Social Cognitive and Affective Neuroscience Advance Access published June 17, 2014

doi:10.1093/scan/nsu060 SCAN (2014) 1 of 8

# Characterizing switching and congruency effects in the Implicit Association Test as reactive and proactive cognitive control

Joseph Hilgard,[1] Bruce D. Bartholow,[1] Cheryl L. Dickter,[2] and Hart Blanton[3]
[1]Department of Psychological Sciences, University of Missouri, Columbia, MO 65211-2500, [2]Department of Psychology, College of William and Mary, Williamsburg, VA 23187-8795, and [3]Department of Psychology, University of Connecticut, Storrs, CT 06269-1020, USA

**Recent research has identified an important role for task switching, a cognitive control process often associated with executive functioning, in the Implicit Association Test (IAT). However, switching does not fully account for IAT effects, particularly when performance is scored using more recent *d*-score formulations. The current study sought to characterize multiple control processes involved in IAT performance through the use of event-related brain potentials (ERPs). Participants performed a race-evaluative IAT while ERPs were recorded. Behaviorally, participants experienced superadditive reaction time costs of incongruency and task switching, consistent with previous studies. The ERP showed a marked medial frontal negativity (MFN) 250–450 ms post-stimulus at midline fronto-central locations that were more negative for incongruent than congruent trials but more positive for switch than for no-switch trials, suggesting separable control processes are engaged by these two factors. Greater behavioral IAT bias was associated with both greater switch-related and congruency-related ERP activity. Findings are discussed in terms of the Dual Mechanisms of Control model of reactive and proactive cognitive control.**

**Keywords:** implicit attitudes; cognitive control; event-related potentials; racial bias

The idea of implicit measures is quite attractive (see De Houwer *et al.*, 2009). Because research participants may not always want to or be capable of reporting their true attitudes about socially sensitive issues, measures that could reveal the types and strengths of evaluative associations without reliance on self-report promise to significantly advance the scientific study of attitudes. A primary concern with most measures based on self-report, and even measures based on observable behavior, is that participants can exert control over their responses in a way that obscures underlying evaluations. Thus, the ideal for an implicit measure is that it be structured so as to avoid the influence of control-related processes, whereby responses are based purely on automatically activated evaluations and associations.

The Implicit Association Test (IAT; Greenwald *et al.*, 1998) was designed with this goal in mind. The IAT is meant to assess the basic evaluative associations people hold toward attitude objects without relying on self-disclosure (Greenwald *et al.*, 2009). In the original race-evaluative IAT, respondents are asked to sort names according to the racial group they imply (e.g. 'Tyrell' is more common among Black men than White men) and words according to their positive/negative valence using only two response keys. Congruent trials are those in which stereotypically associated race categories and word valences share a response key (i.e. Black with negative; White with positive); incongruent trials are those in which race categories are paired with counter-stereotypic word valences (i.e. Black with positive; White with negative). Dozens of studies have shown that performance tends to be better [i.e. faster response time (RT)] for congruent than for incongruent trial blocks (see Greenwald *et al.*, 2009), a finding generally interpreted as evidence of an implicit bias against Blacks (or a preference for Whites). Moreover, individual differences in the size of this so-called 'IAT effect' (incongruent block RT minus congruent block RT) are often taken to indicate individual differences in the strength of underlying implicit biases (see Greenwald *et al.*, 2009).

Traditionally, performance on the IAT has been purported to reflect exclusively automatic processes (see De Houwer *et al.*, 2009), with responses 'under the control of automatically activated evaluations' (Greenwald *et al.*, 1998: 1464). In recent years this view has been challenged on both theoretical and empirical grounds (see Amodio and Mendoza, 2010; Sherman *et al.*, 2010; Teige-Mocigemba *et al.*, 2010). For example, using multinomial modeling techniques that attempt to estimate the extent to which various automatic and controlled processes contribute to behavioral responses, Sherman and colleagues (see Conrey *et al.*, 2005; Sherman *et al.*, 2008) have shown that control-related processes contribute significantly to IAT performance.

In particular, Klauer and colleagues (e.g. Mierke and Klauer, 2001; Klauer and Mierke, 2005; Klauer *et al.*, 2010) have demonstrated that a specific cognitive control ability—task switching—is critical for IAT performance. Task switching is required in any behavioral task in which respondents must use response rules that vary from one trial to the next according to some stimulus feature. For example, on each trial of the number–letter task (Rogers and Monsell, 1995), a number–letter pair (e.g. 7G) is presented above or below a central line. Respondents must classify the number as odd or even when the pair appears above the line, but they must classify the letter as consonant or vowel when the pair appears below the line. Task switching is required in the IAT because respondents must switch between semantically categorizing attitude objects on some trials (e.g. classifying names as Black or White) and evaluatively categorizing words (as good or bad) on other trials. When such switching effects are modeled, studies often find that performance on a typical race-evaluative IAT is predicted by an interaction of congruency and task switching, such that the poorest performance (i.e. slowest RTs) occurs on switch trials in the incongruent block (e.g. Mierke and Klauer, 2001, 2003).

From the perspective of attempting to understand control-related factors that contribute to IAT effects, such findings illustrate two critical issues. First, task switching alone does not account for the IAT

Received 20 February 2013; Revised 7 March 2014; Accepted 28 April 2014

This work was supported in part by a Life Sciences Graduate Fellowship to Joseph Hilgard and by Grant R01 AA020970 from the National Institute on Alcohol Abuse and Alcoholism to Bruce D. Bartholow.

Correspondence should be addressed to Bruce D. Bartholow, Department of Psychological Sciences, University of Missouri, 210 McAlester Hall, Columbia, MO 65211, USA. E-mail: BartholowB@missouri.edu.

effect; if it did, one would expect the typical main effect of congruency to be supplanted by a main effect of switching, and the two factors would not interact. Second, although switches occur in both the congruent and incongruent blocks, switching effects appear to operate differently in the two blocks. Together, these issues suggest that additional control-related processes not directly involved in switching but important for the congruency effect also contribute to IAT effects. Consistent with this idea, analysis of performance generally shows that responses are slower in the incongruent relative to the congruent block even on non-switching trials (see Klauer and Mierke, 2005).

To the extent that multiple control processes are involved in the IAT, Braver's (2012) Dual Mechanisms of Control model (DMC) could provide a useful conceptual framework for understanding these processes. The central thesis of the DMC framework is that cognitive control operates via two distinct operating modes. The first, 'proactive control', is the sustained maintenance of goal information in working memory that serves to bias information processing in a goal-congruent manner. The second, 'reactive control', is considered a late correction mechanism for dealing with cognitive and behavioral conflict as it arises (De Pisapia and Braver, 2006; Braver, 2012). An important assumption of the DMC model is that people determine a control strategy weighting proactive and reactive modes of control according to situational factors, particularly factors indicating the degree to which conflict or interference can be anticipated. Specifically, proactive control will be heightened when a high degree of conflict is expected, to maintain task goals in the face of challenging stimulus–response mappings. In contrast, unpredictable conflict should be associated with less proactive control (to conserve resources) but greater reactive control, so as to permit momentary adjustments in the face of conflict when it occurs. A number of studies involving both behavioral and neural measures have provided support for these basic assumptions (see Speer *et al.*, 2003; Burgess and Braver, 2010; West and Bailey, 2012).

Given the structure of the IAT, in which congruent and incongruent trials are presented in separate blocks, both of which involve switching between tasks, congruency and switching effects in the IAT would seem to map onto this proactive–reactive distinction, respectively. Relative to the congruent block, performance during the incongruent block requires heightened vigilance and sustained goal maintenance (i.e. proactive control), similar to what other researchers have observed when manipulating the proportion of congruent to incongruent trials in a Stroop task (e.g. Bailey *et al.*, 2010) or high and low interference in a working memory task (Burgess and Braver, 2010). In contrast, task switches occur in both IAT blocks, and thus, any conflict associated with a task switch likely relies on reactive control processes. Thus, it could be that task switching largely relates to reactive control within the IAT, whereas the congruency effect relates primarily to proactive control.

Event-related brain potentials (ERPs) provide an excellent means for testing the extent to which switching and congruency effects in the IAT are associated with distinguishable control processes as outlined in the DMC. Recent research using paradigms involving congruency manipulations similar to the IAT has linked the amplitude of a medial frontal negativity (MFN) in the ERP to neural processes supporting proactive control (see Bailey *et al.*, 2010; West and Bailey, 2012; West *et al.*, 2012). For example, West and Bailey (2012) found that MFN amplitude was greater on incongruent than congruent trials in a mostly incongruent trial block but not in a mostly congruent trial block of the counting Stroop task. The MFN is similar in scalp distribution and time course to the N2 or N200, which is also sensitive to conflict (see Kopp *et al.*, 1996) but which generally shows an opposing pattern in response to context, being larger for incongruent trials when most trials within a block are congruent (i.e. when conflict is a low-probability event and causes greater preparation of an inappropriate response; see Kopp *et al.*, 1996; Nieuwenhuis *et al.*, 2003). Both the MFN and N2 have been linked to activity in the anterior cingulate cortex and neighboring regions (see van Veen and Carter, 2002; West and Bailey, 2012), structures consistently implicated in cognitive control (see Botvinick *et al.*, 2001; Yeung *et al.*, 2004; Braver, 2012; Shenhav *et al.*, 2013). Thus, it seems likely that the congruency manipulation of the IAT, requiring proactive control, will influence MFN amplitude, such that MFN is larger (more negative) for the incongruent than for the congruent block.

Task switching also has been linked with specific neural responses as measured via ERP. In particular, researchers consistently report that switch trials elicit a positive voltage deflection over fronto-central scalp locations, often termed 'D-pos', which appears to reflect processes involved in retrieval of the appropriate task set (Karayanidis *et al.*, 2003; Nicholson *et al.*, 2005; Jamadar *et al.*, 2010). To date, the vast majority of ERP studies of task switching have involved presentation of visual cues, signaling whether an upcoming trial will be switch or no-switch, and D-pos is typically observed following cue onset rather than target onset. Still, in one experiment, Nicholson *et al.* (2005) found that, in the absence of a predictive cue, significant D-pos activity was observed 200–400 ms following presentation of the target stimulus. In the IAT paradigm, no cue is presented before target onset to signal whether a task switch will take place. Thus, any switching-related neural activity can be considered 'reactive', representing a just-in-time shift in response set supporting the execution of the appropriate response. Given that both the timing (200–400 ms post-stimulus) and scalp location (fronto-central midline) of D-pos overlap with the MFN, switching-related D-pos activity in the IAT paradigm is likely to manifest as a positive deflection in the MFN.

Some previous research (Karayanidis *et al.*, 2003; Kieffaber and Hetrick, 2005; Lavric *et al.*, 2008) indicates that greater D-pos activity is associated with better switching ability and, thus, reduced switch costs. However, those studies have all measured D-pos elicited by preparatory cues, and thus have examined activity associated with preparation for, rather than execution of, a switch. In addition to the lack of preparatory cues, another important difference between typical switching tasks and the IAT is that individual differences in implicit attitudes influence the necessity of switching. Klauer *et al.* (2010) referred to the phenomenon of 'task-switch neglect', which can occur in the congruent block if implicit associations allow participants to classify all items according to the attribute task, obviating the need to switch tasks. Moreover, more biased individuals might experience larger switch-related neural activity in the incongruent block due to increased task set interference resulting from stereotype-related response mappings.

The current research had two primary goals. First, we sought to characterize the cognitive control processes engaged during performance of the race-evaluative IAT by measuring ERP responses associated with switching and congruency effects, here characterized as representing reactive and proactive control processes as outlined in the DMC (Braver, 2012). In accordance with prior research and theory linking congruency proportion manipulations with proactive control (see Speer *et al.*, 2003; Burgess and Braver, 2010) and linking proactive control with MFN amplitude (see Bailey *et al.*, 2010; West and Bailey, 2012; West *et al.*, 2012), we predicted that incongruent trials would be associated with increased negativity, relative to congruent trials, during the interval associated with the MFN (~250–450 ms post-stimulus). In addition, and consistent with research linking task switching with a positive voltage deflection (i.e. D-pos) roughly 200–400 ms following stimulus onset (Nicholson *et al.*, 2005), we predicted that switch trials would be associated with relative positivity, compared with no-switch trials, during this same interval (i.e. in the

same deflection representing the MFN). Such findings would suggest that congruency and switching effects are associated with distinct neurocognitive processes, here hypothesized to reflect proactive and reactive cognitive control, respectively (Braver, 2012).

The second primary goal of this research was to examine whether these distinct neural responses are independently associated with IAT performance, providing further support for the idea that multiple control processes are invoked by the IAT. In theory, more biased individuals require more proactive control, relative to their less-biased peers, to respond correctly during the incongruent block. If so, the magnitude of the IAT effect in behavior should be negatively associated with the congruency effect in the MFN, such that more biased participants (larger positive IAT effects) experience larger, more negative incongruent-block MFN amplitudes. Additionally, increased bias could make overcoming task-set interference more difficult, resulting in enhanced positivity in the ERP associated with task switching as IAT score increases. Thus, participants with increased latent bias should demonstrate greater psychophysiological recruitment of proactive and reactive control during task performance.

## METHOD

### Participants

Twenty-nine undergraduates (24 female and 5 male) between the ages of 18 and 21 years ($M = 19.0$) at a major public university participated for partial course credit. All participants had normal or corrected-to-normal vision, had never suffered a head injury resulting in loss of consciousness for >3 min and were predominantly right-handed.

### Experimental task

Participants performed a race-evaluative IAT. On each trial, participants were presented with one of six stimulus types (positive words, negative words, stereotypically Black male names, stereotypically Black female names, stereotypically White male names or stereotypically White female names) and two category labels.[1] They were asked to categorize each stimulus as quickly as possible using one of two response keys. The IAT consisted of a slightly modified standard structure (Greenwald *et al.*, 2003) including three non-critical blocks with 20 trials each and two critical blocks with 120 trials each (300 total trials). These blocks were presented in the following order: a 20-trial practice block of names, a 20-trial practice block of positive and negative words, a 120-trial critical block of both names and words using a congruent mapping, a 20-trial practice block of words using a reversed mapping and a 120-trial critical block of names and words using an incongruent mapping. During the congruent critical block, participants saw all four stimulus types and were required to categorize each as either White or positive (using one key) or Black or negative (using the other). The incongruent critical block used the reverse categorization pairings (i.e. White or negative *vs* Black or positive). Label side was counterbalanced across participants. Within each block, stimuli were presented at random rather than strictly alternating between names and words. Response labels remained on screen (upper right and upper left) throughout each block, and targets (names and words) remained onscreen until a response was made on each trial; responses were followed by a 500 ms intertrial interval.

[1] We used race names, rather than pictures of faces, so that racial category cues and evaluative stimuli would all be words, thereby ensuring that any ERP differences associated with these different stimulus types would not be confounded by stimulus modality. Information presented on the Project Implicit webpage (https://implicit.harvard.edu/implicit/demo/background/faqs.html#faq17) indicates that effects from IAT versions using faces, those using names have produced highly similar effects (for one such comparison, see Nosek *et al.*, 2002).

### Electrophysiological recording and scoring

The electroencephalogram (EEG) was recorded from 28 tin electrodes embedded in a nylon cap (Electro-Cap, International, Eaton, OH, USA) and placed according to the expanded 10–20 system (Sharbrough *et al.*, 1991). Vertical and horizontal eye movements were measured using additional bipolar electrodes placed just above and below the left eye and ~2 cm from the outer canthus of each eye, respectively. EEG was sampled at 250 Hz and filtered online at 0.1–30 Hz. Impedance at all electrodes was kept below 10 kΩ. Scalp electrodes were referenced online to the right mastoid; an average mastoid reference was derived offline. Eye movement artifacts were removed from the EEG signal using a regression-based procedure (Semlitsch *et al.*, 1986). Epochs were created extending to 1000 ms post-stimulus onset with a 200 ms pre-stimulus baseline. All epochs were baseline corrected, after which trials with peaks exceeding 100 μV were excluded. Remaining epochs were averaged according to stimulus conditions and electrodes.

Visual inspection of the grand average waveforms indicated a negative-going deflection 250–450 ms post-stimulus prominent at frontal and frontocentral scalp sites that appeared sensitive to congruency and task-switching effects, thereby resembling the superposition of MFN and D-pos. Mean voltage over this epoch was calculated at frontal (F3, Fz, F4) and frontocentral electrodes (FC3, FCz, FC4) for each participant.

## RESULTS

### Analytic approach

Data from the first four participants were excluded because a stimulus timing error was discovered after their data were collected. Data from another six participants were excluded owing to excessive EEG artifact, leaving a final sample size of 19. All retained participants had at least 28 trials in each quantified waveform (median = 49 trials per condition).

Primary analyses of the ERP data were carried out using mixed hierarchical linear models (HLM). Multivariate approaches such as HLM have several advantages over univariate repeated-measures analysis of variance (ANOVA) for analyzing psychophysiological data (see Gratton, 2007; Vasey and Thayer, 1987), particularly when sample size is modest (see Luck, 2005). First, unlike univariate approaches, multivariate models do not assume the data meet the criterion of sphericity (that is, that the variances of the differences between any two factor levels are equal), an assumption that is frequently violated in psychophysiological data (Jennings and Wood, 1976). Corrections for violating this assumption within ANOVA (e.g. Greenhouse-Geisser or Huynh-Feldt *P*-value adjustments) result in loss of statistical power. Second, interindividual variability in both baseline and stimulus-elicited EEG activity often is greater than variability attributable to variables of interest (see Gratton, 2007), contributing to inflated error variance estimates in ANOVA that also reduce power. The present approach includes an intercept for each electrode within each subject, reducing these error variance estimates. Finally, multivariate statistics are robust to missing values, allowing bad electrodes to be excluded on an individual subject basis rather than excluding subjects with missing data or interpolating missing values. Here, the data were modeled as 24 observations (every trial type at six frontal electrodes) within 19 individuals, including random intercepts of subject and of electrodes within subjects.[2]

[2] Note that denominator degrees of freedom for *F*-tests derived from HLM often differ substantially from those used in repeated-measures ANOVA. This is because degrees of freedom in HLM are derived from the products of numbers of participants and numbers of explanatory variables (in this case, 19 cases × 6 electrodes × 2 Congruency × 2 Switching).

## Behavior

### *Reaction times*

Log-transformed RTs from correct-response trials were submitted to a 2 (Congruency: congruent blocks, incongruent blocks) × 2 (Task Switching: task switch, task no-switch) mixed model with a random intercept of subject. This model showed main effects of Congruency [$F(1, 54) = 132.92$, $P < 0.001$] and Switching [$F(1, 54) = 40.44$, $P < 0.001$], and a Congruency × Switching interaction, $F(1, 54) = 20.38$, $P < 0.001$. All pairwise contrasts between cells were significant at $P < 0.05$. RTs were slower for trials in the incongruent block ($M = 906$ ms; s.d. = 129 ms) than in the congruent block ($M = 670$ ms; s.d. = 233 ms) and slower for switch trials ($M = 853$ ms; s.d. = 254 ms) than for no-switch trials ($M = 723$ ms; s.d. = 254 ms). These effects were superadditive, with incongruent switch trials ($M = 1018$ ms; s.d. = 242 ms) eliciting considerably slower responses than all other trials (*M*s = 651, 689 and 795 ms; s.ds = 130, 129 and 163 ms for congruent no-switch, congruent switch and incongruent no-switch trials, respectively).

To address a potential concern that the inclusion of female names might alter the IAT effect, we tested whether log-RT was predicted by the three-way Gender × Race × Congruency interaction. No parameter involving Gender was significant in this model. The largest of these parameters was a Gender × Congruency interaction, $F(1, 278) = 3.47$, $P = 0.064$, suggesting that the congruency effect was slightly larger for male names ($M = 166$ ms) than for female names ($M = 120$ ms). Despite the slight difference in the size of the effect for male and female names, the effect was still robust and significant in both cases (*b*s = 0.23, 0.17; $P < 0.0001$ for male and female names, respectively).

### *IAT scores*

IAT scores were calculated according to the conventional IAT scoring algorithm ($C_1$; Greenwald *et al.*, 1998) and the updated '*d*-score' scoring algorithm ($D_2$; Greenwald *et al.*, 2003). For each metric, higher scores are generally interpreted as revealing relatively more negative implicit evaluations of Blacks relative to Whites. Both IAT measures were significantly greater than 0, revealing significant bias [$C_1$: $M = 0.263$, s.d. = 0.11, $t(18) = 10.45$, $P < 0.0001$; $D_2$: $M = 0.773$, s.d. = 0.259, $t(18) = 13.02$, $P < 0.0001$]. $C_1$ and $D_2$ IAT were strongly correlated, but not as strongly as might be expected for two indices of the same construct based on the same data ($r = 0.61$, $P < 0.0001$).

## ERP data

Figure 1 presents ERP waveforms recorded at frontal and fronto-central scalp locations, time-locked to the onset of the stimulus, as a function of congruency and switching. Amplitude values measured from the 250–450 ms post-stimulus window (the negative-going deflection corresponding to the superposition of MFN and D-Pos) were tested using a 2 (Congruency: congruent block, incongruent block) × 2 (Task Switching: switch trials, no-switch trials) mixed model. The HLM included random intercepts of subject and of electrodes within subjects. The analysis showed significant main effects of Congruency, $F(1, 339) = 18.78$, $P < 0.0001$, and Switching, $F(1,339) = 45.07$, $P < 0.0001$. Incongruent-block trials elicited more negative amplitudes than did congruent-block trials (*M*s = −0.57 and 0.07 μV, respectively). Switch trials, however, elicited more positive amplitudes than did no-switch trials (*M*s = 0.25 and −0.74 μV, respectively). The Congruency × Switching interaction was not significant ($F < 1$).

## Associating IAT performance with neural responses

Bivariate correlations among the ERP and behavioral measures of interest were used to investigate potential associations between switching and congruency effects in neural response and patterns of behavioral responses. Correlation coefficients and their *P*-values are given in Table 1 and reveal several noteworthy patterns. First, and consistent with previous reports (e.g. Back *et al.*, 2005), RT switch cost was marginally associated with $C_1$ IAT scores but not associated with $D_2$ IAT scores. In addition, congruency and switching effects in RT were strongly associated, suggesting either that bias plays a role in determining both or that the control processes they elicit are correlated. Of greater interest for the current report, the congruency effect in the ERP was reliably associated with IAT scores (both scoring methods), as would be expected for a neural measure sensitive to the congruency manipulation on which those scores are based. Finally, although not large enough to be significant in the current sample, the switching effect in the ERP was associated with RT switch cost in a predictable manner (i.e. more positive switching-related ERP amplitude was associated with larger RT switch cost).

To more directly test the hypothesis that more biased participants (as measured by the IAT) experience greater need of proactive and reactive control in performing the task, IAT score and all possible interactions (i.e. IAT score × Congruency; IAT score × Switching; and the three-way) were added to the previous model in predicting ERP amplitude. When $C_1$ IAT score and its interactions were added to the model, a significant Congruency × $C_1$ interaction emerged, $F(1, 336) = 52.36$, $P < 0.0001$, such that higher (more biased) $C_1$ scores were associated with more negative amplitudes in the incongruent block relative to the congruent block (Figure 2.A1). Thus, the effect of Congruency increased with $C_1$ score [$b = -7.57$, $t(336) = -4.15$, $P < 0.0001$]. As $C_1$ score is a continuous predictor, parameter *b* represents the change in slope of $C_1$ between the congruent and incongruent blocks, rather than a pairwise comparison of means. No other interactions were significant. The Switching × $C_1$ interaction was not significant (Figure 2.A2), nor was the Switching × Congruency × $C_1$ interaction (*F*s < 1). This model provided significantly better fit than a model without $C_1$ IAT and its associated interactions, $X^2$ (4) = 64.9, $P < 0.0001$.

When $D_2$ IAT score and its associated interactions were added to the model instead, significant interactions of Congruency × $D_2$ [$F(1, 336) = 45.28$, $P < 0.0001$] and Switching × $D_2$ [$F(1, 336) = 7.46$, $P = 0.007$] emerged. These interactions were such that the magnitude of the effects of Congruency and Switching in the ERP each increased with increasing $D_2$ score. Participants with more biased $D_2$ scores had more negative incongruent-block amplitudes, compared with their congruent-block amplitudes [$b = -3.35$, $t(336) = -4.33$, $P < 0.0001$] (Figure 2.B1), and also had more positive switch-trial amplitudes as compared with their no-switch amplitudes [$b = 7.46$, $t(336) = 2.36$, $P = 0.019$] (Figure 2.B2). The Congruency × Switching × $D_2$ interaction was not significant. Again, as with the model including $C_1$ and its interactions, this model fit significantly better than the model without $D_2$ and its associated interactions, $X^2$ (4) = 57.3, $P < 0.0001$.

## DISCUSSION

Although initially argued to be a relatively pure measure of automatic associations that circumvents control-related processes (see De Houwer *et al.*, 2009; Greenwald *et al.*, 1998), performance on the IAT has been shown in recent years to be significantly related to cognitive control (e.g. Mierke and Klauer, 2003; Conrey *et al.*, 2005; Klauer and Mierke, 2005; Sherman *et al.*, 2008, 2010; Klauer *et al.*, 2010; see also Amodio and Mendoza, 2010), as have other implicit measures of attitudes (see Klauer *et al.*, 1997; Amodio *et al.*, 2004; Payne, 2005; Ferreira *et al.*, 2006; Klauer and Tiege-Mocigemba, 2007; Bartholow *et al.*, 2009; Ito *et al.*, submitted for publication). The purpose of the current research was to examine whether

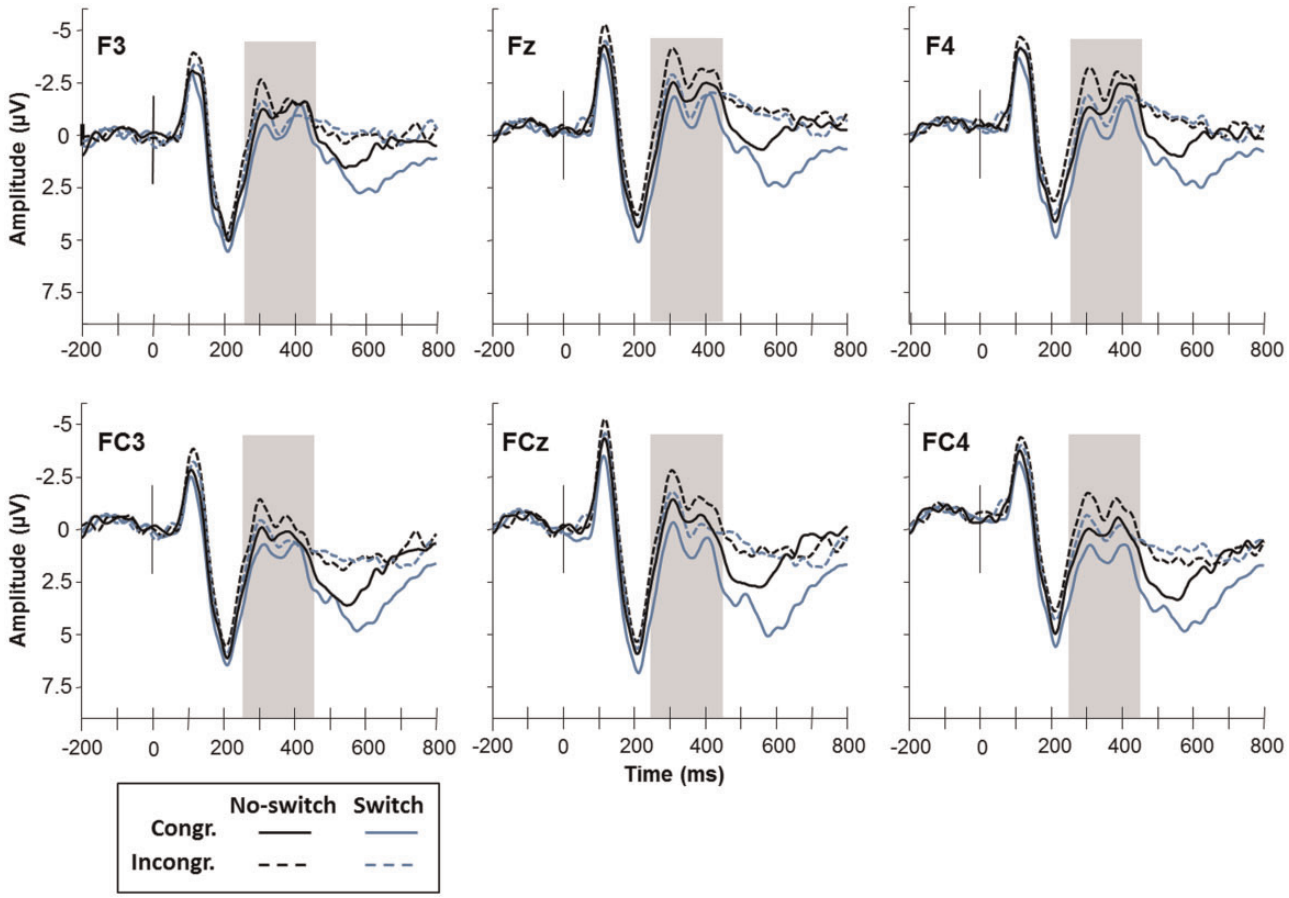


**Fig. 1** ERP waveforms recorded at frontal and fronto-central electrodes as a function of block [Congruent (Congr.) and Incongruent (Incongr.)] and task switching (No-Switch and Switch). Stimulus onset occurred at 0 ms (also indicated by the vertical line on the waveforms). The shaded area indicates the latency window (250–450 ms) used for measuring mean MFN/D-pos amplitude.

**Table 1** Correlations among ERP and behavioral measures of interest

| | ERP Congr. | ERP Switch | RT Congr. | RT Switch | $C^1$ IAT |
|---|---|---|---|---|---|
| ERP Congr. | — | | | | |
| ERP Switch | 0.10 | — | | | |
| RT Congr. | *−0.39* | 0.12 | — | | |
| RT Switch | −0.10 | 0.34 | **0.66** | — | |
| $C_1$ IAT | **−0.58** | 0.01 | **0.82** | *0.38* | — |
| $D_2$ IAT | **−0.55** | 0.29 | 0.34 | −0.11 | **0.61** |

*Note*: Coefficients in boldface are significant, $P < 0.05$; coefficients in italics are marginally non-significant, $P \leq 0.10$.
ERP Congr. = ERP amplitude difference between incongruent and congruent block trials, corresponding the MFN differences; ERP Switch = ERP amplitude difference between switch and no-switch trials, corresponding to D-pos differences; RT Congr. = RT difference between incongruent and congruent block trials; RT Switch = RT switch cost (i.e. RT difference between switch and no-switch trials); $C_1$ IAT = conventionally scored IAT score (Greenwald *et al.*, 1998); $D_2$ IAT = updated, *d*-score approach IAT score (Greenwald *et al.*, 2003).

distinguishable control-related processes as described in the DMC (Braver, 2012) are invoked during the IAT, and whether these processes are associated with IAT performance.

The pattern of findings seen here suggests that IAT performance elicits two distinct control processes: one influenced by the blocked congruency manipulation, reflected in the MFN, and one influenced by trial-to-trial task-switching demands, reflected in D-pos. The amplitude of the negative-going ERP deflection emerging 250–450 ms post-stimulus was independently influenced by congruency and switching and in predictably opposing directions. As argued previously, from the perspective of the DMC (Braver, 2012), the predictable blocked structure of the IAT is perfectly suited to elicit block-level differences in proactive control, whereas the unpredictable trial-level task switching in this version of the IAT should elicit the kind of just-in-time conflict resolution process associated with reactive control. Previous ERP research has linked proactive control with the MFN (Bailey *et al.*, 2010; West and Bailey, 2012; West *et al.*, 2012), and task switching with positivity in voltage generally occurring during this same post-stimulus epoch (i.e. D-pos; see Nicholson *et al.*, 2005). In addition, recent research by Forbes *et al.* (2012) showed a conceptually similar pattern of greater negativity in the ERP, in an epoch similar to when the MFN emerged here, for the incongruent relative to the congruent block in a gender-stereotype version of the IAT. Williams and Themanson (2010) reported a similar pattern of increased negativity for incongruent relative to congruent word trials in a gay-straight IAT. However, the current results are the first to highlight that performing the IAT engages neural circuits involved in both proactive and reactive cognitive control as outlined in the DMC (Braver, 2012).

Moreover, the current results suggest that engagement of these two control processes differs according to IAT performance, such that individual differences in bias might contribute to participants' experience of the need to engage control during the task. This idea is consistent with previous arguments specifically focused on the need to engage in task switching in the IAT. As noted by Klauer and Mierke (2005), it is possible to respond accurately in the congruent IAT block without consistently switching between tasks. For example, responding to 'Tyrell' on the basis of its category membership (Black name) or on

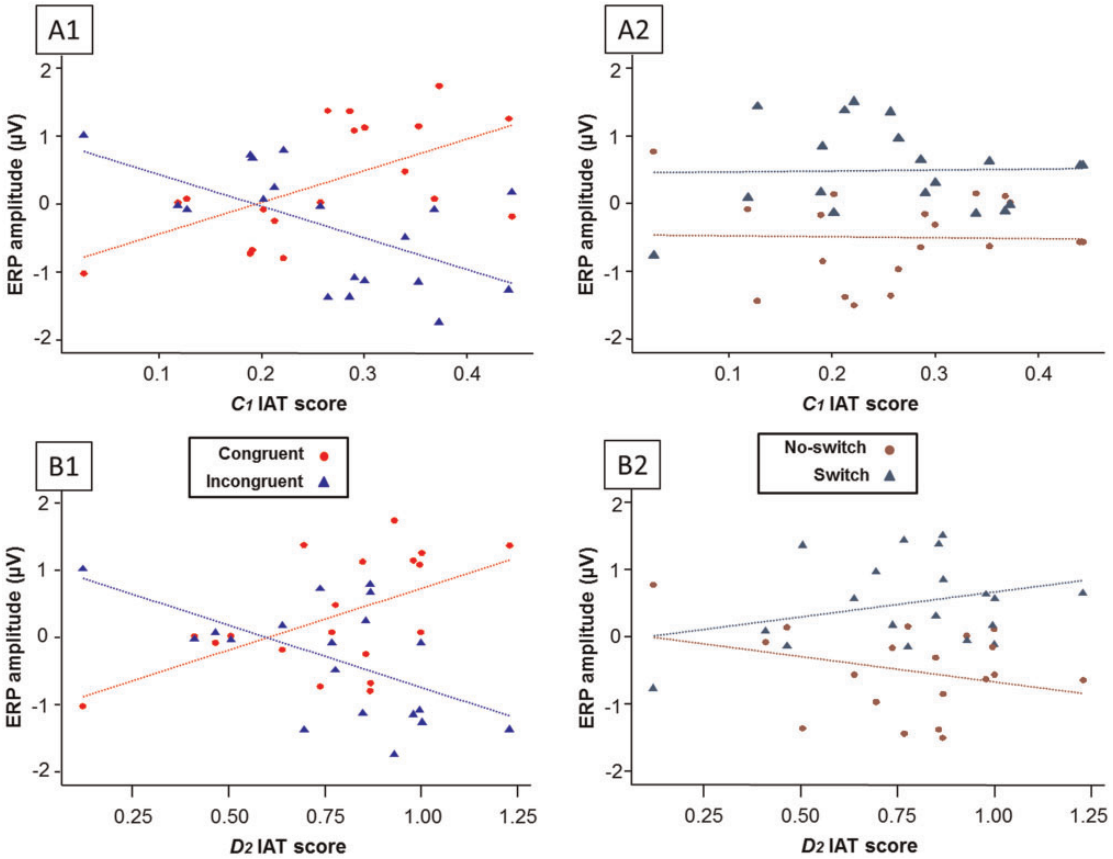


**Fig. 2** Scatterplots (with regression lines) depicting associations between ERP amplitudes and IAT scores as a function of congruency and switching factors. Panels A1 and A2 depict associations with $C_1$ IAT scores; panels B1 and B2 depict associations with $D_2$ IAT scores. ERP amplitudes depicted here are residual values after regressing out variance owing to subject and electrodes within subjects. Moving left-to-right on the graph, IAT score increases, as do the differences between switch and no-switch trials and congruent and incongruent trials.

the basis of it being negatively evaluated (due to racial stereotypes) would lead to the same response during the congruent block. In theory, this should be particularly true for more biased individuals, who are therefore more likely than their less-biased peers to experience the congruent block as a single-task block in which the response mapping component of task switching does not change, even though the decision rule component (i.e. substituting the race category judgment for the word valence judgment) should change (see Meiran *et al.*, 2000). In contrast, accurate responding in the incongruent block requires executing both components of each and every task switch, leading this block to be experienced as a so-called mixed-task block. Compared with single tasks, mixed tasks demand more ongoing monitoring (i.e. proactive control; Braver *et al.*, 2003), as participants must determine both which decision rule to implement and which response channel to use (Koch *et al.*, 2005; Rubin and Meiran, 2005; Yehene and Meiran, 2007). Because ERPs were predicted by two-way interactions (i.e. $D_2 \times$ Congruency; $D_2 \times$ Switching) but not three-way interactions ($D_2 \times$ Congruency $\times$ Switching), it may be the case that participants experience the whole block as a mixed task rather than specifically experiencing particular conflict on switch trials.

At first blush, these associations may seem inconsistent with the results of previous studies in which increased amplitude of certain other control-related ERP components has been associated with reduced expression of bias in other implicit bias tasks (e.g. Amodio *et al.*, 2004, 2008; Bartholow *et al.*, 2006, 2012). However, several differences between those previous reports and the current work suggest explanations for these apparently diverging patterns. First, all prior reports linking control-related ERP responses and performance on implicit bias tasks have examined accuracy bias. Although the $D_2$ scoring algorithm does include a penalty for errors, the IAT does not impose a response deadline. Thus, accuracy is largely irrelevant to task performance. Participants generally make few errors in the IAT, and bias is reflected instead in RT.

Second, and related to the previous point, the ERP responses examined in previous reports—the error-related negativity (ERN; Amodio *et al.*, 2004, 2008; Bartholow *et al.*, 2012) and negative slow wave (Bartholow *et al.*, 2006)—reflect neural processes subserving arguably different control-related functions than were investigated here. Consider the findings of Amodio *et al.* (2004, 2008), who showed that making bias-related errors (i.e. making 'stereotype-congruent' responses when the task calls for a stereotype-incongruent response) in the Weapons Identification Task (WIT; Payne, 2001) elicited a large ERN, which correlated with greater involvement of control in performance of the task overall. In essence, this finding indicates that failure to overcome stereotypic bias generates a neural response that, in theory, contributes to the individual's ability to overcome bias, and that this neural response is larger among participants who are motivated to suppress bias (Amodio *et al.*, 2008; although they still have some activated bias; see Devine, 1989) and smaller among participants impaired by alcohol (Bartholow *et al.*, 2012). In contrast, bias in the IAT is determined by the latency required to correctly make 'stereotype-incongruent' responses, and as with numerous other laboratory RT tasks (e.g. classic Stroop or flanker tasks), trials that require more control elicit much slower responses than trials that require less control. Presumably, this occurs because of an association between the degree to which making cognitively incongruous responses is difficult

and the degree to which relevant neural control circuits must be engaged to enact those responses. Consistent with this idea, Kopp *et al.* (1996) found that the amplitude of the N2 elicited by incompatible flanker trials varied along with the degree to which those trials elicited activation of incorrect responses in motor cortex, as did RT. In other words, the amplitude of neural control responses akin to the MFN increase as a function of the extent to which control is needed to perform correctly in a high-conflict situation. In the IAT, such conflict appears determined by individual variability in underlying evaluative bias.

Finally, it is important to recognize that not all bias tasks tap the same underlying dimensions of bias, often resulting in relatively low correlations across measures (see Cunningham *et al.*, 2001; Ito *et al.*, submitted for publication). For example, whereas the WIT largely reflects bias associated with semantic associations (Blacks are more strongly linked with guns owing to the contents of anti-Black stereotypes), the race-evaluative IAT primarily reflects biased affective evaluations. As numerous scholars have pointed out (e.g. Devine, 1989), virtually all people in a given culture experience automatic activation of semantic stereotypes on encountering a member of a stereotyped group, but individuals vary considerably in their evaluative biases. It follows, then, that to the extent one has less evaluative bias, control should be less necessary in a task like the IAT. For all of these reasons, the control functions underlying performance in the IAT and other bias tasks such as the WIT cannot be assumed to be the same, and neither should relevant associations between neural control responses and behavioral expression of bias.

The current results have implications for understanding contributions of control-related processes to performance of the IAT. Previous reports (Back *et al.*, 2005; Klauer *et al.*, 2010) have shown that the influence of task switching on IAT scores is significantly reduced when IAT performance is scored according to the *d*-score method (i.e. $D_2$) outlined by Greenwald *et al.* (2003). Hence, it might be tempting to conclude that somehow this newer scoring method eliminates (or considerably reduces) concerns over the role of switching in the IAT. Of course, changing how task behavioral responses are scored does nothing to alter the inherent structure of the task and how it is experienced by respondents, and therefore should not change the cognitive and neural processes it elicits. Moreover, the current findings build on previous work (Mierke and Klauer, 2001, 2003; Conrey *et al.*, 2005; Sherman *et al.*, 2008) by situating different facets of control within a broader, cognitive neuroscience-based framework for understanding control and the partially separable neural structures supporting those facets (Braver, 2012), thereby more directly linking knowledge about IAT performance with what is known about performance on a host of other laboratory tasks more often used in cognitive neuroscience research (e.g. Stroop and flanker tasks; working memory and inhibition tasks; switching tasks), as well as a vast literature on the behavioral and neural manifestations of control (e.g. Botvinick *et al.*, 2001; Miller and Cohen, 2001; Karayanidis *et al.*, 2003; Koechlin *et al.*, 2003; Nicholson *et al.*, 2005; O'Reilly, 2006; Braver *et al.*, 2009).

Nevertheless, the advances made in this study must be understood within the context of a number of limitations. First, the sample size ultimately used for analyses ($n = 19$) was modest, especially for analyses aimed at examining individual differences. Thus, caution is clearly warranted in drawing firm conclusions from the current data. Although investigations of brain–behavior relationships using functional magnetic resonance imaging routinely use even smaller samples to draw inferences regarding the meaning of neural responses, it will be critical for future studies to use larger samples to determine whether the patterns reported here will replicate. Second, the version of the IAT used in this study differed in some ways from the way the task is often administered: participants performed five blocks rather than seven; stimuli included both male and female names; and the task (semantic classification *vs* evaluation) switched randomly between trials, rather than switching on every trial. Still, the fact that participants performed many more critical trials in the current study (120 each of congruent and incongruent) than is typical (40 of each) arguably represents an advantage in terms of ensuring stable patterns of response.

In conclusion, the current study builds on a number of recent reports highlighting a role for cognitive control in IAT performance (e.g. Conrey *et al.*, 2005; Klauer and Mierke, 2005; Klauer *et al.*, 2010; Teige-Mocigemba *et al.*, 2010), but significantly extends prior work by identifying neurocognitive responses linking the control-related processes engaged by the IAT to specific facets of control as outlined in the DMC theory (Braver, 2012), and further characterizing differences in the control requirements of the task in the congruent and incongruent blocks. At the same time, the current work highlights the role of underlying racial bias in determining the extent to which these control-related processes are necessitated; after all, for an unbiased participant, the congruent and incongruent blocks of the task should be functionally identical and thus require the same limited engagement of cognitive control for task switching. That the ERP responses in this study were sensitive to interactions involving manipulated congruency and switching factors with IAT scores suggests that proactive and reactive control processes are recruited to the extent that cognitive effort is required to overcome the biased associations the IAT is intended to measure (see Sherman *et al.*, 2008). The exercise of control, then, may be an important cognitive process in IAT performance and a meaningful source of variance in IAT scores, even while variability in the efficiency of control is a source of contamination (Mierke and Klauer, 2003; Back *et al.*, 2005).

## REFERENCES


Amodio, D.M., Devine, P.G., Harmon-Jones, E. (2008). Individual differences in the regulation of intergroup bias: the role of conflict monitoring and neural signals for control. *Journal of Personality and Social Psychology*, *94*, 60–74.

Amodio, D.M., Harmon-Jones, E., Devine, P.G., Curtin, J.J., Hartley, S.L., Covert, A. (2004). Neural signals for the control of unintentional race bias. *Psychological Science*, *15*, 88–93.

Amodio, D.M., Mendoza, S. (2010). Implicit intergroup bias: cognitive, affective, and motivational underpinnings. In: Gawronski, B., Payne, B.K., editors. *Handbook of Implicit Social Cognition*. New York: Guilford Press, pp. 353–74.

Back, M.D., Schmukle, S.C., Egloff, B. (2005). Measuring task-switching ability in the Implicit Association Test. *Experimental Psychology*, *52*, 167–79.

Bailey, K.M., West, R., Anderson, C.A. (2010). A negative association between video game experience and proactive cognitive control. *Psychophysiology*, *47*, 34–42.

Bartholow, B.D., Dickter, C.L., Sestir, M.A. (2006). Stereotype activation and control of race bias: cognitive control of inhibition and its impairment by alcohol. *Journal of Personality and Social Psychology*, *90*, 272–87.

Bartholow, B.D., Henry, E.A., Lust, S.A., Saults, J.S., Wood, P.K. (2012). Alcohol effects on performance monitoring and adjustment: affect modulation and impairment of evaluative cognitive control. *Journal of Abnormal Psychology*, *121*, 173–86.

Bartholow, B.D., Riordan, M.A., Saults, J.S., Lust, S.A. (2009). Psychophysiological evidence of response conflict and strategic control of responses in affective priming. *Journal of Experimental Social Psychology*, *45*, 655–66.

Botvinick, M.M., Braver, T.S., Carter, C.S., Barch, D.M., Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–52.

Braver, T.S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*, 106–13.

Braver, T.S., Paxton, J.L., Locke, H.S., Barch, D.M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*, 7351–6.

Braver, T.S., Reynolds, J.R., Donaldson, D.I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, *39*, 713–26.

Burgess, G.C., Braver, T.S. (2010). Neural mechanisms of interference control in working memory: effects of interference expectancy and fluid intelligence. *PLoS One*, *5*, e12861.

Conrey, F., Sherman, J., Gawronski, B., Hugenberg, K., Groom, C. (2005). Separating multiple processes in implicit social cognition: the quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*, 469–87.

Cunningham, W.A., Preacher, K.J., Banaji, M.R. (2001). Implicit attitude measures: consistency, stability, and convergent validity. *Psychological Science*, *12*, 163–70.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., Moors, A. (2009). Implicit measures: a normative analysis and review. *Psychological Bulletin*, *135*, 347–68.

De Pisapia, N., Braver, T.S. (2006). A model of dual control mechanisms through anterior cingulate and prefrontal cortex interactions. *Neurocomputing*, *69*, 1322–6.

Devine, P.G. (1989). Stereotypes and prejudice: their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5–18.

Ferreira, M.B., Garcia-Marques, L., Sherman, S.J., Sherman, J.W. (2006). Automatic and controlled components of judgment and decision making. *Journal of Personality and Social Psychology*, *91*, 797–813.

Forbes, C.E., Cameron, K.A., Grafman, J., et al. (2012). Identifying temporal and causal contributions of neural processes underlying the Implicit Association Test (IAT). *Frontiers in Human Neuroscience*, *6*(320), 1–18.

Gratton, G. (2007). Biosignal processing. In: Cacioppo, J., Tassinary, L., Berntson, G., editors. *Handbook of Psychophysiology*. New York: Cambridge University Press, pp. 900–23.

Greenwald, A., McGhee, D., Schwartz, J. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, *74*, 1464–80.

Greenwald, A., Nosek, B., Banaji, M. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–215.

Greenwald, A.G., Poehlman, T.A., Uhlmann, E., Banaji, M.R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*, 17–41.

Jamadar, S., Hughes, M., Fulham, W.R., Michie, P.T., Karayanidis, F. (2010). The spatial and temporal dynamics of anticipatory preparation and response inhibition in task-switching. *Neuroimage*, *51*, 432–49.

Jennings, J.R., Wood, C.C. (1976). The e-adjustment procedure for repeated-measures analyses of variance. *Psychophysiology*, *13*, 277–8.

Karayanidis, F., Coltheart, M., Michie, P.T., Murphy, K. (2003). Electrophysiological correlates of anticipatory and poststimulus components of task switching. *Psychophysiology*, *40*, 329–48.

Kieffaber, P.D., Hetrick, W.P. (2005). Event-related potential correlates of task switching and switch costs. *Psychophysiology*, *42*, 56–71.

Klauer, K.C., Mierke, J. (2005). Task-set inertia, attitude accessibility, and compatibility-order effects: new evidence for a task-set switching account of the Implicit Attitude Test. *Personality and Social Psychology Bulletin*, *31*, 208–17.

Klauer, K.C., Rossnagel, C., Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 246–55.

Klauer, K.C., Schmitz, F., Teige-Mocigemba, S., Voss, A. (2010). Understanding the role of executive control in the implicit association test: why flexible people have small IAT effects. *Quarterly Journal of Experimental Psychology*, *63*, 595–619.

Klauer, K.C., Tiege-Mocigemba, S. (2007). Controllability and resource dependence in automatic evaluation. *Journal of Experimental Social Psychology*, *43*, 648–55.

Koch, I., Prinz, W., Allport, A. (2005). Involuntary retrieval in alphabet-arithmetic tasks: task-mixing and task-switching costs. *Psychological Research*, *69*, 252–61.

Koechlin, E., Ody, C., Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*, 1181–5.

Kopp, B., Rist, F., Mattler, U. (1996). N200 in the flanker task as a neurobehavioral tool for investigating executive control. *Psychophysiology*, *33*, 282–94.

Lavric, A., Mizon, G.A., Monsell, S. (2008). Neurophysiological signature of effective anticipatory task-set control: a task-switching investigation. *European Journal of Neuroscience*, *28*, 1016–29.

Luck, S.J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

Meiran, N., Chorev, Z., Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology*, *41*, 211–53.

Mierke, J., Klauer, K.C. (2001). Implicit association measurement with the IAT: evidence for effects of executive control processes. *Experimental Psychology*, *48*, 107–22.

Miller, E., Cohen, J. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Mierke, J., Klauer, K.C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology*, *85*, 1180–92.

Nicholson, R., Karayanidis, F., Poboka, D., Heathcote, A., Michie, P.T. (2005). Electrophysiological correlates of anticipatory task-switching processes. *Psychophysiology*, *42*, 540–54.

Nieuwenhuis, S., Yeung, N., Van den Wildenberg, W., Ridderinkhof, K.R. (2003). Electrophysiological correlates of anterior cingulate function in a Go/NoGo task: effects of response conflict and trial-type frequency. *Cognitive, Affective and Behavioral Neuroscience*, *3*, 17–26.

O'Reilly, R.C. (2006). Biologically based computational models of high-level cognition. *Science*, *314*, 91–4.

Payne, B. (2001). Prejudice and perception: the role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, *81*, 181–92.

Payne, B. (2005). Conceptualizing control in social cognition: how executive functioning modulates the expression of automatic stereotyping. *Journal of Personality and Social Psychology*, *89*, 488–503.

Rogers, R.D., Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology*, *124*, 207–31.

Rubin, O., Meiran, N. (2005). On the origins of the task mixing cost in the cuing task-switching paradigm. *Journal of Experimental Psychology*, *31*, 1477–91.

Semlitsch, H., Anderer, P., Schuster, P., Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology*, *23*, 695–703.

Sharbrough, F., Chatrian, G.E., Lesser, R.P., Lüders, H., Nuwer, M., Picton, T.W. (1991). American Electroencephalographic Society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *8*, 200–2.

Shenhav, A., Botvinick, M.M., Cohen, J.D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. *Neuron*, *79*, 217–40.

Sherman, J., Gawronski, B., Gonsalkorale, K., Hugenberg, K., Allen, T., Groom, C. (2008). The self-regulation of automatic associations and behavioral impulses. *Psychological Review*, *115*, 314–35.

Sherman, J.W., Klauer, C.K., Allen, T.J. (2010). Mathematical modeling of implicit social cognition. In: Gawronski, B., Payne, B.K., editors. *Handbook of Implicit Social Cognition*. New York: Guilford Press, pp. 156–75.

Speer, N.K., Jacoby, L.L., Braver, T.S. (2003). Strategy-dependent changes in memory: effects on behavior and brain activity. *Cognitive, Affective, and Behavioral Neuroscience*, *3*, 155–67.

Teige-Mocigemba, S., Klauer, C.K., Sherman, J.W. (2010). A practical guide to implicit association tests and related tasks. In: Gawronski, B., Payne, B.K., editors. *Handbook of Implicit Social Cognition*. New York: Guilford Press, pp. 117–39.

Van Veen, V., Carter, C. (2002). The timing of action-monitoring processes in the anterior cingulate cortex. *Journal of Cognitive Neuroscience*, *14*, 593–602.

Vasey, M.W., Thayer, J.F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: a multivariate solution. *Psychophysiology*, *24*, 479–86.

West, R., Bailey, K. (2012). ERP correlates of dual mechanisms of control in the counting Stroop task. *Psychophysiology*, *49*, 1309–18.

West, R., Bailey, K., Tiernan, B.N., Boonsuk, W., Gilbert, S. (2012). The temporal dynamics of medial and lateral frontal neural activity related to proactive cognitive control. *Neuropsycholgia*, *50*, 3450–60.

Williams, J.K., Themanson, J.R. (2010). Neural correlates of the implicit association test: evidence for semantic and emotional processing. *Social Cognitive and Affective Neuroscience*, *6*, 468–76.

Yehene, E., Meiran, N. (2007). Is there a general task switching ability? *Acta Psychologica*, *126*, 169–95.

Yeung, N., Botvinick, M.M., Cohen, J.D. (2004). The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931–59.