Response conflict and frontocingulate dysfunction in unmedicated participants with major depression

Avram J. Holmes, Diego A. Pizzagalli

Department of Psychology, Harvard University, 1220 William James Hall, 33 Kirkland Street, Cambridge, MA 02138, USA

Article info
Article history:
Received 17 March 2008
Received in revised form 8 May 2008
Accepted 29 May 2008
Available online 4 June 2008

Keywords:
Depression
Action monitoring
Anterior cingulate cortex
Dorsolateral prefrontal cortex
Cognitive control
Executive function

Abstract
Individuals with major depressive disorder (MDD) often exhibit impaired executive function, particularly in experimental tasks that involve response conflict and require adaptive behavioral adjustments. Prior research suggests that these deficits might be due to dysfunction within frontocingulate pathways implicated in response conflict monitoring and the recruitment of cognitive control. However, the temporal unfolding of conflict monitoring impairments in MDD remains poorly understood. To address this issue, we recorded 128-channel event-related potentials while 20 unmedicated participants with MDD and 20 demographically matched, healthy controls performed a Stroop task. Compared to healthy controls, MDD subjects showed larger Stroop interference effects and reduced N2 and N450 amplitudes. Source localization analyses at the time of maximal N450 activity revealed that MDD subjects had significantly reduced dorsal anterior cingulate cortex (dACC; Brodmann area 24/32) and left dorsolateral prefrontal cortex (Brodmann area 10/46) activation to incongruent relative to congruent trials. Consistent with the heterogeneous nature of depression, follow-up analyses revealed that depressed participants with the lowest level of conflict-related dACC activation 620 ms post-stimulus were characterized by the largest Stroop interference effects (relatively increased slowing and reduced accuracy for incongruent trials). Conversely, MDD participants with relatively stronger dACC recruitment did not differ from controls in terms of interference effects. These findings suggest that for some, but not all individuals, MDD is associated with impaired performance in trials involving competition among different response options, and reduced recruitment of frontocingulate pathways implicated in conflict monitoring and cognitive control.

1. Introduction
Major depressive disorder (MDD) is characterized by impairments in executive function, particularly in situations requiring behavioral adjustments and adaptive action monitoring (Austin, Mitchell, & Goodwin, 2001; Nitschke & Mackiewicz, 2005; Paradiso, Lamberty, Garvey, & Robinson, 1997; Porter, Gallagher, Thompson, & Young, 2003). For example, MDD has been associated with increased sensitivity to mistakes in performance (e.g., Beats, Sahakian, & Levy, 1996; Holmes & Pizzagalli, 2008; Steffens, Wagner, Levy, Horn, & Krishnan, 2001) and negative feedback (e.g., Elliott, Sahakian, Herrod, Robbins, & Paykel, 1997). These deficits might be related to symptoms of indecisiveness or diminished ability to concentrate, which characterize the clinical presentation of MDD subjects (American Psychiatric Association, 1994).

In addition to hyper-responsiveness to errors and negative feedback, evidence indicates that depressed subjects might display conflict monitoring dysregulation in paradigms generating competition among response options (see Ottowitz, Dougherty, & Savage, 2002 for review). Studies using the Stroop tasks, for example, in which the prepotent tendency to read a word competes with the task demand of naming the color, have described impaired performance in depressed subjects (i.e., increased slowing and reduced accuracy during response conflict; Lemelin, Baruch, Vincent, Everett, & Vincent, 1997; Moritz et al., 2002; Ottowitz et al., 2002; Paradiso et al., 1997; Trichard et al., 1995; but see Austin et al., 1999). Notably, these deficits predicted poor treatment outcome (Sneed et al., 2007), persisted after symptom remission (Trichard et al., 1995), and were seen in individuals with subclinical depressive symptomatology (Holmes & Pizzagalli, 2007), indicating that conflict monitoring dysfunctions are a promising marker of dysfunctional executive function in depression.
findings showing that depressed subjects display abnormal activation within top-down prefrontal cortex (PFC) and anterior cingulate cortex (ACC) during tasks involving conflict monitoring (e.g., George et al., 1997; Mitterschiffthaler et al., 2007; Wagner et al., 2006; for review see Davidson, Pizzagalli, Nitschke, & Putnam, 2002; Nitschke & Mackiewicz, 2005; Rogers et al., 2004). These findings are intriguing since theories regarding the nature of the action monitoring system have proposed a distributed executive control system, primarily centered on the ACC and PFC (Botvinick, 2007; Gehring & Willoughby, 2002; Holroyd & Coles, 2002). According to these theories, one role of this system is the implementation of the cognitive control necessary to monitor and adjust for the occurrence of response conflict (Carter & van Veen, 2007). In line with this assumption, research indicates that (1) the dorsolateral PFC (DLPFC) is critical for the implementation of top-down attentional control (Vanderhasselt, De Raedt, Baeken, Leyman, & D’haenen, 2006; MacDonald, Cohen, Stenger, & Carter, 2000; Miller & Cohen, 2001); and (2) ACC activity during response conflict predicts DLPFC recruitment and subsequent behavioral adjustments (Kerns et al., 2004). Together, these findings suggest that conflict monitoring impairments in depression might be linked to dysfunctions within frontocingulate pathways. This assumption is further supported by computational modeling of depressed participant’s Stroop task performance, which has shown that these deficits can be accounted for by disrupted prefrontal/ACC activity and associated decrease in cognitive control (Siegle, Steinhaeuer, & Thase, 2004).

While MDD has been linked to reduced conflict monitoring performance and dysregulated frontocingulate activation in tasks probing cognitive control, relatively little is known about the temporal unfolding of brain mechanisms implicated in these dysfunctions, which in turn may offer important insights into the source of executive impairments in depression. Event-related potential (ERP) techniques are ideally suited for investigating this important issue. Specifically, prior studies using the Stroop or related tasks have described two ERP components – N2 and N450 – that appear to be related to conflict monitoring processes. The N450 component, in particular, a negative voltage deflection beginning ~400 ms following the presentation of an incongruent trial, has been consistently linked to the Stroop interference effect and is assumed to index conflict detection, most likely at the response window (Hanslmayr et al., 2008; Rebai, Bernard, & Lannou, 1997; West, 2003; West & Alain, 2000b; West, Jakubek, Wymbus, Perry, & Moore, 2005). The N2, a negative frontocentral deflection showing greater amplitudes for incongruent than congruent trials in interference task (e.g., Gehring, Gratton, Coles, & Donchin, 1992; Kopp, Mattler, Goertz, & Rist, 1996), has been also associated with conflict detection, although not as consistently as the N450 component (West, Krompinger, Bowry, & Doll, 2004; West et al., 2005). In agreement with the conflict monitoring theory of ACC functioning (Botvinick, Cohen, & Carter, 2004; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004), source localization analyses have identified regions within the ACC as the potential generator of both the N2 (van Veen & Carter, 2002) and N450 (West & Alain, 2000b). In addition, recent ERP studies have highlighted the role of PFC regions (Markela-Lerenc et al., 2004) and functional coupling between ACC and PFC regions (Hanslmayr et al., 2008) during response conflict.

Based on prior literature we hypothesized that, relative to healthy control subjects, unmedicated participants with MDD would show reduced conflict monitoring abilities, which would be manifested as (1) decreased performance in incongruent (i.e., high-conflict), but not congruent, trials; (2) decreased scalp N2 and N450 amplitudes; and (3) reduced ACC and DLPFC activation following incongruent trials. Because the N450 has been more strongly associated with the Stroop interference effects than the N2 (e.g., West et al., 2005), the primary hypotheses focused on the N450 component. To test these hypotheses, we performed novel analyses on a recently published dataset, in which we previously investigated error processing dysfunction in depression (Holmes & Pizzagalli, 2008). While our prior analyses focused on response-locked ERPs and error processing, the current study considers ERPs time-locked to the Stroop stimuli, giving us the opportunity to investigate conflict monitoring dysfunction in depression, a topic not explored in our prior work.

2. Materials and Methods

2.1. Participants

The participant recruitment, assessment, and clinical characterization of this sample have been previously described in detail (Holmes & Pizzagalli, 2008). Briefly, 45 right-handed participants between the age of 18 and 55 years and with normal or corrected vision were recruited from the Boston area. For MDD subjects, inclusion criteria included: meeting DSM-IV diagnosis for current MDD (American Psychiatric Association, 1994), as established by a Structured Clinical Interview for DSM-IV, Patient Edition (SCID-P; First, Spitzer, Gibbon, & Williams, 2002); absence of any other Axis I comorbidity, with the exception of anxiety disorders (simple phobia n = 1); absence of psychotic medication usage within 2 weeks of the initial session (4 weeks for neuroleptics and benzodiazepines, 6 weeks for fluoxetine, and 6 months for dopaminergic drugs); no evidence of current or past psychiatric symptomatology, and no history of electroconvulsive therapy, seizures, and/or head injuries resulting in loss of consciousness. Healthy comparison participants were included if they had no current or past psychopathology, neurological disorders, and/or head injuries. Five participants were lost due to discovery of exclusionary criteria at the SCID interview (n = 4) or non-compliance (n = 1). The final sample consisted of 20 MDD subjects and 20 healthy controls. MDD and control subjects did not differ with respect to gender [% females: 35% vs. 30%; χ²(1) = 0.92, p > 0.34, age [36.00 ± 12.16 vs. 28.80 ± 9.87 years; t(18) = 0.51, p > 0.15, ethnicity [% Caucasian; 80% vs. 70%; χ²(1) = 0.53, p > 0.46], or education [15.65 ± 1.87 vs. 15.65 ± 1.93 years; t(38) = 0.00, p > 0.99] as expected the MDD participants reported significantly increased levels of depressive symptoms, as assessed by the Beck Depression Inventory-II score (BDI-II; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) [22.55 ± 9.23 vs. 24.5 ± 3.31, [F(38) = 9.17, p = 0.005].

After receiving a detailed study description, participants provided written informed consent to a research protocol approved by the Committee on the Use of Human Subjects at Harvard University. Participants were compensated $10/h for their participation.

2.2. Task and Procedure

After study eligibility was assessed, participants were invited for an experimental session that occurred within 1 week of the clinical interview and involved collection of behavioral and electrophysiological data while participants performed a modified Stroop task. The task consisted of pseudo-random presentations of three words: (RED, GREEN, and BLUE) printed in one of three colors of ink (red, green, and blue). Trials were either congruent (i.e., the word and the color matched) or incongruent. Participants were instructed to use their index, middle, and ring fingers of their right hand to respond through a button press, as quickly and accurately as possible, to each probe’s ink color. Trials began with the presentation of a fixation cross (250 ms), followed by a Stroop probe (150 ms) and a jittered inter-trial interval (ITI; 1850–1950 ms).

Before the task participants completed two practice blocks (24 trials each). Reaction times (RT) from the second practice block were used to determine a threshold for late responses (see below). For the actual task, feedback regarding performance was added to reduce potential confounds related to group differences in error awareness. When participants responded correctly within the individually titrated response window (equal to 85% of each participant’s mean RT during the second practice block), positive feedback (a schematic smiling face) was presented for 250 ms. Negative feedback (a schematic frowning face) was presented for 250 ms if participants responded outside of the response window and/or made an incorrect response. To account for possible performance changes over time, the response window threshold was recalculated at the middle and end of each block. To reduce the likelihood that physiological activity associated with the previous trial would interfere with the current trial, participants were instructed to think about something different between trials.

In this condition, participants responded as quickly and accurately as possible using their index, middle, and ring fingers of their right hand to respond through a button press, as quickly and accurately as possible, to each probe’s ink color. Trials began with the presentation of a fixation cross (250 ms), followed by a Stroop probe (150 ms) and a jittered inter-trial interval (ITI; 1500–1600 ms).

Before the task participants completed two practice blocks (24 trials each). Reaction times (RT) from the second practice block were used to determine a threshold for late responses (see below). For the actual task, feedback regarding performance was added to reduce potential confounds related to group differences in error awareness. When participants responded correctly within the individually titrated response window (equal to 85% of each participant’s mean RT during the second practice block), positive feedback (a schematic smiling face) was presented for 250 ms. Negative feedback (a schematic frowning face) was presented for 250 ms if participants responded outside of the response window and/or made an incorrect response. To account for possible performance changes over time, the response window threshold was recalculated at the middle and end of each block. To reduce the likelihood that physiological activity associated with the previous trial would interfere with the current trial, participants were instructed to think about something different between trials.
128-channel ERPs were recorded using the Geodesic Sensor Net system (Electrical Geodesic, Inc., OR). Impedances were kept below 50 kΩ and a 250-Hz sampling rate (bandwidth: 0.01–100 Hz) was used with the vertex electrode (Cz) as the recording reference.

2.4. Data reduction

2.4.1. Behavioral data

Only trials in which participants made a response were considered. To reduce the potential effect of outliers, trials with RTs (after In transformation) beyond individual mean ± 3SD for each trial type were excluded (on average, 0.28 ± 0.24%)

The main analyses focused on behavioral adjustments related to the occurrence of response-conflict. To this end, the Stroop and Gratton effects were calculated. The Stroop effect is a measurement of interference elicited by the incongruent trials, relative to congruent. It is calculated as: \[ \text{SRT}_{\text{incongruent}} - \text{SRT}_{\text{congruent}}\]

and Accuracy, where Accuracy congruent trials = Accuracy incongruent trials. 

2.4.2. LORETA data

Data were obtained with Brain Vision software (Brain Products GmbH, Gilching, Germany). Artifacts were removed through independent component analyses (Jung et al., 2000). Individual channels with corrupted signal were replaced through spatially weighted linear interpolations. Subsequent semi-automatic artifact detection was performed to identify remaining artifacts (maximal amplitude: ±75 μV; within-segment absolute amplitude difference: 150 μV; gradients: 50 μV). Stimulus-locked ERPs were computed 200 ms prior to and 924 ms following the presentation of a Stroop probe. Mirroring the behavioral data analyses, ERPs were computed only for trials in which participants made a correct response. Data were then band-pass filtered (0–30 Hz, 12 dB/octave), baseline-corrected (−200 ms to −100 ms pre-probe onset), and re-referenced to an average reference. Grand-mean ERP waveforms were calculated by averaging across data conditions and groups.

Based on prior studies using the Stroop or related tasks, analyses focused on the N2 and, particularly, N450 component, which were empirically defined using a space-oriented bootstrapping segmentation procedure (Koenig & Lehmann, 1996). This procedure was used to define the start and end points of “microstates”, which are periods of stable field configurations assumed to index specific brain functions. This was accomplished by calculating, at each time frame, the Global Map Dissimilarity (GMD) index. GMD is a reference-free, single-value variable that assesses the difference in distribution between two successive maps (Lehmman & Skrandies, 1984), whose values can vary between 0 (when two successive maps are identical) and 2 (when two maps have identical topography but reversed polarity). The resulting dissimilarity peaks were used to locate the occurrence of a microstate. For each microstate, Global Field Power (GFP) peaks were then identified. GFP is computed as the average standard deviation within a given surface map (Lehmman & Skrandies, 1984); GFP peaks are hypothesized to represent points of maximal neuronal activity, and thus offer optimal signal-to-noise ratio.

The resulting microstates were confirmed through visual inspection of the surface data. Next, N2 and N450 amplitudes and latencies were extracted from sensors showing maximal deflections (FC1, FC2, FCz, C1, Cz, C2, C1p, C2p, and Cp2). The N2 and N450 amplitudes were calculated as the average voltage amplitude 136–240 ms (N2) and 340–692 ms (N450), respectively, following a congruent and incongruent trial.

This choice was based on prior recommendations suggesting the use of mean amplitudes to characterize ERP waveforms, particularly those showing sustained unfolding (Luck, 2005, 234–235). The MDD and control group did not differ in the number of segments available for the ERP analyses (incongruent: 239.25 ± 40.43 vs. 253.9 ± 29.42, \( t(38) = 1.31, p > 0.19 \); congruent: 467.10 ± 61.28 vs. 467.65 ± 53.82, \( t(38) = 0.52, p > 0.60 \).

2.4.3. LORETA data

In case of significant scalp group differences, Low-Resolution Electromagnetic Tomography (LORETA; Pascual-Marqui, Michel, & Lehmann, 1994; Pascual-Marqui et al., 1999) was used to estimate the three-dimensional intracranial current density at the times of maximal response-conflict activity (see Pizzagalli, 2007 for a review of the theoretical assumptions, mathematical implementation, and cross-modal validation of the LORETA algorithm against hemodynamic neuroimaging techniques).

For each participant and condition, LORETA solutions were computed within the solution space (2394 voxels with a 7-mm³ resolution) at the times of maximal Global Field Power (GFP; Lehmann & Skrandies, 1984) within the N2 and N450 time windows. Prior to statistical analyses, LORETA activity was normalized to the total current density of 1 and log-transformed.

2.5. Statistical analyses

2.5.1. Behavioral data

Exploratory analyses revealed no significant effects of gender or ethnicity; therefore, these variables were not further considered. For accuracy and RT scores, a separate mixed 2 x 2 analysis of variance (ANOVA) with Group (MDD subjects, controls) as a between-subject factor and Condition (incongruent, congruent) as repeated measure was conducted to investigate Stroop effects. For the Gratton effect, the performance for incongruent trials following a congruent relative to incongruent trial was entered. As stated above, only trials following correct trials were considered.

2.5.2. Scalp ERP data

For the N2, a mixed 2 x 2 x 3 ANOVA with Group, Condition (incongruent, congruent), Group x Condition (incongruent, congruent) and latency as factors was significant across the entire ERP latency range of 340–692 ms. Post hoc Newman–Keuls tests were performed in case of significant ANOVA findings. Effect sizes are reported in the form of partial eta squared (\( \eta^2 \)) values.

2.5.3. LORETA data

Source localization analyses were conducted to follow-up significant findings from the scalp analyses. Two sets of analyses were performed. In the first set, analyses focused on activation at the time of GPF peaks, which yield optimal signal-to-noise ratio. For each identified GPF peak (N2: 212 ms; N450: 388 and 620 ms), voxelwise 2 x 2 ANOVAs with Group and Condition (incongruent, congruent) were performed on current density values using in-house matlab software. The output was thresholded at \( p < 0.01 \), uncorrected for multiple comparisons with a minimum cluster size of five contiguous voxels. To further avoid Type I errors, only findings hypothesizing regions (i.e., ACC, DLPFC) were considered. In the second set, LORETA analyses were repeated using extended time windows identical to the ones used for the scalp data.

3. Results

3.1. Behavioral data

3.1.1. Stroop effects

For accuracy, the main effect of Condition was significant \[ F(1, 38) = 44.17, p < 0.001; \] partial \( \eta^2 = 0.54 \), due to the expected higher accuracy for congruent (0.93 ± 0.05) relative to incongruent (0.86 ± 0.08) trials. The main effect of Group \[ F(1, 38) = 0.81, p = 0.37; \] partial \( \eta^2 = 0.02 \) and the Group x Condition \[ F(1, 38) = 2.61, p = 0.12; \] partial \( \eta^2 = 0.06 \) interaction were not significant. Accordingly, MDD (0.89 ± 0.06) and comparison subjects (0.90 ± 0.06) did not differ in their overall accuracy, confirming that a comparable number of data were available for the ERP computation between the groups.

For RT scores, the main effect of Condition was significant \[ F(1, 38) = 61.42, p < 0.001; \] partial \( \eta^2 = 0.62 \), due to the expected shorter RT for congruent (467.37 ± 56.48 ms) than incongruent (536.44 ± 102.44 ms) trials. Of primary importance for the current hypotheses, this effect was qualified by a significant Group x Condition interaction \[ F(1, 38) = 6.01, p < 0.02; \] partial \( \eta^2 = 0.14 \). Post hoc Newman–Keuls tests clarified that this interaction was due to significantly longer RT for MDD relative to control subjects for incongruent \( p < 0.001 \), but not congruent \( p > 0.60 \), trials (Table 1A). Although both groups showed a significant Stroop effect (incongruent > congruent; \( p < 0.002 \), the significant interaction indicates that MDD subjects had significantly larger Stroop effects compared to control subjects \[ 90.67 ± 70.29 ms vs. 47.47 ± 35.69; \( t(38) = 2.45, p < 0.03 \) \]. The main effect of Group was not significant \[ F(1, 38) = 1.33, p > 0.25; \] partial \( \eta^2 = 0.03 \), strengthening the interpretation of selective deficits in MDD participants.

3.1.2. Conflict-adaptation (Gratton) effects

For both RT and accuracy, the main effect of Condition was significant \[ RT: F(1, 38) = 22.91, p < 0.001; \] partial \( \eta^2 = 0.38 \); accu-
racy: \( F(1, 38) = 6.10, p < 0.02 \); partial \( \eta^2 = 0.14 \), with participants responding less accurately and more slowly for incongruent trials following congruent trials (RT: 349.35 \( \pm \) 71.25 ms; accuracy: 0.90 \( \pm \) 0.05) than incongruent trials following incongruent trials (RT: 331.23 \( \pm \) 69.43 ms; accuracy: 0.91 \( \pm \) 0.06). No other effects emerged. Taken together, the behavioral findings indicate that the current paradigm elicited the intended behavioral effects, and highlight RT slowing in MDD subjects specific to high-conflict (incongruent) trials.

3.2. Scalp ERP analyses

3.2.1. N2

Table 1B summarizes all significant effects emerging from the \( \text{Group} \times \text{Condition} \times \text{Laterality} \times \text{Caudality} \) ANOVA considering averaged N2 amplitudes. For the sake of brevity, only effects involving Group and Condition will be presented in detail (other findings are available upon request). Briefly, the main effect of Group \( [F(1, 38) = 4.88, p < 0.035; \text{partial } \eta^2 = 0.114] \) was significant, due to overall more negative N2 amplitudes for control \((−0.05 \pm 0.40 \mu V)\) than MDD \((0.86 \pm 0.46 \mu V)\) subjects. This effect was qualified by significant \( \text{Group} \times \text{Condition} \times \text{Laterality} \ [F(2, 76) = 3.70, p < 0.038; \text{partial } \eta^2 = 0.089]\) and \( \text{Group} \times \text{Laterality} \times \text{Caudality} \ [F(4, 152) = 2.81, p < 0.045; \text{partial } \eta^2 = 0.069]\) interactions. Lower-order ANOVAs were performed to clarify these effects.

For the \( \text{Group} \times \text{Condition} \times \text{Laterality} \) interaction, follow-up \( \text{Group} \times \text{Condition} \) ANOVAs revealed that the main effect of Group was significant for left, central, and right sensors (all \( ps < 0.05 \)), whereas the \( \text{Group} \times \text{Condition} \) interaction emerged only for the right hemisphere \( [F(1, 38) = 4.77, p < 0.035; \text{partial } \eta^2 = 0.112]\). Post hoc Newman–Keuls tests revealed that this interaction was due to more negative N2 to incongruent than congruent stimuli for control \((p < 0.013)\) but not MDD \((p > 0.63)\) subjects (Table 1B). Moreover, for either stimulus, controls had significantly larger N2 than MDD subjects \((ps < 0.0002; \text{Fig. 1A})\). For the \( \text{Group} \times \text{Laterality} \times \text{Caudality} \) interaction, separate \( \text{Group} \times \text{Caudality} \) ANOVAs performed on N2 value for each laterality level revealed no effects involving Group or Condition.

3.2.2. N450

Table 2B lists all significant effects emerging from the \( \text{Group} \times \text{Condition} \times \text{Laterality} \times \text{Caudality} \) ANOVA on the N450 amplitudes. As above, only effects involving Group and Condition are reported. Findings of interest include a main effect of Condition \( [F(1, 38) = 25.60, p < 0.001; \text{partial } \eta^2 = 0.403]\) due to more

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Summary of behavioral, ERP, and LORETA findings for control ((n = 20)) and depressed ((n = 20)) subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control subjects</td>
</tr>
<tr>
<td>(A) Behavioral performance</td>
<td></td>
</tr>
<tr>
<td>Accuracy incongruent</td>
<td>0.88 ( \pm ) 0.05</td>
</tr>
<tr>
<td>Accuracy congruent</td>
<td>0.93 ( \pm ) 0.05</td>
</tr>
<tr>
<td>Stroop effect accuracy</td>
<td>−0.05 ( \pm ) 0.029</td>
</tr>
<tr>
<td>RT incongruent</td>
<td>511.61 ( \pm ) 69.50</td>
</tr>
<tr>
<td>RT congruent</td>
<td>464.14 ( \pm ) 47.23</td>
</tr>
<tr>
<td>Stroop effect RT</td>
<td>47.47 ( \pm ) 35.69</td>
</tr>
<tr>
<td>(B) Scalp ERP data</td>
<td></td>
</tr>
<tr>
<td>N2 incongruent</td>
<td>0.29 ( \pm ) 0.89</td>
</tr>
<tr>
<td>N2 congruent</td>
<td>0.46 ( \pm ) 0.93</td>
</tr>
<tr>
<td>N450 incongruent</td>
<td>2.38 ( \pm ) 1.67</td>
</tr>
<tr>
<td>N450 congruent</td>
<td>2.91 ( \pm ) 1.60</td>
</tr>
<tr>
<td>(C) LORETA data ((620 \text{ ms}))</td>
<td></td>
</tr>
<tr>
<td>dACC incongruent</td>
<td>−3.70 ( \pm ) 0.17</td>
</tr>
<tr>
<td>dACC congruent</td>
<td>−3.75 ( \pm ) 0.19</td>
</tr>
<tr>
<td>( \Delta (\text{dACC}) )</td>
<td>0.05 ( \pm ) 0.11</td>
</tr>
<tr>
<td>Left DLPFC incongruent</td>
<td>−3.31 ( \pm ) 0.17</td>
</tr>
<tr>
<td>Left DLPFC congruent</td>
<td>−3.36 ( \pm ) 0.18</td>
</tr>
<tr>
<td>( \Delta (\text{left DLPFC}) )</td>
<td>0.50 ( \pm ) 0.13</td>
</tr>
</tbody>
</table>

n.s. = non-significant.
| *The difference score (incongruent – congruent) was calculated at the time of maximal GFP \((620 \text{ ms})\) within the N450 time window peak.|

Fig. 1. Stimulus-locked grandmean waveforms for congruent and incongruent probes. In (A) the waveforms were averaged across electrodes FC2, C2, Cp2 to reflect the N2 ANOVA findings. In (B) waveforms were averaged across FC1, FC2, FC2, C1, Cz, C2, Cp1, CP2, and Cp2 to reflect the N450 ANOVA findings.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Summary of ANOVA findings for the (A) N2 and the (B) N450 component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrast</td>
<td>F</td>
</tr>
<tr>
<td>(A) N2</td>
<td></td>
</tr>
<tr>
<td>Grp</td>
<td>4.881</td>
</tr>
<tr>
<td>Cond</td>
<td>3.826</td>
</tr>
<tr>
<td>Lat</td>
<td>23.170</td>
</tr>
<tr>
<td>Grp \times Lat \times Cond</td>
<td>3.702</td>
</tr>
<tr>
<td>Grp \times Lat \times Caud</td>
<td>2.810</td>
</tr>
<tr>
<td>(B) N450</td>
<td></td>
</tr>
<tr>
<td>Cond</td>
<td>25.601</td>
</tr>
<tr>
<td>Lat</td>
<td>25.601</td>
</tr>
<tr>
<td>Caud</td>
<td>19.89</td>
</tr>
<tr>
<td>Grp \times Cond</td>
<td>4.131</td>
</tr>
<tr>
<td>Con \times Lat</td>
<td>3.494</td>
</tr>
<tr>
<td>Lat \times Caud</td>
<td>4.127</td>
</tr>
<tr>
<td>Grp \times Lat \times Caud</td>
<td>3.292</td>
</tr>
</tbody>
</table>

*Grp = Group; Caud = Caudality; Lat = Laterality; Cond = Condition.*
negative N450 amplitude for incongruent (2.64±1.41 μV) than congruent (3.02±1.38 μV) trials, a significant Group × Condition interaction [F(1, 38) = 4.13, p < 0.05; partial η² = 0.098], and a significant Group × Laterality × Caudality interaction [F(4, 152) = 3.29, p < 0.021; partial η² = 0.080].

For the Group × Condition interaction, post hoc tests revealed that, consistent with the N2 findings, MDD subjects failed to show differentiation between conditions (p > 0.09), whereas controls showed more negative N450 waveforms for incongruent than congruent trials (p < 0.001). Moreover, for both congruent (p < 0.045) and incongruent (p < 0.001) trials, control subjects had significantly more negative N450 than MDD subjects (Table 1B), an effect that was particularly evident for the incongruent condition (Fig. 1B).

For the Group × Laterality × Caudality interaction, lower-order Group × Laterality ANOVAs were performed for each level of Caudality separately. However, no further significant effects involving Group emerged.

A closer evaluation of the ERP waveforms reveals that two peaks were present within the N450 microstates (Fig. 1B). Based on this observation, scalp analyses were repeated within an “early” and “late” N450 window. The early window (340–436 ms) was anchored (±48 ms) around the first GFP peak (388 ms), whereas the late window encompassed the remaining time period (436–692 ms). For the earlier N450 peak, no effects involving Group emerged (all Fs < 1.76, all ps > 0.18). For the later N450 peak, the Group × Condition interaction was replicated [F(1, 38) = 4.19, p < 0.048].

3.3. LORETA data

3.3.1. N2

No effects involving Group emerged from a priori regions when performing voxel-by-voxel Group × Condition ANOVA on current density computed at the time of the maximal N2 GFP peaks (212 ms). Similarly, no significant effects emerged when considering the extended N2 time window used for the surface scalp analyses (136–240 ms).

3.3.2. N450

As above, no effects involving Group emerged from a priori regions when considering the early N450 GFP peak (388 ms). However, consistent with our hypothesis, the analysis for the later GFP peak (620 ms) revealed a significant Group × Condition effect within the dACC [BA24/32; 10 voxels; F(1, 38) = 9.54, p < 0.004; partial η² = 0.20; Fig. 2A], indicating that the two groups differed significantly in their activation to incongruent relative to congruent trials. Post hoc testing confirmed that the MDD and control groups showed opposite patterns of dACC activation. As shown in Fig. 2A, controls showed a trend for higher current density for incongruent than congruent trials (p < 0.06), whereas MDD subjects showed a trend in the opposite direction (p < 0.09). Moreover, compared to controls, MDD subjects displayed decreased dACC current density for incongruent (p < 0.07), but not congruent (p > 0.11) trials (Table 1C).

The only other finding emerging was a highly significant Group × Condition interaction in a left DLPFC cluster [BA10/46; F(1, 38) = 8.79, p < 0.006; partial η² = 0.18; Fig. 2B], which however included only three voxels, and thus missed the cluster threshold. While this finding should be interpreted tentatively due to the limited cluster size, this region-of-interest (ROI) was explored further in light of a priori hypotheses concerning DLPFC dysfunction in depression. Post hoc tests revealed that this effect was due to significantly lower current density in response to incongruent trials for MDD compared to control subjects (p < 0.001; Fig. 2B). Moreover, unlike controls, MDD subjects showed an unexpected pattern of significantly increased current density for the congruent than incongruent trials (p < 0.009). To test the specificity of these findings in terms of laterality, we extracted current density from the homologous right DLPFC region, and performed a Group × Condition × Hemisphere ANOVA. This omnibus ANOVA confirmed a significant Group × Condition interaction [F(1, 38) = 11.78, p < 0.001], which was driven by a left hemispheric current density reduction in the MDD group for incongruent trials. Although no significant effects involving Group emerged when considering the right DLPFC cluster (all Fs < 2.59, all ps > 0.12), it is important to emphasize that the Group × Condition × Hemisphere was not significant [F(1, 38) = 0.93, p > 0.34], indicating that the DLPFC findings were not specific to the left hemisphere.

Finally, to ensure that the N450 LORETA findings were not confounded by potential group differences in N450 latency and to maximize comparability between the scalp and LORETA analyses, control analyses were performed using the time frame utilized in the scalp analyses. Mirroring null findings for the early GFP peak, no significant clusters emerged for the early N450 window (340–432 ms). However, when considering the late N450 window (436–692 ms), the Group × Condition effect was confirmed for both the dorsal ACC [F(1, 38) = 6.49, p < 0.015; partial η² = 0.15] and DLPFC [F(1, 38) = 9.91, p < 0.003; partial η² = 0.21].

3.4. Correlation between behavioral and LORETA data

For controls, dACC current density to incongruent stimuli correlated with incongruent accuracy (Pearson r = 0.619, p < 0.004; Fig. 3A), suggesting that stronger dACC recruitment was associated with better performance on high-conflict trials. For the MDD group, this correlation was not significant (Pearson r = 0.239, p > 0.31; Fig. 3B), although a Fisher test revealed that correlations between groups were not significantly different (z = 1.40; p > 0.05). No correlations emerged when considering the left DLPFC cluster for either group.

3.5. Behavioral performance as a function of dACC activation

Based on the current ERP and prior fMRI findings indicating that ACC activation during high-conflict trials is associated with adaptive behavioral adjustments (Kerns et al., 2004), we reasoned that participants with MDD showing the strongest dACC activation 620 ms post-conflict would display the smallest conflict monitoring deficit. To test this hypothesis, difference scores were calculated for the dACC ROI emerging from the N450 GFP peak (incongruent – congruent). A median-split procedure was then applied to identify control and MDD participants who displayed the highest and lowest dACC activation (MDD low: −0.14 ± 0.07; MDD high: 0.03 ± 0.05; control low: −0.05 ± 0.05; control high: 0.14 ± 0.06). Next, independent sample t-tests were conducted to compare the Stroop interference effect [RTt(18) = 1.40; p > 0.29]. Interestingly, MDD subjects showing the highest dACC activation 620 ms post-conflict did not differ from either the high t(18) = 1.08, p > 0.30] or low t(18) = 0.21, p > 0.84] dACC control sub-groups.

For RT, MDD subjects with the lowest dACC activation displayed a significantly higher Stroop interference relative to MDD subjects with high dACC activation [t(18) = 4.38, p < 0.001; Fig. 4A]. These low dACC MDD participants also showed significantly higher interference effect compared to both control sub-groups [low dACC: t(18) = 4.21, p < 0.003; high dACC: t(18) = 3.44, p < 0.003; Fig. 4A], which did not differ from each other [t(18) = 1.09, p > 0.29]. Interestingly, MDD subjects showing the highest dACC activation 620 ms post-conflict did not differ from either the high t(18) = 1.08, p > 0.30] or low t(18) = 0.21, p > 0.84] dACC control sub-groups.
Fig. 2. (A) dACC cluster [BA24/32; 10 voxels; peak voxel MNI coordinates: $x = -10, y = 31, z = 29; F(1, 38) = 9.54, p < 0.004; \text{partial } \eta^2 = 0.20$], and (B) left DLPFC cluster [BA10/46; three voxels; peak voxel MNI coordinates: $x = -45, y = 45, z = 15; F(1, 38) = 8.79, p < 0.0006; \text{partial } \eta^2 = 0.18$] emerging from the Group $\times$ Condition interaction 620 ms following the presentation of the Stroop Probe. Mean (and S.E.) current density within the ROI is shown for the MDD ($n = 20$) and control ($n = 20$) participants.
Fig. 3. Scatter plot between the current density (averaged across voxels) within the dACC cluster 620 ms following the presentation of an incongruent probe and incongruent trial accuracy for the (A) control subjects ($r = 0.619, p < 0.004$) and (B) MDD subjects ($r = 0.239, p > 0.31$).

Similar findings emerged when accuracy was considered. MDD subjects with low dACC activation displayed a significantly higher Stroop effect relative to the MDD subjects with high dACC activation ($[t (18) = 2.24, p < 0.04$; Fig. 4B), as well as both control sub-groups [low dACC: $t (18) = 2.14$; high dACC: $t (18) = 2.19, ps < 0.05$; Fig. 4B]. As above, MDD subjects with high dACC activation did not differ from either the high $[t (18) = 0.31, p > 0.76]$ or low $[t (18) = 0.38, p > 0.71]$ dACC control sub-group. Finally, no differences emerged between the high and low dACC control sub-groups $[t (18) = 0.09, p > 0.93]$. Importantly, these effects were not due to differences in depression severity, since the high and low dACC MDD groups did not differ in their BDI scores $[21.30 \pm 7.50$ vs. $23.80 \pm 3.31; t (18) = 0.60, p > 0.56]$.

Table 3
Summary of unpaired t-tests assessing behavioral performance in subjects with relatively low vs. high dACC activation 620 ms post-conflict

<table>
<thead>
<tr>
<th></th>
<th>Stroop effect RT</th>
<th>Stroop effect accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>MDD subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low ACC ($n = 10$)</td>
<td>38.86 ± 67.73$^{a,b}$</td>
<td>$-0.05 \pm 0.03^b$</td>
</tr>
<tr>
<td>High ACC ($n = 10$)</td>
<td>41.49 ± 21.65$^a$</td>
<td>$-0.04 \pm 0.04^a$</td>
</tr>
<tr>
<td>Control subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low ACC ($n = 10$)</td>
<td>56.09 ± 36.81$^b$</td>
<td>$-0.05 \pm 0.03^b$</td>
</tr>
<tr>
<td>High ACC ($n = 10$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the dACC cluster, a difference score was calculated (incongruent – congruent). Stroop effects were calculated as: (RT$_{\text{incongruent trials}}$ – RT$_{\text{congruent trials}}$) and (Accuracy$_{\text{congruent trials}}$ – Accuracy$_{\text{incongruent trials}}$). Sub-groups differ at: $^a p < 0.001$; $^b p < 0.003$; $^c p < 0.003$; $^d p < 0.04$; $^e p < 0.05$; $^f p < 0.05$.

4. Discussion

The goal of the present study was to examine behavioral and electrophysiological correlates of response conflict deficits in unmedicated subjects with major depression. The following findings emerged. First, depressed subjects were characterized by significantly increased RT interference effects. Additional analyses clarified that this impairment was due to RT slowing specific to incongruent trials, and emerged in the context of no group differences in accuracy. Thus, depressed subjects had slowed performance exclusively in high-conflict trials featuring the presence of competing response tendencies. The present behavioral findings join prior observations highlighting increased interference effects in MDD for both emotional (Mitterschiffthaler et al., 2007) and non-emotional (Ottowitz et al., 2002) Stroop tasks.

Second, unlike controls, MDD subjects failed to show larger negative deflections for incongruent than congruent trials at both early (N2) and later (N450) stages of the information processing flow. In fact, patients showed no differentiation between incongruent and congruent trials, and were characterized by significantly reduced N2 and N450 amplitudes, relative to controls, suggesting that response conflict in depression might be impaired at both the
stimulus processing (N2) and response stages (N450). These findings are intriguing, particularly since prior ERP studies have shown that the N2 and N450 are attenuated in populations where the ability to inhibit competing word information on incongruent trials is compromised (Mayes, Molfese, Key, & Hunter, 2005; McNeely, West, Christensen, & Alain, 2003; West & Alain, 2000a). Moreover, since in healthy controls the amplitude of the N450 varies as a function of task difficulty (West & Alain, 2000b), the present data might reflect a failure on the part of the MDD group to adequately recruit the cognitive control necessary to account for changing task contingencies.

Third, consistent with current conceptualizations of executive control system implicated in the detection of and adjustments to contingencies (Carter & van Veen, 2007), and in line with previous fMRI and ERP research (e.g., Hanslmayr et al., 2008; Kerns et al., 2004; Liotti, Woldorff, Perez, & Mayberg, 2000), source localization analyses indicated that MDD subjects had reduced activation within dACC and left DLPFC regions 620 ms after stimulus presentation. Follow-up analyses suggested that this relative difference was driven by reduced activation for MDD subjects in response to incongruent trials. Reduced recruitment of dACC and left DLPFC regions 620 ms after presentation of a Stroop stimulus is interesting, particularly in light of recent findings of increased functional coupling between the ACC and left PFC ~600 ms following the presentation of incongruent Stroop stimuli in healthy controls (Hanslmayr et al., 2008). In the present study, stronger dorsal ACC recruitment 620 ms after presentation of incongruent trials correlated with better performance in control subjects, a pattern that was absent in patients (correlations for control and MDD subjects were, however, not significantly different). Because the mean RT for control subjects was 847.87 ms, it is likely that dACC activation at the N450 time point reflected sustained recruitment needed to successfully respond during high-conflict trials.

Finally, MDD subjects showing the lowest level of relative dACC activation to incongruent trials were characterized by the largest Stroop interference effects (relatively increased slowing and decreased accuracy for incongruent trials). Notably, MDD subjects with relatively stronger dACC recruitment did not differ from controls in terms of their interference effects. These findings are interesting, particularly since the two MDD subgroups had very similar depression severity (BDI) scores (high dACC: 21.30 ± 7.50; low dACC: 23.80 ± 3.31), and did not differ on any other self-report, clinical (e.g., number of prior episodes, duration of current episode), or demographic measure. Altogether, these data support the hypothesis that major depression is characterized by reduced response conflict abilities, likely coupled with impaired recruitment of cognitive control, and dysfunction within frontocingulate pathways implicated in action monitoring and executive functioning. Moreover, initial evidence indicates that response conflict dysfunctions might be restricted to a sub-group of MDD subjects showing the most pronounced dACC blunting to response conflict. Although the present data highlight the heterogeneous nature of MDD, further research will be necessary to examine what aspects of depressive symptomatology might differentiate these sub-groups.

While the findings emerging from the current analyses are consistent with prior neuroimaging studies that have described decreased dACC and left DLPFC activation during various executive tasks in depressed subjects (Elliott, Baker, et al., 1997; Elliott, Sahakian, et al., 1997; George et al., 1997; Okada, Okamoto, Morinobu, Yamawaki, & Yokota, 2003; for review see Davidson et al., 2002), it is important to emphasize that depression has been associated with both frontocingulate hypo as well as hyperactivity during executive tasks. A closer look at prior findings reveals, however, that the direction of frontocingulate dysfunction might be related to task performance. Specifically, studies reporting higher activation in the left DLPFC (Harvey et al., 2005; Matsuo et al., 2007; Wagner et al., 2006; Walter, Wolf, Spitzer, & Vasic, 2007) and dACC (Harvey et al., 2005; Mitterschiffthaler et al., 2007; Rose, Simonotto, & Ebmeier, 2006; Wagner et al., 2006) in depression did not find any group differences in behavioral performance. Accordingly, it is possible that greater recruitment of frontocingulate regions is required to achieve behavioral performance equivalent to control subjects (Killgore, Gruber, & Yurgelun-Todd, 2007; Wagner et al., 2006). Conversely, decreased frontocingulate activation has emerged in studies, in which depressed subjects showed impaired performance (the present study; Audenaert et al., 2002; Elliott, Baker, et al., 1997; Elliott, Sahakian, et al., 1997; Okada et al., 2003; but see Harvey et al., 2005; Hugdahl et al., 2004).

While these data provide evidence for a dysregulated conflict monitoring system in MDD, several limitations should be noted. First, although groups differed in their Stroop effects, no differences emerged when considering the Gratton effects. One explanation for this null finding is that the presentation of task-relevant feedback interfered with the temporally sensitive nature of this effect. Second, due to the small size of the present sample and the fact that all patients were unmedicated, we were unable to address the potential effect of depression subtypes and/or psychotropic medication usage on the action monitoring. Since there have been inconsistent findings of action monitoring deficits in MDD, possibly due to the diagnostic heterogeneity and/or pharmacological treatment effects (Markela-Lerenc, Kaiser, Fiedler, Weisbrod, & Mundt, 2006), further studies in this area will be necessary. Finally, it is important to emphasize that while the DLPFC finding was consistent with the hypotheses, the resulting region was smaller than the minimum cluster threshold and a formal laterality test revealed that DLPFC dysfunctions in MDD were not specific to the left hemisphere. Thus, caution should be exerted in interpreting these findings and implications in future studies are warranted.

In spite of these limitations, the present behavioral and electrophysiological findings confirm that depression is characterized by executive dysfunction and dysregulation within frontocingulate pathways critically implicated in conflict monitoring and cognitive control. Of note, prior analyses of this dataset revealed hyperactivation in rostral ACC regions to errors in depressed subjects (Holmes & Pizzagalli, 2008), emphasizing the presence of a multifaceted dysfunction of action monitoring system in depression, as well as dissociable roles for the rostral and dorsal subdive.
sions of the cingulate (Bush, Luu, & Posner, 2000). Understanding relations among clinical phenomenology, executive function, and functional/structural integrity of frontocingulate pathways should remain an important goal of future studies.

Disclosure

Dr. Pizzagalli has received research support from GlaxoSmithKline and Merck and Co., Inc. for studies unrelated to the present research. Mr. Holmes reports no competing interests.

Acknowledgments

Supported by a Sackler Scholar in Psychobiology Research Grant, a National Institute of Health Training Grant 1 F31MH078346 (AJH), and NIMH Research Grant R01MH68376 (DAP). The authors wish to thank Trinity Urban, Daniel G. Dillon, Diane L. Santesso, Ryan Bogdan, and Christen M. Deveney for their contributions and assistance with various aspects of this research.

References


