knowledge is only one component of MDD, despite being a core feature. Behavioral abnormalities may not be sensitive and specific for MDD, since they are affected by non-clinical factors such as personality traits (114–117). Thus, more evidence is needed to confirm the behavioral abnormalities identified in the current review.

Third, although we reviewed various investigations that focused on abnormal self-knowledge in depression, a classical review is relatively less objective compared with a systematic meta-analysis.

FUTURE STUDIES

In the exploration of self-knowledge in depression, there are still many unanswered questions. First, although the discrepancy between explicit and implicit self-esteem in depression has been confirmed by several previous studies (11), and the neural mechanism of explicit self-esteem has been richly explored (8, 10, 23, 108), little is known about the neural basis of implicit self-esteem in depression, suggesting the need for further research.

Second, it remains unclear whether the pattern of self-knowledge in patients with depression would be different in a cross-cultural context. For instance, collectivism of eastern Asia, relative to individualism in Western culture, allows individuals to view themselves as dynamic entities that are continually defined by their social context and relationships (118). Thus, in Eastern cultures, judgment by important others about oneself, which is currently ignored in self-related studies, plays a critical role in the quality of one's self-view (119). Indeed, the development of self-knowledge relies not only on one's reflection of the self, but also on how important others evaluate the individual (22, 58, 119–121).

Third, some previous neurogenetic research explored the association between different gene types and abnormal self-knowledge in depression, and found that the serotonin transporter promoter polymorphism (5-HTTLPR) played a crucial role in susceptibility to developing depression (122). In a recent study, Ma et al. (21) reported a modulation effect of the 5-HTTLPR polymorphism in brain activities associated with negative self-knowledge in depression. It was suggested that the s allele of 5-HTTLPR could possibly be a risk factor for individuals vulnerable to depression (21). However, differences in abnormal self-knowledge due to genetic variation between healthy and depressed populations remains unconfirmed, calling for further research.

Fourth, to repair discrepant self-esteem found in depression, which involves low explicit and high implicit self-esteem, the development of cognitive and/or medical approaches is needed to enhance explicit self-attitudes. A previous study indicated that depression can be prevented or reduced by interventions that improve explicit self-esteem (123–126). For example, researchers have utilized positive self-images (127) and mindfulness (128, 129) to realize an improvement of both explicit and implicit self-esteem. It is possible that these methods can also be used to diminish the discrepancy of self-esteem found in depression. Furthermore, since the s allele of 5-HTTLPR may elevate the risk of developing depression (21), it is reasonable to consider whether the use of selective serotonin reuptake inhibitors (SSRIs) could enhance self-satisfaction (130–132).

Finally, to conquer complex diseases such as MDD, the National Institute of Mental Health (NIMH) has raised the importance of identifying clinically useful biomarkers and behavioral indicators that predict change across the trajectory of illnesses (19). However, the most fundamental challenge is to identify these diseases effectively. In the diagnosis of MDD, the most commonly used measurements are structured interviews and/or depression inventories (133), which are relatively subjective and require researchers to be professionally trained. To facilitate the identification of objective criteria for MDD diagnosis, it must be determined whether abnormal self-knowledge can be used as an objective and specific marker for identifying MDD. For this purpose, patterns of abnormal self-knowledge should be compared between MDD and other mental disorders, such as bipolar disorder.

CONCLUSION

MDD is a main cause of disease burden worldwide (6, 7), and abnormal self-knowledge is one of the cardinal symptoms of this disorder. Through a review of previous studies that measured abnormal self-knowledge in individuals with clinical MDD, several abnormalities that distinguish patients with MDD as well as those at risk of the illness were revealed. We also pointed out several possible directions for future clinical studies based on previous findings. Overall, this review extends our understanding of the relationship between self-knowledge and depression.

AUTHOR CONTRIBUTIONS

YLo wrote the paper. YLe supervised the review and assisted in paper revision. YM assisted in paper revision. PL assisted in paper revision. HL assisted in paper writing and funding supports. All authors were involved in revising the manuscript critically for important intellectual content and approved the final version of the manuscript.

FUNDING

This work was supported by the Shenzhen Peacock Plan [Grant No. KQTD2015033016104926] and the Guangdong Pearl River Talents Plan Innovative and Entrepreneurial Team [Grant No. 2016ZT06S220].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2019.00130/full#supplementary-material

REFERENCES

1. Fava M, Kenneth KS. Major depressive disorder. *Neuron.* (2000) 28:335. doi: 10.1016/S0896-6273(00)00112-4
2. Chesney E, Goodwin GM, Fazel S. Risks of all-cause and suicide mortality in mental disorders: a meta-review. *World Psychiatry Offic J World Psychiatr Assoc.* (2014) 13:153. doi: 10.1002/wps.20128
3. Klonsky ED, May AM, Saffer BY. Suicide, suicide attempts, and suicidal ideation. *Ann Rev Clin Psychol.* (2016) 12:307. doi: 10.1146/annurev-clinpsy-021815-093204
4. Smith K. Mental health: a world of depression. *Nature.* (2014) 515:181. doi: 10.1038/515180a
5. Vos T, Abajobir AA, Abbafati C, Abbas KM, Abate KH, Abd-Allah F, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet.* (2017) 390:1211–59. doi: 10.1016/S0140-6736(17)32154-2
6. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* (2012) 380:2163–96. doi: 10.1016/s0140-6736(12)61729-2
7. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet.* (2012) 380:2197–223. doi: 10.1016/s0140-6736(12)61689-4
8. Auerbach RP, Stanton CH, Proudfit GH, Pizzagalli DA. Self-referential processing in depressed adolescents: a high-density event-related potential study. *J Abnorm Psychol.* (2015) 124:233–45. doi: 10.1037/abn0000023
9. Kiang M, Farzan F, Blumberger DM, Kutas M, McKinnon MC, Kansal V, et al. Abnormal self-schema in semantic memory in major depressive disorder: evidence from event-related brain potentials. *Biol Psychol.* (2017) 126:41–7. doi: 10.1016/j.biopsycho.2017.04.003
10. Li Y, Li M, Wei D, Kong X, Du X, Hou X, et al. Self-referential processing in unipolar depression: distinct roles of subregions of the medial prefrontal cortex. *Psychiatry Res.* (2017) 263:8–14. doi: 10.1016/j.pscychresns.2017.02.008
11. Smeijers D, Vrijsen JN, van Oostrom I, Isaac L, Speckens A, Becker ES, et al. Implicit and explicit self-esteem in remitted depressed patients. *J Behav Ther Exp Psychiatry.* (2017) 54:301–6. doi: 10.1016/j.jbtep.2016.10.006
12. Auerbach RP, Ho M, Kim JC. Identifying cognitive and interpersonal predictors of adolescent depression. *J Abnorm Child Psychol.* (2014) 42:913–24. doi: 10.1007/s10802-013-9845-6
13. Derry PA, Kuiper NA. Schematic processing and self-reference in clinical depression. *J Abnorm Psychol.* (1981) 90:286. doi: 10.1037/0021-843X.90.4.286
14. Beck AT, Steer RA, Brown GK. Beck depression inventory - second edition manual. *Psihologijski Mjerni Instrumenti.* (1996) 93:88.
15. Rosenberg M. *Society and the Adolescent Self-Image (rev. ed.).* Princeton, NJ: Princeton University Press (1965).
16. Ramel W, Goldin PR, Eyler LT, Brown GG, Gotlib IH, McQuaid JR. Amygdala reactivity and mood-congruent memory in individuals at risk for depressive relapse. *Biol Psychiatry.* (2007) 61:231–9. doi: 10.1016/j.biopsych.2006.05.004
17. Northoff G, Heinzel A, De GM, Bermpohl F, Dobrowolny H, Panksepp J. Self-referential processing in our brain–a meta-analysis of imaging studies on the self. *Neuroimage.* (2006) 31:440–57. doi: 10.1016/j.neuroimage.2005.12.002
18. Caudek C. Individual differences in cognitive control on self-referenced and other-referenced memory. *Conscious Cogn.* (2014) 30:169–83. doi: 10.1016/j.concog.2014.08.017
19. Auerbach RP, Bondy E, Stanton CH, Webb CA, Shankman SA, Pizzagalli DA. Self-referential processing in adolescents: Stability of behavioral and ERP markers. *Psychophysiology.* (2016) 53:1398–406. doi: 10.1111/psyp.12686

20. Ji JL, Grafton B, MacLeod C. Referential focus moderates depression-linked attentional avoidance of positive information. *Behav Res Ther.* (2017) 93:47–54. doi: 10.1016/j.brat.2017.03.004
21. Ma Y, Li B, Wang C, Shi Z, Sun Y, Sheng F, et al. 5-HTTLPR polymorphism modulates neural mechanisms of negative self-reflection. *Cereb Cortex.* (2014) 24:2421–9. doi: 10.1093/cercor/bht099
22. Yang J, Xu X, Chen Y, Shi Z, Han S. Trait self-esteem and neural activities related to self-evaluation and social feedback. *Sci Rep.* (2016) 6:20274. doi: 10.1038/srep20274
23. Lemogne C, le Bastard G, Mayberg H, Volle E, Bergouignan L, Lehericy S, et al. In search of the depressive self: extended medial prefrontal network during self-referential processing in major depression. *Soc Cogn Affect Neurosci.* (2009) 4:305–12. doi: 10.1093/scan/nsp008
24. Caudek C, Monni A. Do you remember your sad face? *Roles Negat Cogn Style Sad Mood Memory*. (2013) 21:891. doi: 10.1080/09658211.2013.765893
25. Janoffbulman R. Assumptive worlds and the stress of traumatic events: applications of the schema construct. *Soc Cogn.* (1989) 7:113–36. doi: 10.1521/soco.1989.7.2.113
26. Ryff CD, Keyes C. The structure of psychological well-being revisited. *J Pers Soc Psychol.* (1995) 69:719–27. doi: 10.1037/0022-3514.69.4.719
27. Franck E, De Raedt R, Dereu M, Van den Abbeele D. Implicit and explicit self-esteem in currently depressed individuals with and without suicidal ideation. *J Behav Ther Exp Psychiatry.* (2007) 38:75–85. doi: 10.1016/j.jbtep.2006.05.003
28. Kitayama S, Uchida Y. Explicit self-criticism and implicit self-regard: Evaluating self and friend in two cultures. *J Exp Soc Psychology.* (2003) 39:476–82. doi: 10.1016/S0022-1031(03)00026-X
29. Yamaguchi S, Greenwald AG, Banaji MR, Murakami F, Chen D, Shiomura K, et al. Apparent universality of positive implicit self-esteem. *Psychol Sci.* (2007) 18:498–500. doi: 10.1111/j.1467-9280.2007.01928.x
30. Kim YH, Chiu CY, Peng S, Cai H, Tov W. Explaining east-west differences in the likelihood of making favorable self-evaluations: the role of evaluation apprehension and directness of expression. *J Cross Cult Psychol.* (2010) 41:62–75. doi: 10.1177/0022022109348921
31. Kudo E, Numazaki M. Explicit and direct self-serving bias in Japan. *J Cross Cult Psychol.* (2003) 34:511–21. doi: 10.1177/0022022103256475
32. Grundy JG, Benarroch MF, Lebarr AN, Shedden JM. Electrophysiological correlates of implicit valenced self-processing in high vs. *Low Self Esteem Individ Soc Neurosci.* (2015) 10:100–12. doi: 10.1080/17470919.2014.965339
33. Randenborgh AV, Pawelzik M, Quirin M, Kuhl J. Bad roots to grow: deficient implicit self-evaluations in chronic depression with an early onset. *J Clin Psychol.* (2016) 72:580–90. doi: 10.1002/jclp.22275
34. Roberts JE, Porter A, Vergara-Lopez C. Implicit and explicit self-esteem in previously and never depressed individuals: Baseline differences and reactivity to rumination. *Cogn Ther Res.* (2015) 40:164–72. doi: 10.1007/s10608-015-9732-2
35. Greenwald AG, Mcghee DE, Schwartz JL. Measuring individual differences in implicit cognition: the implicit association test. *J Personal Soc Psychol.* (1998) 74:1464. doi: 10.1037/0022-3514.74.6.1464
36. Greenwald AG, Nosek BA, Banaji MR. Understanding and using the implicit association test: I. An improved scoring algorithm *J Personal Soc Psychol.* (2003) 85:197. doi: 10.1037/0022-3514.85.2.197
37. Greenwald AG, Banaji MR. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychol Rev*. (1995) 102:4–27. doi: 10.1037/0033-295X.102.1.4
38. Nuttin JM. Narcissism beyond Gestalt and awareness: the name letter effect. *Eur J Soc Psychol.* (1985) 15:353–61. doi: 10.1002/ejsp.2420150309
39. Risch AK, Buba A, Birk U, Morina N, Steffens MC, Stangier U. Implicit self-esteem in recurrently depressed patients. *J Behav Ther Exp Psychiatry.* (2010) 41:199–206. doi: 10.1016/j.jbtep.2010.01.003
40. Romero N, Sanchez A, Vázquez C, Valiente C. Explicit self-esteem mediates the relationship between implicit self-esteem and memory biases in major depression. *Psychiatry Res.* (2016) 242:336–44. doi: 10.1016/j.psychres.2016.06.003
41. Franck E, De Raedt R, De Houwer J. Activation of latent self-schemas as a cognitive vulnerability factor for depression: the potential role of implicit self-esteem. *Cogn Emot.* (2008) 22:1588–99. doi: 10.1080/02699930801921271
42. Franck E, De Raedt R, De HJ. Implicit but not explicit self-esteem predicts future depressive symptomatology. *Behav Res Ther.* (2007) 45:2448. doi: 10.1016/j.brat.2007.01.008
43. Fleischhauer M, Strobel A, Diers K, Enge S. Electrophysiological evidence for early perceptual facilitation and efficient categorization of self-related stimuli during an Implicit Association Test measuring neuroticism. *Psychophysiology.* (2014) 51:142–51. doi: 10.1111/psyp.12162
44. Wu L, Gu R, Cai H, Zhang J. Electrophysiological evidence for executive control and efficient categorization involved in implicit self-evaluation. *Soc Neurosci.* (2016) 11:153–63. doi: 10.1080/17470919.2015.1044673
45. Hoorens V, Takano K, Franck E, Roberts JE, Raes F. Initial and noninitial name-letter preferences as obtained through repeated letter rating tasks continue to reflect (different aspects of) self-esteem. *Psychol Assess.* (2015) 27:905. doi: 10.1037/pas0000092
46. Jones JT, Pelham BW, Mirenberg MC, Hetts JJ. Name letter preferences are not merely mere exposure: implicit egotism as self-regulation. *J Exp Soc Psychol.* (2002) 38:170–7. doi: 10.1006/jesp.2001.1497
47. Stieger S, Burger C. More complex than previously thought: new insights into the optimal administration of the initial preference task. *Self Ident.* (2013) 12:201–16. doi: 10.1080/15298868.2012.655897
48. Koole SL, Dijksterhuis A, Van KA. What's in a name: implicit self-esteem and the automatic self. *J Personal Soc Psychol.* (2001) 80:669. doi: 10.1037/0022-3514.80.4.669
49. Fischer C, Dailler F, Morlet D. Novelty P3 elicited by the subject's own name in comatose patients. *Clin Neurophysiol Offic J Int Feder Clin Neurophysiol.* (2008) 119:2224. doi: 10.1016/j.clinph.2008.03.035
50. Koole SL, Pelham BW. On the nature of implicit self-esteem: the case of the Name Letter Effect. In: Spencer S, Fein S, Zanna MP, editors. *Motivated Social Perception*: *The Ontario Symposium.* Hillsdale: NJ: Lawrence Erlbaum (2003), p. 93–116.
51. Fujii T. Does the name-letter effect emerge in Korea? *Japan J Res Emot.* (2015) 23:32–7. doi: 10.4092/jsre.23.32
52. Hoorens V, Nuttin JM, Herman IE, Pavakanun U. Mastery pleasure versus mere ownership: a quasi-experimental cross-cultural and cross-alphabetical test of the name letter effect. *Eur J Soc Psychol.* (1990) 20:181–205. doi: 10.1002/ejsp.2420200302
53. Kitayama S, Karasawa M. Implicit self-esteem in Japan: name letters and birthday numbers. *Personal Soc Psychol Bull.* (1997) 23:736–42. doi: 10.1177/0146167297237006
54. Chen Y, Zhong Y, Zhou H, Zhang S, Tan Q, Fan W. Evidence for implicit self-positivity bias: an event-related brain potential study. *Exp Brain Res.* (2014) 232:985–94. doi: 10.1007/s00221-013-3810-z
55. Mezulis AH, Abramson LY, Hyde JS, Hankin BL. Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychol Bull.* (2004) 130:711–47. doi: 10.1037/0033-2909.130.5.711
56. Pahl S, Eiser JR. Valence, comparison focus and self-positivity biases: does it matter whether people judge positive or negative traits? *Exp Psychol.* (2005) 52:303. doi: 10.1027/1618-3169.52.4.303
57. Watson LA, Dritschel B, Obonsawin MC, Jentzsch I. Seeing yourself in a positive light: brain correlates of the self-positivity bias. *Brain Res.* (2007) 1152:106. doi: 10.1016/j.brainres.2007.03.049
58. Zhou A, Li S, Herbert C, Xia R, Xu K, Xu Q, et al. Perspective taking modulates positivity bias in self-appraisals: behavioral and event-related potential evidence. *Soc Neurosci.* (2013) 8:326. doi: 10.1080/17470919.2013.807873
59. Leary MR. Motivational and emotional aspects of the self. *Ann Rev Psychol.* (2007) 58:317. doi: 10.1146/annurev.psych.58.110405.085658
60. Brand-Gothelf A, Yoeli-Bligh N, Gilboa-Schechtman E, Benaroya-Milshtein N, Apter A. Perceptions of self, mother and family and behavior of prepubertal depressed children. *Eur Psychiatry.* (2015) 30:69–74. doi: 10.1016/j.eurpsy.2014.05.005
61. Ruehlman LS, West SG, Pasahow RJ. Depression and evaluative schemata. *J Personal.* (2010) 53:46–92. doi: 10.1111/j.1467-6494.1985.tb00888.x

62. Shestyuk AY, Deldin PJ. Automatic and strategic representation of the self in major depression: trait and state abnormalities. *Am J Psychiatry.* (2010) 167:536–44. doi: 10.1176/appi.ajp.2009.06091444
63. Shestyuk AY, Deldin PJ, Brand JE, Deveney CM. Reduced sustained brain activity during processing of positive emotional stimuli in major depression. *Biol Psychiatry.* (2005) 57:1089–96. doi: 10.1016/j.biopsych.2005.02.013
64. Tripković I, Roje R, Krnić S, Nazor M, Karin Ž, Capkun V. (2015). Depression and self-esteem in farly adolescence. *Centr Eur J Public Health.* 23:166–9. doi: 10.21101/cejph.a4017
65. Watson LA, Dritschel B, Jentzsch I, Obonsawin MC. Changes in the relationship between self-reference and emotional valence as a function of dysphoria. *Br J Psychol.* (2008) 99:143–52. doi: 10.1348/000712607X248689
66. Alexander B, Brewin CR, Vearnals S, Wolff G, Leff J. An investigation of shame and guilt in a depressed sample. *Psychol Psychother Theor Res Pract.* (2011) 72:323–38. doi: 10.1348/000711299160031
67. Braet C, Wante L, Van Beveren ML, Theuwis L. Is the cognitive triad a clear marker of depressive symptoms in youngsters? *Eur Child Adolesc Psychiatry.* (2015) 24:1261. doi: 10.1007/s00787-015-0674-8
68. Thew GR, Gregory JD, Roberts K, Rimes KA. The phenomenology of self-critical thinking in people with depression, eating disorders, and in healthy individuals. *Psychol Psychother.* (2017) 90:751–69. doi: 10.1111/papt.12137
69. Gara MA, Woolfolk RL, Allen LA. Social cognitive complexity and depression: cognitive complexity moderates the correlation between depression self-ratings and global self-evaluation. *J Nerv Mental Dis.* (2002) 190:670. doi: 10.1097/00005053-200210000-00003
70. Orchard F, Reynolds S. The combined influence of cognitions in adolescent depression: biases of interpretation, self-evaluation, and memory. *Br J Clin Psychol.* (2018) 57:420–35. doi: 10.1111/bjc.12184
71. Gildengers AG, Houck PR, Mulsant BH, Dew MA, Aizenstein HJ, Jones BL, et al. Trajectories of treatment response in late-life depression: psychosocial and clinical correlates. *J Clin Psychopharmacol.* (2005) 25(4 Suppl. 1):S8. doi: 10.1097/01.jcp.0000161498.81137.12
72. Woolfolk RL, Gara MA, Ambrose TK, Williams JE, Allen LA, Irvin SL, et al. Self-complexity and the persistence of depression. *J Nerv Ment Dis.* (1999) 187:393–9. doi: 10.1097/00005053-199907000-00001
73. Hur JW, Kim WJ, Kim YK. The mediating effect of psychosocial factors on suicidal probability among adolescents. *Arch Suicide Res.* (2011) 15:327–36. doi: 10.1080/13811118.2011.615701
74. Wetzel RD, Reich T. The cognitive triad and suicide intent in depressed in-patients. *Psychol Rep.* (1989) 65:1027–32. doi: 10.2466/pr0.1989.65.3.1027
75. Kelly MA, Roberts JE, Bottonari KA. Non-treatment-related sudden gains in depression: the role of self-evaluation. *Behav Res Ther.* (2007) 45:737–47. doi: 10.1016/j.brat.2006.06.008
76. Orth U, Robins RW, Roberts BW. Low self-esteem prospectively predicts depression in adolescence and young adulthood. *J Personal Soc Psychol.* (2008) 95:695–708. doi: 10.1037/0022-3514.95.3.695
77. Orth U, Robins RW, Widaman KF, Conger RD. Is low self-esteem a risk factor for depression? Findings from a longitudinal study of Mexican-origin youth. *Dev Psychol.* (2014) 50:622–33. doi: 10.1037/a0033817
78. Steiger AE, Allemand M, Robins RW, Fend HA. Low and decreasing self-esteem during adolescence predict adult depression two decades later. *J Pers Soc Psychol.* (2014) 106:325–38. doi: 10.1037/a0035133
79. Disner SG, Shumake JD, Beevers CG. Self-referential schemas and attentional bias predict severity and naturalistic course of depression symptoms. *Cogn Emot.* (2017) 31:632–44. doi: 10.1080/02699931.2016.1146123
80. Dobson KS, Shaw BF. Specificity and stability of self-referent encoding in clinical depression. *J Abnorm Psychol.* (1987) 96:34. doi: 10.1037/0021-843X.96.1.34
81. De Houwer J. What are implicit measures and why are we using them? *Rwwiers*. (2006) 11–28. doi: 10.4135/9781412976237.n2
82. Gemar MC, Segal ZV, Sagrati S, Kennedy SJ. Mood-induced changes on the implicit association test in recovered depressed patients. *J Abnorm Psychol.* (2001) 110:282. doi: 10.1037/0021-843X.110.2.282
83. Kesting ML, Mehl S, Rief W, Lindenmeyer J, Lincoln TM. 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 Res.* (2011) 186:197–202. doi: 10.1016/j.psychres.2010.08.036
84. Lemmens LH, Roefs A, Arntz A, van Teeseling HC, Peeters F, Huibers MJ. The value of an implicit self-associative measure specific to core beliefs of depression. *J Behav Ther Exp Psychiatry*. (2014) 45:196–202. doi: 10.1016/j.jbtep.2013.10.006
85. Raedt RD, Schacht R, Franck E, Houwer JD. Self-esteem and depression revisited: implicit positive self-esteem in depressed patients? *Behav Res Ther.* (2006) 44:1017–28. doi: 10.1016/j.brat.2005.08.003
86. van Tuijl LA, Glashouwer KA, Bockting CL, Tendeiro JN, Penninx BW, de Jong PJ. Implicit and explicit self-esteem in current, remitted, recovered, and comorbid depression and anxiety Disorders: the NESDA study. *PLoS ONE.* (2016) 11:e0166116. doi: 10.1371/journal.pone.0166116
87. Jabben N, de Jong PJ, Kupka RW, Glashouwer KA, Nolen WA, Penninx BW. Implicit and explicit self-associations in bipolar disorder: a comparison with healthy controls and unipolar depressive disorder. *Psychiatry Res.* (2014) 215:329–34. doi: 10.1016/j.psychres.2013.11.030
88. Creemers DHM, Scholte RHJ, Engels RCME, Prinstein MJ, Wiers RW. Damaged self-esteem is associated with internalizing problems. *Front Psychol.* (2013) 4:152. doi: 10.3389/fpsyg.2013.00152
89. Dimaro LV, Roberts NA, Moghaddam NG, Dawson DL, Brown I, Reuber M. Implicit and explicit self-esteem discrepancies in people with psychogenic nonepileptic seizures. *Epilepsy Behav.* (2015) 46:109–17. doi: 10.1016/j.yebeh.2015.03.032
90. Leeuwis FH, Koot HM, Creemers DH, van Lier PA. Implicit and explicit self-esteem discrepancies, victimization and the development of late childhood internalizing problems. *J Abnor Child Psychol.* (2015) 43:909–19. doi: 10.1007/s10802-014-9959-5
91. Schröder-Abé M, Rudolph A, Schütz A. High implicit self-esteem is not necessarily advantageous: discrepancies between explicit and implicit self-esteem and their relationship with anger expression and psychological health. *Eur J Pers.* (2007) 21:319–39. doi: 10.1002/per.626
92. Valiente C, Cantero D, Vazquez C, Sanchez A, Provencio M, Espinosa R. Implicit and explicit self-esteem discrepancies in paranoia and depression. *J Abnorm Psychol.* (2011) 120:691–9. doi: 10.1037/a0022856
93. Eberl C, Winkler I, Pawelczack S, Tröbitz E, Rinck M, Becker ES, et al. Self-esteem consistency predicts the course of therapy in depressed patients. *PLoS ONE.* (2018) 13:e0199957. doi: 10.1371/journal.pone.0199957
94. Poulsen C, Luu P, Crane SM, Quiring J, Tucker DM. Frontolimbic activity and cognitive bias in major depression. *J Abnorm Psychol.* (2009) 118:494–506. doi: 10.1037/a0015920
95. Waters AC, Tucker DM. Principal components of electrocortical activity during self-evaluation indicate depressive symptom severity. *Soc Cogn Affect Neurosci.* (2016) 11:1335–43. doi: 10.1093/scan/nsw046
96. Dainer-best J, Trujillo LT, Schnyer DM, Beevers CG. Sustained engagement of attention is associated with increased negative self-referent processing in major depressive disorder. *Biol Psychol.* (2017) 129:231–41. doi: 10.1016/j.biopsycho.2017.09.005
97. Speed BC, Nelson BD, Auerbach RP, Klein DN, Hajcak G. Depression risk and electrocortical reactivity during self-referential emotional processing in 8 to 14 year-old girls. *J Abnorm Psychol.* (2016) 125:607–19. doi: 10.1037/abn0000173
98. West WC, Holcomb PJ. Imaginal, Semantic, and Surface-Level Processing of Concrete and Abstract Words: An Electrophysiological Investigation. *J. Cogn. Neurosci.* (2000) 12:1024–37. doi: 10.1162/08989290051137558
99. Kutas M, Federmeier KD. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Ann Rev Psychol.* (2011) 62:621. doi: 10.1146/annurev.psych.093008.131123
100. White KR, Crites SL, Taylor JH, Corral G. Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Soc Cogn Affect Neurosci.* (2009) 4:191. doi: 10.1093/scan/nsp004
101. Yang J, Hong L, Jiang Q, Qinglin Z. Effects of self evaluation on late ERP components.*自然科学进展 Natural in progress.* (2007) B07:106–9.
102. Naumann E, Bartussek D, Diedrich O, Laufer ME. Assessing cognitive and affective information processing functions of the brain by means of the late positive complex of the event-related potential. *J Psychophysiol.* (1992) 63:285–98.

103. Drevets WC, Price JL, Furey ML. Brain structural and functional abnormalities in mood disorders: implications for neurocircuitry models of depression. *Brain Struct Funct.* (2008) 213:93–118. doi: 10.1007/s00429-008-0189-x
104. Morel N, Villain N, Rauchs G, Gaubert M, Piolino P, Landeau B, et al. Brain activity and functional coupling changes associated with self-reference effect during both encoding and retrieval. *PLoS ONE.* (2014) 9:e90488. doi: 10.1371/journal.pone.0090488
105. Gillihan SJ, Farah MJ. Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychol Bull.* (2005) 131:76–97. doi: 10.1037/0033-2909.131.1.76
106. Grimm S, Ernst J, Boesiger P, Schuepbach D, Hell D, Boeker H, et al. Increased self-focus in major depressive disorder is related to neural abnormalities in subcortical-cortical midline structures. *Hum Brain Mapp.* (2009) 30:2617–27. doi: 10.1002/hbm.20693
107. Mor N, Winquist J. Self-focused attention and negative affect: a meta-analysis. *Psychol Bull.* (2002) 128:638. doi: 10.1037/0033-2909.128.4.638
108. Yoshimura S, Okamoto Y, Onoda K, Matsunaga M, Ueda K, Suzuki S, et al. Rostral anterior cingulate cortex activity mediates the relationship between the depressive symptoms and the medial prefrontal cortex activity. *J Affect Disord.* (2010) 122:76–85. doi: 10.1016/j.jad.2009.06.017
109. Bradley KA, Colcombe S, Henderson SE, Alonso CM, Milham MP, Gabbay V. Neural correlates of self-perceptions in adolescents with major depressive disorder. *Dev Cogn Neurosci.* (2016) 19:87–97. doi: 10.1016/j.dcn.2016.02.007
110. Kuhn S, Gallinat J. Resting-state brain activity in schizophrenia and major depression: a quantitative meta-analysis. *Schizophr Bull.* (2013) 39:358–65. doi: 10.1093/schbul/sbr151
111. Pavlickova H, Turnbull OH, Bentall RP. Discrepancies between explicit and implicit self-esteem and their relationship to symptoms of depression and mania. *Psychol Psychother Theor Res Prac.* (2014) 87:311–23. doi: 10.1111/papt.12015
112. Grimm S, Boesiger P, Beck J, Schuepbach D, Bermpohl F, Walter M, et al. Altered negative BOLD responses in the default-mode network during emotion processing in depressed subjects. *Neuropsychopharmacology.* (2009) 34:932. doi: 10.1038/npp.2008.81
113. Masten CL, Eisenberger NI, Borofsky LA, Mcnealy K, Pfeifer JH, Dapretto M. Subgenual anterior cingulate responses to peer rejection: a marker of adolescents' risk for depression. *Dev Psychopathol.* (2011) 23:283. doi: 10.1017/s0954579410000799
114. Moritz D, Roberts JE. Self-other agreement and metaperception accuracy across the Big Five: examining the roles of depression and self-esteem. *J Personal.* (2017) 86:296–307. doi: 10.1111/jopy.12313
115. Ranjith G, Farmer A, Mcguffin P, Cleare AJ. Personality as a determinant of social functioning in depression. *J Affect Disord.* (2005) 84:73–6. doi: 10.1016/j.jad.2004.10.008
116. Roberts SB, Kendler KS. Neuroticism and self-esteem as indices of the vulnerability to major depression in women. *Psychol Med.* (1999) 29:1101–9. doi: 10.1017/S0033291799008739
117. Schmitz N, Kugler J, Rollnik J. On the relation between neuroticism, self-esteem, and depression: results from the National Comorbidity Survey. *Comprehens Psychiatry.* (2003) 44:169–76. doi: 10.1016/S0010-440X(03)00008-7
118. Chiao JY, Harada T, Komeda H, Li Z, Mano Y, Saito D, et al. Neural basis of individualistic and collectivistic views of self. *Hum Brain Mapp.* (2009) 30:2813–20. doi: 10.1002/hbm.20707
119. Bosson JK, Brown RP, Zeigler-Hill V, Swann WB. Self-enhancement tendencies among people with high explicit self-esteem: the moderating role of implicit self-esteem. *Self Identity.* (2003) 2:169–87. doi: 10.1080/15298860309029
120. Leary MR, Tambor ES, Terdal SK, Downs DL. Self-esteem as an interpersonal monitor: the sociometer hypothesis. *J Pers Soc Psychol.* (1995) 68:270–4. doi: 10.1037/0022-3514.68.3.518
121. Pfeifer JH, Lieberman MD, Dapretto M. "I Know You Are But What Am I?!": neural bases of self- and social knowledge retrieval in children and adults. *J Cogn Neurosci.* (2007) 19:1323–37. doi: 10.1162/jocn.2007.19.8.1323
122. Hauser J, Leszczynska A, Samochowiec J, Czerski PM, Ostapowicz A, Chlopocka M, et al. Association analysis of the insertion/deletion polymorphism in serotonin transporter gene in patients with affective disorder. *Eur Psychiatry J Assoc Eur Psychiatr.* (2003) 18:129. doi: 10.1016/S0924-9338(03)00026-9
123. Dinger U, Ehrenthal JC, Nikendei C, Schauenburg H. Change in self-esteem predicts depressive symptoms at follow-up after intensive multimodal psychotherapy for major depression. *Clin Psychol Psychother.* (2017) 24:1040–6. doi: 10.1002/cpp.2067
124. Orth U, Robins RW. Understanding the Link between low self-esteem and depression. *Curr Direct Psychol Sci.* (2013) 22:455–60. doi: 10.1177/0963721413492763
125. Sowislo JF, Orth U. Does low self-esteem predict depression and anxiety? A meta-analysis of longitudinal studies. *Psychol Bull.* (2013) 139:213–40. doi: 10.1037/a0028931
126. Wegener I, Geiser F, Alfter S, Mierke J, Imbierowicz K, Kleiman A, et al. Changes of explicitly and implicitly measured self-esteem in the treatment of major depression: evidence for implicit self-esteem compensation. *Comprehensive Psychiatry.* (2015) 58:57–67. doi: 10.1016/j.comppsych.2014.12.001
127. Hulme N, Hirsch C, Stopa L. Images of the self and self-esteem: do positive self-images improve self-esteem in social anxiety? *Cogn Behav Ther.* (2012) 41:163. doi: 10.1080/16506073.2012.664557
128. Keng SL, Seah STH, Tong EMW, Smoski M. Effects of brief mindful acceptance induction on implicit dysfunctional attitudes and concordance between implicit and explicit dysfunctional attitudes. *Behav Res Ther.* (2016) 83:1–10. doi: 10.1016/j.brat.2016.05.004
129. Remmers C, Topolinski S, Koole SL. Why being mindful may have more benefits than you realize: Mindfulness improves both explicit and implicit mood regulation. *Mindfulness.* (2016) 7:829–37. doi: 10.1007/s12671-016-0520-1
130. Di SM, Norbury R, Harmer CJ. Short-term antidepressant administration reduces negative self-referential processing in the medial prefrontal cortex in subjects at risk for depression. *Mol Psychiatry.* (2012) 17:503–10. doi: 10.1038/mp.2011.16
131. Dubini A, Bosc M, Polin V. Noradrenaline-selective versus serotonin-selective antidepressant therapy: differential effects on social functioning. *J Psychopharmacol.* (1997) 11:17–23.
132. Matthews SC, Simmons AN, Strigo IA, Arce E, Stein MB, Paulus MP. Escitalopram attenuates posterior cingulate activity during self-evaluation in healthy volunteers. *Psychiatry Res.* (2010) 182:81–7. doi: 10.1016/j.pscychresns.2010.02.003
133. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch General Psychiatry.* (1961) 4:561–71. doi: 10.1001/archpsyc.1961.01710120031004

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

II

ALTERED TEMPOROPARIETAL JUNCTION ACTIVITY DURING REFLECTED SELF-EVALUATION IN SUB-CLINICAL DEPRESSION

by

Yixue Lou, Shengdong Chen, Guoqiang Hu, Piia S. Astikainen, Susannah C.S.A. Otieno-Leppänen, Yi Lei, Paavo H.T. Leppänen, 2023

Cerebral cortex, 33(12), 7553–7563

Available online: https://doi.org/10.1093/cercor/bhad060

Cerebral Cortex, 2023, 1–11
https://doi.org/10.1093/cercor/bhad060
Original Article

Altered temporoparietal junction activity during reflected self-evaluation in sub-clinical depression

Yixue Lou[1,2,3], Shengdong Chen[4], Guoqiang Hu[5], Piia Astikainen[2], Susannah C.S.A. Otieno-Leppänen[2], Yi Lei[1,*], Paavo H.T. Leppänen[2]

[1]Institute for Brain and Psychological Sciences, Sichuan Normal University, Chengdu, 610066, China,
[2]Department of Psychology, University of Jyvaskyla, Jyväskylä 40014, Finland,
[3]School of Psychology, Shenzhen University, Shenzhen 518060, China,
[4]School of Psychology, Qufu Normal University, Qufu 273165, China,
[5]School of Biomedical Engineering, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China

*Corresponding author: Institute for Brain and Psychological Sciences, Sichuan Normal University, Jing'an Road 5, Jinjiang District, Chengdu 610066, P.R. China. E-mail: leiyi821@vip.sina.com

Negative self-schema is a core symptom of depression. According to social psychological theories, two types of self-evaluations play important roles in forming the negative self-view: direct self-evaluation (that is, evaluating the self directly through one's first-person perspective introspection) and reflected self-evaluation (which requires theory of mind (ToM) ability, and is evaluating the self through reflecting on a third person's perspective). Although many previous studies have investigated the processing of the direct self-evaluation in depression, few have extended research on the reflected self-evaluation. In the current study, functional magnetic resonance imaging scans were acquired in 26 dysphoric (individuals with elevated number of depressive symptoms) and 28 control participants during both direct and reflected self-evaluation tasks. Two regions of interest were defined within bilateral temporoparietal junctions (TPJs) because their significant role in ToM. Results showed that the dysphoric participants evaluated themselves more negatively than the control participants, regardless of whose perspective they were taking. More importantly, the enhanced TPJs' activations were observed in the control group during the reflected self-evaluation task versus the direct self-evaluation task, whereas no such difference was observed in the dysphoric participants. The results are interpreted in the framework of impaired ToM ability in sub-clinical depression.

General Scientific Summary (GSS)
Negative self-schema is one of the core symptoms of depression. This study suggests that the negative self-schema reflects not only in direct self-evaluation (i.e. evaluating the self via one's own introspection) but also in reflected self-evaluation (i.e. evaluating the self via others' perspective). Importantly, altered TPJ activity was found during a reflected self-evaluation task among individuals with depressive symptoms. These changes in brain function might be associated with impaired ToM ability in sub-clinical depression.

Key words: depressive symptoms; dysphoria; self-view; reflected self-evaluation; temporoparietal junction.

Introduction

An abnormal negative self-view has been associated with depression (for a review, see Lou et al. 2019). For example, depressed individuals often consider themselves worthless and unfavored (Beck et al. 1979), even though these beliefs are not always an objective reflection of reality. However, it is unclear how these individuals develop such a negative self-view. According to the classic theories of social psychology, there are at least two plausible factors that can shape how individuals see themselves (Cooley 1902; Mead 1934; Markus 1977). First, this view may develop through direct life experiences, where the self is evaluated via a first-person perspective introspection (Markus 1977). Secondly, the self may be evaluated via a third person perspective; that is, how other people "see" us (Cooley 1902; Mead 1934). In empirical research, the first approach is commonly termed "direct self-evaluation/self-appraisal," whereas the second approach is termed "reflected self-evaluation/self-appraisal" (Ochsner et al. 2005; Tice and Wallace 2005; Pfeifer et al. 2009; Jankowski et al. 2014; Xu et al. 2015; Pfeifer et al. 2017; Van der Cruijsen et al. 2019). To date, depression-related social neuroscience studies have mainly focused on the processing of direct self-evaluation (e.g. Yoshimura et al. 2010; Wagner et al. 2015; Disner et al. 2017; Li et al. 2017; Liu et al. 2020), but few have extended the research on reflected self-evaluation.

According to previous studies, reflected self-evaluation mainly involves the theory of mind (ToM) ability (e.g. perspective-taking) (D'Argembeau et al. 2007; Pfeifer et al. 2009), which is the ability to reason about others' mental states and predict their behaviors based on reasoning (Saxe and Kanwisher 2003; Bora and Berk 2016). For example, in order to make a reflected self-evaluation, people often project themselves via/through others' "eyes" and estimate what others might think of them (D'Argembeau et al. 2007; Pfeifer et al. 2017). Based on recent meta-analysis, the ToM ability is impaired in major depressive disorder as the depressed patients significantly underperformed in comparison with the healthy controls in a variety of tasks measuring ToM (Bora and Berk 2016; Nestor et al. 2022), and such underperformance is significantly related to the severity of the patients' depressive symptoms (Bora and Berk 2016). Besides, the meta-analysis has also suggested that the ToM deficits in depression might be related to functional abnormalities of brain areas/networks that are specifically important for the ToM, rather than other basic information processing (such as visual perception) (Bora and Berk 2016).

Received: April 10, 2022. **Revised:** February 9, 2023. **Accepted:** February 10, 2023

Neuroimaging studies have shown that the temporoparietal junction (TPJ), which is roughly characterized as an area at the border between the temporal and parietal lobes surrounding the ends of the Sylvian fissure (Schurz et al. 2014), is activated in the ToM ability measuring tasks (Denny et al. 2012; Mahy et al. 2014; Schurz et al. 2014; Richardson and Saxe 2020). For example, Saxe and her colleagues found that the bilateral TPJs, especially the right TPJ, is more activated in thinking about others' mental state (such as others' thoughts and false beliefs) relative to non-mental thinking (such as others' bodily sensation and physical attributes), among both adults (Saxe and Kanwisher 2003; Saxe and Powell 2006; Saxe 2010) and children (Gweon et al. 2012; Richardson and Saxe 2020). Furthermore, patient studies have found that lesions of the TPJ could result in selective deficits in the patients' ability to infer to someone else's belief (e.g. someone else's false belief), but such deficits could not be simply explained by the patients' basic cognitive problems (such as sentence and picture perception problems) or executive problems (attention shifting and behavioral inhibition problems, for instance) as most of the patients' cognitive and executive functions were intact (Apperly et al. 2004; Samson et al. 2004). In the field of self-related studies, researchers have also observed stronger activation in TPJ among healthy adolescent and young adults during reflected versus direct self-evaluation (Pfeifer et al. 2009; Veroude et al. 2014; Van der Cruijsen et al. 2019). Therefore, these findings suggest that the TPJ plays a critical role in reasoning about others' mental states.

In the present study, with bilateral TPJs as regions of interest, we investigated whether the bilateral TPJs show aberrant responses during reflected self-evaluation in depression? For this purpose, we designed a reflected self-evaluation task that asked participants to evaluate, using a scale from 1 (not at all) to 4 (very well), the extent to which a series of personality-trait adjectives (brave, confident, selfish, cowardly, etc.) could be used to describe them, either according to their own perspective or to others' (such as a friend or a classmate). The functional magnetic resonance imaging (fMRI) scans were acquired during the evaluation task in a sub-clinical depression group (that is, those with an elevated amount of depressive symptoms although they did not have clinical diagnosis, here labeled as dysphoric participants) and a non-depressed control group. Considering the influence of negative self-schema (Beck 1967; Beck et al. 1979), we expected that the dysphoric group would exhibit more negative behavioral ratings than the control group. At the brain level, we expected, in line with the previous findings (Pfeifer et al. 2009; Van der Cruijsen et al. 2019), increased TPJ activation in the reflected versus direct self-evaluation for the control group. For the dysphoric group, however, the TPJ activation might be less significant during the reflected self-evaluation, in comparison to the direct self-evaluation, because of the suggested ToM deficits in depression (Bora and Berk 2016; Nestor et al. 2022).

Materials and methods

Participants

Before formal recruitment, university students from Beijing city who were interested in participating in the study volunteered to fill the Beck Depression Inventory-II (BDI-II) online (Beck et al. 1996). The BDI-II is a 21-item instrument that measures depressive symptoms, with its total score ranging from 0 to 64. There were 275 respondents. Based on the standardized cutoffs of the BDI-II (Beck et al. 1996), all respondents who scored 14 or above were then invited to participate in the dysphoric group ($n = 30$), whereas respondents who matched the dysphoric participants in relation to gender and age and scored below 13 in the BDI-II measurement were invited to participate in the control group ($n = 30$). The sample size was determined based on the reference of previous fMRI studies that investigated similar topics (e.g. (Pfeifer et al. 2017; Shiota et al. 2017)). None of the participants reported current or had a history of physiological, neurological, or psychiatric disorders. A convenient sub-clinical sample (that is, the dysphoric group) was recruited to avoid potential confounds related to the use of anti-depression medications, which have been thought to possibly affect the brain structures that are involved in self-related processing (Delaveau et al. 2016). Similar sub-clinical samples have been used in some previous depression-related self-evaluation studies as well (e.g. Shiota et al. 2017; Lou et al. 2021).

Of the 60 original participants, data from four dysphoric participants and two control participants were excluded due to excessive head movement (≥ 2 mm) during the collection of the fMRI data, leaving a final sample of 54 participants (Dysphoric group = 26, Control group = 28; Table 1). There was no significant difference in age (t (52) = 0.98, $P = 0.33$) or gender (Pearson χ^2 (1) = 0.30, $P = 0.59$) between the two groups. All participants were native speakers of Chinese, right-handed, and had normal or corrected-to-normal vision. The study was approved by the Review Board for Human Participant Research of Shenzhen University, and each subject had signed an informed consent form before the experiment.

Stimuli

Forty personality-trait adjectives (in Chinese, 20 positive and 20 negative) were selected from a pool of 562 personality-trait adjectives that had been developed by Huang and Zhang (1992) (see illustration list in Supplementary Data I, Table 1). According to our pilot pre-measurement ($n = 25$), there were no significant differences between the positive adjectives and the negative adjectives in terms of numbers of character strokes, meaningfulness, familiarity, or arousal ratings, and the positive adjectives were rated as more desirable than the negative adjectives (see Fig. 1). All stimuli consisted of two Chinese characters and were presented on a gray background in black Song font.

Experimental protocol

Before the experiment, the participants were asked to offer two names of their acquaintances, one of whom they personally knew well (that is, "close other," such as one of their best friends), another of an acquaintance who was not familiar or close ("unclose other," such as an unfamiliar classmate). These two people were selected to be at a similar age and of the same gender as the participant. During the experiment, the participants were asked to evaluate the extent to which a series of personality-trait adjectives described themselves according to their own opinion (Self to Self-condition, S2S), the close other's perspective (Close other to Self, C2S), and the unclose other's perspective (Unclose other to Self, U2S). We also asked participants to evaluate how these adjectives described the close other (Self to Close other, S2C) and the unclose other (Self to Unclose other, S2U). The baseline condition was to evaluate the desirability of each adjective. All of the evaluations were made using four-point Likert scales and the set of adjectives was consistent across all the conditions. Instructions and rating scales for each condition are illustrated in Table 2.

The procedure was block-designed, with each block consisting of one of the six conditions. As the positive and the negative adjectives were presented in separate blocks, the total amount

Table 1. Demographic description of participants in both dysphoric and control groups.

Description	Units	Dysphoric	Control
N participants	N (females)	26 (18)	28 (11)
Age	M ± SD (range) years	21.12 ± 2.86 (18 ~ 28)	21.86 ± 2.69 (18 ~ 28)
BDI-II	M ± SD (range) points	19.85 ± 4.95 (14 ~ 32)	2.11 ± 2.17 (0 ~ 8)

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

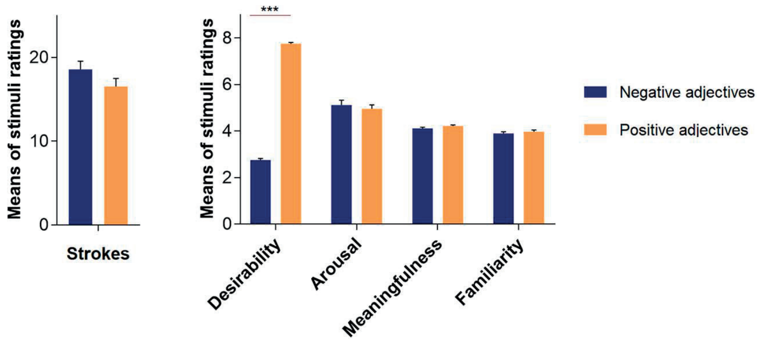

Fig. 1. Pilot experiment characteristics of the stimuli. Left-hand panel: mean number of the characters' stroke for positive (orange) and negative (blue) adjectives. Right-hand panel: Ratings of desirability (1 = very bad; 5 = not good and not bad; 9 = very good), arousal (1 = very calm or relaxing; 5 = no feelings; 9 = very exciting or alerting), meaningfulness (1 = not understandable at all; 5 = very understandable), and familiarity (1 = not familiar at all; 5 = very familiar). Error bars represent standard errors. ***$P < 0.001$.

Table 2. Illustration of instructions and rating scales in each condition.

Condition	Instruction	Rating scale
S2S	Please evaluate, in your opinion, how appropriate each of the following words is to describe you?	1—not at all
C2S	Please evaluate, according to 'friend's name', how appropriate each of the following words is to describe you?	2—a little
U2S	Please evaluate, according to 'classmate's name', how appropriate each of the following words is to describe you?	3—very well
S2C	Please evaluate, in your opinion, how appropriate each of the following words is to describe 'friend's name'?	4—completely
S2U	Please evaluate, in your opinion, how appropriate each of the following words is to describe 'classmate's name'?	
Baseline	Please evaluate, in general, how does each of the following words describe a desirable personality-trait?	1—very bad 2—bad 3—good 4—very good

Note. S2S = self to self; C2S = close other (e.g. one of the participant's best friend) to self; U2S = unclose other (e.g. one of the participant's unfamiliar classmate) to self; S2C = self to close other (e.g. the best friend); S2U = self to unclose other (e.g. the classmate); Baseline = evaluating the desirability of each stimulus word

of the blocks was 12. To avoid fatigue, the data were collected in four scanning runs so that the participants could take a break between each run. Each run consisted of a complete block set (that is, 12 blocks—6 positive blocks and 6 negative blocks). Each block consisted of five trials, with each trial containing one personality-trait adjective. The same adjectives were used across all the positive/negative blocks (e.g. if the participants evaluated in Block A "how appropriate the word 'smart' is to describe yourself", they would also need to evaluate in Block B "how appropriate the word 'smart' is to describe your friend XXX". Similarly, the word 'smart' would also appear in other positive blocks). Within a run, all 12 blocks were presented in complete random order, whereas the order of the four runs was counterbalanced by Latin-square design among participants.

Prior to the onset of each block, an introduction was displayed on the center of the screen for six seconds, indicating whose perspective the participants should take to evaluate whom (for instruction details, see Table 2). Within a trial, a personality-trait adjective was presented during the initial two seconds, followed by a four-point Likert Scale that lasted for three seconds. Participants were asked to make their evaluation during the presentation of the Likert Scale (for rating details, see Table 2). Once a block was finished, a fixation cross appeared on the center of the screen as a jitter and lasted for either six, eight, or 10 seconds. The next block initiated with a presentation of another instruction (see Fig. 2). Within a block, the adjectives were presented in a random order.

Data analysis

Behavioral data analyses

Closeness ratings

The participants' subjective closeness with the close other (e.g. the friend) and the unclose other (e.g. the classmate) were assessed using the Inclusion of Other in the Self scale, a one-item graphic measure that depicts the closeness between the respondent and other person (Aron et al. 1992). The ratings were then entered into a two-way mixed analysis of variance (ANOVA) using SPSS 25.0, with group (dysphoric vs. control) as a between-subjects variable, and closeness (close other vs. unclose other) as a within-subjects variable. This assessment was conducted to test whether the dysphoric participants, in comparison with the control participants, perceive others as less close to themselves. By conducting this analysis, we were able

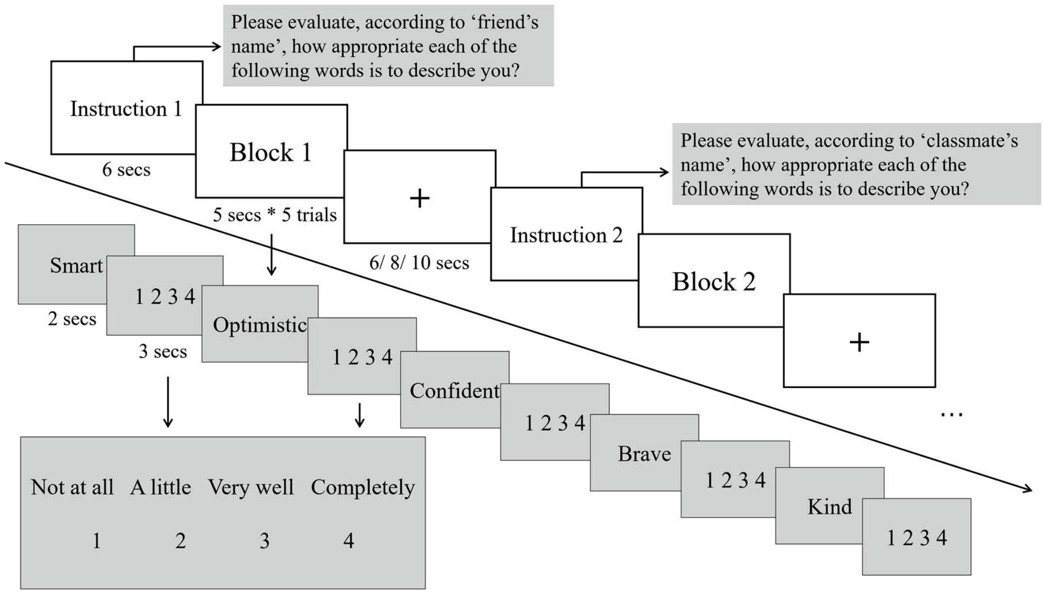

Fig. 2. Illustration of one scanning run. Each run consisted of 12 blocks. As illustrated, "block 1" represented the Close other to Self (C2S) – Positive block, whereas "block 2" represented the Unclose other to Self (U2S) – Positive block. The close others were set as one of the participants' best friends, whereas the unclose others were set as one of their unfamiliar classmates. Other blocks all followed the same setting. All the adjectives consist of two simplified Chinese characters (for fully translated examples: see Supplementary Data I, Table 1).

to minimize the possibility that the potential group difference in reflected self-evaluation might be caused by the participant group's different closeness perceptions.

Desirability ratings

We examined whether there was any significant difference between the dysphoric group and the control group in desirability ratings for positive and negative adjectives. Means of the participants' adjective desirability ratings, which were collected from the baseline condition, were entered into a two-way mixed ANOVA with group (dysphoric vs. control) as the between-subject variable, and valence (positive vs. negative) as the within-subject variable. This analysis was intended to test whether the dysphoric participants would evaluate the positive adjectives as less positive and the negative adjectives as more negative than the control group did. By conducting this analysis, we were able to minimize the possibility that the potential group difference in reflected self-evaluation might be caused by their preference of the adjectives.

Evaluation ratings

To test the possible group difference in both direct and reflected self-evaluation at the behavioral level, the means of the participants' evaluation ratings in each condition were entered into a three-way mixed ANOVA with group (dysphoric vs. control) as a between-subjects variable, and valence (positive vs. negative) and conditions (S2S, S2C, S2U, C2S, U2S) as within-subjects variables.

Imaging data analysis

Imaging data acquisition and preprocessing

Image acquisition was performed with a 3.0 Tesla MRI scanner (Magnetom Prisma, Siemens, Erlangen, Germany). The imaging parameters were as follows: Task-based fMRI data were acquired using an EPI sequence with TR = 2000 ms, TE = 30 ms, flip angle = 90°, FOV = 224 × 224 mm^2, matrix size = 64 × 64, slice thickness = 2 mm, voxel size = 2 × 2 × 2 mm^3, and slice number = 64. T1-weighted structural images were acquired using a sagittal 3D-magnetization prepared rapid acquisition gradient echo (MPRAGE) sequence: TR = 2530 ms, TE = 2.98 ms, inversion time = 1100 ms, FA = 7°, FOV = 256 × 256 mm^2, slice thickness = 1 mm, voxel size = 0.5 × 0.5 × 1 mm^3, and sagittal planes = 192. Statistical Parametric Mapping 12 (https://www.fil.ion.ucl.ac.uk/spm/software/spm12/) was used for the fMRI data analysis with regular preprocessing steps of realignment, volume registration, spatial normalization (resampled into 2-mm isotropic voxels), and spatial smoothing with a Gaussian kernel of 6 mm full width at half maximum. Head movement estimates derived from the realignment step were included as nuisance regressors in subsequent general linear model (GLM) to diminish the impact of movement-related effects.

fMRI data analysis

For the fMRI data analysis, the self-to-other evaluation conditions (namely the S2C and the S2U conditions) were not included. This was because the analysis of the brain imaging data was to investigate brain activity differences between direct self-evaluation (that is, self-to-self-evaluation, such as the S2S condition) versus reflected self-evaluation (that is, other-to-self-evaluation, such as the C2S and U2S conditions) among dysphoric participants.

In the first-level individual analysis, we created three contrasts for the self-evaluation conditions separately for positive and negative valence. The activation in the baseline condition was subtracted from the activation in the self-evaluation conditions, with the corresponding valence (i.e. S2S(pos) > Baseline(pos), C2S(pos) > Baseline(pos), U2S(pos) > Baseline(pos), S2S(neg) > Baseline(neg), C2S(neg) > Baseline(neg), U2S(neg) > Baseline(neg)) for each participant. Two regions of interest (ROIs), consisting of 8-mm spheres—the left TPJ [MNI coordinates: $x = -53$, $y = -59$, $z = 20$] and the right TPJ [$x = 56$, $y = -56$, $z = 18$]—were then defined based on recent brain imaging studies as well as meta-analyses in

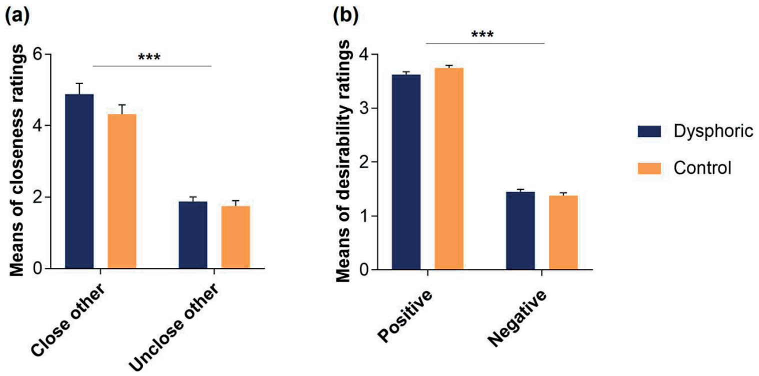

Fig. 3. (**a**) Means of the participants' closeness ratings toward the close other (e.g. one of their best friends) and the unclose other (e.g. an unfamiliar classmate). (**b**) Means of desirability ratings for both positive and negative adjectives. Error bars represent standard errors. ***$P < 0.001$.

which personality-trait word stimuli were involved to investigate brain responses of ToM and reflected self-evaluation (Schurz et al. 2014; Van der Cruijsen et al. 2019). Besides the TPJs, activity of the midline brain structures, in particular the mPFC, was also frequently reported in previous studies that investigated the representation of the self (see examples at Lemogne et al. 2012; Nejad et al. 2013; Jankowski et al. 2014; Li et al. 2017; Pfeifer et al. 2017; Shiota et al. 2017; Van der Cruijsen et al. 2019; Liu et al. 2020). Therefore, one additional ROI within the mPFC area was defined and analyzed with an exploratory purpose of examining the role of the mPFC in the current study (see the analysis and results at Supplementary Data II). For each of the ROIs, the mean percent signal changes (PSCs) of each participant were extracted using the Marsbar toolbox (Brett et al. 2002) and exported to the SPSS. The contrasts above were then submitted to the second-level group analyses in three-way mixed ANOVAs with group (dysphoric vs. control) as a between-subjects variable, and valence (positive vs. negative) as well as conditions (S2S, C2S, U2S) as within-subjects variables.

For both the behavioral and the brain data analyses, the significance levels were set at 0.05, while the degrees of freedom of the F-ratio were corrected for violations of spherical assumptions using the Greenhouse–Geisser method. The Bonferroni correction was used for the post hoc comparisons if significant main or interaction effects were found. Partial eta squared (η_p^2) values were calculated and reported to demonstrate the effect size of significant ANOVA results.

Results

Behavioral results

Closeness ratings

There was no other significant main or interaction effect except for a significant main effect of closeness [F (1, 52) = 262.63, $P < 0.001$, $\eta_p^2 = 0.84$]. Participants rated the close others (e.g. friends; $M = 4.59$, SD = 1.45) as being closer to them than the unclose others (such as classmates; $M = 1.81$, SD = 0.73), regardless of their depressive symptoms (see Fig. 3a). The lack of group difference suggests that the potential group difference in relation to reflected self-evaluation should not be caused by the difference in the participant group's closeness perception toward the others.

Desirability ratings

There was no other significant main or interaction effect apart from a significant main effect of valence [F (1, 52) = 1236.87, $P < 0.001$, $\eta_p^2 = 0.96$]. Participants rated the positive words ($M = 3.69$, SD = 0.28) as more positive than negative words ($M = 1.41$, SD = 0.25; see Fig. 3b). The lack of group difference suggests that the potential group difference regarding reflected self-evaluation should not be caused by the possibility that the dysphoric group, relative to the control group, might perceive the stimulus words in a more negative way.

Evaluation ratings

Results showed a significant valence × conditions × group interaction [F (4, 208) = 6.37, $P < 0.001$, $\eta_p^2 = 0.11$]. The post hoc analysis investigating this interaction showed significant condition × group interactions in both the positive [F (4, 208) = 4.49, $P = 0.006$, $\eta_p^2 = 0.08$] and the negative [F (4, 208) = 6.57, $P < 0.001$, $\eta_p^2 = 0.11$] valence. For the positive valence, there were significant group differences in the S2S (t (52) = 3.13, $P = 0.003$, Cohen's $d = 0.85$), the C2S (t (52) = 2.33, $P = 0.023$, Cohen's $d = 0.63$), and the U2S (t (52) = 2.05, $P = 0.046$, Cohen's $d = 0.54$) conditions. According to these group differences, the dysphoric participants rated themselves less positively than the control participants did, regardless of whose perspective the participants took (for means and standard deviations, see Table 3 and Fig. 4). However, the group difference was not significant in the S2C (t (52) = 0.88, $P = 0.384$) and the S2U (t (52) = −.76, $P = 0.450$) conditions, suggesting that the dysphoric participants rated the others as positively as the control group did. For the negative valence, the group difference was significant in all the conditions (S2S (t (52) = −7.94, $P < 0.001$, Cohen's $d = 2.13$); S2C (t (52) = −4.51, $P < 0.001$, Cohen's $d = 1.23$); S2U (t (52) = −3.34, $P = 0.002$, Cohen's $d = 0.88$); C2S (t (52) = −7.45, $P < 0.001$, Cohen's $d = 1.99$); U2S (t (52) = −5.00, $P < 0.001$, Cohen's $d = 1.34$)). The dysphoric participants rated everyone more negatively than the control group did, regardless of whose perspective they took (for means and standard deviations, see Table 3 and Fig. 4).

On the other hand, with the positive valence, there were significant condition effects in both the dysphoric group [F (4, 100) = 4.48, $P = 0.009$, $\eta_p^2 = 0.15$] and the control group [F (4, 108) = 7.22, $P < 0.001$, $\eta_p^2 = 0.21$]. For the dysphoric participants, the S2S evaluation was significantly less positive than the S2C evaluation, whereas there was no difference between the S2S and the S2U evaluation. For the control participants, however, the S2S evaluation was significantly more positive than the S2U evaluation, while there was no difference between the S2S and the S2C evaluation (see all the statistical data in detail at Table 3). With the negative valence, the condition effect was also

Table 3. Means and standard deviations for each condition and post hoc comparisons investigating the condition effect for the positive and the negative valence among the dysphoric and the control groups.

Group	Valence	Condition	*M* (SD)	*t* (df)				
				S2S	S2C	S2U	C2S	U2S
Dysphoric	Positive	S2S	2.59(0.46)	—				
		S2C	2.92(0.50)	−3.28(25)**	—			
		S2U	2.83(0.63)	−1.76(25)	0.77(25)	—		
		C2S	2.81(0.49)	−3.82(25)**	1.10(25)	0.19(25)	—	
		U2S	2.55(0.50)	0.54(25)	3.02(25)**	2.22(25)*	3.32(25)**	—
	Negative	S2S	2.07(0.46)	—				
		S2C	1.60(0.39)	6.24(25)***	—			
		S2U	1.57(0.48)	4.43(25)***	0.33(25)	—		
		C2S	1.75(0.38)	6.43(25)***	−2.77(25)**	−1.96(25)	—	
		U2S	1.73(0.52)	3.71(25)**	−1.56(25)	−1.33(25)	0.31(25)	—
Control	Positive	S2S	2.99(0.48)	—				
		S2C	3.04(0.50)	−0.59(27)	—			
		S2U	2.71(0.53)	2.88(27)**	3.38(27)**	—		
		C2S	3.09(0.40)	−1.90(27)	−0.72(27)	−3.98(27)***	—	
		U2S	2.83(0.53)	2.30(27)*	2.24(27)*	−1.33(27)	4.27(27)***	—
	Negative	S2S	1.32(0.19)	—				
		S2C	1.22(0.20)	2.15(27)*	—			
		S2U	1.25(0.19)	1.45(27)	−0.47(27)	—		
		C2S	1.18(0.14)	4.01(27)***	1.09(27)	1.64(27)	—	
		U2S	1.21(0.18)	3.04(27)**	0.37(27)	1.13(27)	−0.61(27)	—

Note. S2S = self to self; S2C = self to close other (e.g. one of their best friends); S2U = self to unclose other (e.g. an unfamiliar classmate); C2S = close other to self; U2S = unclose other to self; *M* = means; SD = standard deviations. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

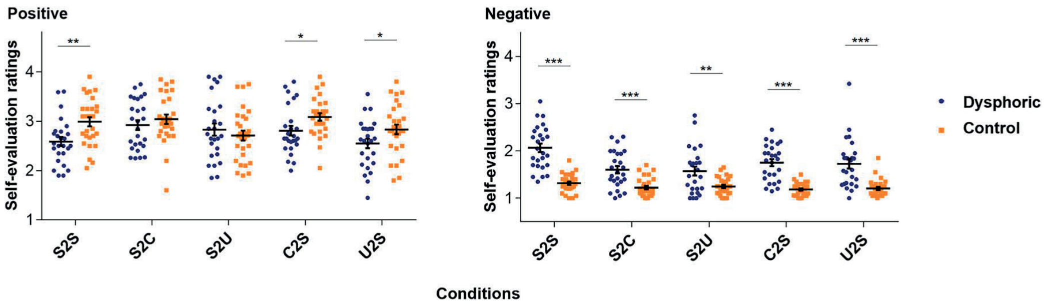

Fig. 4. Scatter plots of self-evaluation ratings in each of the conditions among the dysphoric group and the control group. S2S = self to Self; S2C = Self to Close other (e.g. one of their best friends); S2U = Self to Unclose other (e.g. an unfamiliar classmate); C2S = Close other to Self; U2S = Unclose other to Self. Horizontal lines with error bars represent means and standard errors, separately. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

significant in both the dysphoric group [F (4, 100) = 10.55, $P < 0.001$, $\eta_p^2 = 0.30$] and the control group [F (4, 108) = 3.09, $P = 0.028$, $\eta_p^2 = 0.10$]. For the dysphoric participants, the S2S evaluation was more negative than all of the four other evaluations, and for the control participants, the S2S evaluation showed no difference compared to the S2U evaluation, but it was also more negative than other evaluations (all of the statistical data were provided in detail in Table 3).

fMRI results

Although the valence × condition × group interaction was not significant [F (2, 104) = 1.78, $P = 0.175$], there was a significant condition × group interaction in the left TPJ [F (2, 104) = 7.05, $P = 0.001$, $\eta_p^2 = 0.12$]. The post hoc analysis of this interaction showed a significant condition effect in the control group [F (2, 54) = 21.73, $P < 0.001$, $\eta_p^2 = 0.45$] but not in the dysphoric group [F (2, 50) = 0.003, $P = 0.997$]. For the control participants, brain activations in the C2S ($M = 0.14$, SD = 0.11; t (27) = −4.81, $P < 0.001$, Cohen's $d = 0.97$) and in the U2S ($M = 0.17$, SD = 0.14; t (27) = −5.80, $P < 0.001$, Cohen's $d = 1.08$) conditions were significantly greater than in the S2S condition ($M = 0.04$, SD = 0.11), whereas the difference between the C2S and the U2S conditions was not significant (t (27) = −1.60, $P = 0.24$; see Fig. 5a).

For the right TPJ, the interaction of valence, condition and group was not significant [F (2, 104) = 0.09, $P = 0.910$], but there was also a significant condition × group interaction [F (2, 104) = 3.61, $P = 0.032$, $\eta_p^2 = 0.07$]. The post hoc analysis of this interaction showed a significant condition effect, again, in the control group [F (2, 54) = 6.99, $P = 0.003$, $\eta_p^2 = 0.21$] but not in the dysphoric group [F (2, 50) = 0.060, $P = 0.938$]. For the control participants, brain activation in the U2S condition ($M = 0.13$, SD = 0.10) was significantly greater than in the S2S condition ($M = 0.03$, SD = 0.15; t (27) = −4.03, $P < 0.001$, Cohen's $d = 0.80$). Additionally, the difference between the C2S condition ($M = 0.09$, SD = 0.11) and the S2S condition was

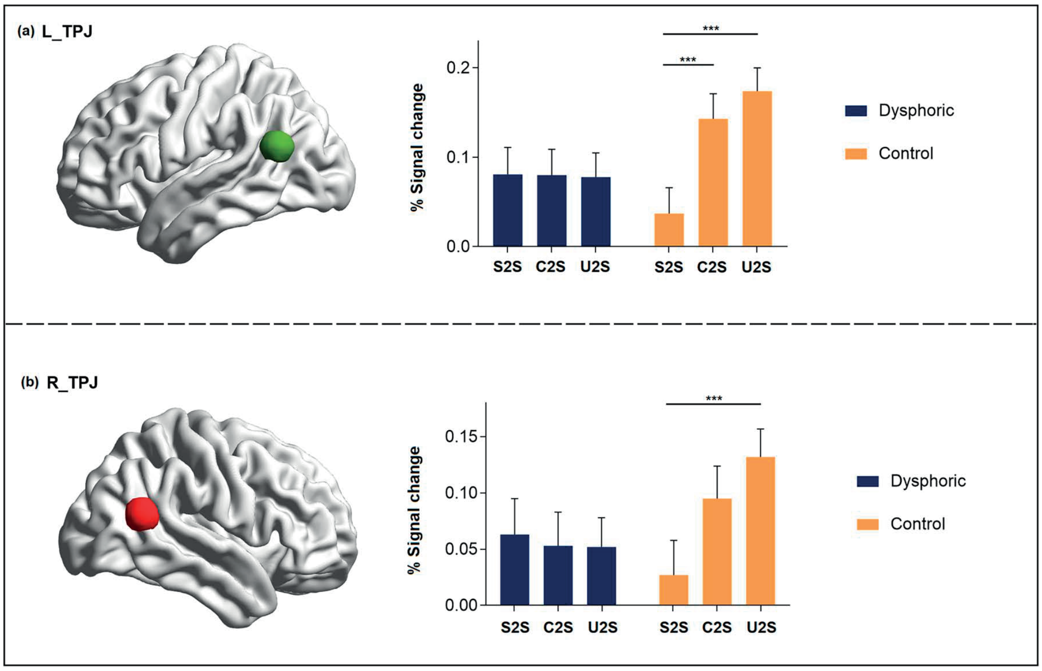

Fig. 5. Mean PSCs of each condition relative to baseline in (**a**) the L_TPJ (left temporoparietal junction; the ROI is illustrated by the green ball) and (**b**) the R_TPJ (right temporoparietal junction; the ROI is illustrated by the red ball). S2S = Self to Self; C2S = Close other (e.g. one of their best friends) to Self; U2S = Unclose other (e.g. an unfamiliar classmate) to Self. Error bars represent standard errors, $^{***}P < 0.001$.

marginally significant (t (27) = −2.13, P = 0.084), while the difference between the C2S and the U2S conditions was not significant (t (27) = −1.36, P = 0.370; see Fig. 5b).

Correlation between behavioral and brain responses

Based on the above results, a correlation analysis was conducted with the whole participant sample to investigate whether the brain responses are associated with the amount of participants' depressive symptoms. Pearson's correlation coefficient was calculated between the BDI-II scores and brain activations (means of self-evaluation condition activation minus baseline condition; that is, S2S − baseline, C2S − baseline, and U2S − baseline) in the bilateral TPJ. The results showed a significant negative correlation between the BDI-II scores and the activations in the C2S condition, as well as in the U2S condition, in both the left and the right TPJ. Table 4 provided the resultant correlation matrix, and Fig. 6 provided the scatterplot of the correlations. These results suggest that the more depressive symptoms the participant had, the smaller the bilateral TPJ activation was when taking others' perspective to evaluate the self.

Additional whole brain analysis

In addition to studying the TPJ ROIs, a whole-brain analysis was exploratorily conducted to see all possible statistically significant group effects. This analysis was performed using the general linear model (GLM), and the multiple comparison was corrected by a level of significance of $P < 0.05$ (FWE-small volume corrected). Both the group × valence interaction and the group × condition interaction were entered in a factorial design module during the second-level analysis, and the main effect of group was also examined. However, the results showed that no region had passed the significant threshold, either for the interactions or the main effect.

Discussion

The aim of the current study was to explore reflected self-evaluation in dysphoric and control participants by utilizing behavioral and brain activity measurements. To achieve this goal, we asked participants to evaluate themselves through either their own opinions or the opinions of others' (such as the participants' close friends and unclose classmates) perspectives. The behavioral ratings and the fMRI scans were acquired when the participants were performing the evaluation tasks. According to our results, the dysphoric group exhibited in general more negative behavioral ratings than the control group. Moreover, an increased TPJ activation was observed, as expected, during the reflected versus direct self-evaluation for the control group, and such a difference was absent in the dysphoric group. Next, we discuss the results in detail.

At the behavioral level, on the one hand, the dysphoric group, in comparison to the control group, exhibited higher self-evaluation ratings for the negative personality trait adjectives, and lower self-evaluation ratings for the positive personality trait adjectives, regardless of whose perspective they took. One the other hand, the dysphoric participants rated themselves as less positive than others, whereas the control participants rated themselves more positive than others. In general, this finding is in line with Beck's cognitive theory of depression (Beck 1967), which suggests that individuals with enhanced depressive symptoms usually view themselves negatively because of the influence of negative self-schema. In previous studies, researchers have reported that the depressed patients, in comparison with non-depressed healthy individuals, are prone to use more negative as well as less

Table 4. Correlations among depressive state and brain activation in the bilateral TPJ.

		BDI-II	L_TPJ			R_TPJ		
			S2S	C2S	U2S	S2S	C2S	U2S
	BDI-II	1						
L_TPJ								
	S2S	0.09	1					
	C2S	−0.32*	0.53**	1				
	U2S	−0.34*	0.40**	0.52**	1			
R_TPJ								
	S2S	0.09	0.73**	0.37**	0.20	1		
	C2S	−0.29*	0.26	0.62**	0.21	0.40**	1	
	U2S	−0.33*	0.28*	0.38**	0.71**	0.41**	0.45**	1

Note. BDI-II = Beck Depression Inventory-II; L_TPJ = left temporoparietal junction; R_TPJ = right temporoparietal junction; S2S = self to self; C2S = close other (e.g. one of their best friends) to self; U2S = unclose other (e.g. an unfamiliar classmate) to self. $^{*}P < 0.05$, $^{**}P < 0.01$

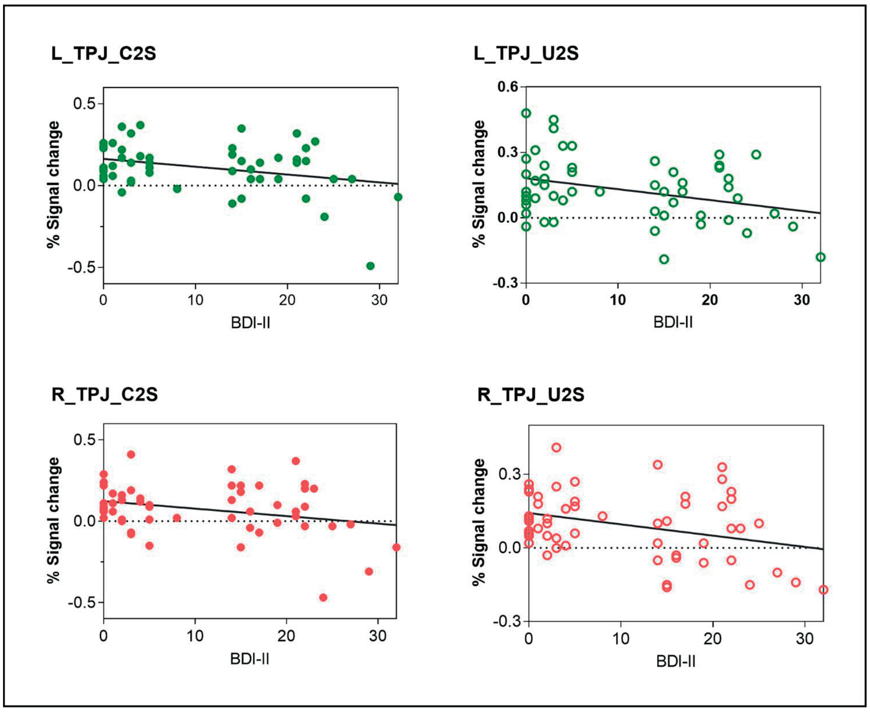

Fig. 6. Correlation between participants' BDI-II scores and their brain activities in bilateral TPJ. Brain activities are reflected by mean percentage signal changes of each condition relative to the baseline. Green color represents data in left TPJ (L_TPJ) and red color represents data in right TPJ (R_TPJ); filled circles represent data in Close-to-Self (C2S) conditions, whereas hollow circles represent data in Unclose-to-Self conditions (U2S).

positive adjectives to evaluate themselves during the direct self-evaluation (Dainer-best et al. 2017; Kiang et al. 2017). Results of the current study suggest that such a self-negativity bias exists also in the sub-clinical depression population, and could be observed not only during the direct self-evaluation, but also the reflected self-evaluation that requires people to evaluate the self through others' opinion.

At the brain level, among the control group, we observed significantly greater activations in the bilateral TPJs during reflected self-evaluation, including both the close-other-to-self condition and the unclose-other-to-self condition, versus direct self-evaluation. This result is in line with results in previous studies. For instance, by using an experimental design that is similar to ours, Pfeifer et al. (2009) observed greater TPJ activations during reflected self-evaluation versus direct self-evaluation among healthy adolescent and young adults. The engagement of TPJ in reflected self-evaluation among young adults has also been reported in Jankowski and colleagues' work (Jankowski et al. 2014). In another study, researchers found that because of the influence of the collectivism culture, Chinese young adults exhibited greater TPJ activation during reflected self-evaluation that is made from their peers than American participants did (Pfeifer et al. 2017). Taken together, the current result supports the existing notion that the TPJ plays an important role in reflecting about others' thoughts toward the self (Pfeifer et al. 2009, 2017; Jankowski et al. 2014).

Interestingly, such enhanced TPJ activations in the reflected self-evaluation, in comparison to the direct self-evaluation, were absent in the dysphoric group. The result suggests a lack of TPJ engagement during the processing of third-person perspective taking, in relation to first-person thinking, in this population. Moreover, we also found that the TPJ activations during the reflected self-evaluation task was negatively correlated with the amount of depressive symptoms, suggesting that the more depressive symptoms the participants had, the less the TPJ engagement there was when the participants evaluated themselves through others' perspectives.

The lack of TPJ engagement here might be explained by the impaired ToM ability in sub-clinical/mild depressed population (Lee et al. 2005; Manstead et al. 2013; Erle et al. 2019). Specifically, by including both clinically depressed individuals who came from outpatients, as well as sub-clinical depressed individuals that came from a wider community, Lee et al. (2005) reported a significantly impaired ToM ability in the entire depressed population compared with a non-depressed healthy control group. Importantly, in Lee's study, the severely depressed participants (most of whom were clinically depressed patients) and moderately depressed participants (most of whom were sub-clinically individuals) did not differ from each other in the ToM performance (Lee et al. 2005), suggesting that the ToM ability may be impaired in both clinically and sub-clinically depressed people. Moreover, by using the same ToM task as Lee did, Manstead et al. (2013) also observed a worse ToM performance among sub-clinically depressed participants relative to non-depressed healthy controls. In another study, by using a convenient sample that consisted of community individuals who showed mild depressive symptoms, Erle et al. (2019) found that individuals exhibiting high levels of depressive symptoms relative to non-depressed healthy controls were impaired on ToM tasks. Therefore, for the dysphoric group in the current study, the lack of TPJ engagement during the reflected self-evaluation could probably be associated with the impaired ToM ability.

According to previous studies, impaired ToM might be related to the excessively self-focused attention (Erle et al. 2019). To be specific, the ToM tasks, such as the reflected self-evaluation task, usually require people to decenter from their own perspective, in order to perceive others' thoughts (Bukowski and Lamm 2017; Erle et al. 2019). In other words, to make an accurate mental state reasoning, people need to avoid simply attributing their own thoughts to others, as such thoughts and beliefs rarely align with those of others (Steinbeis 2016). However, the excessively self-focused attention, namely a maladaptively increased attention directed inwardly to the self as opposed to the external world, is considered one of the core symptoms in both clinical depression (Brockmeyer et al. 2015) and sub-clinical dysphoria (Erle et al. 2019). We assume that, with the symptom, it might be difficult for the dysphoric people to reason others mind, and to make a self-evaluation based on the others' opinions.

Additionally, the current study did not observe significant valence effects in the TPJ activation. One possible reason is that the TPJ is not related to valence processing. Specifically, researchers have investigated the valence-related brain activation during the processing of emotional stimuli, identifying several brain areas as valence sensitive regions. For example, the orbitofrontal cortex (Anderson et al. 2003; Lewis et al. 2006), medial prefrontal cortex (Viinikainen et al. 2012), and dorsolateral prefrontal cortex (Nejati et al. 2021) have been found sensitive to valence changes of emotional stimuli. However, to our knowledge, none of the research has reported the valence-related activation within the TPJ area, although few have observed the TPJ's activity in differentiating the arousal level of emotional stimuli (Kensinger and Schacter 2006). In the current study, considering that the positive and negative stimulus words did not differ from each other in means of arousal ratings, it is reasonable that the TPJ responded to them with similar intensity.

Caution should be exercised when adopting the interpretations and generalizing the findings for the following reasons. First, considering how little we know about the function of the TPJ, we cannot rule out that the TPJ activity here is specific to social cognition. It is known that the TPJ is activated also in other tasks, for example, the right TPJ is found activated in spatial attention tasks (Krall et al. 2015; Krall et al. 2016; Dugué et al. 2018; Käsbauer et al. 2020). Second, considering the relatively limited sample size in the present study, the replicability of the current findings should be further tested in bigger sample sizes. Finally, although we interpret the absence of TPJ activation during the reflected self-evaluation as reduced ToM performance among the dysphoric participants, future studies are needed to directly investigate the ToM abilities in reflected self-evaluation with the same participants.

CRediT authors statement

Yixue Lou (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review and editing), Shengdong Cheng (Formal analysis, Methodology, Software), Guoqiang Hu (Methodology), Piia Astikainen (Writing—original draft, Writing—review and editing), Susannah C.S.A. Otieno-Leppänen (Writing—original draft), Yi Lei (Conceptualization, Funding acquisition, Project administration, Supervision, Writing—original draft, Writing—review and editing), Paavo Leppänen (Project administration, Supervision, Writing—original draft, Writing—review and editing).

Funding

The study was funded by the National Natural Science Foundation of China [NSFC32271142 and NSFC31871130]; Ministry of Education Key Projects of Philosophy and Social Sciences Research [grant number 21JZD063]; and Shenzhen Science and Technology Research Funding Program [JCYJ20200109144801736].

Conflict of interest statement: The authors declare that there have been no involvements that might raise the question of bias in the work reported or in theconclusions, implications, or opinions stated.

Data availability

The data that support the findings of this study are available from the corresponding author, Yi Lei, upon reasonable request.

References

Anderson AK, Christoff K, Stappen I, Panitz D, Ghahremani DG, Glover G, Gabrieli JDE, Sobel N. Dissociated neural representations of intensity and valence in human olfaction. *Nat Neurosci*. 2003:6(2):196–202. https://doi.org/10.1038/nn1001.

Apperly IA, Samson D, Chiavarino C, Humphreys GW. Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *J Cogn Neurosci*. 2004:16(10): 1773–1784. https://doi.org/10.1162/0898929042947928.

Aron A, Aron EN, Smollan D. Inclusion of other in the self scale and the structure of interpersonal closeness. *J Pers Soc Psychol*. 1992:63(4):596–612. https://doi.org/10.1037/0022-3514.63.4.596.

Beck, AT. *Depression: clinical, experimental, and theoretical aspects*. New York: Hoeber Medical Division, Harper and Row; 1967. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2237012/

Beck AT, Rush AJ, Shaw BF, Emery G. *Cognitive therapy of depression*. New York: Guilford Press; 1979.

Beck AT, Steer RA, Brown GK. *Beck Depression Inventory*. 2nd ed. New York: Psychological Corporation; 1996.

Bora E, Berk M. Theory of mind in major depressive disorder: a meta-analysis. *J Affect Disord*. 2016:191:49–55. https://doi.org/10.1016/j.jad.2015.11.023.

Brett M, Anton JL, Valabregue R, Poline J-B. Region of interest analysis using the MarsBar toolbox for SPM 99. *NeuroImage*. 2002:16(2):S497.

Brockmeyer T, Zimmermann J, Kulessa D, Hautzinger M, Bents H, Friederich HC, Herzog W, Backenstrass M. Me, myself, and I: self-referent word use as an indicator of self-focused attention in relation to depression and anxiety. *Front Psychol*. 2015:6:1564. https://doi.org/10.3389/fpsyg.2015.01564.

Bukowski H, Lamm C. Temporoparietal junction. In Zeigler-Hill V, Shackelford TK (Eds.), *Encyclopedia of personality and individual differences* (pp. 1–5). New York: Springer International Publishing. 2017. https://doi.org/10.1007/978-3-319-28099-8_863-1.

Cooley CH. *Human nature and the social order*. New York: Charles Scribner's sons; 1902.

Dainer-best J, Trujillo LT, Schnyer DM, Beevers CG. Sustained engagement of attention is associated with increased negative self-referent processing in major depressive disorder. *Biol Psychol*. 2017:129:231–241. https://doi.org/10.1016/j.biopsycho.2017.09.005.

D'Argembeau A, Ruby P, Collette F, Degueldre C, Balteau E, Luxen A, Maquet P, Salmon E. Distinct regions of the medial prefrontal cortex are associated with self-referential processing and perspective taking. *J Cogn Neurosci*. 2007:19(6):935–944. https://doi.org/10.1162/jocn.2007.19.6.935.

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

Denny BT, Kober H, Wager TD, Ochsner KN. A meta-analysis of functional neuroimaging studies of self- and other judgments reveals a spatial gradient for mentalizing in medial prefrontal cortex. *J Cogn Neurosci*. 2012:24(8):1742–1752. https://doi.org/10.1162/jocn_a_00233.

Disner SG, Shumake JD, Beevers CG. Self-referential schemas and attentional bias predict severity and naturalistic course of depression symptoms. *Cogn Emot*. 2017:31(4):632–644. https://doi.org/10.1080/02699931.2016.1146123.

Dugué L, Merriam EP, Heeger DJ, Carrasco M. Specific visual subregions of TPJ mediate reorienting of spatial attention. *Cereb Cortex*. 2018:28(7):2375–2390. https://doi.org/10.1093/cercor/bhx140.

Erle TM, Barth N, Topolinski S. Egocentrism in sub-clinical depression. *Cognit Emot*. 2019:33(6):1239–1248. https://doi.org/10.1080/02699931.2018.1552120.

Gweon H, Dodell-Feder D, Bedny M, Saxe R. Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Dev*. 2012:83(6):1853–1868. https://doi.org/10.1111/j.1467-8624.2012.01829.x.

Huang X, Zhang S. Desirability, meaningfulness and familiarity ratings of 562 personality-trait adjectives. *Psychol Sci (in Chinese)*. 1992:5:17–22

Jankowski KF, Moore WE, Merchant JS, Kahn LE, Pfeifer JH. But do you think I'm cool? Developmental differences in striatal recruitment during direct and reflected social self-evaluations. *Dev Cogn Neurosci*. 2014:8:40–54. https://doi.org/10.1016/j.dcn.2014.01.003.

Käsbauer AS, Mengotti P, Fink GR, Vossel S. Resting-state functional connectivity of the right temporoparietal junction relates to belief updating and reorienting during spatial attention. *J Cogn Neurosci*. 2020:32(6):1130–1141. https://doi.org/10.1162/jocn_a_01543.

Kensinger EA, Schacter DL. Processing emotional pictures and words: effects of valence and arousal. *Cogn Affect Behav Neurosci*. 2006:6(2):110–126. https://doi.org/10.3758/CABN.6.2.110.

Kiang M, Farzan F, Blumberger DM, Kutas M, McKinnon MC, Kansal V, Rajji TK, Daskalakis ZJ. Abnormal self-schema in semantic memory in major depressive disorder: evidence from event-related brain potentials. *Biol Psychol*. 2017:126:41–47. https://doi.org/10.1016/j.biopsycho.2017.04.003.

Krall SC, Rottschy C, Oberwelland E, Bzdok D, Fox PT, Eickhoff SB, Fink GR, Konrad K. The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Struct Funct*. 2015:220(2):587–604. https://doi.org/10.1007/s00429-014-0803-z.

Krall SC, Volz LJ, Oberwelland E, Grefkes C, Fink GR, Konrad K. The right temporoparietal junction in attention and social interaction: a transcranial magnetic stimulation study. *Hum Brain Mapp*. 2016:37(2):796–807. https://doi.org/10.1002/hbm.23068.

Lee L, Harkness KL, Sabbagh MA, Jacobson JA. Mental state decoding abilities in clinical depression. *J Affect Disord*. 2005:86(2):247–258

Lemogne C, Delaveau P, Freton M, Guionnet S, Fossati P. Medial prefrontal cortex and the self in major depression. *J Affect Disord*. 2012:136(1–2):e1–e11. https://doi.org/10.1016/j.jad.2010.11.034.

Lewis P, Critchley H, Rotshtein P, Dolan R. Neural correlates of processing valence and arousal in affective words. *Cereb Cortex*. 2006:17(3):742–748. https://doi.org/10.1093/cercor/bhk024.

Li Y, Li M, Wei D, Kong X, Du X, Hou X, Qiu J. Self-referential processing in unipolar depression: distinct roles of subregions of the medial prefrontal cortex. *Psychiatry Res Neuroimaging*. 2017:263:8–14. https://doi.org/10.1016/j.pscychresns.2017.02.008.

Liu P, Vandermeer MRJ, Joanisse MF, Barch DM, Dozois DJA, Hayden EP. Neural activity during self-referential processing in children at risk for depression. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 2020:5(4):429–437. https://doi.org/10.1016/j.bpsc.2019.12.012.

Lou Y, Lei Y, Mei Y, Leppänen PHT, Li H. Review of abnormal self-knowledge in major depressive disorder. *Front Psychiatry*. 2019:10:130. https://doi.org/10.3389/fpsyt.2019.00130.

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

Mahy CE, Moses LJ, Pfeifer JH. How and where: theory-of-mind in the brain. *Dev Cogn Neurosci*. 2014:9:68–81. https://doi.org/10.1016/j.dcn.2014.01.002.

Manstead ASR, Dosmukhambetova D, Shearn J, Clifton A. The influence of dysphoria and depression on mental state decoding. *J Soc Clin Psychol*. 2013:32(1):116–133. https://doi.org/10.1521/jscp.2013.32.1.116.

Markus H. Self-schemata and processing information about the self. *J Pers Soc Psychol*. 1977:35(2):63–78. https://doi.org/10.1037/0022-3514.35.2.63.

Mead GH. *Mind, self and society*. Chicago: University of Chicago Press; 1934.

Nejad AB, Fossati P, Lemogne C. Self-referential processing, rumination, and cortical midline structures in major depression. *Front Hum Neurosci*. 2013:7:666. https://doi.org/10.3389/fnhum.2013.00666.

Nejati V, Majdi R, Salehinejad MA. The role of dorsolateral and ventromedial prefrontal cortex in the processing of emotional dimensions. *Sci Rep*. 2021:11(1):1971. https://doi.org/10.1038/s41598-021-81454-7.

Nestor BA, Sutherland S, Garber J. Theory of mind performance in depression: a meta-analysis. *J Affect Disord*. 2022:303:233–244. https://doi.org/10.1016/j.jad.2022.02.028.

Ochsner KN, Beer JS, Robertson ER, Cooper JC, Gabrieli JD, Kihsltrom JF, D'Esposito M. The neural correlates of direct and reflected self-knowledge. *NeuroImage*. 2005:28(4):797–814. https://doi.org/10.1016/j.neuroimage.2005.06.069.

Pfeifer JH, Masten CL, Borofsky LA, Dapretto M, Fuligni AJ, Lieberman MD. Neural correlates of direct and reflected self-appraisals in adolescents and adults: when social perspective-taking informs self-perception. *Child Dev*. 2009:80(4):1016–1038. https://doi.org/10.1111/j.1467-8624.2009.01314.x.

Pfeifer JH, Mahy CEV, Merchant JS, Chen C, Masten CL, Fuligni AJ, Lieberman MD, Lessard J, Dong Q, Chen C. Neural systems for reflected and direct self-appraisals in Chinese young adults: exploring the role of the temporal-parietal junction. *Cultur Divers Ethnic Minor Psychol*. 2017:23(1):45–58. https://doi.org/10.1037/cdp0000122.

Richardson H, Saxe R. Chapter 21 - early signatures of and developmental change in brain regions for theory of mind. In Rubenstein P, Rakic B, Chen B, Kwan KY (Eds.), *Neural circuit and cognitive development*. 2nd ed. (pp. 467–484). Cambridge: Academic Press. 2020 https://doi.org/10.1016/B978-0-12-814411-4.00021-4.

Samson D, Apperly IA, Chiavarino C, Humphreys GW. Left temporoparietal junction is necessary for representing someone else's belief. *Nat Neurosci*. 2004:7(5):499–500. https://doi.org/10.1038/nn1223.

Saxe R. The right temporo-parietal junction: a specific brain region for thinking about thoughts. In: Leslie A, German T (Eds.), *Handbook of theory of mind*. (pp. 1–35). London (UK): Taylor and Francis; 2010. pp. 1–35

Saxe R, Kanwisher N. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *NeuroImage*. 2003:19(4):1835–1842. https://doi.org/10.1016/s1053-8119(03)00230-1.

Saxe R, Powell LJ. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol Sci*. 2006:17(8): 692–699. https://doi.org/10.1111/j.1467-9280.2006.01768.x

Schurz M, Radua J, Aichhorn M, Richlan F, Perner J. Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev*. 2014:42:9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009.

Shiota S, Okamoto Y, Okada G, Takagaki K, Takamura M, Mori A, Yokoyama S, Nishiyama Y, Jinnin R, Hashimoto RI, et al. Effects of behavioural activation on the neural basis of other perspective self-referential processing in subthreshold depression: a functional magnetic resonance imaging study. *Psychol Med*. 2017:47(5): 877–888. https://doi.org/10.1017/S0033291716002956.

Steinbeis N. The role of self-other distinction in understanding others' mental and emotional states: neurocognitive mechanisms in children and adults. *Philos Trans R Soc B*. 2016:371(1686):20150074. https://doi.org/10.1098/rstb.2015.0074.

Tice DM, Wallace HM. *The reflected self: creating yourself as (you think) others see you*. New York: Guilford; 2005.

Van der Cruijsen R, Peters S, Zoetendaal KPM, Pfeifer JH, Crone EA. Direct and reflected self-concept show increasing similarity across adolescence: a functional neuroimaging study. *Neuropsychologia*. 2019:129:407–417

Veroude K, Jolles J, Croiset G, Krabbendam L. Sex differences in the neural bases of social appraisals. *Soc Cogn Affect Neurosci*. 2014:9(4):513–519. https://doi.org/10.1093/scan/nst015.

Viinikainen M, Kätsyri J, Sams M. Representation of perceived sound valence in the human brain. *Hum Brain Mapp*. 2012:33(10): 2295–2305. https://doi.org/10.1002/hbm.21362.

Wagner G, Schachtzabel C, Peikert G, Bär K-J. The neural basis of the abnormal self-referential processing and its impact on cognitive control in depressed patients. *Hum Brain Mapp*. 2015:36(7): 2781–2794. https://doi.org/10.1002/hbm.22807.

Xu X, Chen Y, Zhao Y, Yang J. The neural evidence of the reflected appraisal process as a main path for learning about one's self-knowledge of personality. *Biol Psychol*. 2015:110:85–90. https://doi.org/10.1016/j.biopsycho.2015.06.017.

Yoshimura S, Okamoto Y, Onoda K, Matsunaga M, Ueda K, Suzuki S, Shigetoyamawaki. Rostral anterior cingulate cortex activity mediates the relationship between the depressive symptoms and the medial prefrontal cortex activity. *J Affect Disord*. 2010:122(1–2): 76–85. https://doi.org/10.1016/j.jad.2009.06.017.

III

BRAIN RESPONSES OF DYSPHORIC AND CONTROL PARTICIPANTS DURING A SELF-ESTEEM IMPLICIT ASSOCIATION TEST

by

Yixue Lou, Yi Lei, Piia S. Astikainen, Weiwei Peng, Suzanne C.S.A. Otieno,
Paavo H.T. Leppänen, 2021

Psychophysiology, 58(4), e13768

Available online: https://doi.org/10.1111/psyp.13768

Received: 7 April 2020 | Revised: 29 September 2020 | Accepted: 23 December 2020

DOI: 10.1111/psyp.13768

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

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

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

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

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

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

Correspondence
Yi Lei, Institute for Brain and Psychological Sciences, Sichuan Normal University, Jing'an Road 5, Jinjiang District, Chengdu 610066, China.
Email: leiyi821@vip.sina.com

Funding information
Major Program of Guangdong, Grant/Award Number: 2016KZDXM009; Shenzhen Basic Research Scheme, Grant/Award Number: JCYJ20150729104249783; Innovative Team Program in Higher Education of Guangdong, Grant/Award Number: 2015KCXTD009; National Natural Science Foundation of China, Grant/Award Number: 31571153, 31871130 and 31671150; Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions, Grant/Award Number: 2019SHIBS0003; Shenzhen Peacock Plan, Grant/Award Number: KQTD2015033016104926

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 400–1,000 ms poststimulus 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 toward 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, event-related potentials (ERPs), implicit association test (IAT), implicit self-esteem, late positive component (LPC)

1 | INTRODUCTION

1.1 | Research background

Self-esteem refers to a person's overall attitude toward 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 (IAT);

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 et al., 2019). However, only few studies have used behavioral paradigms to investigate implicit self-esteem in depression (Leeuwis et al., 2015; Roberts et al., 2015; Smeijers et al., 2017; van Randenborgh et al., 2016).

The IAT (Greenwald et al., 1998), which is a categorization task in nature, is the most commonly used paradigm for measuring a person's implicit attitude (Bosson et al., 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 (Chen et al., 2014; Mezulis et al., 2004; Pahl & Eiser, 2005; Watson et al., 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 nondepressed controls have longer RTs in the incongruent relative to the congruent condition, but this difference between RTs in incongruent versus 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 nondepressed controls. These results have been interpreted to indicate lowered implicit self-esteem among individuals with depressive symptoms. Franck et al. (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 et al., 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 et al., 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 et al., 2005). Larger LPC amplitude has been observed in emotional versus neutral stimuli (Schupp et al., 2006), in high-arousal versus low-arousal stimuli (Rozenkrants & Polich, 2008), and in rare versus 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 poststimulus. 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 to 1,000 ms poststimulus) 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. 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 (Chen et al., 2018; 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 nondepressed healthy participants (Wu et al., 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 et al., 2014; Wu et al., 2016; Xiao et al., 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 nondysphoric 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 et al., 2014); occipital P1: (Fleischhauer et al., 2014); occipital N170: (Ibáñez et al., 2010); P2: (Grundy et al., 2015; Healy et al., 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 poststimulus 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 S1.

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 et al., 1996). Individuals with scores distributed in the top 5% 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 5% 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 preclinical 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

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)

Abbreviations: BDI-II, Beck depression inventory-II; *N*, number; *M* ± *SD*, means ± standard deviations.

estimated based on a standard medium effect size ($f = 0.25$) (Cohen, 1988). Power analysis, conducted with G*Power 3 (Faul et al., 2009), showed a requirement of 27 participants in each group (dysphoric and control) with a statistical power of $(1 - \beta) = .95$ and a significance level of $\alpha = .05$. 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.

2.2 | Experimental protocol

2.2.1 | Implicit association test

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, etc.). Press the *J* key when you see a *not-me*-related word (e.g., other, his, they, etc.). 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, etc.). Press the *J* key when you see a *negative* describing word (e.g., fool, coward, dishonest, etc.). 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, etc.) OR *positive* describing words (e.g., smart, brave, honest, etc.). Press the *J* key when you see either *not-me* related words (e.g., other, his, they, etc.) OR *negative* describing words (e.g., fool, coward, dishonest, etc.).

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 1,000 and 2,000 ms. A stimulus word was then presented for 1,000 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.

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 S2, Table S2.1). Most of these words were selected from a pool of 562

(a)

Phase sequence	1 Practice (10 trials)	2 Practice (10 trials)	3 Practice (20 trials)	4 Data-collection (320 trials)
Task instruction	Sorting *me* vs. *not-me* words	Sorting *positive* vs. *negative* words	Sorting mixed words from phase 1 and 2	Sorting stimulus words as learnt in phase 3
Self-positivity block	F me not-me J	F positive negative J	F me F positive not-me J negative J	F me F positive not-me J negative J
Self-negativity block	F me not-me J	positive J F negative	F me positive J not-me J F negative	F me positive J not-me J F negative

(b)

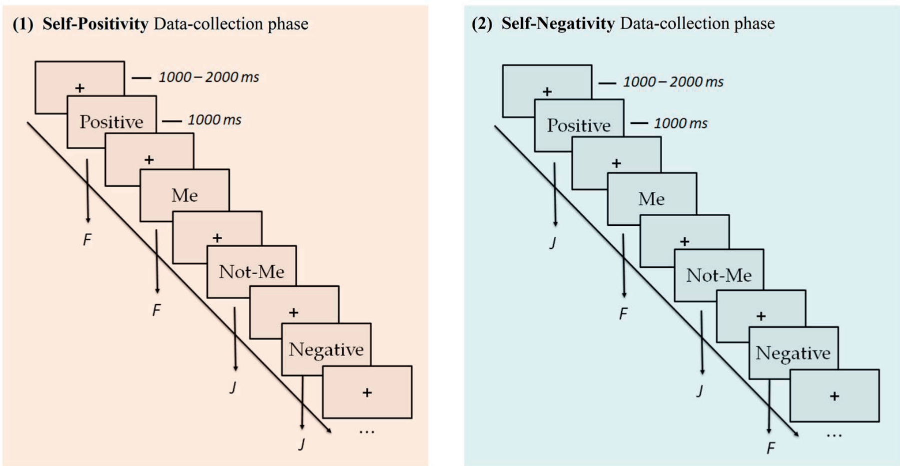

FIGURE 1 (a) Illustration of the four phases during the self-positivity and the self-negativity blocks. The labeled black dots indicate the correct responses in each phase. The order of the two blocks and the assignment of the *F* key and *J* key were both counter-balanced between subjects. (b). Illustration 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 S2, Table S2.1)

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 et al. (2008). According to our pilot preexperiment ($N = 25$), 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°.

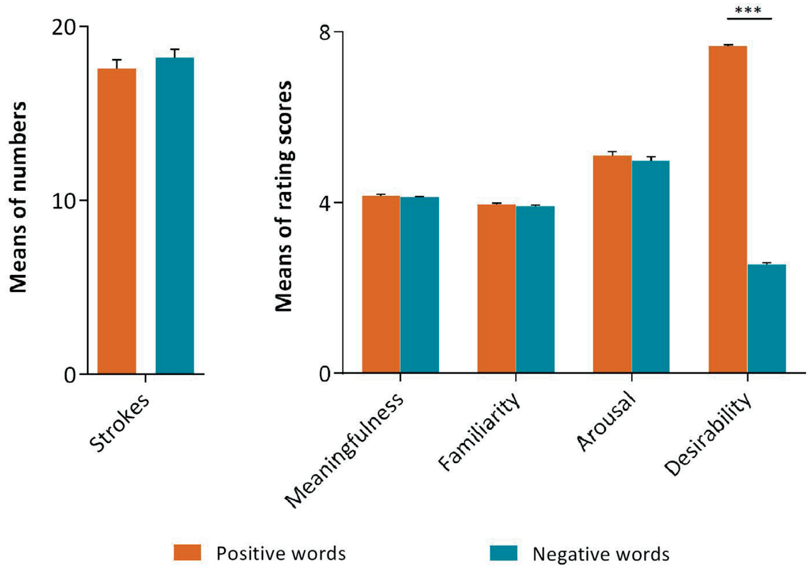

FIGURE 2 Means of the characters' stroke numbers, meaningfulness, familiarity, arousal, and desirability as rated in the pilot preexperiment for positive and negative stimulus words. Error bars represent standard errors. ***$p < .001$

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 midline 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 interelectrode impedances were maintained below 5 kΩ for on-line recording.

During off-line preprocessing, EEG signals were re-referenced to the average signal at the mastoid electrodes (Luck, 2005) and a low-pass filter was applied (30 Hz; 24 dB/octave). A semiautomatic 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 prestimulus baseline and a poststimulus duration of 1,000 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 et al., 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 toward 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 poststimulus 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 600–1,000 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 S3.

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, $p < .001$, $\eta_p^2 = .77$]. 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, $p < .001$, $\eta_p^2 = .49$]. The accuracy was higher in the self-positivity condition (M = .96, SD = .02) than that in the self-negativity condition (M = .93, SD = .04; see Figure 3a-2).

D-*score (the IAT effect for reaction time)*

No significant group difference was observed for *D*-scores (t (56) = .13, p = .90, Cohen's d = .004; see Figure 3a-3). The *D*-scores were significantly higher than zero for both the dysphoric group (M = .45, SD = .28; t (27) = 8.60, $p < .001$; Cohen's d = 1.62) and the control group (M = .46, SD = .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 = .005; Cohen's d = .76). The dysphoric

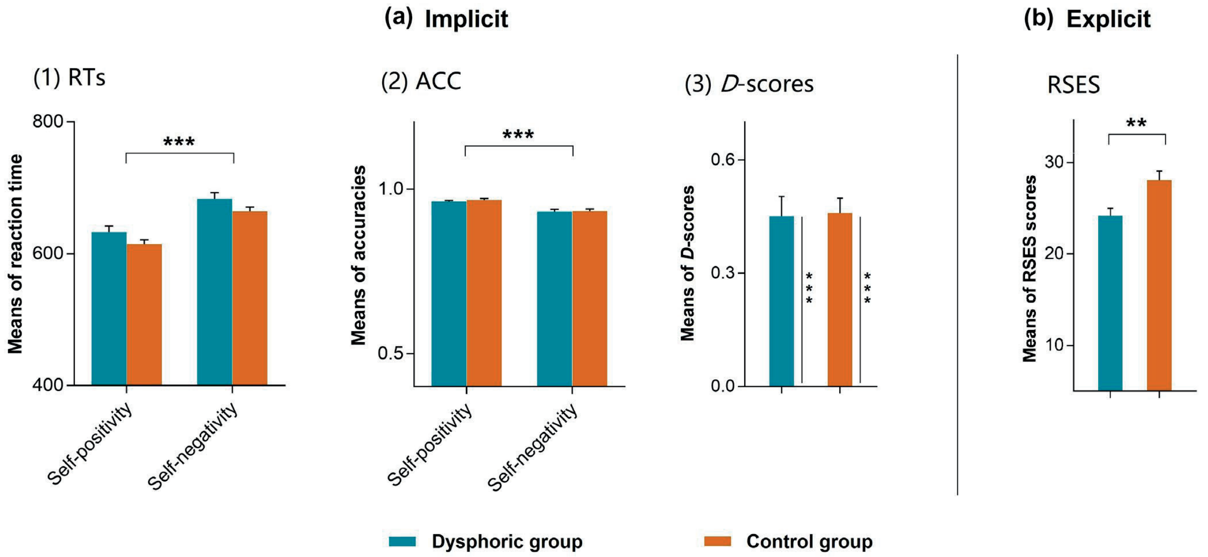

FIGURE 3 Indices of (a). Implicit self-esteem: (1). Means of reaction times (RTs), (2). Means of accuracy (ACC), and (3) Means of the *D*-score (calculated as the division of the difference of self-negativity RT means minus self-positivity RT means by the standard deviation of all RTs in the two conditions)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

group (M = 24.18, SD = 4.27) exhibited lower RSES scores than the control group (M = 28.03, SD = 5.71; see Figure 3b).

3.2 | ERP results

3.2.1 | LPC (300–400 ms)

The results showed neither a significant effect of self-association [F (1, 56) = .57, p = 1.00, η_p^2 = .01] nor group [F (1, 56) = 1.53, p = .66, η_p^2 = .03]. The self-association × group interaction was also nonsignificant [F (1, 56) = 4.36, p = .12, η_p^2 = .07].

3.2.2 | LPC (400–600 ms)

There was a significant interaction effect of self-association and group [F (1, 56) = 18.58, p < .001, η_p^2 = .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, η_p^2 = .33] and the dysphoric group [F (1, 28) = 5.18, p = .03, η_p^2 = .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, η_p^2 = .07] but not in the self-positivity condition [F (1, 56) = .09, p = .77, η_p^2 = .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).

3.2.3 | LPC (600–1,000 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, η_p^2 = .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, η_p^2 = .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) = .51, p = .48, η_p^2 = .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 = .02, η_p^2 = .09], but not in the self-positivity condition [F (1, 56) = .28, p = .60, η_p^2 = .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).

3.3 | Correlation analysis

According to these results, opposite LPC response patterns were observed for the dysphoric group and the control group

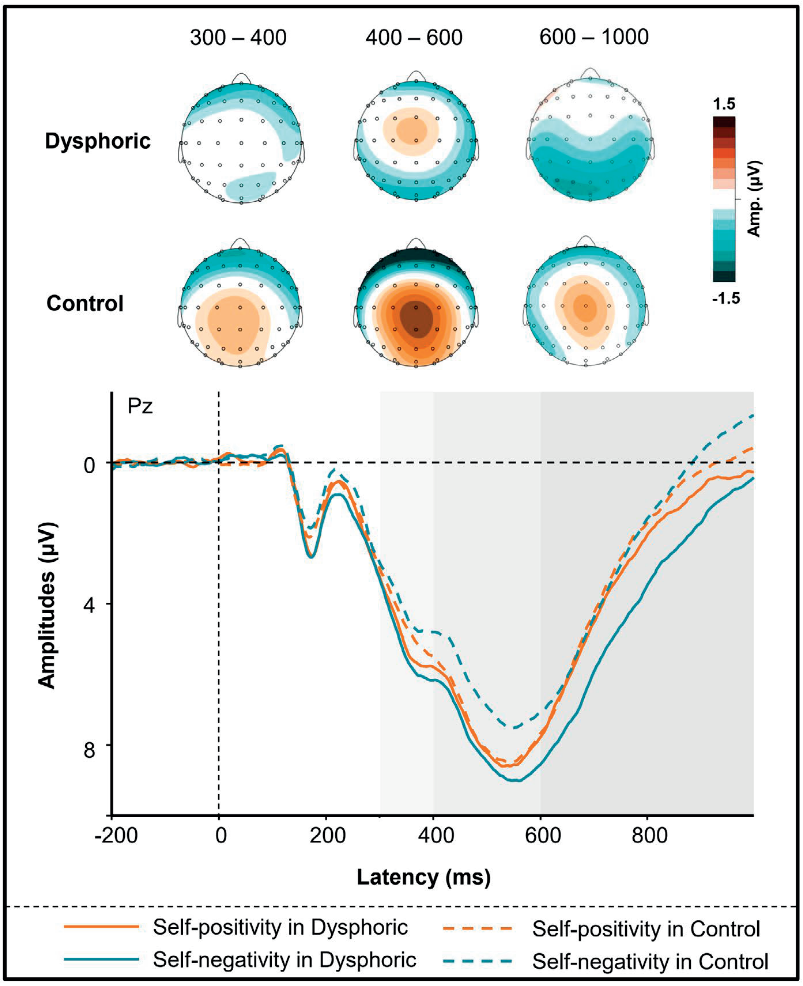

FIGURE 4 Illustration of LPC waveforms within 300–1,000 ms (presented separately for time windows of 300–400 ms, 400–600 ms, and 600–1,000 ms) in the dysphoric group (solid lines) and the control group (dashed lines) during the self-positivity (orange lines) and the self-negativity (blue lines) conditions. 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

within both 400–600 ms and 600–1,000 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 600–1,000 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) = −.32, p = .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) = −.03, p = .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) = .29, p = .03) and 600–1000 ms (r (58) = .28, p = .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.

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

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 600–1,000 ms time windows

				LPC Amp.	LPC Amp.
Variables	**BDI-II**	**RSES**	***D*-score**	***Self-positivity***	***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***	–
600–1,000 ms					
LPC Amp. *Self-positivity*	0.03	−0.08	0.01	–	
LPC Amp. *Self-negativity*	0.28*	−0.14	0.19	0.86***	–

Abbreviations: Amp., amplitude; BDI-II, Beck depressive inventory-II; IAT, implict association task; LPC, late positive component; RSES, Rosenberg self-esteem scale.
*$p < .05$; ***$p < .001$.

performing IAT. As we expected, the significant group-related differences were observed during 400–1,000 ms poststimulus (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–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 (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 1,000 ms poststimulus 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 et al., 1979; Clak et al., 1999) and plays an important role in the development, maintenance, and relapse of depressive disorder (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 et al., 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 = .96, SD = .02) and the self-negativity condition (M = .93, SD = .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 preclinical 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.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China [Grant NO. 31871130, 31571153, and 31671150], the Innovative Team Program in Higher Education of Guangdong, China [Grant NO. 2015KCXTD009], the Major Program of Guangdong, China [Grant NO. 2016KZDXM009], the Shenzhen Basic Research Scheme [Grant NO. JCYJ20150729104249783], and the Shenzhen-Hong Kong Institute of Brain Science-Shenzhen Fundamental Research Institutions [Grant No. 2019SHIBS0003], and the Shenzhen Peacock Plan [Grand No: KQTD2015033016104926].

AUTHOR CONTRIBUTION

Yixue Lou: Conceptualization; Data curation; Investigation; Methodology; Software; Visualization; Writing-original draft; Writing-review & editing. **Yi Lei:** Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing-review & editing. **Piia Astikainen:** Methodology; Writing-review & editing. **Weiwei Peng:** Writing-review & editing. **Suzanne Otieno:** Writing-review & editing. **Paavo HT Leppänen:** Methodology; Supervision; Writing-review & editing.

ORCID

Yixue Lou https://orcid.org/0000-0002-0261-8344
Yi Lei https://orcid.org/0000-0002-8874-923X
Weiwei Peng https://orcid.org/0000-0002-4600-0036

REFERENCES

Beck, A. T. (1967). *Depression: Clinical, experimental, and theoretical aspects*. Hoeber Medical Division, Harper and Row.

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

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory* (2nd ed.). 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., Zhu, H., & 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. 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*. 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*. New Jersey Lawrence Erlbaum Associates.

Delaveau, P., Jabourian, M., Lemogne, C., Allaïli, N., Choucha, W., Girault, N., Lehericy, S., Laredo, J., & 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., & Van den Abbeele, 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*, *5*, 17–22.(in Chinese)

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., de Jong, P. J., 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. *Frontiers in 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 S. J. Luck (Ed.), *An introduction to the event-related potential technique* (Vol. *1*, pp. 2–48). MIT Press.

Kok A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38* (3), 557–577. https://doi.org/10.1017/s0048577201990559.

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), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

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

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

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

Rozenkrants B., Polich J. (2008). Affective ERP processing in a visual oddball task: Arousal, valence, and gender. *Clinical Neurophysiology*, *119* (10), 2260–2265. http://doi.org/10.1016/j.clinph.2008.07.213

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), 2786–2791. https://doi.org/10.1073/pnas.1515828113

Schupp, H. T., Flaisch, T., Stockburger, J., & Junghöfer, M. (2006). Emotion and attention: event-related brain potential studies. *Progress in Brain Research*, *156*, 31–51.

Smeijers, D., Vrijsen, J. N., van Oostrom, I., 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 Randenborgh, A., 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

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

Verleger R., Jaśkowski P., Wascher E. (2005). Evidence for an Integrative Role of P3b in Linking Reaction to Perception. *Journal of Psychophysiology*, *19* (3), 165–181. http://doi.org/10.1027/0269-8803.19.3.165

Verleger R., Śmigasiewicz K. (2016). Do rare stimuli evoke large P3s by being unexpected? A comparison of oddball effects between standard-oddball and prediction-oddball tasks. *Advances in Cognitive Psychology*, *12* (2), 88–104. https://doi.org/10.5709/acp-0189-9

Vianello, M., & Bar-Anan, Y. (2020). Can the implicit association test measure automatic judgment? The validation continues. *Perspectives on Psychological Science*, Advance online publication. 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*, *22*(8), 608–612.(in Chinese)

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

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), 468–476. https://doi.org/10.1093/scan/nsq065

Wu, L., Cai, H., Gu, R., Luo, Y. L. L., Zhang, J., Yang, J., Shi, Y., & 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*, Advance online publication. https://doi.org/10.1179/174313209x431138

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

Data S1. Additional early components

Table S1 .1. Means (M) and Standard deviations (SD) of the ERP amplitudes in each analysis electrodes for N1, P1, N170, and P2 components

Figure S1 .1. Illustrations for (a). selection of the analysis electrodes for frontal N1 (gray clusters), occipital P1 (blue cluster), temporo-occipital N170 (yellow cluster), and fronto-central distributed P2 (red cluster), and (b). the within-subject and the between-subject factors that we included in the repeated measures ANOVA analysis of each component.

Figure S1 .2. Illustration of the ERP waveforms for N1, P1, N170, and P2 components. For each component, both left and right sites were illustrated. The solid lines indicated the dysphoric group, whereas the dash lines indicated the control group. The orange lines indicated the self-positivity condition, while the blue lines indicated the self-negativity condition.

Data S2. Illustration of stimulus words

Table S2 .1. Illustration of the word stimuli used in the current study

Data S3. Topographic effects of the late positive component (LPC)

How to cite this article: Lou Y, Lei Y, Astikainen P, Peng W, Otieno S, Leppänen PHT. Brain responses of dysphoric and control participants during a self-esteem implicit association test. *Psychophysiology*. 2021;58:e13768. https://doi.org/10.1111/psyp.13768