Social Cognitive and Affective Neuroscience Advance Access published November 22, 2012

doi:10.1093/scan/nss118 SCAN (2012) 1 of 9

Moral concerns increase attention and response monitoring during IAT performance: ERP evidence

Félice van Nunspeet,[1,2] Naomi Ellemers,[1,2] Belle Derks,[1,2] and Sander Nieuwenhuis[2,3]
[1]Social and Organizational Psychology Unit, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands, [2]Leiden Institute for Brain and Cognition, Postzone C2-S P.O. Box 9600, 2300 RC Leiden, The Netherlands, and [3]Cognitive Psychology Unit, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands

Previous research has revealed that people value morality as a more important person characteristic than competence. In this study, we tested whether people adjust their less explicit behavior more to moral than competence values. Participants performed an Implicit Association Test (IAT) that was either framed as a test of their morality or as a test of their competence. The behavioral results revealed a smaller IAT effect (i.e. a weaker negative implicit bias toward Muslims) in the morality condition than in the competence condition. Moreover, event-related potentials indicated increased social categorization of faces (as indexed by the N1 and P150) and enhanced conflict- and error monitoring (N450 and error-related negativity) in the morality condition compared to the competence condition. These findings indicate that an emphasis on morality can increase attentional and motivational processes that help to improve people's task performance.

Keywords: morality; social categorization; conflict monitoring; error-related negativity

INTRODUCTION

We tend to evaluate people's personal characteristics and behavior along two dimensions: one concerning morality (i.e. how we should behave) and one concerning competence (i.e. how we are able to behave). Behaving according to these dimensions is differentially diagnostic for who we are and how we are perceived: Skowronski and Carlston (1987) showed that for morality negative behaviors are perceived as more diagnostic than positive behaviors, whereas for competence positive behaviors are more diagnostic than negative behaviors. In contrast to behaving incompetently, behaving immorally thus seems to be more indicative of who we are.

Recent research has shown that for people's self-views and the positive evaluation of the group to which they belong moral characteristics are perceived as more important than characteristics concerning competence or sociability [as these are distinct dimensions of social judgment (Leach *et al.*, 2007), in contrast to warmth in which traits concerning both morality and sociability are included (Fiske *et al.*, 2007)]. Moreover, when people form an impression of a person or a group, they are more interested in information concerning morality traits than traits concerning competence and sociability (Brambilla *et al.*, 2011a,b). Indeed, when people form a first impression within milliseconds, they are more efficient in making inferences about trustworthiness than in making inferences about competence or likeability (Willis and Todorov, 2006).

People seem to be aware that moral judgments are important. For instance, Ellemers *et al.* (2008) demonstrated that people are inclined to adapt their choice to increase outcomes for the self or for the group to what other group members see as moral than to what other group members see as competent. Moreover, people anticipate being respected by their group members when they adjust their behavior to what the group considers moral (Pagliaro *et al.*, 2011). These findings suggest that morality is of great importance for impression formation and deliberate impression management. We argue that people might also be more inclined to adjust their less deliberate actions (i.e. their implicit behavior) to what is considered moral than to what is considered competent.

In the current study, we examine this prediction using an Implicit Association Test (IAT) that is framed as a test of either individual morality or competence. The IAT (Greenwald *et al.*, 1998) has been used to measure implicit attitudes toward particular social groups, for example, people with dark/white skin or, as in the current study, Muslim/non-Muslim women (see Supplementary data for details). Targets in an IAT consist of stimuli representing social groups that are associated with positive and negative attributes. When people associate stimuli that represent their own (in-)group with positivity and stimuli that represent another (out-)group with negativity, they should respond more quickly and easily to trials that are congruent with these implicit associations than to incongruent combinations (e.g. ingroup stimuli and negative attributes). The IAT assesses the degree to which this is the case, as an indicator of implicit bias.

Whether IAT performance is really implicit and thus uncontrollable is much debated, however. There is much research showing the malleability of implicit attitudes, for example by repeated exposure to admired and disliked individuals (Dasgupta *et al.*, 2009), emotions (Dasgupta and Greenwald, 2001), and several self and social motives (for a review, see Blair, 2002). Moreover, it has been shown that the IAT effect is enhanced under stereotype threat (Frantz *et al.*, 2004), but can be diminished when participants have a strategy that helps them to reduce their bias (Fiedler and Bluemke, 2005). In this research, we take advantage of the malleability of the IAT effect: We emphasize the social implications of participants' task performance (i.e. concerning their morality or their competence) and expect that participants to whom the implications concerning morality are emphasized will reduce their negative bias toward Muslim women. More specifically, we hypothesize that these participants will try to inhibit their implicit associations between Muslims and negative attributes, resulting in increased reaction times on congruent trials and thus a smaller IAT effect [consistent with the research of Fiedler and Bluemke (2005)].

Received 22 March 2012; Accepted 30 September 2012

We thank Ilona Domen, Suzanne Cederhout, Reinier Lagerwerf, Piarella Rodriguez, Lenny van den Beukel, Jelle van Hasselt and Bart van Wingerde for their help with the data collection, and David Amodio, Guido Band, Stephen Brown, Eveline Crone and Henk van Steenbergen for their advice. This work was supported by the SNS-REAAL KNAW-Merian prize and the NWO Spinoza prize awarded to N.E. by the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO), and a NWO VENI grant (451-08-022) awarded to B.D.

Correspondence should be addressed to Félice van Nunspeet, Social and Organizational Psychology Unit, Institute of Psychology, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: nunspeetfvan@fsw.leidenuniv.nl

Moreover, we are interested in the cognitive processes underlying the motivation to be moral and thus the inhibition of a negative bias on the IAT. Are intentions to behave in line with moral values associated with control of undesirable behavior, or do they influence selective attention that facilitates correct behavior? The current research addresses these questions using event-related brain potentials (ERPs) associated with perceptual processing and conflict- and error monitoring.

Perceptual attention

ERPs that are associated with early perceptual processing and more specifically, with selective attention and social categorization are the N1, P150 and N2. These components are associated with attention in such a way that increased amplitudes reflect the extent to which attention is directed toward a particular stimulus (Ito and Urland, 2003). Moreover, research has shown that this attention differs between different social stimuli. For instance, the N1—a negative deflection occurring ~100 ms after a stimulus is presented—is often larger when viewing stimuli resembling outgroup compared to ingroup members [i.e. black *vs* white faces (Ito and Urland, 2003; Kubota and Ito, 2007); however, see Ito and Urland, 2005 for the reversed pattern]. The P150, a positive peak that occurs somewhat later (~150–250 ms post-stimulus, therefore also referred to as the P200), is also larger in amplitude for outgroup than for ingroup members (Ito and Urland, 2003; Kubota and Ito, 2007). In contrast to the N1 and P150, the N2—a negative deflection ~200 ms post-stimulus—is found to be greater for stimuli representing the ingroup compared to the outgroup (Ito and Urland, 2003, 2005; Dickter and Bartholow, 2007). Examination of these components can thus show whether the emphasis on morality attracts greater attention to the group membership of the faces presented in the IAT (which is of importance when the test is said to measure participants' moral values concerning egalitarianism, but not when the test is said to measure competence). Moreover, components related to selective attention and social categorization can also be associated with motivated perception (e.g. the P150/P200; Amodio, 2010). We propose that emphasizing morality increases the motivation to suppress bias toward the outgroup. Although this could lead to diminished social categorization, we hypothesize that social categorization is actually enhanced: People's focus on the different group members should be increased to be sure to respond in line with egalitarian values (i.e. to be able to control implicit bias, as is also seen in research by Amodio, 2010). In other words, we expect to find stronger group-related modulations of the N1, P150 and N2 in the morality condition than in the competence condition.

Conflict- and response monitoring

Since we expect that emphasizing morality motivates people to inhibit their bias, we are also interested in ERPs associated with control. More specifically, conflict- and response monitoring. To assess conflict monitoring, we measure the N450. This is a negative modulation of the ERP signal, typically occurring ~400 ms post-stimulus, when subjects perform incongruent trials. The N450 modulation has been proposed to reflect the occurrence of response conflict (Nigam *et al.*, 1992; Rebai *et al.*, 1997), and is also evident in incongruent IAT trials (Williams and Themanson, 2011). Since the importance of trial congruency in the IAT may be more evident in the moral than in the competence IAT, and because we expect that control is increased when morality is made salient, we predict that the N450 modulation is larger in the morality compared to the competence condition.

To examine error-monitoring, we assess the error-related negativity (ERN; Gehring *et al.*, 1993; Nieuwenhuis *et al.*, 2001). The ERN is a negative peak occurring within 100 ms after an erroneous response. The amplitude of the ERN is sensitive to the significance of errors. Hajcak *et al.* (2005) showed, for example, that the ERN amplitude was greater on error trials when fast and accurate responses were associated with a large reward, and when participants' performance was being evaluated by a research assistant. In the current study, we hypothesize that subjects will be more motivated to prevent errors in the morality condition than in the competence condition, because the former might be viewed as a sign of immoral behavior, which is seen as more diagnostic for people's impression formation than incompetent behavior (Skowronski and Carlston, 1987). We therefore predict that erroneous responses will be associated with larger ERN modulations in the morality than in the competence condition.

We conducted two studies to test these predictions. In Study 1, we examined our hypothesis that social bias in the IAT is reduced when the test is said to measure participants' morality as opposed to their competence. In Study 2, we examined the cognitive processes associated with this reduced bias, as manifested in the ERP components discussed above.

STUDY 1

Method

Participants

Sixty-six non-Muslim students from Leiden University (24 males, $M_{age} = 20.2$ years, s.d. = 1.8) participated in this study for money or course credits.

Procedure

After providing written informed consent, participants performed five blocks of the IAT (Greenwald *et al.*, 1998). Stimuli representing the target concepts consisted of 10 pictures of women without a headscarf (i.e. ingroup pictures) and 10 pictures of women with a headscarf (i.e. outgroup pictures; for details concerning the pretest of these stimuli, see Supplementary data). Stimuli that represented positive and negative attributes consisted of five pictures of positive scenes, and five pictures of negative scenes, selected from the International Affective Picture System (Lang *et al.*, 2005). The stimuli were selected based on the scores for pleasure (i.e. negative pictures with scores <4 and positive pictures with scores >7).

In a block of congruent trials, ingroup pictures shared the same response key as positive pictures, and outgroup pictures the same response key as negative pictures. In a block of incongruent trials, this was the case for ingroup and negative pictures, and outgroup and positive pictures. The order of the congruent and incongruent blocks was counterbalanced across participants. Training blocks (IAT steps 1, 2 and 4) consisted of 26 trials, test blocks (steps 3 and 5) of 156 trials each. Every trial started with a fixation point (with a duration that varied between 500 and 1500 ms), followed by stimulus presentation (680 ms), and a feedback screen (500 ms). This screen indicated whether participants' response was correct (i.e. green check mark), incorrect (i.e. red cross) or "too late". Participants could not correct their incorrect responses.

Morality vs competence task instruction. Participants were randomly assigned to an instruction condition. In the morality condition, participants read that the test would indicate their 'values' concerning equal treatment of different people. In the competence condition, participants read that the test would indicate how well they are 'able' to process new information (for the complete instructions, see Supplementary data). All participants were instructed to respond as quickly and accurately as possible. The test implications were repeated before the start of each test block.

Checks. To check that the perceived validity of the IAT did not differ between the conditions, we asked participants after they finished the test to respond to the statement: 'My test score can assess what kind of person I am'. Furthermore, two items measured participants' task engagement: 'I think it is important to perform well on this test' and 'It does not matter to me what my test score is' (reverse coded) ($r = 0.62$, $P < 0.001$). Participants could respond to each statement on a seven-point scale ranging from 'completely disagree' (1) to 'completely agree' (7). The experiment took ~1 h after which participants were debriefed and thanked.

The IAT effect

The dependent measure was the IAT effect, indicated by the *D*-score. Based on the scoring algorithm described by Greenwald *et al.* (2003), this was calculated as the difference in reaction times on incongruent and congruent trials divided by a pooled s.d. of all correct trials. We included all trials, replaced error latencies with a replacement value ($M + 2$ s.d.$_{correct}$) and replaced latencies exceeding the maximum response time with the maximum response time of 680 ms.

Results and discussion

Checks

As intended, participants in the morality and competence condition did not think differently about the perceived validity of the test [M (morality) = 3.12, s.d. = 1.65; M (competence) = 3.24, s.d. = 1.60; $F(1,64) < 1$]. Neither did they differ in their self-reported task engagement: M (morality) = 4.14, s.d. = 1.00; M (competence) = 4.24, s.d. = 1.16; $F(1,64) < 1$.

IAT effect

Participants showed the standard IAT effect: A negative implicit bias toward the outgroup (i.e. women with a headscarf) [$t(65) = 4.72$, $P < 0.001$]. However, this bias was stronger in the competence condition [$t(32) = 5.40$, $P < 0.001$] than in the morality condition [$t(32) = 1.77$, $P = 0.09$]. More importantly, an analysis of variance (ANOVA) predicting the *D*-score from instruction conditions and order of test blocks revealed that the bias was reduced in the morality, compared to the competence condition [M (morality) = 0.13, s.d. = 0.43; M (competence) = 0.34, s.d. = 0.36; $F(1,62) = 4.56$, $P = 0.04$, $\eta^2 = 0.07$]. The reduced IAT effect was caused by a smaller difference between response times on incongruent and congruent trials in the morality condition: Consistent with previous research (Fiedler and Bluemke, 2005), participants in the morality condition responded somewhat more slowly on congruent correct trials than participants in the competence condition [$F(1,64) = 3.24$, $P = 0.08$] (Figure 1).[1] The percentages of errors did not differ between conditions; M (morality) = 8.81, s.d. = 6.03; M (competence) = 7.73, s.d. = 4.98; $F(1,64) < 1$. These behavioral results confirmed our hypothesis that task performance is adjusted when morality is made salient. To test which cognitive processes were modulated to produce the corresponding reduction in IAT score, we conducted Study 2.

STUDY 2

Method

Participants

Forty-four, healthy, right-handed, non-Muslim students from Leiden University (5 males, $M_{age} = 20.4$, s.d. = 4.3) provided written informed consent and participated in this study for money or course credits. One participant (morality condition) was excluded from the study due to an outlying IAT score; two participants (morality condition) had to be excluded from EEG analyses because of technical problems. Two more participants (one in each condition) were excluded from statistical analyses of the ERN because they did not make enough errors to reliably quantify this component (<15).

Procedure

Participants performed the IAT as described in Study 1, with the following modifications: We inserted a blank screen after the stimulus presentation to ensure that the ERN modulation occurred before the feedback. Each trial thus consisted of a fixation point (500 ms), a stimulus (680 ms), a blank screen (500 ms) and a feedback screen (750 ms). We also increased the number of congruent and incongruent trials from 156 to 300 to enhance the possibility that participants made enough errors to compute a reliable average ERN.

Participants' task engagement was measured with the items from Study 1 ($r = 0.59$, $P < 0.001$), and we checked whether participants in the morality condition were—as intended—more concerned about the social implications of their performance than participants in the competence condition (i.e. 'I am concerned about the impression people might get of me, if they know how I performed on this test'). Moreover, we assessed the internal motivation scale (IMS) to respond without prejudice developed by Plant and Devine (1998; 5 items, $\alpha = 0.73$; e.g. 'I attempt to act in non-prejudiced ways toward women who wear a headscarf because it is personally important to me'; 7-point scale: 1 'completely disagree' and 7 'completely agree'). Previous research has shown that this internal motivation influences people's ability to regulate biased behavior by conflict-monitoring processes associated with the ERN (Amodio *et al.*, 2008). Thus, to test our prediction that conflict- and error monitoring is enhanced in the morality compared to the competence condition, we controlled for individual differences in IMS. The total experiment lasted 90 min, after which participants were debriefed and thanked.

EEG acquisition

The electroencephalogram was recorded from 19 Ag/AgCl scalp electrodes, and from the left and right mastoids, using a 19-channel BioSemi active-electrode recording system (sampling rate 256 Hz). To assess horizontal and vertical eye movements, electrodes were placed on the outer canthi of the left and right eyes and ~1 cm above and below the right eye. EEG activity was recorded using ActiView software, offline data analyses were performed using BrainVision Analyzer (BVA), and the experiment was controlled by E-Prime (version 2.0). The EEG signal was referenced off-line to the average mastoid signal, corrected for ocular and eye-blink artifacts using the method of Gratton *et al.* (1983), and filtered (1–15 Hz). Single-trial stimulus- and response-locked epochs were extracted, ranging from −300 to 1000 ms after the event. These epochs were subjected to artifact rejection, then averaged and baseline-corrected by subtracting the average signal value between 200 and 0 ms pre-stimulus or between 300 and 50 ms prior to the response. Separate stimulus-locked ERP epochs were created for correct trials with outgroup and ingroup pictures, separately for the congruent and incongruent blocks. Separate response-locked ERP epochs were created for correct and incorrect responses. In an initial analysis, we found no effect of congruency on the ERN. Since participants made few errors on congruent trials, we pooled the congruent and incongruent trials to increase the number of trials averaged for each participant and thus the number of participants included in the ERN analysis.

[1] We did not find decreased response times on incongruent trials (which could be expected based on conflict monitoring theory; e.g. Botvinick *et al.*, 2001) because participants had a limited response time.

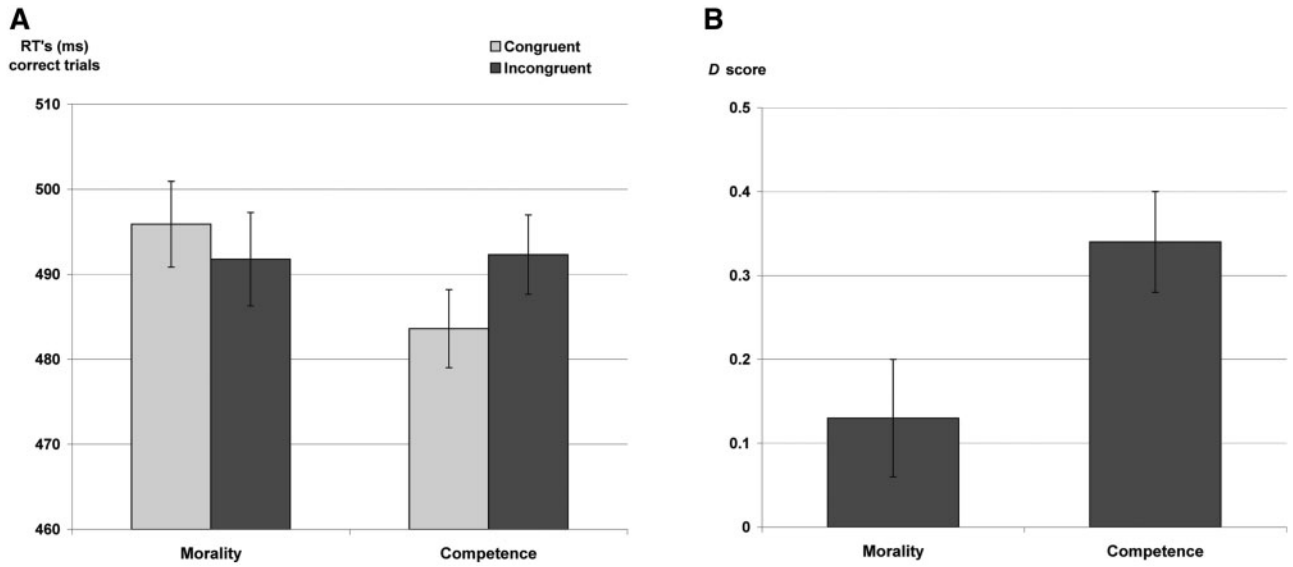

Fig. 1 Reaction times (RT's, in milliseconds) on correct congruent and incongruent trials (**A**) and the IAT effect (D score), in which error and missed trials are included after they are given a replacement value (**B**). Note that the reaction times on incongruent trials are quite fast relative to other IAT studies. This is caused by the limited the presentation time of the stimuli (i.e. participants had to respond within 680 ms).

ERP analyses

Visual inspection of the data indicated that the N1, P150, N2 and N450 potentials were most evident at the midline electrode sites Fz, FCz, Cz, CPz and Pz. These ERP components were quantified as the maximum peak amplitude within a time window (N1, 90–110 ms; P150, 100–250 ms; N2, 200–300 ms; N450, 325–500 ms). To test the main effects of social categorization and conflict monitoring, we submitted the peak-amplitude values to a 5 (electrode site) × 2 (picture type: ingroup/outgroup pictures) × 2 (congruency: congruent/incongruent trials) mixed-model ANOVA.

Visual inspection indicated that the ERN was largest at electrodes Fz, FCz and Cz. To quantify the ERN, we determined the maximal (peak) amplitude of the signal between −50 and 150 ms around the response, separately for correct and incorrect trials. All peak amplitudes were submitted to a 3 (electrode site) × 2 (accuracy: correct/error) mixed-model ANOVA.

Since modulations of the task effects by the instruction manipulation were subtle, subsequent analyses focused on the electrode at which the interaction was most pronounced. The resulting peak-amplitude values were submitted to a mixed-model ANOVA with instruction condition as between-subjects variable and the relevant task factors as within-subject variables. Moreover, to control for individual differences in internal motivation to respond without prejudice, we included IMS score as a covariate in each analysis.[2]

Results and discussion

Checks

As in Study 1, participants in the morality and competence condition did not differ in task engagement [M (morality) = 4.84, s.d. = 0.88; M (competence) = 4.63, s.d. = 0.94; $F(1,41) < 1$]. Nor did they differ in their internal motivation to respond without prejudice [M (morality) = 4.89, s.d. = 0.82; M (competence) = 5.01, s.d. = 0.66; $F(1,41) < 1$]. As expected, participants in the morality condition did report to be more concerned about the social implications of their performance than participants in the competence condition [M (morality) = 3.18, s.d. = 1.68; M (competence) = 1.91, s.d. = 1.02; $F(1,41) = 8.34$, $P = 0.006$, $\eta^2 = 0.17$].

Behavioral results

Overall, participants showed the standard IAT effect (i.e. a negative implicit bias toward women with a headscarf) [$t(42) = 5.04$, $P < 0.001$]. Moreover, this bias was evident in both conditions [morality: $t(20) = 2.52$, $P = 0.02$; competence: $t(21) = 4.68$, $P < 0.001$]. More importantly, an ANOVA with the D-score based on the first 156 trials in each block as dependent variable, the instruction condition and the order of test blocks as independent variables, and IMS as covariate revealed a difference in the IAT effect between the instruction conditions: As in Study 1, the effect was smaller for participants in the morality condition than for participants in the competence condition [M (morality) = 0.13, s.d. = 0.40; M (competence) = 0.42, s.d. = 0.36; $F(1,39) = 5.86$, $P = 0.02$, $\eta^2 = 0.13$]. As can be seen in Figure 2, this effect was caused by a smaller difference between response times on incongruent and congruent trials in the morality condition than in the competence condition. More specifically, (and similar to Study 1), participants in the morality condition responded somewhat more slowly on congruent trials than participants in the competence condition [$F(1,41) = 3.06$, $P = 0.09$]. The percentages of errors did not differ between conditions [M (morality) = 12.36, s.d. = 7.13; M (competence) = 14.25, s.d. = 9.80; $F(1,41) < 1$].

When we included all trials from each test block (a doubling of trials was needed for computing ERPs), the effect of condition was marginally significant [M (morality) = 0.15, s.d. = 0.27; M (competence) = 0.29, s.d. = 0.29; $F(1,39) = 3.05$, $P = 0.09$]. This was caused by a training effect: Participants in both conditions responded faster and made fewer errors on the last 144 trials of each test block, resulting in a similar IAT performance. Although both analyses showed a main effect of the order of test blocks [$F(1,39) = 23.28$, $P < 0.001$ and $F(1,39) = 35.73$, $P < 0.001$, respectively], this factor did not interact with instruction condition (F's < 1).

ERP results

Social categorization. N1. We found the intended main effect of social categorization: The N1 was larger for outgroup pictures (M =

[2]Inclusion of the IMS score only changed the results concerning the ERN, as is mentioned in the Results section.

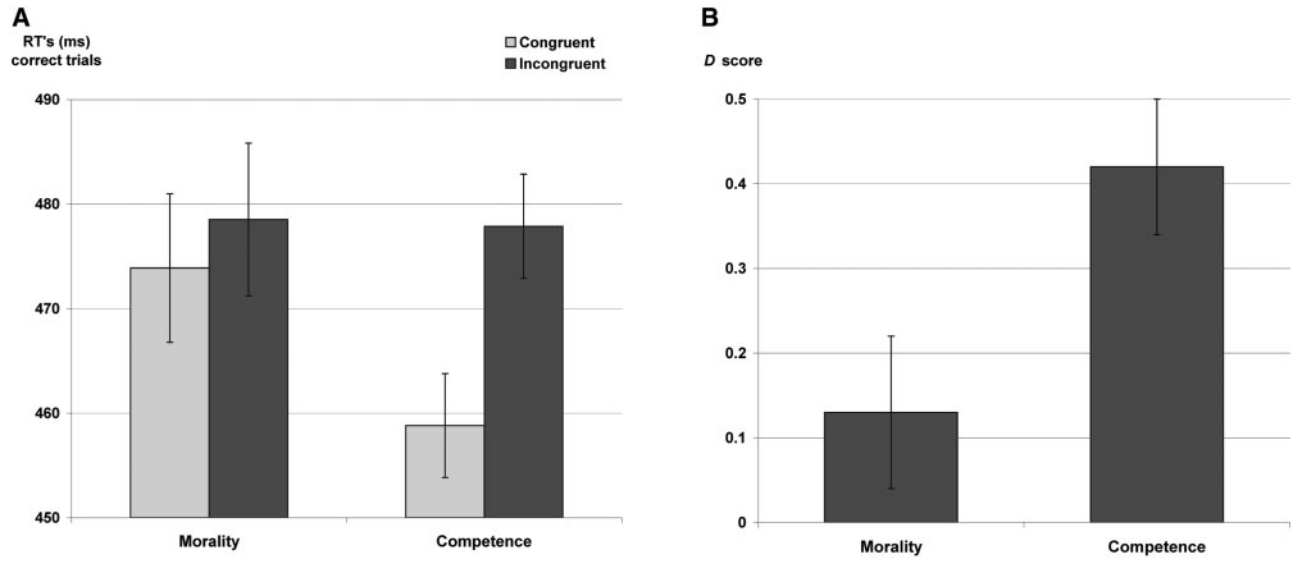

Fig. 2 Reaction times (RT's, in milliseconds) on correct congruent and incongruent trials (**A**) and the IAT effect (D score) in which error and missed trials are included after they are given a replacement value (**B**). Note that the reaction times on incongruent trials are quite fast because of the limited presentation time of the stimuli.

−5.58 μV, s.e. = 0.32) than for ingroup pictures ($M = -5.26$ μV, s.e. = 0.30) [$F(1,38) = 6.86$, $P = 0.012$, $\eta^2 = 0.15$]. Analyses for the FCz electrode confirmed the predicted interaction between instruction condition and picture type [$F(1,38) = 4.11$, $P = 0.050$, $\eta^2 = 0.10$] (Figure 3). The difference between the N1 elicited by outgroup and ingroup pictures was significant in the morality condition [$F(1,38) = 4.69$, $P = 0.04$, $\eta^2 = 0.11$], but not in the competence condition [$F(1,38) < 1$].

P150. As anticipated, the P150 was larger for outgroup pictures ($M = 5.22$ μV, s.e. = 0.52) than for ingroup pictures ($M = 4.23$ μV, s.e. = 0.52) [$F(1,38) = 39.95$, $P < 0.001$, $\eta^2 = 0.51$]. Analyses at Cz showed that, as predicted, there was an interaction effect between instruction condition and picture type [$F(1,38) = 5.12$, $P = 0.029$, $\eta^2 = 0.12$] (Figure 3). The difference in P150 amplitude between outgroup and ingroup pictures was more pronounced in the morality condition [$F(1,38) = 33.75$; $P < 0.001$, $\eta^2 = 0.47$], than in the competence condition [$F(1,38) = 8.51$, $P = 0.006$, $\eta^2 = 0.18$].

N2. The N2 was, as intended, larger for ingroup pictures ($M = -5.52$ μV, s.e. = 0.50) than for outgroup pictures ($M = -4.99$ μV, s.e. = 0.47) [$F(1,38) = 6.93$, $P = 0.012$, $\eta^2 = 0.15$]. However, there was no interaction between picture type and instruction condition [$F(1,38) = 1.08$, $P = 0.31$].

Conflict- and error monitoring. N450. Overall, the N450 was larger for incongruent trials ($M = -2.22$ μV, s.e. = 0.39) than for congruent trials ($M = -1.45$ μV, s.e. = 0.34) [$F(1,38) = 12.51$, $P = 0.001$, $\eta^2 = 0.24$]. Analyses for the CPz electrode confirmed our prediction: Instruction condition interacted with congruency [$F(1,38) = 4.79$, $P = 0.035$, $\eta^2 = 0.11$] (Figure 4). The difference in N450 amplitude between incongruent and congruent trials was significant in the morality condition [$F(1,38) = 16.12$, $P < 0.001$, $\eta^2 = 0.30$], but not in the competence condition [$F(1,38) = 1.20$, $P = 0.28$].

ERN. As anticipated, the ERN was larger for error trials ($M = -6.83$ μV, s.e. = 0.77) than for correct trials ($M = 1.00$ μV, s.e. = 0.53) [$F(1,36) = 129.08$, $P < 0.001$, $\eta^2 = 0.78$]. Moreover, accuracy interacted with IMS score; $F(1,36) = 4.03$, $P = 0.05$, $\eta^2 = 0.10$: A higher internal motivation to respond without prejudice was associated with larger ERN modulations ($B = -1.46$, $P = 0.09$; Amodio *et al.*, 2008). However, more relevant to our current predictions, analyses at Cz showed a marginally significant interaction between accuracy and instruction condition [$F(1,36) = 3.49$, $P = 0.070$, $\eta^2 = 0.09$] (Figure 5).[3] The difference in ERN amplitude between error and correct trials was somewhat larger in the morality condition [$M = -11.22$ μV, s.e. = 1.17; $F(1,36) = 94.17$, $P < 0.001$, $\eta^2 = 0.72$] than in the competence condition [$M = -8.38$ μV, s.e. = 1.08; $F(1,36) = 59.74$, $P < 0.001$, $\eta^2 = 0.62$].

The ERP results are consistent with our expectations that stressing moral implications of the IAT increases social categorization of stimuli and conflict monitoring during the test. More specifically, the emphasis on morality moderates the attention toward outgroup but not ingroup faces (as indexed by increased N1 and P150, but not N2 modulations), and increases the neural response to response conflict and errors in the IAT (as reflected in increased N450 and ERN modulations), suggesting that erroneous responses were perceived as more significant in the morality than in the competence condition.

GENERAL DISCUSSION

Previous research has shown that morality is more important than competence for people's personal and social identity (Leach *et al.*, 2007), and that morality guides explicit strategic behavior (Ellemers *et al.*, 2008). The present studies extend prior research by showing that morality also impacts on non-explicit aspects of task behavior: People inhibited their negative bias toward Muslim women on an IAT when the test was said to be indicative of their morality (instead of their competence). Our findings thus reveal that participants are able to reduce their implicit bias when given the opportunity to reveal their moral side. This complements prior observations that implicit bias is exacerbated when participants are identified as potential racists (Frantz *et al.*, 2004), and is consistent with research showing that moral appeals induce different physiological and behavioral responses, depending on whether these are framed as ideals or as obligations (Does *et al.* 2011, 2012).

[3]The analysis without IMS as a covariate revealed the same pattern of moderation, but resulted in a non-significant interaction [$F(1,37) = 2.57$, $P = 0.12$]. Moreover, as was put forward by an anonymous reviewer, the ERN results were sensitive to changes in the EEG processing settings. For example, shortening the baseline correction period (from 300–50 to 200–50 ms prior to the response) reduced the interaction effect between the ERN modulation and instruction [$F(1,36) = 2.72$, $P = 0.11$, $\eta^2 = 07$], whereas lowering the cutoff score for the high-pass filter (from 1 to 0.1 Hz) made this interaction significant [$F(1,36) = 4.97$, $P = 0.03$, $\eta^2 = 0.12$].

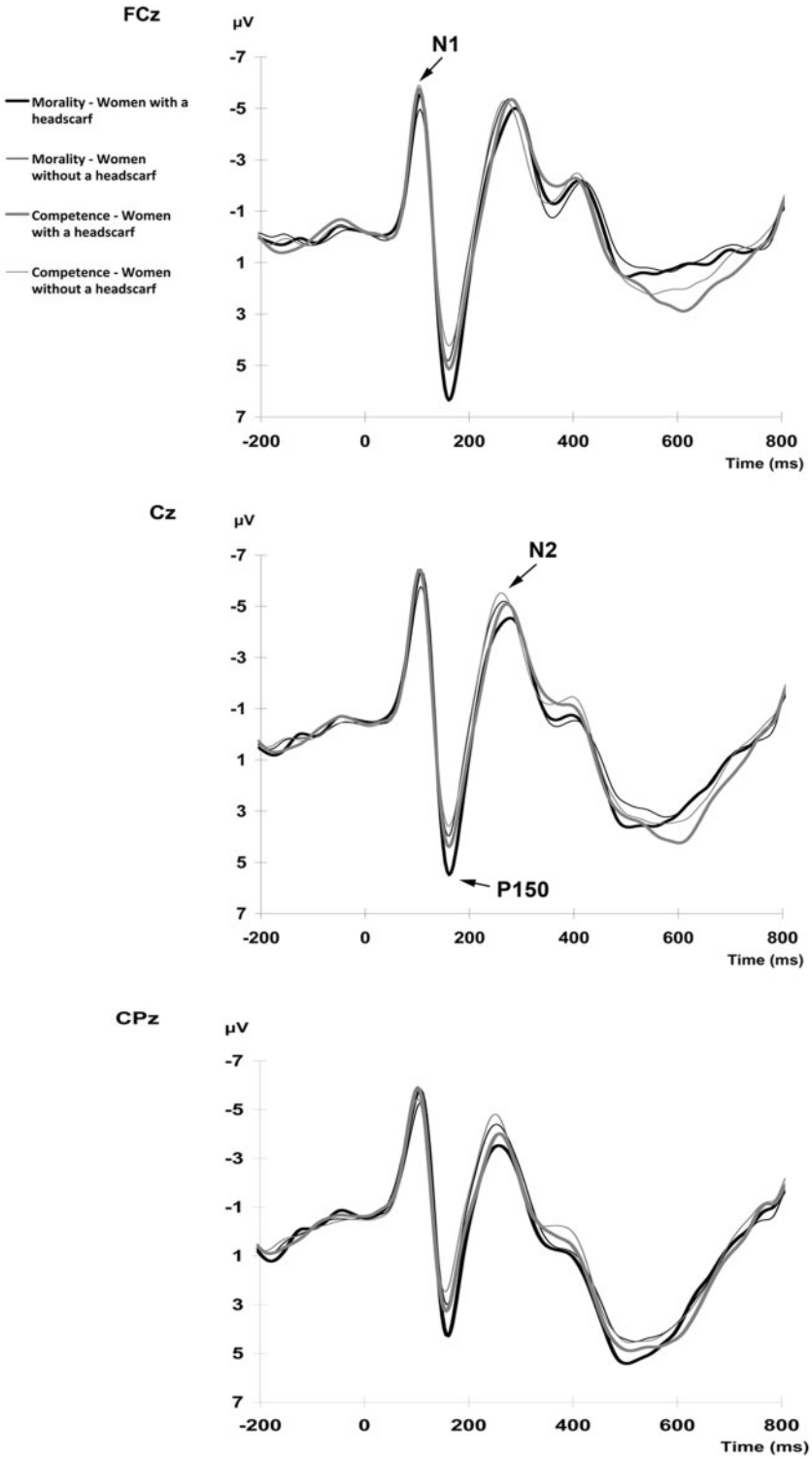

Fig. 3 The N1, P150 and N2 modulations for outgroup and ingroup pictures at three central electrodes. The interaction with instruction condition was significant at FCz for the N1, and at Cz for the P150. The interaction did not reach significance for the N2.

Importantly, the current research provides insight into the neurobiological mechanisms underlying the differential performance on the moral and competence IAT. Previous research has shown that performance on tasks designed to measure implicit attitudes is associated with (increased) motivated perception (Amodio, 2010) and response monitoring (Amodio *et al.*, 2008). Additionally, this study reveals that these cognitive processes are activated or enhanced when people's morality is emphasized. More specifically, when morality is emphasized as opposed to competence, people engage in increased social categorization of outgroup faces, and in enhanced conflict- and response. Since these processes have previously been associated with motivational states (Hajcak *et al.*, 2005; Amodio, 2010) and because morality has been shown to be more important than competence for impression formation and impression management, we interpret these

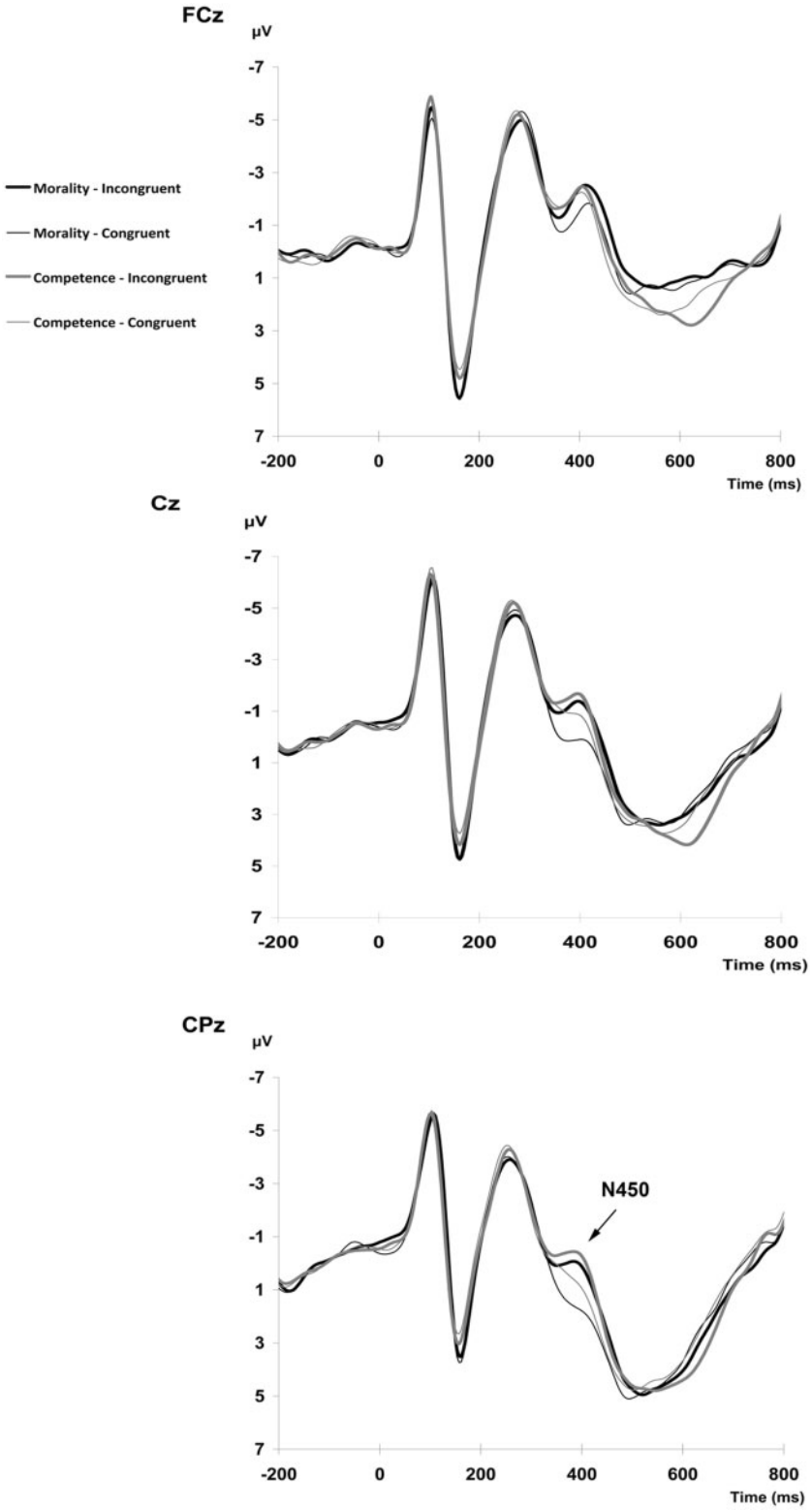

Fig. 4 The N450 modulations for incongruent and congruent trials at three central electrodes. The interaction with instruction condition was significant at CPz.

findings as indicating increased motivation of participants in the morality condition to control their bias on the IAT.

The findings concerning increased conflict- and error monitoring during a moral IAT also extend research showing that low levels of implicit bias (often revealed by people with high internal and low external motivation to avoid prejudice) are associated with successful response monitoring (Amodio *et al.*, 2008; Gonsalkorale *et al.*, 2011). The current results additionally indicate that, regardless of individual differences in internal motivation to respond without prejudice, emphasizing moral values successfully reduces displays of implicit bias. Moreover, our results indicate that emphasizing morality affects not only corrective processes like error monitoring, but also influences

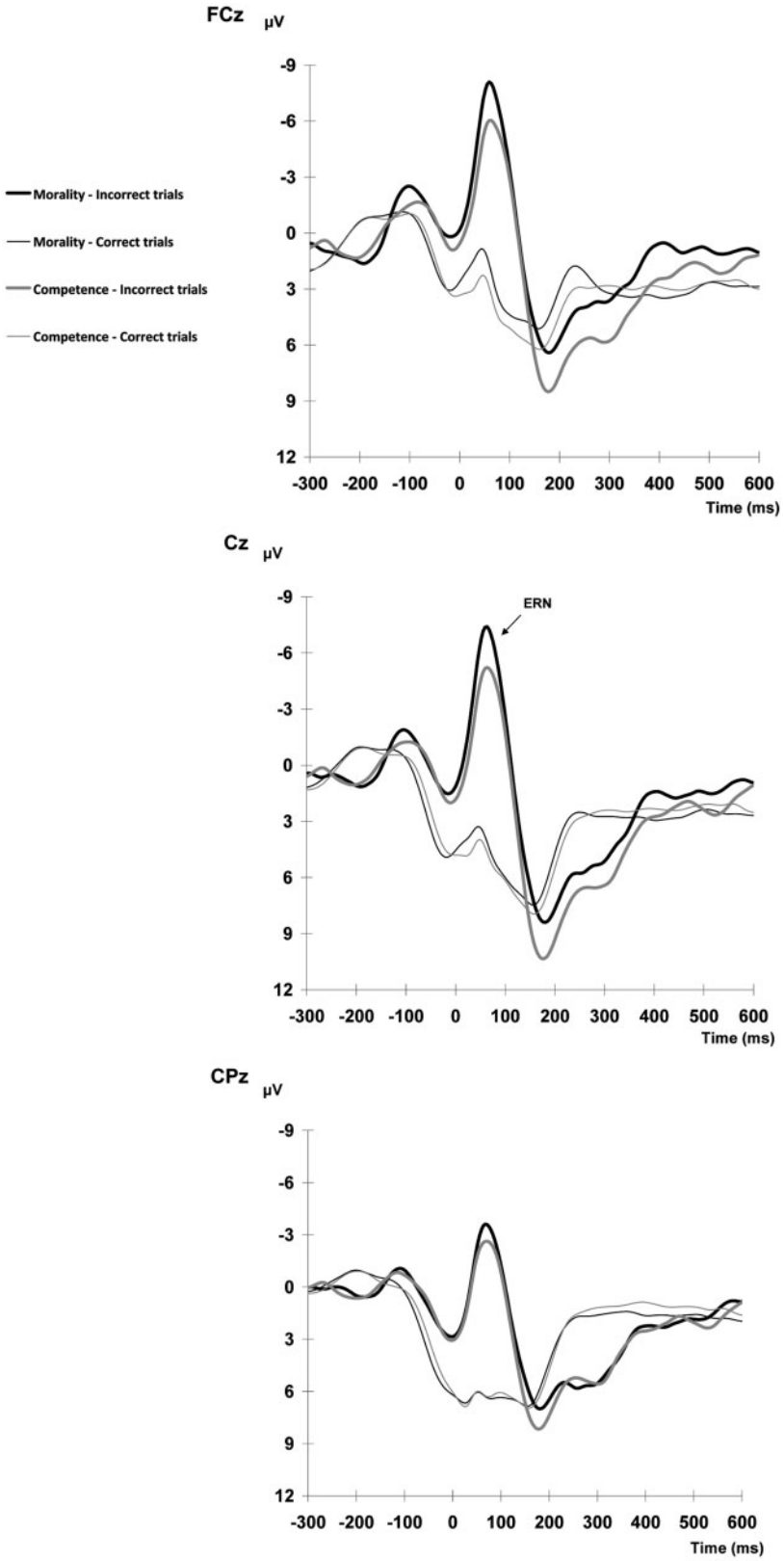

Fig. 5 The ERN modulations for correct and incorrect trials at three central electrodes. The interaction with instruction condition was marginally significant at Cz.

performance through processes involved in the attention to social stimuli before responses are given.

Although the current research broadens the knowledge of the importance of morality for people's self-identity, we also mentioned that morality is more important than competence for people's social identity, and their behavior in groups (Leach *et al.*, 2007; Ellemers *et al.*, 2008). The question thus remains whether our findings would be affected by for example social evaluation. Further research could address

this question by examining whether the emphasis on morality influences people's task performance in the presence of other people and whether this differs between evaluations of ingroup compared to outgroup members.

CONCLUSION

Our findings extend previous research that demonstrates the importance of morality over competence for people's self-view. In particular, our findings show that people control their implicit responses during a moral task, and reveal how they do that: Emphasizing morality facilitates people's task performance by increasing perceptual attention and conflict- and error monitoring.

SUPPLEMENTARY DATA

Supplementary data are available at *SCAN* online.

REFERENCES

Amodio, D.M. (2010). Coordinated roles of motivation and perception in the regulation of intergroup responses: frontal cortical asymmetry effects on the P2 event-related potential and behavior. *Journal of Cognitive Neuroscience*, *22*, 2609–17.

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.

Blair, I.V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*, 242–61.

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

Brambilla, M., Rusconi, P., Sacchi, S., Cherubini, P. (2011a). Looking for honesty: the primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology*, *41*, 135–43.

Brambilla, M., Sacchi, S., Rusconi, P., Cherubini, P., Yzerbyt, V.Y. (2011b). You want to give a good impression? Be honest! Moral traits dominate group impression formation. *British Journal of Social Psychology*, *50*, 1–18.

Dasgupta, N., Greenwald, A.G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology*, *81*, 800–14.

Dasgupta, N., DeSteno, D., Williams, L.A., Hunsinger, M. (2009). Fanning the flames of prejudice: the influence of specific incidental emotions on implicit prejudice. *Emotion*, *9*, 585–91.

Does, S., Derks, B., Ellemers, N. (2011). Thou shalt not discriminate: how emphasizing moral ideals rather than obligations increases Whites' support for social equality. *Journal of Experimental Social Psychology*, *47*, 562–71.

Does, S., Derks, B., Ellemers, N., Scheepers, D. (2012). At the heart of egalitarianism: how morality framing shapes cardiovascular challenge versus threat in Whites. *Social Psychological and Personality Science*, *3*, 747–53.

Dickter, C.L., Bartholow, B.D. (2007). Racial ingroup and outgroup attention biases revealed by event-related brain potentials. *Social Cognitive and Affective Neuroscience*, *2*, 189–98.

Ellemers, N., Pagliaro, S., Barreto, M., Leach, C.W. (2008). Is it better to be moral than smart? The effects of morality and competence norms on the decision to work at group status improvement. *Journal of Personality and Social Psychology*, *95*, 1397–410.

Fiedler, K., Bluemke, M. (2005). Faking the IAT: aided and unaided response control on the implicit association tests. *Basic and Applied Social Psychology*, *27*, 307–16.

Fiske, S.T., Cuddy, A.J.C., Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*, 77–83.

Frantz, C.M., Cuddy, A.J.C., Burnett, M., Ray, H., Hart, A. (2004). A threat in the computer: the race implicit association test as a stereotype threat experience. *Personality and Social Psychology Bulletin*, *30*, 1611–24.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–90.

Gonsalkorale, K., Sherman, J.W., Allen, T.J., Klauer, K.C., Amodio, D.M. (2011). Accounting for successful control of implicit racial bias: the roles of association activation, response monitoring, and overcoming bias. *Personality and Social Psychology Bulletin*, *37*, 1534–45.

Gratton, G., Coles, M.G., Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–84.

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

Greenwald, A.G., Nosek, B.A., Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*, 197–216.

Hajcak, G., Moser, J.S., Yeung, N., Simons, R.F. (2005). On the ERN and the significance of errors. *Psychophysiology*, *42*, 151–60.

Ito, T.A., Urland, G.R. (2003). Race and gender on the brain: electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, *85*, 616–26.

Ito, T.A., Urland, G.R. (2005). The influence of processing objectives on the perception of faces: an ERP study of race and gender perception. *Cognitive, Affective, and Behavioral Neuroscience*, *5*, 21–36.

Lang, P.J., Bradley, M.M., Cuthbert, B.N. (2005). International Affective Picture System (IAPS): digitized photographs, instruction manual and affective ratings. *Technical Report A-6*. Gainesville, FL: University of Florida.

Leach, C.W., Ellemers, N., Barreto, M. (2007). Group virtue: the importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, *93*, 234–49.

Kubota, J.T., Ito, T.A. (2007). Multiple cues in social perception: the time course of processing race and facial expression. *Journal of Experimental Social Psychology*, *43*, 738–52.

Nieuwenhuis, S., Ridderinkhof, K.R., Blom, J., Band, G.P.H., Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: evidence from an antisaccade task. *Psychophysiology*, *38*, 752–60.

Nigam, A., Hoffman, J.E., Simons, R.F. (1992). N400 to semantically anomalous pictures and words. *Journal of Cognitive Neuroscience*, *4*, 15–22.

Pagliaro, S., Ellemers, N., Barreto, M. (2011). Sharing moral values: anticipated ingroup respect as a determinant of adherence to morality-based (but not competence-based) group norms. *Personality and Social Psychology Bulletin*, *37*, 1117–29.

Plant, E.A., Devine, P.G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, *75*, 811–32.

Rebai, M., Bernard, C., Lannou, J. (1997). The Stroop's test evokes a negative brain potential, the N400. *International Journal of Neuroscience*, *91*, 85–94.

Skowronski, J.J., Carlston, D.E. (1987). Social judgment and social memory: the role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, *52*, 689–99.

Williams, J.K., Themanson, J.R. (2011). Neural correlates of the implicit association test: evidence for semantic and emotional processing. *Social Cognitive and Affective Neuroscience*, *6*, 468–76.

Willis, J., Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*, 592–8.