# Tracking the implicit self using event-related potentials

Yvonne Egenolf · Maria Stein · Thomas Koenig ·
Martin Grosse Holtforth · Thomas Dierks ·
Franz Caspar

**Abstract** Negative biases in implicit self-evaluation are thought to be detrimental to subjective well-being and have been linked to various psychological disorders, including depression. An understanding of the neural processes underlying implicit self-evaluation in healthy subjects could provide a basis for the investigation of negative biases in depressed patients, the development of differential psychotherapeutic interventions, and the estimation of relapse risk in remitted patients. We thus studied the brain processes linked to implicit self-evaluation in 25 healthy subjects using event-related potential (ERP) recording during a self-relevant Implicit Association Test (sIAT). Consistent with a positive implicit self-evaluation in healthy subjects, they responded significantly faster to the congruent (self–positive mapping) than to the incongruent sIAT condition (self–negative mapping). Our main finding was a topographical ERP difference in a time window between 600 and 700 ms, whereas no significant differences between congruent and incongruent conditions were observed in earlier time windows. This suggests that biases in implicit self-evaluation are reflected only indirectly, in the additional recruitment of control processes needed to override the positive implicit self-evaluation of healthy subjects in the incongruent sIAT condition. Brain activations linked to these control processes can thus serve as an indirect measure for estimating biases in implicit self-evaluation. The sIAT paradigm, combined with ERP, could therefore permit the tracking of the neural processes underlying implicit self-evaluation in depressed patients during psychotherapy.

**Keywords** Cognitive control · Implicit self-esteem · Depression · ERP · Implicit Association Test (IAT)

How we see ourselves contributes markedly to our emotional experience of the events in our surroundings. Accordingly, previous research has linked high self-esteem to happiness, better psychological adjustment (Baumeister, Campbell, Krueger, & Vohs, 2003), and higher positive and lower negative affect (Orth, Robins, & Widaman, 2012). Low self-esteem, in contrast, has been associated with various psychological symptoms, such as depression (Franck, De Raedt, Dereu, & Van den Abbeele, 2007; Orth, Robins, Trzesniewski, Maes, & Schmitt, 2009), social anxiety (de Jong, 2002; Ginsburg, La Greca, & Silverman, 1998; Tanner, Stopa, & De Houwer, 2006), and bulimia (Cockerham, Stopa, Bell, & Gregg, 2009; Vohs et al., 2001). Furthermore, longitudinal studies suggest that low self-esteem predicts depression, and is therefore thought to be a key factor in the development and maintenance of depression (Ormel, Oldehinkel, & Vollebergh, 2004; Sowislo & Orth, 2013). According to cognitive theories (Beck, Rush, Shaw, & Emery, 1979; Beck, 1995; Ingram, 1984; Teasdale, 1988), depression is characterized by negatively biased information processing grounded in dysfunctional self-schemata, and by related negative assumptions about oneself that are assumed to originate from negative emotional experiences in the past (Beck, 1967). These self-schemata are thought to be only partly accessible to conscious reflection (Bosson, Swann, & Pennebaker, 2000), and thus at best are partially accessible by explicit self-

Y. Egenolf (✉) · M. Stein · F. Caspar
Department for Clinical Psychology and Psychotherapy,
University of Bern, Gesellschaftsstrasse 49,
3012 Bern, Switzerland
e-mail: yvonne.egenolf@psy.unibe.ch

M. Stein · T. Koenig · T. Dierks
Department of Psychiatric Neurophysiology,
University Hospital of Psychiatry, Bern, Switzerland

M. Grosse Holtforth
Department for Clinical Psychology, University of Zürich,
Zürich, Switzerland

reports such as questionnaires. Therefore, implicit methods are an attractive approach to assess implicit self-evaluations as part of subjects' individual self-schemata.

The Implicit Association Test containing self-related stimuli (sIAT) has been developed for measuring biases in implicit self-evaluation (Greenwald & Farnham, 2000). In contrast to self-reports based on conscious, and therefore explicit self-evaluations, the sIAT is a computerized categorization task that relies on differences in reaction times (RTs) to measure the individual implicit association strength between two concept categories (self/not self) and two attribute categories (positive attributes/negative attributes). The rationale behind the sIAT is that the mapping of two closely associated categories to the same response key allows faster responses than does the mapping of two weakly associated categories. Healthy subjects usually display faster responses when self and positive attributes are mapped to the same button. Therefore, we refer to self–positive mapping as the *congruent* condition, whereas self–negative mappings represent the *incongruent* condition.

To date, different explanations for IAT RT differences have been proposed. The original explanation posited was that the congruent condition is "easier" to perform than the incongruent condition because concepts that are mapped together exhibit stronger associations or semantic connections (Greenwald, McGhee, & Schwartz, 1998). A closely related explanation assumes that the familiarity of a concept makes it more or less salient, and that the observed RT differences are due to a salience asymmetry between congruent and incongruent conditions (Rothermund & Wentura, 2004). Another set of potential explanations has focused on additional cognitive resources that are particularly needed in the incongruent condition. For example, one account suggests that more extensive evaluation processes in the incongruent condition are responsible for the observed IAT RT effect (Brendl, Markman, & Messner, 2001). Similarly, IAT RT differences have been linked to differential costs of task switching in the congruent and incongruent conditions (Mierke & Klauer, 2001). An investigation of the brain processes observed during sIAT performance could help to delimit the cognitive processes underlying the observed behavioral sIAT effect.

Previous behavioral studies using the sIAT have repeatedly demonstrated positive implicit self-evaluations in healthy subjects (De Raedt, Schacht, Franck, & De Houwer, 2006; Franck et al., 2007; Greenwald & Farnham, 2000). In addition, recent studies, including a longitudinal multicenter study, have shown that depressed individuals exhibit more negative implicit self-evaluations than control subjects (Glashouwer & de Jong, 2010; Risch et al., 2010). As earlier studies reported no differences between depressed and healthy subjects (De Raedt, et al., 2006; Franck, De Raedt, & De Houwer, 2008; Gemar, Segal, Sagrati, & Kennedy, 2001), however, some variability is present in the literature.

Taken together, the available evidence suggests that negative implicit self-evaluation could be an important aspect of depression that can be tracked using the sIAT paradigm.

To our knowledge, the sIAT itself has not been investigated using neuroimaging methods to date. However, several event-related potential (ERP) and functional magnetic resonance imaging (fMRI) studies have successfully delineated brain activations in the context of other, not self-referential IAT paradigms investigating, for example, implicit concept associations (Chee, Sriram, Soon, & Lee, 2000), beliefs about race and gender (Knutson, Mah, Manly, & Grafman, 2007), ingroup/outgroup attitudes (Hurtado, Haye, González, Manes, & Ibáñez, 2009; Ibáñez et al., 2010), and implicit moral attitudes (Luo et al., 2006). These studies have mainly focused on activation differences between congruent and incongruent IAT conditions. ERP studies have commonly identified differences in both late (Coates & Campbell, 2010; Hurtado, et al., 2009; Williams & Themanson, 2011) and, to a lesser degree, early (Coates & Campbell, 2010; Forbes et al., 2012; Ibáñez et al., 2010) ERP components as a function of IAT condition. fMRI studies have emphasized an important role of activations in the dorsolateral and ventrolateral prefrontal cortex, as well as in the anterior cingulate cortex, in the control processes that are needed to override the implicit associations in order to correctly perform the IAT on incongruent trials (Chee et al., 2000; Knutson et al., 2007; Luo et al., 2006).

Here, our aim was to characterize the temporal dynamic of brain activity involved in sIAT performance using electroencephalography (EEG) recording. We expected that the initial processing stages (ERPs < 120 ms) would not differ across the sIAT conditions, since the task conditions employed identical stimuli, and therefore did not differ with regard to stimulus characteristics, perceptual processing, or overall emotional valence (Vogel & Luck, 2000). In an intermediate time window around 200 to 400 ms poststimulus, we expected to find enhanced ERP amplitudes for the congruent sIAT condition, on the basis of previous ERP findings using self-relevant stimulus materials. For example, enhanced intermediate ERP components were reported during viewing of one's own, as compared with other, familiar or unfamiliar faces (Keyes, Brady, Reilly, & Foxe, 2010), as well as during processing of sentences referring to the self rather than to others (Fields & Kuperberg, 2012). Our expectation was based on the notion that healthy subjects exhibit positive implicit self-evaluations, reflected in a close association of positive—rather than negative—attributes with the self. In a late time window, from 400 ms poststimulus onward, we expected to find variations in the electrical field topography between congruent and incongruent sIAT conditions, mainly on the basis of previous fMRI results showing that additional brain regions are recruited during the performance of incongruent IAT conditions. These additional activations are

thought to be involved in the control processes necessary for overriding implicit associations and generating the required behavioral response.

We supplemented these analyses by applying the sLORETA model for source localization, in an effort to estimate brain regions that might underlie observed amplitude or topographic ERP modulations. The sLORETA model assumes that the voltage changes between adjacent voxels are gradual, and on this basis selects the solution of source magnitudes that is maximally smooth (Pascual-Marqui, 2002). It is important to note that EEG source localization is an ill-posed inverse problem that requires additional assumptions and provides more or less likely approximations, but no provable evidence about source localization. Furthermore, sLORETA is only appropriate for estimating the center of an area of activation, but not for assessing or quantifying the extent of the activated area (Luck, 2005). Nevertheless, this approach offers the opportunity to generate hypotheses about brain regions potentially involved in sIAT performance on the basis of an explicit model and to link our ERP data to the existing fMRI literature.

In an intermediate time window (around 200 ms poststimulus), we expected sources to fall mainly into cortical midline regions that have been linked to self-referential processing (for a review, see Amodio & Frith, 2006; D'Argembeau & Salmon, 2012; Northoff & Bermpohl, 2004; van der Meer, Costafreda, Aleman, & David, 2010), notably the medial prefrontal cortex. Previous neuroimaging studies of self-evaluation processing have generally focused on the neural correlates of explicit self-reflection, and therefore have used paradigms based on explicit evaluations, in which subjects were asked to reflect on their own attitudes, abilities, or personal traits (e.g., Fossati et al., 2003; Johnson et al., 2002; Schmitz, Kawahara-Baccus, & Johnson, 2004). In contrast, the self-evaluations during the sIAT paradigm used in our study were made implicitly, and might well involve distinct neural processes. However, the few studies that have directly compared implicit and explicit evaluations of self-relevance (Moran, Heatherton, & Kelley, 2009) and self-referent evaluations (Rameson, Satpute, & Lieberman, 2010) have observed the recruitment of overlapping cortical midline structures in both implicit and explicit conditions.

In a late time window (400 ms onward), we expected sources to be located mainly in brain regions that have been linked to control processes recruited during incongruent sIAT trials as well as conflict monitoring, most notably the dorso- and ventrolateral prefrontal and the anterior cingulate cortex (Chee et al., 2000; Luo et al., 2006).

Elucidating the activations during sIAT performance in healthy subjects can provide information on the neural processes and networks involved in implicit self-evaluations. This can serve as a baseline for studying alterations in implicit self-evaluation in depressed patients, and generate hypotheses on the development and maintenance of depression. It could further be used to track particular aspects of neural activations during the course of psychotherapy—for example, allowing the identification of patients that are at risk of relapse, despite having reduced depressive symptoms following psychotherapy.

## Materials and method

### Subjects

Data were obtained from 26 healthy subjects. All were native German speakers, and two of the subjects were raised bilingually. One subject had to be excluded from further analysis because of poor quality of the EEG recordings, leaving a sample of 14 female and 11 male subjects, whose mean age was 25 years ($SD = 5.4$, range = 21–48). Of these 25 subjects, 22 were right handed. All subjects were initially screened for depression or other psychiatric disorders using the Structured Clinical Interview for DSM-IV Axis I Disorders, and none of them reported a current psychological disorder or a history of depression. Afterward, they completed a demographic information questionnaire, the Beck Depression Inventory (BDI-II; Beck, Steer, & Brown, 2006), and the Rosenberg Self-Esteem Scale (RSES), which is a widely used ten-item self-report scale measuring global self-esteem (Collani & Herzberg, 2003).

The study was approved by the cantonal ethics committee in Bern, Switzerland, and all subjects provided informed consent before participating.

### Task and stimuli

In the present study, we used a German version of the well-established self-esteem Implicit Association Test (sIAT; after Greenwald & Farnham, 2000).

As in other studies, we used self- and not-self-related pronouns that have been commonly used as concept categories (Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Risch et al., 2010) instead of individually generated self- and not-self-related words, in order to avoid idiosyncratic confounds due to semantic associations (Steffens, Kirschbaum, & Glados, 2008). The lists of positive and negative attributes consisted of ten personal attributes, each selected from the item pool used by Gemar et al. (2001; see the Appendix for all of the sIAT stimuli). All stimuli appeared in the middle of a computer screen in black letters against a white background, and subjects were asked to assign the stimuli as quickly as possible by pressing one of two response keys. The rules of category–response mappings changed from block to block, and a reminder was presented throughout the block in the upper left and right hand corners of the screen. The reminder consisted of the currently active mapping rule (in one corner

["self/positive" or "self/negative"] and in the opposite corner ["others/negative" or "others/positive"]). In the congruent block, subjects had to press the same response button for stimuli referring to "self" and "positive attributes," whereas the other response button was required for "not-self" and "negative attributes." In the incongruent block, "self" and "negative attributes" were assigned to one response button, whereas the other response button had to be pressed for "not self" and "positive attributes." Subjects were not required to correct their incorrect responses, but they received feedback via a red "X" in the middle of the screen.

Before performing the sIAT in combination with EEG measurement, each subject completed a short practice IAT (pIAT) composed of stimuli with minimal self-relevance and emotional valence (farm/traffic, animals/vehicles), so as to become familiar with the structure and the response modalities of the IAT task. The practice IAT contained only four shortened blocks (five trials each), composed of two discrimination blocks, which primed the button mappings for the following congruent (farm–animals, traffic–vehicles) and incongruent (farm–vehicles, traffic–animals) test blocks.

To minimize possible serial-order effects on the sIAT measures, we presented blocks of the congruent (C) condition and the incongruent (I) condition of the sIAT two times each, so that half of the subjects performed blocks in the order "CIIC," whereas the other half performed blocks in "ICCI" order. Thus, sIAT applied here contained ten blocks, with Test Blocks 3, 5, 8, and 10 being proceeded by discrimination blocks (of ten trials each) that primed the button mappings of the following congruent (76 trials) or incongruent (76 trials) blocks (see Table 1). After the first five sIAT blocks, and following the end of the sIAT, the congruent and incongruent test blocks of the pIAT (76 trials each) were presented. The data measured during the pIAT will be reported elsewhere. Our behavioral and ERP analyses focused exclusively on self-

trials (self–positive and self–negative mappings) and excluded the other-trials (other–positive and other–negative mappings), because our intention was to examine attitudes toward the self, and we did not want our findings to be confounded by possible biases in attitudes toward others.

In each trial, the stimulus was presented for 1,500 ms, followed by a blank screen with a randomly jittered duration ranging from 2,000 to 2,200 ms, resulting in a mean stimulus onset asynchrony (SOA) of 3,600 ms. The completion of all blocks required 40 min. Because subjects could take rest periods of self-determined lengths between blocks, the sIAT task lasted between 43 and 55 min. E-Prime (version 2.0, Psychology Software Tools Inc., Pittsburgh, PA, USA) was used to present stimuli and to collect RT data.

Procedure and EEG recording

We placed the EEG cap on the head of each subject and mounted 70 sintered silver chloride ring electrodes on the cap according to the international 10–10 system. Two additional electrodes were placed underneath the eyes in order to record the electrooculogram (EOG). EEG was recorded with a Nihon-Kohden Neurofax 1100 system while subjects completed the sIAT. Fz was the recording reference electrode, and the electrode impedance was kept below 20 kΩ. The EEG was online band-pass filtered (0.3–120 Hz), and the data were digitalized with a sampling rate of 500 Hz.

Data analysis

Analysis of behavioral data

We compared accuracy and RTs for the congruent and incongruent self-trials (self–positive vs. self–negative

**Table 1** Arrangement of the IAT blocks in the case of "CIIC" order

| IAT | Block | Left Label(s) | Right Label(s) | No. of Trials |
|---|---|---|---|---|
| sIAT | 1 Discrimination | self | other | 10 |
| sIAT | 2 Discrimination | positive | negative | 10 |
| sIAT | 3 Test | self/positive | other/negative | 76 |
| sIAT | 4 Discrimination | negative | positive | 10 |
| sIAT | 5 Test | self/negative | other/positive | 76 |
| pIAT | Test | farm/animals | traffic/vehicles | 76 |
| pIAT | Test | traffic/animals | farm/vehicles | 76 |
| sIAT | 6 Discrimination | other | self | 10 |
| sIAT | 7 Discrimination | positive | negative | 10 |
| sIAT | 8 Test | other/positive | self/negative | 76 |
| sIAT | 9 Discrimination | negative | positive | 10 |
| sIAT | 10 Test | other/negative | self/positive | 76 |
| pIAT | Test | farm/vehicles | traffic/animals | 76 |
| pIAT | Test | traffic/vehicles | farm/animals | 76 |

mappings) of the sIAT. To test for significant differences in the amount of incorrect responses between conditions, we applied a paired-samples $t$ test and further calculated the percentages of correct responses in the congruent and incongruent trials.

For the RT analyses, we adapted the improved scoring algorithm by Greenwald, Nosek, and Banaji (2003) to our ERP measurement: That is, we analyzed only trials with RTs above 300 ms and below 3,100 ms, and replaced the RTs of error trials with the mean of the individual RTs of correct trials plus 600 ms. This kind of error penalty has previously shown superior resistance to contamination related to response-speed differences among subjects (Greenwald et al., 2003). Then, we calculated the individual mean RT for congruent and incongruent trials.

We applied a paired-samples $t$ test to test for significant RT differences between the two conditions. Finally, we calculated the individual behavioral IAT effect by subtracting the individual mean RT in congruent trials from the individual mean RT in incongruent trials and by normalizing these differences by the individual standard deviation of response times across the congruent and incongruent trials. Thus, the sIAT effect quantified the behavioral self-evaluation bias, and positive scores denoted positive implicit self-evaluations.

Analysis of EEG data

*Preprocessing and data reduction* Artifact rejection was performed offline using the Brain Vision Analyser software (Brain Products GmbH, Gilching, Germany). Eye movement artifacts were corrected by means of an independent component analysis (ICA; Jung et al., 2000; Tran, Craig, Boord, & Craig, 2004) that was used to identify and subsequently remove components corresponding to blinks or vertical and horizontal eye movements. Channels that contained other artifacts, such as muscle artifacts or artifacts resulting from perspiration, were interpolated by averaging data from adjacent electrodes. (Note that for ten subjects, no channel had to be interpolated; in 12 subjects, one channel, and in three subjects, two channels were interpolated.) After these steps, the data were recomputed against an average reference, and epochs containing further artifacts were excluded manually. A band-pass filter from 0.1 to 20 Hz (Grieder et al., 2012; Khateb, Pegna, Landis, Mouthon, & Annoni, 2010; Wirth et al., 2008) and a 50-Hz notch filter were applied. Individual ERPs were computed by averaging epochs from 0 to 1,000 ms poststimulus separately for congruent and incongruent self-trials. For both the congruent and incongruent conditions, an average of 64 artifact-free segments (ranges among subjects: for congruent, 47–74; for incongruent, 52–74) were included. Finally, the grand mean ERPs for the congruent and the incongruent conditions were computed by averaging the respective individual ERPs.

*Analysis of amplitude differences* To quantify differences in overall amplitude between the two conditions, we used the global field power (GFP; Lehmann & Skrandies, 1980). The significance of GFP differences was established using a simple randomization test that estimated the distribution of GFP differences under the null-hypothesis by randomly permuting the GFP of the two experimental conditions 5,000 times in each subject (Manly, 1997). When the probability that the measured difference was obtained by chance was smaller than 5 %, this was taken as evidence for a quantitative difference in the activation of the generators involved in the processing of congruent as compared with incongruent trials. The test was initially applied separately for each time point of the entire analysis-window therefore, further analyses to protect against false positives due to multiple testing were applied (see below).

*Analysis of topographical differences* The topographic analysis of variance (TANOVA) is a global way to determine topographic differences between conditions. In contrast to channel-wise comparisons, the TANOVA computes the global dissimilarity of whole electrical field topographies between the ERP grand means of the two conditions for each sampling point and determines the significance of these topographic differences with a nonparametric randomization test (Koenig & Melie-García, 2009; Strik, Fallgatter, Brandeis, & Pascual-Marqui, 1998). Before computing the TANOVA, the analyzed data were normalized (GFP = 1) so that significant differences between conditions could only be accounted for by different topographies resulting from different source distribution, and not by different strength of source activation alone. For the computation of the TANOVAs, the RAGU program was used (Koenig, Kottlow, Stein, & Melie-García, 2011) with a single within-subjects factor (congruent/incongruent), 5,000 randomization runs, and a $p$ threshold at .05. As with the GFP, the TANOVA was computed for each time point.

Since the assessment of GFP and topographic differences was computed separately for each time point, our analyses needed some protection against false positive results due to multiple testing. To address this issue, an additional randomization test on the count of consecutive time frames with significant differences in comparison with the count of consecutive significant time frames that would be expected under the null hypothesis was computed (Koenig et al., 2011; Koenig & Melie-García, 2009). TANOVA and GFP results were reported only if the count of consecutive significant time frames was higher than could be expected by chance ($p < .05$).

*Source estimation* Given that differences in topography of the scalp electric field must result from differences in the

underlying neural generators, the inverse resolution method standardized low-resolution brain electromagnetic tomography (sLORETA; Pascual-Marqui, 2002) was applied to approximate brain regions putatively responsible for the observed topographic differences between the two conditions. The sLORETA assumes similar orientations and neuronal strengths of neighboring sources and computes the "smoothest" of all possible intracerebral distributions of the current density that are consistent with the observed scalp electric field topographies. For the time windows that showed significant topographical differences, voxelwise $t$ tests on the normalized and log-transformed current density data were conducted. Since sLORETA was restricted to the time window that had already shown significant topography differences, we applied no further correction for multiple comparisons. In the chosen implementation, sLORETA was based on a boundary element head model and on 6,239 equally spaced voxels (5-mm grid) distributed in the gray matter of the MNI152 average brain template (Pascual-Marqui, 2002).

*Correlation of behavioral and electrophysiological data* In a further analysis, we aimed to identify ERP sources that were activated proportional to the behavioral IAT effect. The rationale for this analysis was that if there were a consistent set of sources showing activation proportional to the IAT effect, this would results in a single topography being added to the ERP in proportion to the IAT effect. This topography can be extracted using covariance mapping (Koenig, Melie-García, Stein, Strik, & Lehmann, 2008). The assessment of the significance of such covariance maps is called TANCOVA and follows randomization procedures that are equivalent to those used for the TANOVA. TANCOVAs with the IAT effect were computed for the ERPs of the congruent and incongruent, and for the ERP difference between these two conditions. As above, TANCOVAs were computed for each time point and an overall test on the duration of significant effects was applied.

In order to retrieve the topography that was modulated proportional to the behavioral IAT effect, we first calculated IAT difference maps by subtracting the electric field topographies (normalized ERPs) of the congruent trials from the one of the incongruent trials. In a next step, we computed the covariance of the behavioral IAT effect and IAT difference map. In periods in which there was evidence for a significant ERP correlate of the IAT effect, the putative localization of the responsible regions was estimated using voxel-wise correlations with the IAT effect in the sLORETA inverse solutions. These sLORETA correlation coefficients were thresholded at ($p < .01$, 23 $df$s, $r = .462$); no further correction for multiple testing was applied since the overall null-hypotheses had already been rejected by the TANCOVA.

## Results

### Behavioral data

We first present accuracy and RT data for our group of 25 subjects, comparing sIAT performance on congruent and incongruent trials.

*Accuracy* Subjects made more incorrect responses in the incongruent ($M = 4.96$, $SD = 3.45$) than in the congruent ($M = 2.20$, $SD = 2.57$) trials. A paired-samples $t$ test yielded a significant difference between the number of incorrect responses in congruent and incongruent trials [$t(24) = 4.872$, $p < .001$]. On average, subjects performed at 97 % accuracy in congruent trials and 93.5 % accuracy in incongruent trials, with all subjects performing above 85 % accuracy in each condition.

*Reaction time* Subjects required less time to react to congruent ($M = 717.53$ ms, $SD = 85.55$) than to incongruent ($M = 847.67$ ms, $SD = 101.93$) trials (see Fig. 1). A paired-samples $t$ test indicated a significant RT difference between congruent and incongruent trials [$t(24) = 10.4$, $p < .001$]. To quantify effect size, we applied the improved scoring algorithm (Greenwald et al., 2003), which revealed a significant IAT effect of medium size ($d = 0.56$, $SD = .25$), with all subjects showing a positive score. These results are thus consistent with the expected faster execution of self-positive as compared to self-negative mappings in healthy subjects.

Explicitly measured self-esteem (RSES) revealed a highly positive explicit self-esteem ($M = 26.95$, $SD = 2.2$, range = 23–30) that was not significantly correlated with the sIAT effect [two-tailed Pearson correlation, $r(20) = .26$, $p = .27$], suggesting that both measures assess different aspects of the self-evaluation.
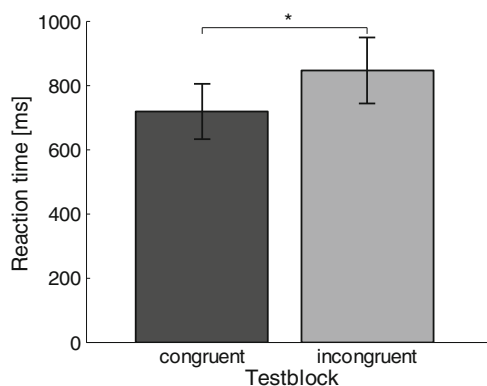


**Fig. 1** Mean reaction times and standard deviations for congruent and incongruent sIAT trials. The *asterisk* denotes statistical significance (paired $t$ test, $p < .05$)

Electrophysiological data

We proceed by presenting ERP results from the group of 25 subjects that were measured during the behavioral task performance. An initial qualitative examination of the average ERPs (see Fig. 2b) revealed a positive bilateral positivity at about 100 ms, corresponding to a P1 potential reflecting the visual evoked activity. At about 150 ms the ERPs showed a N170 with a bilateral occipital negativity and central positivity. Later on, at about 300 ms, a new component developed with a typical P300 field configuration (parietal positivity, frontal negativity). This component had its maximal GFP around 500 ms (see Fig. 2a). Tracking this component, we observed that it appeared to persist longer for incongruent than for congruent trials, as is evident in the ERPs at 700 ms in Fig. 2b.

*Analysis of overall amplitude differences* To quantify differences in ERPs between congruent and incongruent trials, we first examined GFP, which represents momentary signal strength regardless of topographic modulations (see Fig. 2a). The randomization tests (see the Method section) revealed that congruence had no significant effect on GFP modulations (all *p* values > .05).

*Analysis of topographic differences* We next examined the variation in ERP topography across the scalp between congruent and incongruent trials, which we quantified using a topographic analysis of variance (TANOVA). Significant topographic differences occurred only during a time window from 620 to 738 ms (*p* < .05). Note that although we also observed topographic differences at earlier phases of the ERPs, between 136 and 186 ms and between 412 and 462 ms, these effects fell below the significance threshold after a conservative correction for multiple testing using the count of consecutive significant time frames. We concluded that there was a trend toward significance for the early ERP differences, which will not be analyzed further here. We focused further analyses on the statistically significant difference between congruent and incongruent ERPs observed at about 700 ms, which resulted from a variation in ERP topography occurring around this time.

*Source estimation* To approximate brain regions responsible for the significant topographic difference between congruent and incongruent ERPs, we applied sLORETA model to the time window from 620 to 738 ms. A statistical analysis of the individual LORETA density distribution revealed several clusters of voxels that differed in current density estimates between congruent and incongruent trials. We found a number of clusters showing greater current density estimates for incongruent than for congruent trials: Bilaterally, sLORETA differences were localized in the anterior (BA 24) and

subgenual (BA 25) cingulate cortex, as well as in the parahippocampal gyrus (BA 27, 28). Another cluster was located in the right ventrolateral prefrontal cortex (middle frontal gyrus, BA 47). A cluster in the orbitofrontal cortex (BA 25, 11) was more right lateralized. The two clusters yielding greater current density estimates for the congruent than for the incongruent trials were located in the left inferior frontal gyrus (BA45) and the right superior frontal gyrus (BA 6). The Talairach coordinates of the absolute maximal *t* value of each cluster are listed in Table 2.

*Correlation of behavioral and electrophysiological data* Finally, we were interested in identifying brain regions whose activation covaried with the magnitude of the behavioral IAT effect on a subject by subject basis. For this purpose, we used a TANCOVA analysis, which yielded a significant correlation between the IAT difference map and the behavioral IAT effect in a time window between 550 to 710 ms. The covariance map of this time window represented the scalp field of those intracerebral generators that differed in activity between the congruent and incongruent conditions by an amount that was proportional to the behavioral IAT effect. It showed a parietal negativity and an anterior and left temporal positivity that became stronger with a larger behavioral IAT effect, and likewise became weaker with a smaller IAT effect (Fig. 3a). Neither the ERPs of the congruent trials nor the ERPs of the incongruent trials on their own revealed a significant correlation with the behavioral IAT effect.

The significant correlation between the IAT difference map and the behavioral IAT effect indicated that at least one brain region showed changes in activation proportional to the behavioral IAT effect. An sLORETA analysis was thus applied in order to identify the brain regions that showed changes in current density estimates proportional to the IAT effect. For the critical time period, this analysis revealed a positive association between the size of the behavioral IAT effect and the size of the IAT difference in current density estimates located in the right inferior frontal gyrus (BA 44, 45), the bilateral precentral gyrus (BA 6), and the right insula (BA 13), whereas a negative association was located in the left insula (BA 13; see Fig. 3b).

**Discussion**

The main goal of the present study was to study temporal and topographic aspects of ERP activity during implicit self-evaluation using a sIAT paradigm. Behaviorally, in line with previous studies (De Raedt et al., 2006; Franck et al., 2008; Greenwald & Farnham, 2000), a group of healthy subjects responded significantly faster in congruent (self–positive mappings) than in incongruent (self–negative mappings)

trials. All of the subjects showed a positive sIAT effect, which is indicative of the positive implicit self-evaluation expected for healthy individuals. The sIAT-derived measure of self-evaluation was nearly unrelated with an explicit measure of self-esteem based on the Rosenberg self-esteem scale. This is in line with previous results showing no or only weak correlations between implicit and explicit measures of self-evaluation (Bosson et al., 2000; Franck, De Raedt, & De Houwer, 2007; Meites, Deveney, Steele, Holmes, & Pizzagalli, 2008), supporting the notion that implicit and explicit measures reflect distinct aspects of self-evaluation. Whereas explicit reports are conceptualized as being based on conscious reflective and evaluative processes that are thought to be influenced by contextual factors such as social desirability (Tourangeau, Rips, & Rasinski, 1998), implicit measures are thought to reflect "introspectively unidentified effects of the self-attitude" (Greenwald & Banaji, 1995, p. 11) and may thus more faithfully reflect internal biases related to self-evaluation. Thus, we anticipated that the study of brain processes during sIAT performance can provide insights into the neural processes involved in self-evaluation.

As anticipated, we found no systematic ERP differences during the initial period of stimulus processing (up to around 120 ms poststimulus), consistent with the fact that perceptual stimulus processing, attention selection and emotional valence extraction for individual attributes can in fact be expected to be similar between congruent and incongruent trials. During an intermediate period, between about 200 and 400 ms, we had expected, on the basis of previous findings (Fields & Kuperberg, 2012; Keyes et al., 2010), to find that implicit biases in self-evaluation would be reflected in ERP amplitude differences. However, we did not observe any significant ERP amplitude or topographic differences in this time window, although there was a trend toward higher amplitudes in the congruent condition. This suggests that processing in the intermediate period is largely similar between congruent and incongruent sIAT conditions, and may therefore reflect that subjects' inherent implicit self-evaluations were largely independent of which stimulus was presented. Consistent with our initial hypothesis, we did observe topographic ERP differences between congruent (self–positive mappings) and incongruent (self–negative mappings) sIAT trials in a time window from 620 to 738 ms following stimulus onset. In congruent trials, the assignment rule was the same for self and positive attributes, in line with the positive implicit self-evaluation in healthy subjects. It is therefore straightforward for subjects to follow this assignment rule, which corresponds to their implicit association. In contrast, in incongruent trials (self–negative mappings), subjects have to detect the conflict between the implicit self-evaluation and the sIAT assignment rule, and override the response tendency resulting from their implicit association in order to generate the required response. This countermanding of subjects' inherent predispositions requires
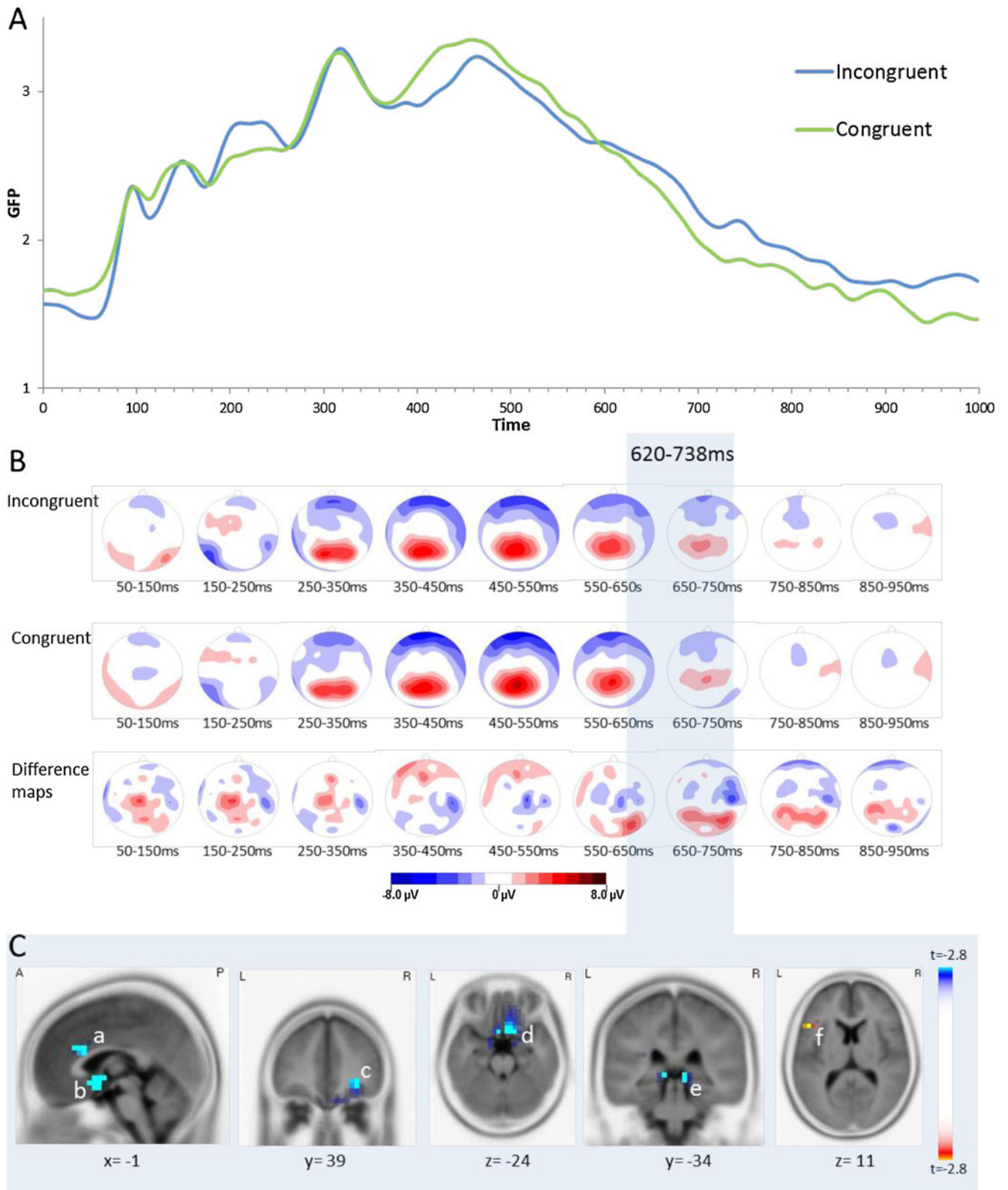
**Fig. 2 a** Global field power (GFP) curves for the grand mean event-related potentials (ERPs) of the congruent and incongruent conditions. **b** Momentary maps and difference maps showing the topographic effect of the congruent and incongruent ERPs, as well as the difference maps (incongruent—congruent). The time frame of the topographic effect is indicated by the vertical bar. **c** Source estimation: Voxelwise *t* values comparing the sLORETA present source density for the time period of 620 to 738 ms between the congruent and incongruent conditions. The incongruent condition showed significantly more activation in clusters in the anterior cingulate gyrus (*a*), the subgenual anterior cingulate gyrus (*b*), the right middle frontal gyrus (*c*), the orbitofrontal cortex (*d*), and bilateral parahippocampal gyrus (*e*), whereas the congruent condition showed more activation in small clusters in the left inferior frontal gyrus (*f*) and the right superior frontal gyrus (not shown). All of the voxels reaching the threshold for *p* < .01 [uncorrected for multiple comparisons, *t*(24) = 2.8] are color coded

cognitive inhibition and control processes. We consider it likely that these inhibitory and control processes are reflected in the observed ERP activations in the late time window, and are probably also responsible for the prolonged RTs on incongruent trials. These findings supports the notion that the recruitment of additional cognitive resources in the incongruent condition underlies the behavioral IAT effect, as has been proposed by several investigators (Brendl et al., 2001; Mierke & Klauer, 2001).

Assuming that countering strong implicit associations requires strong control processes, this result suggests that the extent and magnitude of the ERP activity in the late time window can serve as an index of implicit self-evaluation. Thus, whereas we could not directly detect early ERP activity reflecting implicit association strength, the late ERP activity corresponding to control processes may serve as an indirect measure of implicit self-evaluations in the sIAT paradigm.

The time course of the ERP we observed in our task is generally similar to that found in previous studies investigating IAT paradigms by using ingroup/outgroup face stimuli (Hurtado et al., 2009) as well as homo- and heterosexual couple images (Williams & Themanson, 2011), particularly in terms of the presence of robust congruent/incongruent differences in a late time window. The results relating to early ERP differences are more variable (Ibáñez et al., 2010; O'Tolle & Barnes-Holmes, 2009). For example, one study has reported differences as early as 170 ms post-stimulus-onset that were likely due to the discrimination of configuration aspects of the stimuli (ingroup/outgroup faces and valence) used in this study (Ibáñez et al., 2010). Generally, the available evidence suggests that ERP differences for combined effects (congruent and incongruent associations) tend to be more robust in late than in early time windows, consistent with our present findings showing a nonsignificant trend in early ERPs and robust late ERP effects.

We used source localization to estimate brain structures whose activity might contribute to the ERP topographic differences during the late time window. The results of these analyses are summarized in Table 2. On the basis of previous

**A**

**B**

620-738ms

Incongruent
50-150ms  150-250ms  250-350ms  350-450ms  450-550ms  550-650s  650-750ms  750-850ms  850-950ms

Congruent
50-150ms  150-250ms  250-350ms  350-450ms  450-550ms  550-650ms  650-750ms  750-850ms  850-950ms

Difference maps
50-150ms  150-250ms  250-350ms  350-450ms  450-550ms  550-650ms  650-750ms  750-850ms  850-950ms

-8.0 µV  0 µV  8.0 µV

**C**

t=-2.8

x= -1  y= 39  z= -24  y= -34  z= 11

t=-2.8

fMRI literature, we expected the source localization to identify activations in the ventrolateral (VLPFC) and dorsolateral (DLPFC) prefrontal and the anterior cingulate cortex (ACC). Indeed, the VLPFC and ACC were among a number of brain regions where the sLORETA inverse solution showed stronger activation in the incongruent sIAT condition.

The anterior cingulate cortex (ACC) has been previously associated with conflict monitoring (Botvinick, 2007; Carter

**Table 2** Localization of the to-pographic effects in the sIAT with sLORETA

| Topographic Effect (620–738 ms) | | | Talairach Coordinates | | | t Value |
|---|---|---|---|---|---|---|
| Structures | R/L | BA | x | y | z | |
| incongruent > congruent | | | | | | |
| anterior cingulate cortex (ACC) | | 24 | 0 | 25 | 22 | −3.45 |
| subgenual ACC | | 25 | 0 | 0 | −4 | −4.15 |
| subgenual ACC | | 33 | 0 | 21 | 22 | −3.36 |
| medial frontal gyrus | R | 25 | 5 | 9 | −17 | −3.33 |
| medial frontal gyrus | L | 25 | −5 | 9 | −17 | −3.36 |
| subcallosal gyrus | | 25 | 0 | 4 | −13 | −3.38 |
| subcallosal gyrus | R | 34 | 10 | 4 | −13 | −3.29 |
| subcallosal gyrus | L | 34 | −10 | 4 | −13 | −3.81 |
| rectal gyrus | R | 11 | 5 | 13 | −22 | −3.08 |
| middle frontal gyrus | R | 47 | 30 | 38 | −10 | −2.86 |
| parahippocampal gyrus | R | 27 | 10 | −34 | −2 | −2.95 |
| parahippocampal gyrus | L | 28 | −15 | −5 | 12 | −2.86 |
| superior temporal gyrus | R | 38 | 20 | 8 | −34 | −2.93 |
| congruent > incongruent | | | | | | |
| inferior frontal gyrus | L | 45 | −50 | 20 | 8 | 2.86 |
| superior frontal gyrus | R | 6 | 20 | 3 | 64 | 2.87 |

For each of the significant clusters, the table indicates the hemisphere and Brodmann area (BA) of the most significant voxel in the cluster. The coordinates indicate the absolute maximal t values of the clusters reaching the threshold for $p < .01$ [uncorrected for multiple comparisons, $t(24) = 2.8$] in a voxelwise t-value comparison of the sLORETA current source densities in the congruent and incongruent conditions for the time period of 620–738 ms

& van Veen, 2007; Cohen, Botvinick, & Carter, 2000; Kerns et al., 2004) and the control of goal-directed behavior (Holroyd & Yeung, 2012; Kerns et al., 2004). The ACC tends to show enhanced activity when there is a high level of conflict in the stimulus material regarding the required behavioral response (Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999), and it seems to detect conflict especially at later, response-related levels of processing (van Veen, Cohen, Botvinick, Stenger, & Carter, 2001). This suggests that the enhanced ACC activity that we identified by sLORETA as a potential neural generator during the sIAT task may have resulted from a response conflict. Such a conflict may have

occurred because negative attributes have to be mapped to the self in the incongruent condition, contradicting the implicit positive self-evaluation of the healthy subjects. In conflicting situations, inadequate behavioral response tendencies often need to be suppressed or inhibited, so that the conflict can be resolved and the desired response can be made.

The ventrolateral prefrontal cortex is another brain region that has been implicated in this kind of inhibitory cognitive control of behavior (Aron, Robbins, & Poldrack, 2004; Buchsbaum, Greer, Chang, & Berman, 2005; Chikazoe, 2010; Dillon & Pizzagalli, 2007; Hedden & Gabrieli, 2010). Not surprisingly, many fMRI studies have found coactivation
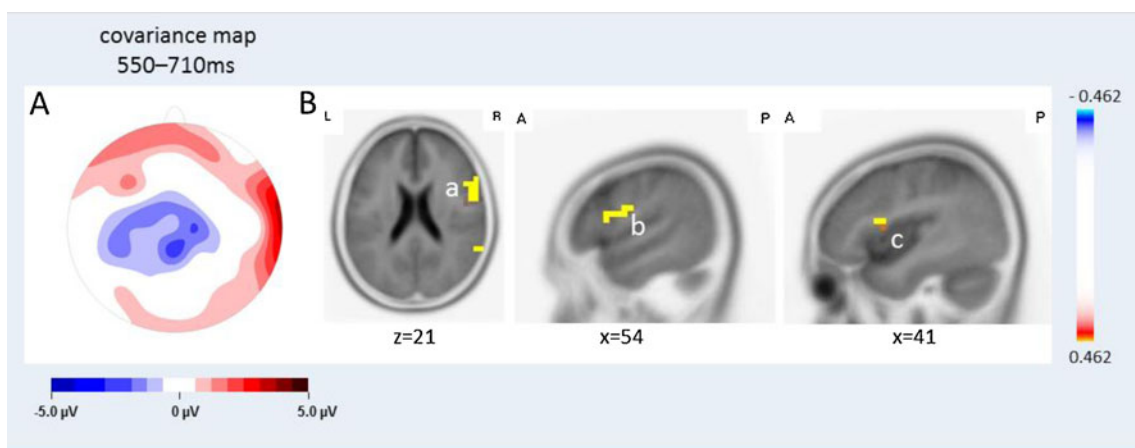


**Fig. 3** **a** Covariance map for the association of the IAT difference map and the behavioral IAT effect. **b** sLORETA source localization for the association between the IAT difference map and the behavioral IAT effect. A threshold of $r = .462$, corresponding to a p value of .01 (uncorrected for multiple comparisons), was chosen. The sources were localized in right inferior frontal gyrus (*a*), bilateral precentral gyrus (*b*), and bilateral insula (*c*)

of the anterior cingulate and prefrontal cortices in tasks involving conflicts and action selection (Nee, Wager, & Jonides, 2007; Niendam et al., 2012). One example is the go/no-go paradigm, in which elevated activity in both the ACC and the vlPFC has been reported in the no-go condition (Liddle, Kiehl, & Smith, 2001; Nakata et al., 2008). In the no-go condition, subjects need to exert inhibitory control to withhold a behavioral response triggered by the stimulus material. Our sLORETA analyses suggest that these control processes, mediated jointly by the ACC and vlPFC, were also engaged in the incongruent sIAT condition in order to inhibit the response tendency resulting from the positive implicit self-evaluation. This finding is consistent with previous fMRI studies that have used various IAT-based paradigms, examining conceptual associations (Chee et al., 2000) or moral prejudice (Luo et al., 2006), and it highlights the robust nature of the cognitive control related activity in these brain areas during the IAT.

Source localization suggested that several additional brain regions were associated with sIAT performance. We discuss the most prominent regions below, while noting that sLORETA-based localization only identifies one of the many possible solutions to the inverse problem, and thus these observations remain somewhat speculative and must be complemented by fMRI or PET experiments. One of the additional brain regions identified by source localization is the orbitofrontal cortex (OFC), which showed enhanced activity during the incongruent IAT condition. This is compatible with fMRI literature implicating the OFC in cognitive control processes (Cunningham, Raye, & Johnson, 2004) such as response selection (Young & Shapiro, 2011) and decision making (Bechara, Damasio, & Damasio, 2000), although previous IAT brain imaging studies have not described any OFC activation during the incongruent condition. We thus consider it possible, that the right hemisphere lateralized OFC activation is attributable to the self-related character of the sIAT: In incongruent trials, negative attributes are mapped to

the self, which may activate internal self-representations that the subjects want to avoid, being called self-related prevention goals. Such an interpretation is in line with findings of a recent study (Eddington et al., 2009) describing right lateralized OFC activity in depressed patients when being confronted with their individual prevention goals. This interpretation is also consistent with previous studies associating right prefrontal structures with processing self-referential information (Kaplan, Aziz-Zadeh, Uddin, & Iacoboni, 2008; Nitschke, Sarinopoulos, Mackiewicz, Schaefer, & Davidson, 2006). Interestingly, source localization estimated two areas of the frontal lobe, the left inferior frontal gyrus and the right superior frontal gyrus, as showing higher activity in the congruent condition. Activity in the left inferior frontal gyrus has been observed during self-referential tasks (Ochsner et al., 2004), and has been linked to inner speech (Morin & Michaud, 2007). Accordingly, we speculate that the activation of the left inferior frontal gyrus in our study could reflect internal verbalizations that may be stronger for positive attributes corresponding to the implicit self-evaluation. The greater activation during congruent trials in the superior frontal gyrus, which is part of the supplementary motor area (SMA), might reflect the earlier start of motor response preparation.

Finally, we used sLORETA to identify brain regions in which, across subjects, activity was correlated with RT differences between congruent and incongruent conditions that we had quantified using the behavioral IAT effect. As was mentioned above, these sLORETA results can only represent possible approximations to the real underlying ERP sources, and must be interpreted with caution. The sLORETA-based sources that correlated with behavioral performance, listed in Table 3, included the precentral gyrus, the inferior frontal gyrus, and the insula bilaterally. The brain areas showing mostly positive correlations with the sIAT effect fall into two classes: The precentral gyrus is part of the motor system and thus the significant correlation could be due to earlier response

**Table 3** sLORETA localization of the timeframe with significant TANCOVA effects

| TANCOVA | | | Talairach Coordinates | | | r Value |
|---|---|---|---|---|---|---|
| Structures | R/L | BA | x | y | z | |
| Inferior frontal gyrus | R | 44 | 45 | 15 | 13 | .48 |
| Inferior frontal gyrus | R | 45 | 59 | 15 | 18 | .49 |
| Inferior frontal gyrus | R | 9 | 59 | 6 | 23 | .49 |
| Precentral gyrus | R | 6 | 59 | 1 | 23 | .49 |
| Precentral gyrus | L | 6 | −45 | −3 | 46 | .48 |
| Insula | R | 13 | 40 | 15 | 13 | .48 |
| Insula | L | 13 | −30 | 20 | 8 | −.50 |
| Supramarginal gyrus | R | 40 | 59 | −48 | 21 | .47 |
| Superior temporal gyrus | R | 22 | 64 | −48 | 21 | .49 |
| Cuneus | L | 23 | −5 | −72 | 13 | .47 |
| Cuneus | R | 18 | 0 | −72 | 13 | .50 |

For each of the significant clusters, the table indicates the hemisphere and Brodmann area (BA) of the most significant voxel in the cluster

execution on congruent trials. The remaining brain areas, notably inferior frontal gyrus and insula, are not closely related to motor preparation or execution, but instead have been linked to selective attention, response selection and inhibitory cognitive control (Aron et al., 2004; Chechko et al., 2012; Chikazoe, 2010). This suggests that in our task, activation in these areas is related to control processes recruited during the incongruent condition, particularly in subjects with strong behavioral sIAT effects that are indicative of highly positive self-evaluations.

Some limitations of the present study should be acknowledged. First, we applied a rather conservative statistical criterion (count of consecutive significant time frames), so the present study might have missed some short lasting effects in the early phase of the ERPs. Furthermore, the subjects in our study were mostly young, highly educated individuals, and all exhibited a highly positive explicit self-esteem, so our results can only be generalized to a limited extent. A final issue concerns possible individual differences in sIAT performance strategies. Since we have no knowledge of their individually applied strategies, some subjects might have focused their response choice more on the rules for the other-attribute mapping, whereas others might have focused mainly on rules for self-attribute mapping. Further studies are necessary to assess such strategies and their impact on sIAT performance and the corresponding neural correlates.

Taken together, we have shown that cognitive control processes that are needed to override response tendencies resulting from implicit associations can be identified and mapped to a number of frontal lobe brain structures. Activity in these structures therefore appears suitable to quantify the strength of implicit self-evaluations on a neural level. The relative strength of these activations might be a useful indicator for tracking the implicit self of patients suffering from depression, potentially providing a quantitative tool for assessing benefits of psychotherapeutic interventions and identifying patients with an elevated risk of relapse. For example, cognitive therapy is assumed to foster reflective correction of negative information processing. It might therefore thus reduce depressive symptoms primarily by focusing on explicit negative self-evaluations, while not fully addressing underlying, negative implicit biases based on self-schemata originating from past emotional experiences. To prevent persistence of the resulting elevated vulnerability for depression, therapeutic interventions are needed that especially target and modify the implicitly represented aspects of the self-schemata.

Emotion-focused therapy (EFT; Greenberg & Watson, 2006), might be an attractive candidate in this context, since the emotions underlying the patients schematic self-representations are activated, differentiated and transformed by adaptive emotional experiences discrepant to the previous maladaptive experiences. The combination of treatment principles from EFT and from exposure-based treatment of trauma (Foa & Rothbaum, 1998), as realized in exposure-based cognitive therapy for depression (EBCT; Hayes, Beevers, Feldman, Laurenceau, & Perlman, 2005), is a promising emerging psychotherapeutic approach. EBCT includes exposure interventions targeting experiential avoidance and preventing avoidant behavior, as well as interventions aimed at facilitating emotional and cognitive processing (Grosse Holtforth et al., 2012). In EBCT, providing information inconsistent with the maladaptive self-associations might trigger corrective cognitive–emotional experiences that initiate therapeutic change in negatively biased implicit self-representations (Castonguay & Hill, 2012). These change processes are assumed to allow for more sustainable schema change and symptom change to take place (Ellison, Greenberg, Goldman, & Angus, 2009), making patients less vulnerable to relapse.

## Appendix

Table 4  sIAT stimuli

| Category (German Word) | | Attributes (German Word) | |
| --- | --- | --- | --- |
| Self | Not-Self | Positive | Negative |
| self (*selbst*) | others (*andere*) | sociable (*gesellig*) | timid (*furchtsam*) |
| I (*ich*) | you (*ihr*) | enthusiastic (*begeistert*) | unattractive (*unattraktiv*) |
| me (*mich*) | you (*euch*) | cheerful (*fröhlich*) | useless (*nutzlos*) |
| my (*mein*) | yours (*eure*) | composed (*gelassen*) | rejected (*angelehnt*) |
| me (*mir*) | yours (*euer*) | lively (*lebhaft*) | unwanted (*unerwünscht*) |
| | | open (*offen*) | helpless (*hilflos*) |
| | | free (*frei*) | weak (*schwach*) |
| | | sincere (*aufrichtig*) | passive (*passiv*) |
| | | calm (*ruhig*) | inferior (*minderwertig*) |
| | | adventurous (*unternehmungslustig*) | shy (*schüchtern*) |

# References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7*, 268–277.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8*, 170–177. doi:10.1016/j.tics.2004.02.010

Baumeister, R. F., Campbell, J., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*, 1–44.

Bechara, A., Damasio, H., & Damasio, A. R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex, 10*, 295–307.

Beck, A. T. (1967). *Depression: Clinical, experimental and theoretical aspects*. New York, NY: Harper and Row.

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

Beck, A. T., Steer, R. A., & Brown, G. K. (2006). *BDI–II—beck depressions-inventar* (2nd ed.). Frankfurt, Germany: Harcourt Test Services GmbH.

Beck, J. S. (1995). *Cognitive therapy: Basics and beyond*. New York, NY: Guilford.

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*, 631–643.

Botvinick, M. M. (2007). Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 356–366. doi:10.3758/CABN.7.4.356

Botvinick, M. M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature, 402*, 179–181. doi:10.1038/46035

Brendl, C. M., Markman, A. B., & Messner, C. (2001). How do indirect measures of evaluation work? evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology, 81*, 760–773.

Buchsbaum, B. R., Greer, S., Chang, W. L., & Berman, K. F. (2005). Meta-analysis of neuroimaging studies of the Wisconsin card-sorting task and component processes. *Human Brain Mapping, 25*, 35–45.

Carter, C. S., & van Veen, V. (2007). Anterior cingulate cortex and conflict detection: An update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 367–379. doi:10.3758/CABN.7.4.367

Castonguay, L. G., & Hill, C. E. (2012). *Transformation in psychotherapy: Corrective experiences across cognitive-behavioral, humanistic, and psychodynamic approaches*. Washington, DC: American Psychological Association.

Chechko, N., Kellermann, T., Zvyagintsev, M., Augustin, M., Schneider, F., & Habel, U. (2012). Brain circuitries involved in semantic interference by demands of emotional and non-emotional distractors. *PLoS One, 7*, e38155. doi:10.1371/journal.pone.0038155

Chee, M. W. L., Sriram, N., Soon, C. S., & Lee, K. M. (2000). Dorsolateral prefrontal cortex and the implicit association of concepts and attributes. *NeuroReport, 11*, 135–140.

Chikazoe, J. (2010). Localizing performance of go/no-go tasks to prefrontal cortical subregions. *Current Opinion in Psychiatry, 23*, 267–272.

Coates, M. A., & Campbell, K. B. (2010). Event-related potential measures of processing during an Implicit Association Test. *NeuroReport, 21*, 1029–1033.

Cockerham, E., Stopa, L., Bell, L., & Gregg, A. (2009). Implicit self-esteem in bulimia nervosa. *Journal of Behavior Therapy and Experimental Psychiatry, 40*, 265–273.

Cohen, J. D., Botvinick, M., & Carter, C. S. (2000). Anterior cingulate and prefrontal cortex: Who's in control? *Nature Neuroscience, 3*, 421–423. doi:10.1038/74783

Collani, G., & Herzberg, P. Y. (2003). Eine revidierte Fassung der deutschsprachigen Skala zum Selbstwertgefühl von Rosenberg. *Zeitschrift für Differenzielle und Diagnostische Psychologie, 24*, 3–7.

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience, 16*, 1717–1729. doi:10.1162/0898929042947919

D'Argembeau, A., & Salmon, E. (2012). The neural basis of semantic and episodic forms of self-knowledge: Insights from functional neuroimaging. *Advances in Experimental Medicine and Biology, 739*, 276–290.

de Jong, P. J. (2002). Implicit self-esteem and social anxiety: Differential self-favouring effects in high and low anxious individuals. *Behavior Research and Therapy, 40*, 501–508.

De Raedt, R., Schacht, R., Franck, E., & De Houwer, J. (2006). Self-esteem and depression revisited: Implicit positive self-esteem in depressed patients? *Behavior Research and Therapy, 44*, 1017–1028.

Dillon, D. G., & Pizzagalli, D. A. (2007). Inhibition of action, thought, and emotion: A selective neurobiological review. *Applied and Preventive Psychology, 12*, 99–114.

Eddington, K. M., Dolcos, F., McLean, A. N., Krishnan, K. R., Cabeza, R., & Strauman, T. J. (2009). Neural correlates of idiographic goal priming in depression: Goal-specific dysfunctions in the orbitofrontal cortex. *Social Cognitive and Affective Neuroscience, 4*, 238–246.

Ellison, J. A., Greenberg, L. S., Goldman, R. N., & Angus, L. (2009). Maintenance of gains following experiential therapies for depression. *Journal of Consulting and Clinical Psychology, 77*, 103–112.

Fields, E. C., & Kuperberg, G. R. (2012). It's all about you: An ERP study of emotion and self-relevance in discourse. *NeuroImage, 62*, 562–574.

Foa, E. B., & Rothbaum, B. O. (1998). *Treating the trauma of rape*. New York, NY: Guilford Press.

Forbes, C. E., Cameron, K. A., Grafman, J., Barbey, A., Solomon, J., Ritter, W., et al. (2012). Identifying temporal and causal contributions of neural processes underlying the Implicit Association Test (IAT). *Frontiers in Human Neuroscience, 6*, 320. doi:10.3389/fnhum.2012.00320

Fossati, P., Hevenor, S. J., Graham, S. J., Grady, C., Keightley, M. L., Craik, F., et al. (2003). In search of the emotional self: An fMRI study using positive and negative emotional words. *The American Journal of Psychiatry, 160*, 1938–1945.

Franck, E., De Raedt, R., & De Houwer, J. (2007). Implicit but not explicit self-esteem predicts future depressive symptomatology. *Behavior Research and Therapy, 45*, 2448–2455.

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

Franck, E., De Raedt, R., 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*, 75–85.

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*, 282–289.

Ginsburg, G. S., La Greca, A. M., & Silverman, W. K. (1998). Social anxiety in children with anxiety disorders: Relation with social and emotional functioning. *Journal of Abnormal Child Psychology, 26*, 175–185.

Glashouwer, K. A., & de Jong, P. J. (2010). Disorder-specific automatic self-associations in depression and anxiety: Results of the Netherlands Study of Depression and Anxiety. *Psychological Medicine, 40*, 1101–1111.

Greenberg, L. S., & Watson, J. C. (2006). *Emotion-focused therapy for depression*. Washington, DC: American Psychological Association.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychology Review, 102*(1), 4–27.

Greenwald, A. G., & Farnham, S. D. (2000). Using the implicit association test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*, 1022–1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

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*, 197–216.

Grieder, M., Crinelli, R. M., Koenig, T., Wahlund, L. O., Dierks, T., & Wirth, M. (2012). Electrophysiological and behavioral correlates of stable automatic semantic retrieval in aging. *Neuropsychologia, 50*, 160–171.

Grosse Holtforth, M., Hayes, A. M., Sutter, M., Wilm, K., Schmied, E., Laurenceau, J.-P., et al. (2012). Fostering cognitive–emotional processing in the treatment of depression: A preliminary investigation in exposure-based cognitive therapy. *Psychotherapy and Psychosomatics, 81*, 259–260. doi:10.1159/000336813

Hayes, A. M., Beevers, C., Feldman, G. C., Laurenceau, J. P., & Perlman, C. (2005). Avoidance and processing as predictors of symptom change and positive growth in an integrative therapy for depression. *International Journal of Behavioral Medicine, 12*, 111–122.

Hedden, T., & Gabrieli, J. D. E. (2010). Shared and selective neural correlates of inhibition, facilitation, and shifting processes during executive control. *NeuroImage, 51*, 421–431.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*, 1369–1385.

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences, 16*, 122–128.

Hurtado, E., Haye, A., González, R., Manes, F., & Ibáñez, A. (2009). Contextual blending of ingroup/outgroup face stimuli and word valence: LPP modulation and convergence of measures. *BMC Neuroscience, 10*, 69. doi:10.1186/1471-2202-10-69

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. doi:10.3389/fnhum.2010.00188

Ingram, R. E. (1984). Toward an information-processing analysis of depression. *Cognitive Therapy and Research, 8*, 443–477.

Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., & Prigatano, G. P. (2002). Neural correlates of self-reflection. *Brain, 125*, 1808–1814.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37*, 163–178.

Kaplan, J. T., Aziz-Zadeh, L., Uddin, L. Q., & Iacoboni, M. (2008). The self across the senses: An fMRI study of self-face and self-

voice recognition. *Social Cognitive and Affective Neuroscience, 3*, 218–223.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*, 1023–1026.

Keyes, H., Brady, N., Reilly, R. B., & Foxe, J. J. (2010). My face or yours? event-related potential correlates of self-face processing. *Brain and Cognition, 72*, 244–254. doi:10.1016/j.bandc.2009.09.006

Khateb, A., Pegna, A. J., Landis, T., Mouthon, M. S., & Annoni, J. M. (2010). On the origin of the N400 effects: An ERP waveform and source localization analysis in three matching tasks. *Brain Topography, 23*, 311–320.

Knutson, K. M., Mah, L., Manly, C. F., & Grafman, J. (2007). Neural correlates of automatic beliefs about gender and race. *Human Brain Mapping, 28*, 915–930.

Koenig, T., Kottlow, M., Stein, M., & Melie-García, L. (2011). Ragu: A free tool for the analysis of EEG and MEG event-related scalp field data using global randomization statistics. *Computational Intelligence and Neuroscience, 2011*(938925), 1–14. doi:10.1155/2011/938925

Koenig, T., & Melie-García, L. (2009). Statistical analysis of multichannel scalp field data. In C. M. Michel, T. Koenig, D. Brandeis, L. R. R. Gianotti, & J. Wackermann (Eds.), *Electrical neuroimaging* (pp. 169–190). Cambridge, UK: Cambridge University Press.

Koenig, T., Melie-García, L., Stein, M., Strik, W., & Lehmann, C. (2008). Establishing correlations of scalp field maps with other experimental variables using covariance analysis and resampling methods. *Clinical Neurophysiology, 119*, 1262–1270.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48*, 609–621.

Liddle, P. F., Kiehl, K. A., & Smith, A. M. (2001). Event-related fMRI study of response inhibition. *Human Brain Mapping, 12*, 100–109.

Luck, S. J. (2005). ERP localization. In S. J. Luck (Ed.), *An introduction to the event-related potential technique* (1st ed., pp. 267–300). Cambridge, MA: MIT Press.

Luo, Q., Nakic, M., Wheatley, T., Richell, R., Martin, A., & Blair, R. J. (2006). The neural basis of implicit moral attitude—an IAT study using event-related fMRI. *NeuroImage, 30*, 1449–1457. doi:10.1016/j.neuroimage.2005.11.005

Manly, B. F. J. (1997). *Randomization, bootstrap, and Monte Carlo methods in biology* (2nd ed.). London, UK: Chapman & Hall.

Meites, T. M., Deveney, C. M., Steele, K. T., Holmes, A. J., & Pizzagalli, D. A. (2008). Implicit depression and hopelessness in remitted depressed individuals. *Behavior Research and Therapy, 46*, 1078–1084.

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift für Experimentelle Psychologie, 48*, 107–122.

Moran, J. M., Heatherton, T. F., & Kelley, W. M. (2009). Modulation of cortical midline structures by implicit and explicit self-relevance evaluation. *Social Neuroscience, 4*, 197–211. doi:10.1080/17470910802250519

Morin, A., & Michaud, J. (2007). Self-awareness and the left inferior frontal gyrus: Inner speech use during self-related processing. *Brain Research Bulletin, 74*, 387–396.

Nakata, H., Sakamoto, K., Ferretti, A., Gianni Perrucci, M., Del Gratta, C., Kakigi, R., et al. (2008). Somato-motor inhibitory processing in humans: An event-related functional MRI study. *NeuroImage, 39*, 1858–1866. doi:10.1016/j.neuroimage.2007.10.041

Nee, D. E., Wager, T. D., & Jonides, J. (2007). Interference resolution: Insights from a meta-analysis of neuroimaging tasks. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 1–17.

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M., Glahn, D. C., & Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective, & Behavioral Neuroscience, 12*, 241–268.

Nitschke, J. B., Sarinopoulos, I., Mackiewicz, K. L., Schaefer, H. S., & Davidson, R. J. (2006). Functional neuroanatomy of aversion and its anticipation. *NeuroImage, 29*, 106–116.

Northoff, G., & Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences, 8*, 102–107.

O'Tolle, C., & Barnes-Holmes, D. (2009). Electrophysiological activity generated during the implicit association test: A study using event-related potentials. *Psychological Record, 59*, 207–220.

Ochsner, K. N., Knierim, K., Ludlow, D. H., Hanelin, J., Ramachandran, T., Glover, G., et al. (2004). Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *Journal of Cognitive Neuroscience, 16*, 1746–1772. doi:10.1162/0898929042947829

Ormel, J., Oldehinkel, A. J., & Vollebergh, W. (2004). Vulnerability before, during, and after a major depressive episode: A 3-wave population-based study. *Archives of General Psychiatry, 61*, 990–996.

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*, 472–478.

Orth, U., Robins, R. W., & Widaman, K. F. (2012). Life-span development of self-esteem and its effects on important life outcomes. *Journal of Personality and Social Psychology, 102*, 1271–1288.

Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology, 24*(Suppl. D), 5–12.

Rameson, L. T., Satpute, A. B., & Lieberman, M. D. (2010). The neural correlates of implicit and explicit self-relevant processing. *NeuroImage, 50*, 701–708. doi:10.1016/j.neuroimage.2009.12.098

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*, 199–206.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the implicit association test: Dissociating salience from associations. *Journal of Experimental Psychology. General, 133*, 139–165. doi:10.1037/0096-3445.133.2.139

Schmitz, T. W., Kawahara-Baccus, T. N., & Johnson, S. C. (2004). Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *NeuroImage, 22*, 941–947.

Sowislo, J. F., & Orth, U. (2013). Does low self-esteem predict depression and anxiety? a meta-analysis of longitudinal studies. *Psychological Bulletin, 139*, 213–240. doi:10.1037/a0028931

Steffens, M. C., Kirschbaum, M., & Glados, P. (2008). Avoiding stimulus confounds in implicit association tests by using the concepts as stimuli. *British Journal of Social Psychology, 47*, 217–243.

Strik, W. K., Fallgatter, A. J., Brandeis, D., & Pascual-Marqui, R. D. (1998). Three-dimensional tomography of event-related potentials during response inhibition: Evidence for phasic frontal lobe activation. *Electroencephalography and Clinical Neurophysiology, 108*, 406–413.

Tanner, R. J., Stopa, L., & De Houwer, J. (2006). Implicit views of the self in social anxiety. *Behavior Research and Therapy, 44*, 1397–1409. doi:10.1016/j.brat.2005.10.007

Teasdale, J. D. (1988). Cognitive vulnerability to persistent depression. *Cognition and Emotion, 2*, 247–274.

Tourangeau, R., Rips, L. J., & Rasinski, K. (1998). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.

Tran, Y., Craig, A., Boord, P., & Craig, D. (2004). Using independent component analysis to remove artifact from electroencephalographic measured during stuttered speech. *Medical & Biological Engineering & Computing, 42*, 627–633.

van der Meer, L., Costafreda, S., Aleman, A., & David, A. S. (2010). Self-reflection and the brain: A theoretical review and meta-analysis of neuroimaging studies with implications for schizophrenia. *Neuroscience and Biobehavioral Reviews, 34*, 935–946. doi:10.1016/j.neubiorev.2009.12.004

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *NeuroImage, 14*, 1302–1308.

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology, 37*, 190–203.

Vohs, K. D., Voelz, Z. R., Pettit, J. W., Bardone, A. M., Katz, J., Abramson, L. Y., . . . Joiner, T. E., Jr. (2001). Perfectionism, body dissatisfaction, and self-esteem: An interactive model of bulimic symptom development. *Journal of Social and Clinical Psychology, 20*, 476–497.

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*, 468–476.

Wirth, M., Horn, H., Koenig, T., Razafimandimby, A., Stein, M., Mueller, T., . . . Strik, W. (2008). The early context effect reflects activity in the temporo-prefrontal semantic system: Evidence from electrical neuroimaging of abstract and concrete word reading. *NeuroImage, 42*, 423–436. doi:10.1016/j.neuroimage.2008.03.045

Young, J. J., & Shapiro, M. L. (2011). The orbitofrontal cortex and response selection. *Annals of the New York Academy of Sciences, 1239*, 25–32. doi:10.1111/j.1749-6632.2011.06279.x