Social Cognitive and Affective Neuroscience Advance Access published March 22, 2013



# The N400 as an index of racial stereotype accessibility

Eric Hehman,[1] Hannah I. Volpert,[2] and Robert F. Simons[3]

[1]Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA, [2]Department of Psychological Sciences, University of Missouri, Columbia, MO 65211, USA, and [3]Department of Psychology, University of Delaware, Newark, DE 19716, USA

**The current research examined the viability of the N400, an event-related potential (ERP) related to the detection of semantic incongruity, as an index of both stereotype accessibility and interracial prejudice. Participants' EEG was recorded while they completed a sequential priming task, in which negative or positive, stereotypically black (African American) or white (Caucasian American) traits followed the presentation of either a black or white face acting as a prime. ERP examination focused on the N400, but additionally examined N100 and P200 reactivity. Replicating and extending previous N400 stereotype research, results indicated that the N400 can indeed function as an index of stereotype accessibility in an interracial domain, as greater N400 reactivity was elicited by trials in which the face prime was incongruent with the target trait than when primes and traits matched. Furthermore, N400 activity was moderated by participants' self-reported explicit bias. More explicitly biased participants demonstrated greater N400 reactivity to stereotypically white traits following black faces than black traits following black faces. P200 activity was additionally associated with participants' implicit biases, as more implicitly biased participants similarly demonstrated greater P200 reactivity to stereotypically white traits following black faces than black traits following black faces.**

**Keywords:** N400; stereotyping; prejudice; P200; intergroup dynamics

## INTRODUCTION

The complexities of human society necessitate the usage of stereotypes as cognitive shortcuts to sort social information. These shortcuts influence decision making regarding the targets of stereotypes (Cuddy *et al.*, 2007), sometimes indirectly contributing to intergroup prejudice and discrimination. Recent work has turned to event-related potentials (ERPs) to study and measure the accessibility of stereotypes (White *et al.*, 2009; Wang *et al.*, 2011), circumventing the limitations of the self-report and behavioral paradigms of previous work by examining underlying processes with greater temporal sensitivity. This research has specifically focused on a negative-going ERP occurring ~400 ms after stimulus onset (the N400) due to its well-understood association with the ease of integrating a semantic stimulus with its current context (Kutas and Hillyard, 1984; Kutas and Federmeier, 2000). The primary goal of this study was to replicate and extend this research by examining N400 responses to negative and positive stereotypes in an intergroup context. In addition, a secondary goal was to explore possible links between neural responses to congruent and incongruent stereotypes with participants' implicit and explicit racial bias.

### The N400

Theoretically rooted in conceptual priming and spreading activation, larger N400 reactivity reflects the difficulty of accessing information stored in semantic memory associated with a meaningful stimulus (Kutas and Federmeier, 2000). For example, the sentence 'Jordan was eaten by a DOORWAY' would elicit a larger N400 than 'Jordan was eaten by a DINOSAUR', as dinosaurs and eating are more semantically associated than doorways and eating. Although initially discovered examining incongruent words in sentences, research has since demonstrated that the N400 is elicited by incongruent word pairings (Bentin *et al.*, 1985), words incongruent with music (Daltrozzo and Schoen, 2008) and incongruent words and images (Nigam *et al.*, 1992). This activity likely originates in the middle superior temporal lobes, associated with the representation of semantic information (Lau *et al.*, 2008).

The N400 has only recently been explored as a method of assessing stereotype accessibility. Individuals who associate certain groups with particular characteristics should demonstrate larger N400s when presented with characteristics incongruent with those groups because they are less associated in memory, compared with congruent, more easily accessible stereotypes. Some evidence indicates that this is the case. For example, in research examining gender stereotypes, pairing 'women' with stereotypically male traits elicited greater N400 reactivity than when 'women' and stereotypically female traits were paired (White *et al.*, 2009). Other work has claimed that N400 variation can function to index prejudice, as shown by larger N400s exhibited by urban Chinese when positive adjectives were paired with disparaged rural outgroup migrant workers relative to when these adjectives were paired with the ingroup (Wang *et al.*, 2011).

### Stereotypes and prejudice

The conclusion that the N400 can function as an index of prejudice should be made tentatively, however, due to how stereotypes and prejudice are traditionally conceptualized. Stereotypes and prejudice are separate constructs, the former referring to associations between a specific group and meaningful behaviors or concepts, whereas the latter refers to evaluative biases regarding a group, typically negative in nature (Dovidio *et al.*, 1986). Awareness of stereotypes does not necessitate biased thoughts or behavior. Indeed, a growing literature suggests that each may depend on separate neural processes and predict different forms of discriminatory behavior (Dovidio *et al.*, 2002; Amodio and Devine, 2006; Amodio, 2008). Should N400 activity be driven only by the difficulty of accessing information stored in semantic memory, then it may be better characterized as indexing stereotype accessibility, rather than negative or positive *evaluations* of an outgroup (i.e. prejudice).

Recent research contrasting sequential with evaluative priming is consistent with this interpretation. Sequential priming is used to investigate how concepts may be associated in memory (Bargh and Chartrand, 2000) and involves a prime quickly followed by a target. Participants make non-evaluative decisions about each target, such as whether it is a word or non-word (Neely, 1991). Participants are typically faster and more accurate in identifying target words primed by related than unrelated words. Evaluative priming, on the other hand,

Received 10 September 2012; Accepted 27 January 2013

The first two authors contributed equally to this work.

Correspondence should be addressed to Eric Hehman, Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH 03755, USA. E-mail: Eric.Hehman@Dartmouth.edu

occurs when negatively or positively valenced primes facilitate evaluative responses to targets congruent in valence. Thus, semantic priming occurs when the meaning of a prime and target are associated, whereas evaluative priming is evident when the prime and target share a valence, but not necessarily a meaning. Although seemingly similar, unique neural responses to each paradigm indicate the processes involved may be distinct; evaluative priming solely influenced the magnitude of the late positive potential (LPP), and not the N400 (Herring *et al.*, 2011). This finding indicates that semantic and evaluative priming may be driven by separable neural mechanisms and that the N400 may not be elicited by the negative and positive evaluations required for prejudice, thus being indicative solely of stereotype accessibility.

### N400 and prejudice

Despite evidence that evaluative biases may not be involved in the N400 directly, the possibility remains that this component can be utilized as an indirect measure of prejudice. Both positively and negatively valenced stereotypes exist for many groups, are learned and subject to reinforcement in memory (Gaertner and McLaughlin, 1983). Thus, an individual who strongly associates a group with negative stereotypes might demonstrate a large N400 to positive associations with that group. A disparity in N400s elicited by negative and positive traits might be indirectly associated with participant behavior, just as stereotypes exert an influence on behavior (Cuddy *et al.*, 2007). A physiological indicator of stereotype accessibility or prejudice would be extremely valuable due to the social desirability difficulties inherent in conscious or self-reported measures of bias (Dienstbier, 1970; Plant and Devine, 1998).

Indeed, some research has attempted to examine the link between N400 reactivity and prejudice (Wang *et al.*, 2011). However, the confounding of the valence and stereotypicality of the target traits utilized in this research limit the conclusions that can be drawn regarding stereotype accessibility. For instance, 'clean', although positively valenced, is more stereotypically associated with urban Chinese than rural migrant workers, as the latter group performs hard labor in 'distasteful jobs that the city residents are unwilling to do' (Wang *et al.*, 2011, p. 104). Should more controlled, negative and positive, stereotypically ingroup and outgroup traits be utilized, the role of negatively or positively valenced incongruities in driving N400 variation might be explored.

Building on the above framework, our secondary goal was to explore links between ERPs and measures of implicit and explicit racial bias. Interpreting brain activity as a psychological phenomenon can be misleading absent corroboration with other measures (Guglielmi, 1999; Amodio, 2008). This issue is particularly problematic for existing N400 research on stereotyping, as N400 activity has not yet been linked to more traditional, implicit and explicit measures of bias. Explicit biases are consciously endorsed beliefs and judgments (Mitchell *et al.*, 2005) predicting consciously controlled behaviors such as verbal bias during intergroup interactions (Dovidio *et al.*, 2002) or decreased political support for President Obama (Hehman *et al.*, 2010). Implicit biases, on the other hand, generally manifest without an individual's awareness (Greenwald and Banaji, 1995) and predict unconscious behavior, such as negative non-verbal gestures or facial expressions during interracial interactions (Dovidio *et al.*, 2002), or seating distance from an outgroup member (Amodio and Devine, 2006).

### The current research

In summary, the question of whether the N400 can be utilized as an index of prejudice is currently unresolved. The current research sought to test this possibility in a semantic priming paradigm, while simultaneously replicating the limited research examining the N400 and stereotype accessibility. Specifically, we examined ERPs following both negative and positive, stereotypically in- and outgroup traits in an inter-racial context.

EEG was recorded while participants completed a semantic priming task, seeing first a black or white face prime, followed by negative and positive words stereotypically characteristic of blacks and whites. ERP analysis focused on the N400, but additionally examined the N100, associated with early attentional processes (Hillyard and Münte, 1984), and the P200, associated with greater attention to negative, threatening stimuli (Bartholow *et al.*, 2003) and implicated in implicit bias (Correll *et al.*, 2006; Payne, 2006; He *et al.*, 2009). ERP components were then examined for relationships with both implicit bias, collected during separate experimental sessions using the Implicit Association Test (IAT; Greenwald *et al.*, 2003) and self-reported explicit bias, as measured by the Attitudes Toward Blacks (ATB) scale (Brigham, 1993).

## METHOD

### Participants and design

Participants included 32 white, right-handed undergraduates (14 male) who participated for partial course credit in a 2(Prime: Black, White) × 2(Valence: Negative, Positive) × 2(Congruence: Congruent, Incongruent) repeated measures design.

### Stimuli

Face primes for the sequential priming task were color photos of 36 whites and 36 blacks with neutral expressions. Negative and positive, stereotypically black and white traits were selected from previous research on stereotype content (Kawakami and Dovidio, 2001; Madon *et al.*, 2001) and included based on pilot testing during which 35 participants evaluated 95 traits first on a stereotypically black = 1 to white = 7 and then a positive = 1 to negative = 7 continuum. Nine traits evaluated as most black/positive (e.g. athletic), black/negative (e.g. hostile), white/positive (e.g. educated) and white/negative (e.g. spoiled) were selected, resulting in a total of 36 traits. Across race, traits were rated as equally negative and positive, $F(1, 34) = 0.26$, $P = 0.612$ and did not vary in length by race $F(1,32) = 0.00$, $P = 1.000$ or valence condition $F(1,32) = 0.51$, $P = 0.482$ (Table 1).

**Table 1** Target traits and mean evaluations of negative to positive valence from pilot

| Black positive | Mean | White positive | Mean |
|---|---|---|---|
| Athletic | 6.29 | Educated | 6.46 |
| Cultural | 5.86 | Industrious | 5.21 |
| Humorous | 6.29 | Managerial | 6.35 |
| Masculine | 6.14 | Rich | 6.04 |
| Muscular | 5.71 | Smart | 6.56 |
| Rhythmic | 5.71 | Trusting | 6.17 |
| Strong | 5.86 | Wealthy | 5.83 |
| Well-built | 6.29 | Well-traveled | 5.00 |
| **Black negative** | **Mean** | **White negative** | **Mean** |
| Armed | 2.67 | Boring | 2.78 |
| Delinquent | 1.43 | Greedy | 2.08 |
| Hostile | 1.71 | Pasty | 2.67 |
| On welfare | 1.14 | Prejudiced | 2.04 |
| Poor | 1.29 | Pretentious | 2.50 |
| Quick tempered | 2.00 | Shallow | 2.28 |
| Unemployed | 1.29 | Spoiled | 2.12 |
| Violent | 1.14 | Whiny | 2.28 |

### Procedure

#### *Electroencephalography collection*

Participants received a general description of the experiment while being fitted with an electrode cap and then completed the sequential priming task. The task was presented on a 17 inch CRT monitor using Presentation (Neurobehavioral Systems). For each trial, the participant would see either a black or white face, followed by a negative or positive trait that was either stereotypically congruent (e.g. a black stereotype following a black face) or incongruent (e.g. a white stereotype following a black face) with the face prime. As explicit tasks increase the magnitude of the N400 (Chwilla *et al.*, 2000), participants rated the traits as negative or positive by response pad, using their left and right index fingers, the assignment of which was counterbalanced across participants. After a 5-trial practice block, participants completed 4 blocks of 72 trials each, for a total of 288 trials. Condition was randomized within block. On each trial, a fixation cross appeared for 500 ms. Facial primes were then presented for 500 ms, followed by a 500 ms blank screen before the target trait was presented for 1000 ms. A 1000 ms blank screen followed, after which the program proceeded to the next trial regardless of response.

#### *Implicit and explicit bias measures*

Participants were pretested before the experimental session to assess their explicit attitudes regarding blacks on the ATB (Brigham, 1993; $\alpha = 0.81$). Higher values indicate more positive feelings toward blacks, or less explicit bias. Following EEG collection ($M = 2.6$ days, s.d. = 4.11 days, range = 13 days), participants completed an evaluative IAT (Gaertner and McLaughlin, 1983) in a different location as part of a presumably unrelated experiment, scored as recommended by Greenwald *et al.* (2003). Participants were randomly presented with stimuli consisting of 6 white faces, six black faces, six 'good' adjectives (e.g. joy), and six 'bad' adjectives (e.g. cancer) that were evaluative in nature. Categories were paired creating either congruent (African American and Bad, European American and Good) or incongruent (African American and Good, European American and Bad) associations. Faces and adjectives presented in the IAT were different from those presented during the sequential priming task. Participants categorized each stimulus via response pad. Order of presentation was counterbalanced. Higher values indicate greater ingroup evaluative bias. Participants were compensated in the form of $5 or course credit. As expected, implicit (IAT D: $M = 0.37$, s.d. = 0.35) and explicit bias ($M = 5.14$, s.d. = 1.00) were not correlated ($r = -0.033$, $P = 0.867$).

#### Data acquisition and reduction

EEG data were collected from 32 Ag/AgCl electrodes embedded in an electrode cap. During recording, all activity was average referenced, while AFz served as the ground site. Electrode impedances were kept below 20 KΩ. Advanced Neuro Technology (ANT) acquisition hardware (Advanced Neuro Technology, Enschede, The Netherlands) was used to amplify, digitize (512 Hz) and filter (bandpass 0.1–30 Hz) the EEG signal. The EEG was corrected for eyeblinks using Advance Source Analysis software from ANT. Trials exceeding ±75 μV were rejected before signal averaging.

To create ERPs, EEG was digitally re-referenced offline to the average of the mastoids. As the stimulus-onset asynchrony (SOA) was 1000 ms, epochs associated with each face–trait pair were time locked to the presentation of the trait stimulus, rather than facial prime. Trials with too many artifacts or with a participant response that did not match the predetermined valence of the trait presented were excluded from further analysis. Subjects with fewer than 20 usable trials in any of the 8 conditions were removed from analysis ($n = 7$). Two additional participants were removed for failure to follow



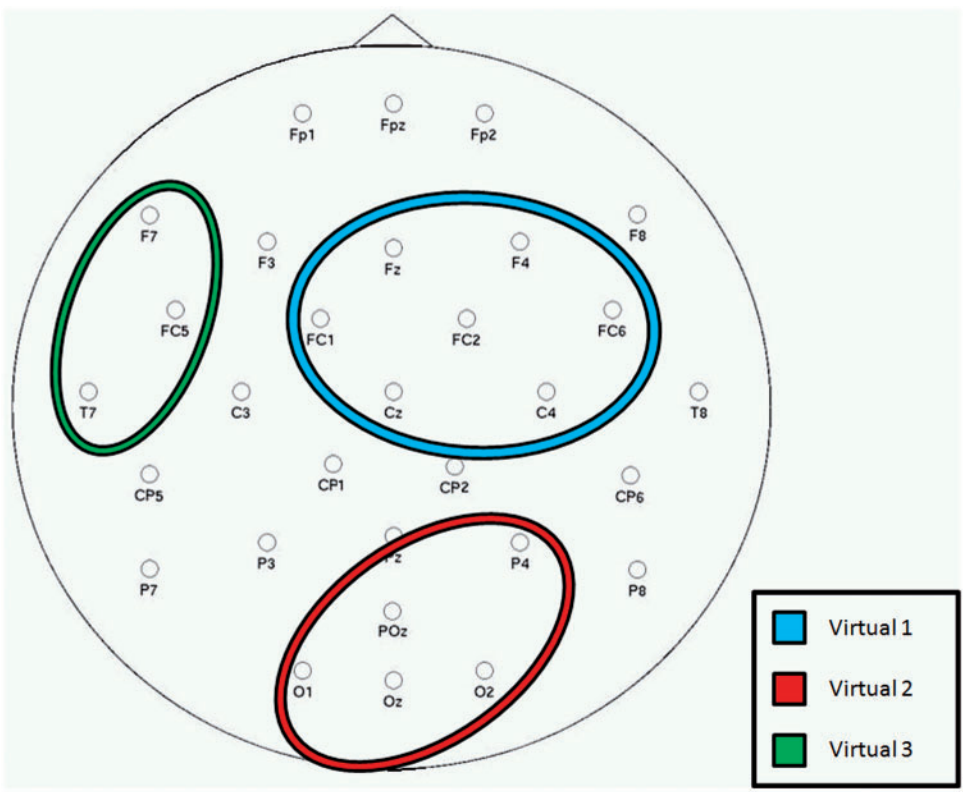


**Fig. 1** Virtual electrode clusters extracted from the spatial PCA.

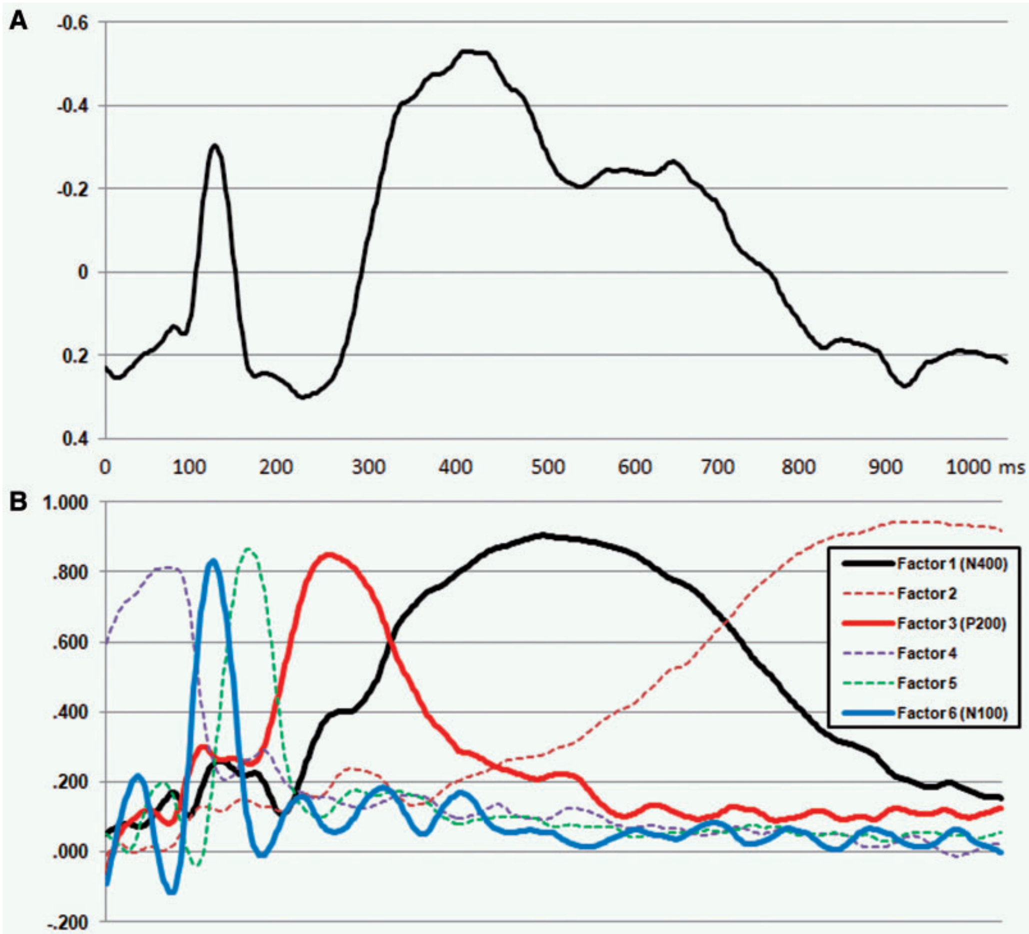


**Fig. 2** ERP at virtual electrode 1 (A) compared with temporal PCA loadings (B) used to identify the N100 (Component 6), the P200 (Component 3) and N400 (Component 1).

directions, resulting in 23 participants (7 male) suitable for data analysis. Each average was baseline corrected by subtracting the average voltage occurring during the 200 ms before stimulus onset from the entire average.

To reduce the dimensionality of the data, a spatial principle components analysis (PCA) was conducted on individual averages of each condition. The spatial PCA identifies and forms virtual electrodes from clusters of highly correlated electrodes and captures the variance uniquely associated with the scalp distribution of the ERPs (Spencer *et al.*, 2001). Four virtual electrode clusters emerged from the spatial PCA accounting for 85.2% of the variance (Figure 1; Cluster 4 was excluded as it closely related to activity at the two mastoids). Next, virtual ERPs were derived from the 'factor scores' for each participant, condition and virtual electrode at all time points. Virtual ERPs were then submitted to a temporal PCA, analyzing the covariance among time points for the 4 spatial factors, 8 experimental conditions and 23 participants to identify distinct components across time. Six temporal components from the PCA were extracted that accounted for 92% of the variance before varimax rotation. The temporal components were visually examined in conjunction with the virtual ERPs (Figure 2) to identify three temporal components of interest: N100 (Component 6), P200 (Component 3) and N400 (Component 1). We statistically analyzed all three temporal components at Virtual Electrode 1, where the amplitude for each was maximal. Components 2 and 4 were not analyzed because they reflected variance prior and subsequent to the epoch containing the ERP, and Component 5 was not analyzed as it did not reflect activity associated with Virtual Electrode 1.

## RESULTS

We first present preliminary analysis regarding behavioral response latencies during the task. We then examine the primary hypotheses regarding N400 variance across conditions, before additionally investigating N100 and P200 variation. Finally, we move on to exploring our secondary hypotheses regarding the relationships between ERP components and measures of implicit and explicit bias. We observed no effects of gender, and the analyses reported below collapse across this dimension.

### Statistical approach

Our primary hypotheses were tested by subjecting response latencies and ERP components to a 2(Prime: Black, White) × 2(Valence of target: Negative, Positive) × 2(Congruence: Congruent, Incongruent) repeated measures analysis of variance (ANOVA).

A different approach was utilized to examine our secondary hypotheses relating ERP components and prejudice measures. ERP difference scores were created to examine the effect of congruence, initially collapsing across valence, for each ERP component. These difference scores were created by subtracting ERP activity for incongruent conditions from congruent conditions, separately across race: (black congruent − black incongruent) and (white congruent − white incongruent]. In other words, an average of the ERPs elicited in trials where black faces were followed by words stereotypically associated with whites were subtracted from an average of the ERPs elicited in trials where black faces were followed by words stereotypically associated with blacks. The same was done for trials with white primes. In separate analyses, implicit and explicit bias scores were then regressed on both (black congruent − black incongruent) and (white congruent − white incongruent] difference scores simultaneously. Significant prediction of these difference scores is analogous to within-subject moderation (Judd *et al.*, 2001). Because these initial difference scores collapsed across the valence of trait to examine the effects of incongruence, additional difference scores were created for select follow-up analyses to explore the role of trait valence in eliciting responses.

### Response latencies

The ANOVA revealed two main effects of response latencies. Participants responded more quickly both to traits primed by black faces, $F(1, 22) = 6.16$, $P = 0.021$, $\eta_p^2 = 0.22$, and to negatively valenced traits, $F(1, 22) = 5.33$, $P = 0.031$, $\eta_p^2 = 0.20$. No other effects or interactions were present. Response latencies were not correlated with ERP activity.

### ERP reactivity across conditions

#### N400

The ANOVA examining N400 reactivity revealed a main effect of congruence, $F(1, 22) = 4.88$, $P = 0.038$, $\eta_p^2 = 0.18$. Larger negative deflections were demonstrated in response to incongruent ($M = -0.808$, s.d. = 0.80) than congruent face-trait pairings ($M = -0.720$, s.d. = 0.80), indicating that counter-stereotypical traits were more difficult to access (see Figure 3 for raw ERPs). This result conceptually replicates and extends previous research finding greater N400 reactivity for incongruent gender stereotypes (White *et al.*, 2009) by demonstrating the effect with a different task, SOA and stereotype group. Importantly, neither the prime × valence × congruence interaction, $F(1, 22) = 1.63$, $P = 0.215$, nor the two-way interactions predicting N400 reactivity were significant. Together, these results indicate that across all participants, stereotypes incongruent with blacks or whites, or stereotypes negative or positive in nature, did not differentially or multiplicatively elicit N400 activity. Rather, consistent with a non-evaluative interpretation of N400 reactivity, only congruence influenced N400 responses.

#### N100

A main effect of race of face prime emerged, $F(1, 22) = 6.48$, $P = 0.018$, $\eta_p^2 = 0.23$. Larger N100s were elicited by traits following black than white faces.

#### P200

A main effect of valence was evident, with greater P200 activity to negative traits, $F(1, 22) = 6.69$, $P = 0.017$, $\eta_p^2 = 0.23$, qualified by a three-way Prime × Valence × Congruence interaction, $F(1, 22) = 9.31$, $P = 0.006$, $\eta_p^2 = 0.30$ (Figure 4). Simple effects revealed that following black face primes, incongruent negative traits elicited greater P200 reactivity than incongruent positive traits, $F(1, 22) = 9.17$, $P = 0.006$, $\eta_p^2 = 0.29$. An opposite effect occurred following white face primes, as *congruent* negative traits elicited greater P200 reactivity than congruent positive traits, $F(1, 22) = 6.12$, $P = 0.022$, $\eta_p^2 = 0.22$. Greater P200 amplitudes to negative *vs* positive stimuli is consistent with previous research (Bartholow *et al.*, 2003), although the differential elicitation of the P200 by incongruent traits following a black prime and congruent traits following a white prime is novel. We return to interpreting this result in the Discussion.

### Relationships with implicit and explicit bias

We next examined our secondary hypotheses regarding how differential N400 reactivity might correlate with more traditional measures of implicit and explicit bias. Each ERP component was examined, but N100 activity was not moderated by implicit or explicit biases.

#### N400

As the N400 is a negative-going deflection, a difference score with a positive value indicates a larger N400 to incongruent compared with congruent face–trait pairings. We first examined whether more explicitly biased participants would demonstrate greater N400 reactivity than low-bias participants. ATB scores were regressed upon both black and white congruence difference scores simultaneously. Explicitly biased participants exhibited larger differences than low-bias participants in N400 activity between black congruent and incongruent trials ($\beta = -0.598$, $P = 0.003$) (Figure 5), but not on white trials ($\beta = -0.038$, $P = 0.837$). In other words, the more explicitly biased the participant, the greater the N400 displayed to stereotypically white traits following a black face (compared with stereotypically black traits following a black face).

The above analysis collapsed across stereotype valence. Thus, to explore whether the effect was particularly driven by N400 reactivity to negative or positive stereotypes, two additional difference scores contrasting N400 reactivity elicited by negative or positive stereotypes were simultaneously regressed on ATB scores. Again, the more



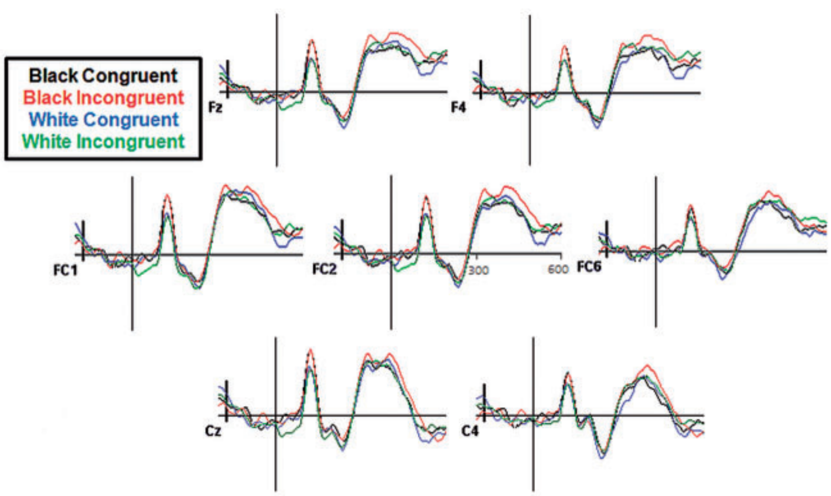


**Fig. 3** Raw ERPs for electrodes comprising virtual electrode 1 by condition.

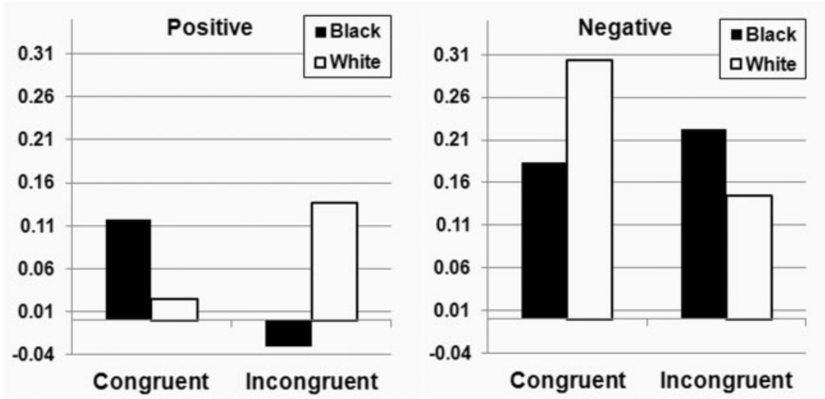


**Fig. 4** P200 factor scores by condition.

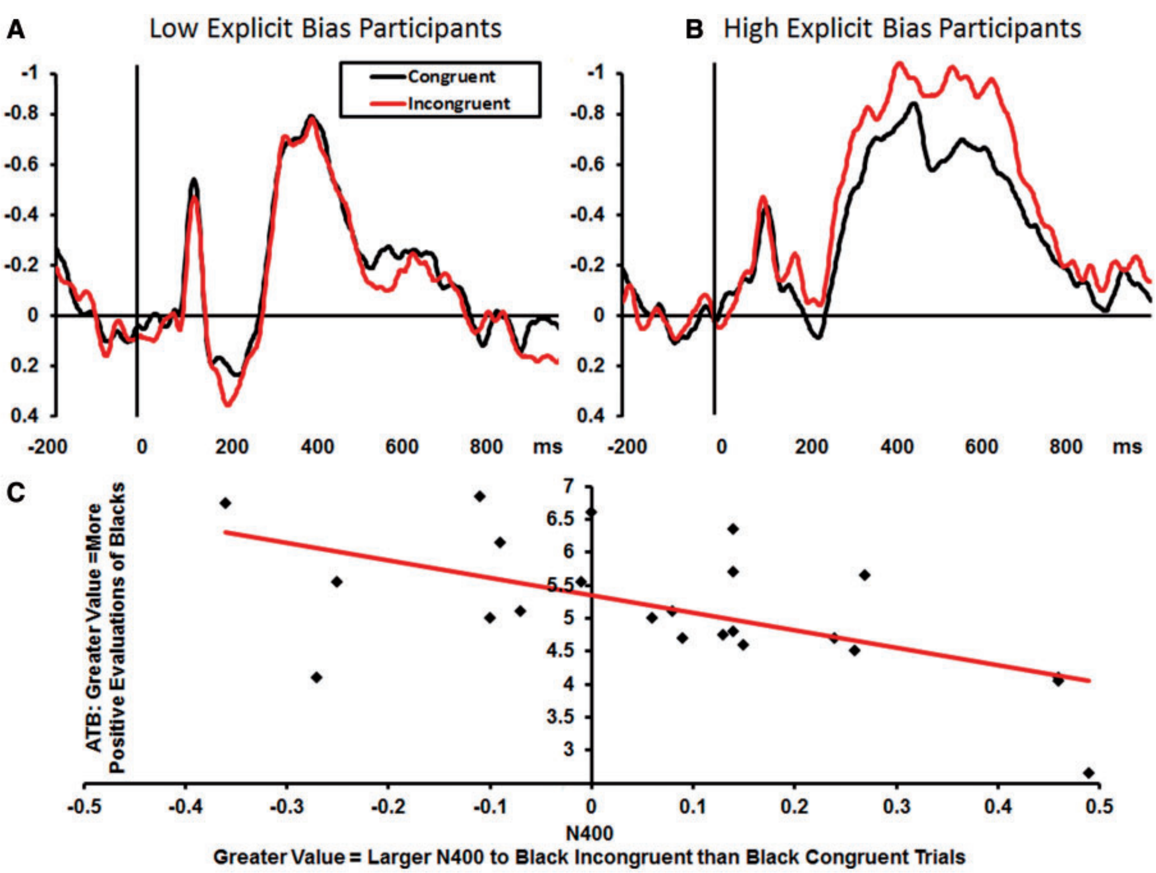


**Fig. 5** (A and B) ERPs at virtual electrode 1. Black congruent conditions contrasted with black incongruent conditions for both low (A; bottom half) and high (B; upper half) explicitly biased participants. (C) Correlation between difference in black congruent *vs* black incongruent conditions and ATB scores across all participants.

explicitly biased the participant, the greater the N400 demonstrated to stereotypically white traits following a black face. Results indicated that N400s elicited by both incongruent positive ($\beta = -0.436$, $P = 0.031$) and incongruent negative traits ($\beta = -0.692$, $P = 0.001$) were contributing uniquely to the relationship with explicit bias, although the effect was larger regarding negative traits.

Consistent with these results, N400 morphology varied between low and high explicitly biased participants (Figure 5). For participants lower in explicit bias, N400 activity was evident for approximately 300 ms, but for participants higher in explicit bias, N400 activity persisted for an additional 250 ms, indicating that individuals with greater explicit bias engaged in more prolonged processing. We return to this unexpected result in the Discussion.

Differential N400 reactivity between congruent and incongruent conditions following black primes robustly explained ~42% of the variance in ATB scores ($R^2 = 0.417$). No other effects regarding ATB scores were evident. As the N400 indexes the difficulty in accessing associated information (Nigam *et al.*, 1992), this result indicates that more explicitly biased participants may have struggled to incorporate both negative and positive stereotypically white traits with the context

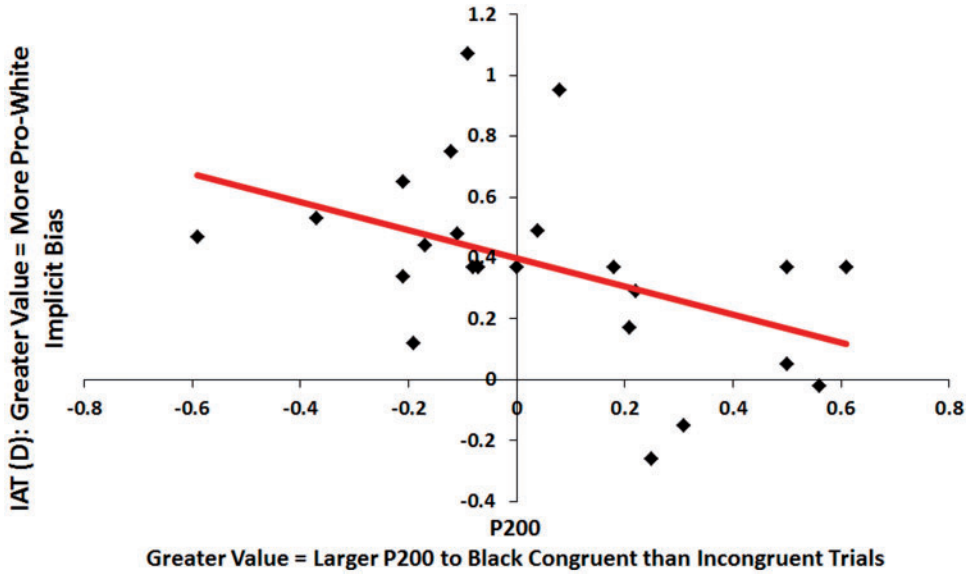


**Fig. 6** Correlation between difference in black congruent *vs* black incongruent conditions and IAT scores across all participants.

of black faces. Implicit bias was not related to N400 activity in any condition.

### *P200*

An analysis identical to that conducted with the N400 was conducted regarding variation on the P200. Implicit and explicit bias scores were regressed on the same series of condition-based difference scores. Unlike N400 reactivity, explicit bias did not predict differential P200 activity between any conditions.

Implicit bias, on the other hand, did predict differential P200 reactivity between conditions. Individuals with greater implicit bias demonstrated greater P200 reactivity differences between congruent and incongruent trials primed by black faces ($\beta = -0.560$, $P = 0.014$), but not trials primed by white faces ($\beta = -0.245$, $P = 0.247$), than low bias participants. Thus, the more implicitly biased the participant, the greater the P200 elicited by stereotypically White traits following a black face (compared with stereotypically black traits following a black face) (Figure 6).

We again simultaneously regressed the difference scores examining the effect of negative or positive valence on IAT scores. Neither relationship was significant ($p$s > 0.1), indicating that stereotype valence was not contributing uniquely. Congruence alone drove the effect. Differential P200 reactivity between congruent and incongruent conditions following black primes explained ~28% of the variance in the IAT ($R^2 = 0.284$). No other relationships with IAT scores were present.

## DISCUSSION

The current research examined the viability of assessing N400 reactivity to congruent and incongruent associations as an index of both stereotype accessibility and interracial prejudice. The current findings replicate previous research demonstrating greater N400 reactivity to stereotype–incongruent than to congruent associations, but with important methodological and theoretical differences. First, results were obtained regarding black and white stereotypes, instead of gendered stereotypes (White *et al.*, 2009). In addition, the results demonstrate that the N400 effect manifests even when stereotypes were unrelated to the task (evaluating words as negative or positive). In previous work, participants explicitly indicated whether target words matched the primes. That N400 reactivity was elicited even when participants were not explicitly associating targets and primes hints at the automatic nature of the association, supporting a spreading activation interpretation of the N400 (Franklin *et al.*, 2007). However, the racial element of this task may have been apparent to participants, and therefore, although it is clear that this task was *less* explicit than previous research, conclusive inferences regarding automaticity cannot be made.

Furthermore, N400 reactivity was demonstrated with a task including a 1000 ms SOA. Research indicates that evaluative priming does not occur with SOAs of greater than ~300 ms (Hermans *et al.*, 2001, 2003; Gawronski *et al.*, 2005; Castner *et al.*, 2007). Thus, the present result reinforces the semantic, rather than evaluative, nature of our paradigm, as well as indicating that separable neural systems may be involved with evaluative and semantic priming effects. The current research did not find differing response latencies for congruent and incongruent trials typically obtained in the above stereotype and evaluative priming research. However, given that the SOAs used in the current paradigm (1000 ms) were longer than is typical in this type of research, this absence may be unsurprising. Regardless, researchers interested in response latencies should note this difference.

In addition, the race-based stereotypes in the current work were sociopolitically charged, and participants would likely have been reluctant or unable to explicitly acknowledge the strength of their associations. The current results theoretically extend previous work by demonstrating that the N400 can be used as a metric of stereotype accessibility even in domains where social desirability is a salient issue. This result highlights the utility of the N400 for researchers interested in examining stereotype-accessibility in various domains.

The secondary goal of this study was to examine whether N400 reactivity could be extended to examine individual prejudices by contrasting N400 reactivity to both positively and negatively valenced stereotypes. Our approach was to correlate differences in individuals' responses to valenced conditions with more traditional and validated measures of interracial prejudice: self-report measures and the IAT. Indeed, a strong relationship was evident between N400 activity and self-reported explicit racial bias, as racially biased participants demonstrated particularly strong reactivity to stereotypically white words following black faces. Interestingly, both negatively and positively valenced stereotypes were independently involved in self-reported bias.

There was a striking difference in N400 morphology related to the explicit bias of the participant. In biased individuals, the N400 elicited by incongruent stereotypes endured for an additional 250 ms (Figure 5). To our knowledge, the only other research documenting such an effect was found comparing N400s elicited by incongruent words appearing as pairs or within a sentence (Van Petten, 1993). Van Petten found more enduring N400 activity (~200 ms longer) when incongruencies were located within sentences, compared with N400 activity following an incongruent word pair, concluding that the longer duration reflected the associated semantic detail retrieved for the word that elicited the N400. In the current research, a similar interpretation suggests that individuals with greater explicit bias might have engaged in longer and more complex analysis when stereotypically white words were primed with black faces, compared with their less biased peers. Thus, white words associated with black faces may be incongruent only for individuals with greater explicit bias, while less biased individuals exhibit an N400 more typical of semantic processing (Kutas and Hillyard, 1984; Nigam *et al*., 1992; Van Petten, 1993). This novel, although unexpected, result is consistent with the original hypotheses. The results of our secondary hypotheses should be viewed with some care, given our moderate sample size and the sometimes ephemeral nature of individual differences. That said, we note the robust relationship between N400 reactivity and self-reported bias ($R^2 = 0.417$) and believe that future work should examine the link between N400 reactivity and individual biases more closely.

Regarding effects involving other ERPs, participants demonstrated larger P200s to negative than positive traits in general, consistent with previous characterizations of P200 activity as primarily demonstrating vigilance to threat (Bartholow *et al*., 2003). In addition, the P200 was sensitive to the race of prime, valence and congruence of trait simultaneously as demonstrated by the three-way interaction (Figure 4). This pattern of activity might be explained as an overall reaction to negative White traits. Participants might have found negative ingroup traits more threatening than positive ingroup traits. In turn, these negative ingroup traits may have been particularly threatening when explicitly associated with the ingroup by the face prime.

An alternative to this explanation is that larger P200s were simply elicited by negative, stereotypically white words. We favor the incongruency explanation, however, given our care in selecting target traits. Negative stereotypically white words were not (i) more stereotypical or (ii) longer than any other trait categories and were not (iii) more negative in valence than negative black stereotypes. Future research may examine whether these traits varied on other dimensions, and how that might account for the current results. Nonetheless, the current results indicate that the P200 may be more sensitive to multiple dimensions than previously believed.

In addition, differential P200 reactivity predicted individuals' implicit bias. This relationship was evident only when contrasting congruent and incongruent conditions following black faces. Unlike the P200 interaction regarding activity across all conditions, neither the presence of a white prime nor valence of trait contributed to this effect. The P200 has previously been linked with behaviors associated with greater implicit bias, such as willingness to shoot or not shoot a target (Correll *et al*., 2006) and IAT scores (He *et al*., 2009). The present effect appears to replicate these patterns of results, although we note that the current methodology varies dramatically from them. Thus, the current results converge with a growing body of research pointing to the P200 as a valuable neurological index of affectively based biases with important behavioral outcomes, but future research should more closely examine the P200-implicit bias link to determine what these different tasks may be similarly capturing.

Although the difference scores utilized contrasts varying traits following the face prime (e.g. Black Congruent – Black Incongruent), an alternative method would have been to contrast conditions where the trait was held constant while varying the race of face prime (e.g. Black Congruent – White Incongruent). This alternative method of incongruence was in fact examined, but no relationships with ERPs were evident. Increased vigilance and attention following black primes may be responsible for this pattern of effects. Supportive of this possibility, faster response latencies and larger N100s were elicited by traits following black than white primes. Previous work has demonstrated that the N100 is sensitive to task factors (Hillyard and Münte, 1984) such as whether targets are armed in a shooter paradigm (Correll *et al*., 2006), indicating that participants may have believed that traits following Black faces were more relevant to the task, or that assessments of prejudice were being made on those trials and thus commanding attention. We had no a priori hypotheses about which contrasts might best predict bias, however, and future work should examine exactly why these particular comparisons were linked with bias. This result highlights the importance of corroborating neurological measures with other, better understood assessments of behavior (Guglielmi, 1999).

A limitation of the current work involves the selected stereotypical traits. Specifically, negative black stereotypes were evaluated as more negative than negative white stereotypes. This result is likely due to the physical danger connotations of the negative black stereotypes. We selected the most negatively or positively valenced traits in each category to maximize participant responsiveness, but more equally matched comparisons would have been preferable. However, it is critical to note that this difference in valence regarding negative stimuli does not account for any of the current effects. Participants generally responded to these different categories of stimuli equally and indeed sometimes demonstrated greater reactivity to negative white stereotypes (e.g. P200). Nonetheless, future research should ensure stimuli are equated on all dimensions.

In summary, the current research extends previous work finding that the N400 can act as a viable index of stereotype accessibility in interracial domains. In addition, we provide preliminary evidence that differential N400 and P200 reactivity to negatively and positively valenced stereotypes may provide a window into individual biases. These results warrant future attention regarding neurological indicators of prejudice and stereotype endorsement. Further research may be able to use such components to shed further light on the neurological roots of intergroup conflict.

## REFERENCES


Amodio, D.M. (2008). The social neuroscience of intergroup relations. *European Review of Social Psychology*, *19*, 1–54.

Amodio, D.M., Devine, P.G. (2006). Stereotyping and evaluation in implicit race bias: evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*, 652–61.

Bartholow, B.D., Pearson, M.A., Gratton, G., Fabiani, M. (2003). Effects of alcohol on person perception: a social cognitive neuroscience approach. *Journal of Personality and Social Psychology*, *85*, 627–38.

Bentin, S., McCarthy, G., Wood, C.C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*, 343–55.

Brigham, J.C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology*, *23*, 1933–67.

Bargh, J.A., Chartrand, T.L. (2000). *The Mind in the Middle: A Practical Guide to Priming and Automaticity Research*. New York, NY: Cambridge University Press.

Castner, J.E., Chenery, H.J., Copland, D.A., Coyne, T.J., Sinclair, F., Silburn, P.A. (2007). Semantic and affective priming as a function of stimulation of the subthalamic nucleus in Parkinson's disease. *Brain*, *130*, 1395–407.

Chwilla, D.J., Kolk, H.H.J., Mulder, G. (2000). Mediated priming in the lexical decision task: evidence from event-related potentials and reaction time. *Journal of Memory and Language*, *42*, 314–41.

Correll, J., Urland, G.R., Ito, T.A. (2006). Event-related potentials and the decision to shoot: the role of threat perception and cognitive control. *Journal of Experimental Social Psychology*, *42*, 120–8.

Cuddy, A.J.C., Fiske, S.T., Glick, P. (2007). The BIAS map: behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*, 631–648.

Daltrozzo, J., Schoen, D. (2008). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience*, *21*, 1882–92.

Dienstbier, R.A. (1970). Positive and negative prejudice: interactions of prejudice with race and social desirability. *Journal of Personality*, *38*, 198–215.

Dovidio, J.F., Evans, N., Tyler, R.B. (1986). Racial stereotypes: the contents of their cognitive representations. *Journal of Experimental Social Psychology*, *22*, 22–37.

Dovidio, J.F., Kawakami, K., Gaertner, S.L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–68.

Franklin, M.S., Dien, J., Neely, J.H., Huber, E., Waterson, L.D. (2007). Semantic priming modulates the N400, N300, and N400RP. *Clinical Neurophysiology*, *118*, 1053–68.

Gaertner, S.L., McLaughlin, J.P. (1983). Racial stereotypes: associations and ascriptions of positive and negative characteristics. *Social Psychology Quarterly*, *46*, 23–30.

Gawronski, B., Deutsch, R., Seidel, O. (2005). Contextual influences on implicit evaluation: a test of additive versus contrastive effects of evaluative context stimuli in affective priming. *Personality and Social Psychology Bulletin*, *31*, 1226–36.

Greenwald, A.G., Banaji, M.R. (1995). Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, *102*, 4–27.

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

Guglielmi, R.S. (1999). Psychophysiological assessment of prejudice: past research, current status, and future directions. *Personality and Social Psychology Review*, *3*, 123–157.

He, Y., Johnson, M.K., Dovidio, J.F., McCarthy, G. (2009). The relation between race-related implicit associations and scalp-recorded neural activity evoked by faces from different races. *Social Neuroscience*, *4*, 426–442.

Hehman, E., Gaertner, S.L., Dovidio, J.F. (2010). Evaluations of presidential performance: race, prejudice, and perceptions of Americanism. *Journal of Experimental Social Psychology*, *47*, 430–5.

Hermans, D., De Houwer, J., Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition and Emotion*, *15*, 143–165.

Hermans, D., Spruyt, A., Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: priming at SOA 300 but not at SOA 1000. *Cognition and Emotion*, *17*, 83–99.

Herring, D.R., Taylor, J.H., White, K.R., Crites, S.L. (2011). Electrophysiological responses to evaluative priming: the LPP is sensitive to incongruity. *Emotion*, *11*, 794–806.

Hillyard, S.A., Münte, T.F. (1984). Selective attention to color and location: an analysis with event-related brain potentials. *Perception & Psychophysics*, *36*, 185–98.

Judd, C.M., Kenny, D.A., McClelland, G.H. (2001). Estimating and testing mediation and moderation in within-subject designs. *Psychological Methods*, *6*, 115–34.

Kawakami, K., Dovidio, J.F. (2001). The reliability of implicit stereotyping. *Personality and Social Psychology Bulletin*, *27*, 212–5.

Kutas, M., Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–3.

Kutas, M., Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–70.

Lau, E.F., Phillips, C., Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*, 921–933.

Madon, S., Guyll, M., Aboufadel, K., Montiel, E., Smith, A., Palumbo, P., Jussim, L. (2001). Ethnic and national stereotypes: the Princeton trilogy revisited and revised. *Personality and Social Psychology Bulletin*, *27*, 996–1010.

Mitchell, J.P., Banaji, M.R., Macrae, C.N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *17*, 1306–15.

Neely, J.H. (1991). Semantic priming effects in visual word recognition: a selective review of current findings and theories. In: Besner, D., Humphreys, G.W., Besner, D., Humphreys, G.W., editors. *Basic Processes in Reading: Visual Word Recognition*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., pp. 264–336.

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

Payne, B.K. (2006). Weapon bias: split second decisions and unintended stereotyping. *Current Directions in Psychological Science*, *15*, 287–91.

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

Spencer, K.M., Dien, J., Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, *38*, 343–58.

Van Petten, C. (1993). A comparison of lexical and sentence-level context effects in event-related potentials. *Language and Cognitive Processes*, *8*, 485–531.

Wang, L., Ma, Q., Song, Z., Shi, Y., Wang, Y., Pfotenhauer, L. (2011). N400 and the activation of prejudice against rural migrant workers in China. *Brain Research*, *1375*, 103–10.

White, K.R., Crites, S.L., Taylor, J.H., Corral, G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Social Cognitive and Affective Neuroscience*, *4*, 191–8.