OXFORD

Social Cognitive and Affective Neuroscience, 2016, 1698–1706

doi: 10.1093/scan/nsw082
Advance Access Publication Date: 21 June 2016
Original article

The mere presence of an outgroup member disrupts the brain's feedback-monitoring system

Nicholas M. Hobson[1] and Michael Inzlicht[1,2]

[1]Department of Psychology, University of Toronto and [2]Rotman School of Management, University of Toronto, Toronto, Ontario, Canada, M5S 3G3

Correspondence should be addressed to Nicholas M. Hobson, Department of Psychology, University of Toronto, 1265 Military Trail Toronto, Ontario, Canada M1C 1A4. E-mail: nick.hobson@utoronto.ca

Abstract

Much of human learning happens in the social world. A person's social identity—the groups to which they belong, the people with whom they identify—is a powerful cue that can affect our goal-directed behaviors, often implicitly. In the present experiment, we explored the underlying neural mechanisms driving these processes, testing hypotheses derived from social identity theory. In a within-subjects design, participants underwent a minimal group manipulation where they were randomly assigned to an arbitrary ingroup. In two blocks of the experiment, participants were asked to complete a task for money while being observed by an ingroup member and outgroup member separately. Results revealed that being observed by an ingroup or outgroup member led to divergent patterns of neural activity associated with feedback monitoring, namely the feedback-related negativity (FRN). Receiving feedback in the presence of an ingroup member produced a typical FRN signal, but the FRN was dampened while receiving feedback in the presence of an outgroup member. Further, this differentiated neural pattern was exaggerated in people who reported greater intergroup bias. Together, the mere presence of a person can alter how the brain adaptively monitors feedback, impairing the reinforcement learning signal when the person observing is an outgroup member.

Key words: intergroup bias; social identity; feedback-related negativity; reward-monitoring

Introduction

Rarely do people learn about their own behavior in privacy. From infancy, we have a preparedness to take in information and feedback from those around us to help guide our behaviors. Throughout life, learning is provided against a backdrop of interpersonal interactions, team performances, romantic and familial bonds and professional partnerships. In short, when it comes to understanding our behaviors, we rely on others for cues. What happens in our brain as we monitor our behavior in the presence of a person who is either a part of our group or not? In this experiment, we ask whether the social context, specifically the presence of ingroup *vs* outgroup members, can alter neural feedback monitoring.

Research has consistently demonstrated that feedback monitoring is instrumental for guiding our performance (e.g. Holroyd and Coles, 2002; Cohen and Ranganath, 2007). But how is this feedback monitoring shaped by the background social context? And does the brain compute feedback differently depending on whether we are around an ingroup member or outgroup member? This experiment examines these questions to explore whether group membership differentially affects the brain's basic feedback monitoring.

Neural monitoring and the social context

Research has shown that social contexts and interpersonal interactions impact neural processing of goal-relevant information. For example, action- and feedback monitoring is amplified in joint player situations where one's performance has a direct effect on others (Koban *et al.*, 2010, 2011; de Bruijn *et al.*, 2011), and it can be altered depending on whether the context is cooperative, competitive or neutral (de Bruijn *et al.*, 2008, 2011; Van Meel and Van Heijningen, 2010; Radke *et al.*, 2011) or when the outcome of one's performance causes another person physical pain (Koban *et al.*, 2013).

Recent work that harkens back to social facilitation theory (e.g. Zajonc, 1965; Baron, 1986; Huguet *et al.*, 1999) has shown

Received: 1 December 2015; **Revised:** 24 May 2016; **Accepted:** 9 June 2016

that even the presence of others can lead to increased activity in the ventral striatum after the delivery of rewarding feedback (Simon *et al.*, 2014). This shows that being watched by others increases activation in one of the regions involved in reward processing, suggesting that even subtle social information like being observed, though irrelevant to the task, can boost the brain's reward signal.

Together, previous work has documented the effects of the social context on neural feedback monitoring, providing evidence that the brain represents others when calculating different goal-directed behaviors for the self. Even a person knowing that others are observing them can enhance the feedback signal (Simon *et al.*, 2014), aligning with previous theory suggesting that groups can influence a person's individual performance (Paulus, 1983). Here, we wondered if it matters who is doing the observing and the nature of the relationship. We ask: How is neural feedback monitoring impacted by intergroup dynamics and a person's social identity?

Social identity, group biases and self-regulatory behaviors

According to social identity theory, people readily categorize themselves into groups (Tajfel *et al.*, 1971; Tajfel, 1974; Turner *et al.*, 1987; Brewer, 1999). In an effort to protect their sense of 'us', people will often direct their behaviors and decision making in a biased manner to bolster their shared social identity (i.e. favoring ingroup members) and minimize potential threats against it (i.e. outgroup discrimination). In short, social identity, as a core value to the self, plays into the regulatory function of a person's behavior.

It is well understood that the regulation of the self is a psychosocial process that relies on a number of social and cultural factors (Markus and Wurf, 1987; Karoly, 1999; DeWall *et al.*, 2008). Longstanding research in social psychology has shown that people are motivated to monitor and learn from their behavior more (or less) depending on who is around them and the group categorizations that are made salient (e.g. Anderson and Glassman, 1996). The relational-self—a mental representation of a person's self-knowledge that is linked with the knowledge about significant others—provides cues for a person's own motivated behavior, reinforcement learning and self-regulation (Anderson and Chen, 2002). For example, it has been shown that negative information is more salient in the presence of an ingroup member than an outgroup member (Festinger, 1957). Similarly, intergroup biases lead people to place more trust in ingroup members than outgroup members and to treat ingroup members as a more valid source of information (Brewer, 1979; Turner *et al.*, 1987).

The presence of either an ingroup or outgroup member can act as a subtle, yet powerful, cue. Since people are less motivated to process information and self-monitor when the social context is tied to an outgroup interaction (Festinger, 1957; Brewer, 1979; Mackie *et al.*, 1990; van Knippenberg and Wilke, 1992; van Knippenberg, 1999), we wonder if the presence of an outgroup member (but not an ingroup member) would impair the brain's feedback-monitoring signal.

This question sits alongside a growing body of research that has begun to uncover the neural substrates of intergroup relations (for reviews, see Molenberghs, 2013; Amodio, 2014; Cikara and Van Bavel, 2014), including the social categorization and motivated evaluation related to intergroup dynamics (e.g. Golby *et al.*, 2001; Van Bavel *et al.*, 2008, 2011; Morrison *et al.*, 2012; Ratner and Amodio, 2012). Finally, even minimal group memberships—groups devoid of any prior history—have been shown to modulate these processes. In one study, researchers used a minimal group paradigm to show that the medial prefrontal cortex, an area of the brain attuned to social categorization, is activated when participants consider belonging to even the simplest of groups (Molenberghs and Morrison, 2014). These minimal and arbitrary social categorizations have also shown to bias the neural processing underlying action-perception systems, with relative increased activation in the inferior parietal lobule during the observation of fellow group members compared to outgroup members (Molenberghs *et al.*, 2012). Critically, some of these automatic and unconscious neural mechanisms have been found to influence broad discriminatory attitudes and behavior, like the failure of empathic responding to outgroup members' misfortunes (e.g. Cikara and Fiske, 2011, 2012).

Taken together then, we predict that if a person's group categorization is a motivationally salient feature that serves to maintain their social identity, then even the simplest intergroup context (e.g. using minimal groups), and the mere presence of an ingroup member or outgroup member, will alter neural feedback monitoring and lead to a failed feedback signal in the presence of an outgroup member in particular. We arrive at this prediction because longstanding research holds that interactions with outgroup members are perceived as relatively less important (e.g. Festinger, 1957). This automatic appraisal, we propose, interferes with basic motivational and regulatory functions, as reflected in disrupted feedback-monitoring signals of the brain. Importantly, this hypothesis would underscore the impact of how a person is less motivated to care about the quality of outgroup interactions, even when doing so comes at a detriment to one's own performance.

The feedback-related negativity—sensitivity to social motivational features

In this study, we examine the feedback-related negativity (FRN), an event-related potential (ERP) that tracks neural activation within 200–500 ms following task feedback and that is thought to reflect the neural reactivity to external feedback (Holroyd and Coles, 2002; Hajcak *et al.*, 2005). Traditional accounts of the FRN suggest that it is generated by the anterior cingulate cortex (ACC), indicating a negative reward-prediction error that facilitates subsequent behavioral adjustment (e.g. Miltner *et al.*, 1997; Gehring and Willoughby, 2002, 2004, Holroyd and Coles, 2002; Luu *et al.*, 2003; Bellebaum and Daum, 2008). These reinforcement models suggest that the FRN is modulated by phasic changes in mesencephalic dopamine activity (Holroyd and Coles, 2002; Bellebaum and Daum, 2008), which cues reinforcement learning and optimizes subsequent behavior (e.g. Nieuwenhuis *et al.*, 2004; Yasuda *et al.*, 2004; Cohen and Ranganath, 2007). Recent evidence suggests, however, that the FRN may also emerge as a positive wave form that is sensitive to gains rather than losses (e.g. Baker and Holroyd, 2011; Carlson *et al.*, 2011; Kujawa *et al.*, 2013). Together then, the FRN likely represents two separate but overlapping processes: a negative waveform that is heightened during aversive feedback and a positive waveform that is heightened during rewarding feedback. As a result, it is an established practice in ERP research to interpret the FRN as a difference-wave score (e.g. Luck, 2005; Holroyd and Krigolson, 2007).

Finally, there is evidence that neural feedback monitoring is also modulated by the motivational inputs of social and interpersonal dynamics (e.g. Ma *et al.*, 2011). Given that an intergroup context can be particularly motivating situation for a person who is strongly identified with a group (e.g. Tajfel, 1982), it is plausible that the FRN will be selectively altered by the mere presence of an ingroup member *vs* outgroup member. More to the point, because people are generally less concerned with the outgroup compared to the ingroup, we expect to see a disrupted FRN signal during outgroup observation. We also note that since we did not specify and pre-register our hypotheses a priori, the present hypotheses should be considered exploratory in nature. Thus, any observed effects in favor of our predictions would need to be replicated in follow-up research with a confirmatory style approach.

The present research

This study investigated online measures of neurophysiological feedback monitoring in an experimentally manipulated minimal group context (Tajfel and Turner, 1979). More specifically, we measured people's FRN activation after they were delivered monetary feedback on a task in the presence of either a minimal ingroup or outgroup member. Importantly, using minimal groups allowed us to control for previous history in established group categories, thus providing a clear and direct test of our hypotheses. That is, any group factors beyond basic categorization would not be able to account for the observed effects in this experiment.

We suspected that the FRN would be moderated by the minimal intergroup context, leading to a disrupted FRN during outgroup presence, relative to the FRN during ingroup presence. Further, we predicted that attitudes of intergroup bias, even biased attitudes with lab-created groups (e.g. Coull *et al.*, 2001) would lead to polarized patterns in the FRN, with biased people eliciting a diminished FRN around an outgroup member, and non-biased people eliciting comparable FRNs around ingroup and outgroup members.

Methods

Participants

Thirty students from the University of Toronto Scarborough participated for course credit and were also told that they could earn extra money as bonus depending on their performance. But due to the nature of the task, all participants received $10 regardless of their performance. Five participants were excluded from all analyses due to computer/hardware malfunction ($n = 1$), high (electroencephalographic) EEG artifact rate (>30% artifacts; $n = 1$) or suspicion of the minimal group manipulation and/or of the confederates' role in the study ($n = 3$). We were left with a total sample of 25 participants (15 females, 10 males; mean age = 19.3, s.d. = 2.6). Participants were recruited online over the introductory psychology web portal. A power analysis using G*Power (Faul *et al.*, 2007) was run on our sample size. Assuming a small to medium effect size (i.e. $r = 0.21$, $d = 0.43$; Richard *et al.*, 2003) and high correlations between repeated measures typical of ERP designs ($r = 0.60$–0.80; Olvet and Hajcak, 2009; Segalowitz *et al.*, 2010), a sample size of 25 participants yields a power value of 0.84.

Procedures

Participants signed up for the study at home and were told that the purpose of the experiment was to measure personality and styles of cognition. Borrowing from social identity theory (e.g. Tajfel, 1974), we used a minimal group manipulation, the dot-estimation task, in which people are placed in arbitrary groups based on whether they are deemed 'under-estimators' or 'over-estimators' (see Figure 1 for a schematic diagram of the procedures). This would serve as the intergroup dynamic in the lab. At home, participants first completed a short survey in which they provided demographic information. At the end of the survey they completed a simple test, which they were led to believe was related to their cognition. Participants were shown three images, each one made up of an array of black dots on a white background, and asked to estimate the total number of dots in the image. They were informed that their responses were being recorded.

Participants then came into the lab where, for the first time, they were told that the experiment would be done in groups of four. For each session, however, there was only one real participant and three confederates, creating two groups of two. The experimenter, ostensibly using the saved data from the participants' earlier dot-estimating responses, told the participants that there are generally two types of people, over-estimators and under-estimators and that the data from before revealed that two of them were over-estimators and the other two, under-estimators. The experimenter asked each student to wear either a red shirt (ingroup with the real participant and one confederate) or a blue shirt (outgroup with two confederates). Importantly, to the one real participant, it appeared as though two groups were created based on the personal dot-estimation data. In addition, the confederates were randomly categorized into groups across all sessions in order to remove any confounds related to the confederates' appearance, gender and/or ethnicity.

The experimenter told the group that since there was only enough time to complete the EEG set-up for one person, that a name would be randomly drawn to determine who would be hooked up for the current testing session; the name of the real participant was drawn in each case. The experimenter brought each member to their own testing room and completed the EEG setup and gelling for the real participant while the confederates waited. The actual testing room (where the task performance/observation happened while recording the EEG) was where the participant was first seated. The confederates were brought into this room separately and sat next to the participant (more details below).

Time estimation task. To test the influence of social context and group membership on participant's reward-monitoring processes, continuous EEG was recorded while the participants received either winning or losing feedback. To incentivize their performance, participants were told that they would have the chance to earn extra bonus money in a cognitive/perception timing task. To the participants then, the task was related to the purpose of the study, namely perceptual styles and timing, but, importantly, the task allowed us to surreptitiously gather their responses to favorable and unfavorable outcomes. The experimenter told the participants that in each block of the task, one person would be performing (receiving feedback) and another observing (watching the other person receive feedback), while sitting in the same room side-by-side.

During the task, each trial began with a fixation cross presented for 250 ms, followed by a blank screen. Participants were instructed to press the spacebar when they believed 1 s had passed since the appearance of the fixation cross. Visual performance feedback was provided 2 s after the initial cue, resulting in an approximately 1 s interval between response and feedback. The feedback stayed on screen for 1 s and was followed by an intertrial interval varying between 1 and 2 s (see Supplementary Material for more detailed procedures).

Following the time-estimation task, participants then completed a filler questionnaire, which they were told would help

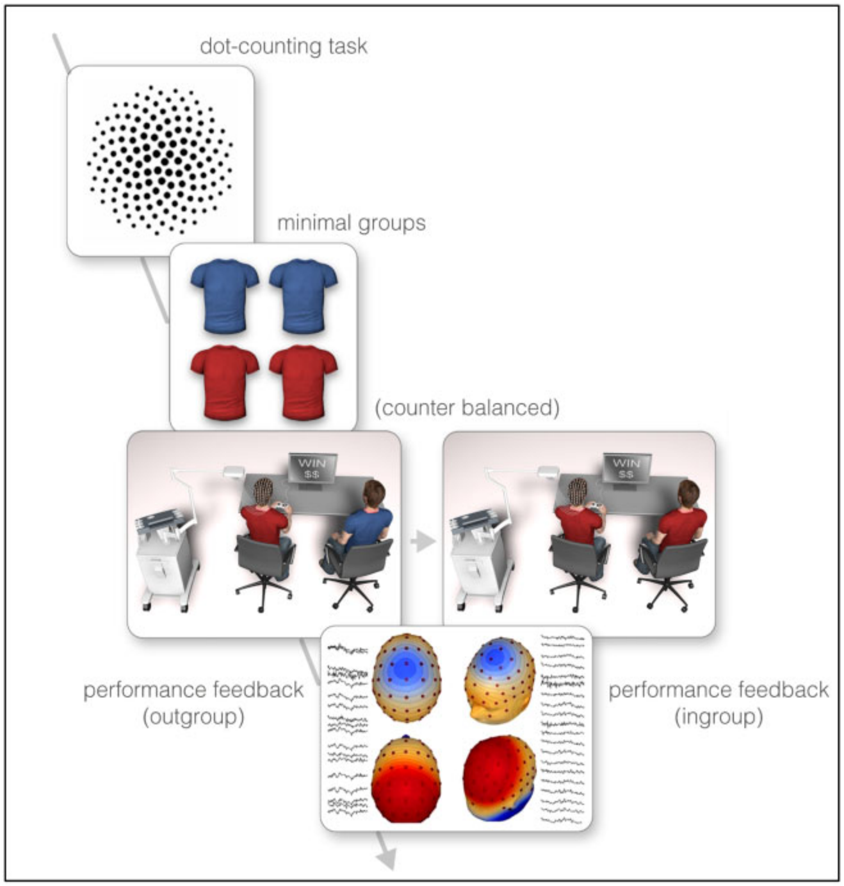
dot-counting task
minimal groups
(counter balanced)
WIN $$
WIN $$
performance feedback (outgroup)
performance feedback (ingroup)

Fig. 1. Schematic diagram of the experimental design. Participants were categorized based on an arbitrary over- or under-estimation of dot counting and then received monetary feedback on a performance task in the presence of both ingroup and outgroup members separately.

researchers understand the connection between personality and cognition. Embedded in the survey were two Feelings Thermometer questions (Esses *et al.*, 1993; 0–10 scale), asking how warm (10) or cold (0) they felt toward the participant with the red shirt (i.e. ingroup member) and the two participants with the blue shirt (i.e. outgroup member). To control for demand effects, we framed the questions around the study's purpose of perception, asking the participant to form an impression of other people based on initial 'person perceptions'.

Neurophysiological recording

Continuous EEG was recorded during the four blocks of the time-estimation task using a stretch Lycra cap embedded with 32 tin electrodes (Electro-Cap International, Eaton, OH), focusing solely on midline electrode sites (Fz, FCz, Cz, CPz, Pz and Oz). The EEG recording was marked independently for the two observation blocks (ingroup *vs* outgroup member observers) in order to analyze the separate FRN signals. Recordings used average ear and a forehead channels as reference and ground, respectively. The continuous EEG was digitized using a sample rate of 512 Hz, and electrode impedances were maintained below 5 kΩ during recording. Offline, EEG was analyzed with Brain Vision Analyzer 2.0 (Brain Products GmbH, Munich, Germany). EEG data were corrected for vertical electrooculogram artifacts (Gratton *et al.*, 1983). An automatic procedure was employed to detect and reject artifacts. The criteria applied were a voltage step of more than 25 μV between sample points, a voltage difference of 150 μV within 150 ms intervals, voltages above 85 μV and below −85 μV and a maximum voltage difference of less than 0.50 μV within 100 ms intervals. These intervals were rejected from individual channels in each trial.

A challenge in measuring the FRN is that its latency partially overlaps with another ERP—the P300—that has different spectral (frequency) features.[1] In order to control for this, we quantified the FRN in the averaged ERP waveforms using the base-to-peak method (e.g. Yeung and Sanfey, 2004; Hajcak *et al.*, 2005; Cohen *et al.*, 2007). In tasks where the FRN and P300 are both generated, the base-to-peak method provides the most reliable and accurate estimate of the FRN signal (e.g. Yeung and Sanfey, 2004). Specifically, the FRN was defined at the central electrode site, Cz, as the difference between the maximum value between 150 ms and 350 ms following feedback presentation and the most negative

[1] Visually inspecting the ERPs, it appears that P300 might also be modulated by intergroup observation. We therefore conducted similar analyses looking at the P300 as defined as the average mean amplitude between 200 and 400 ms following the presentation of the feedback stimulus. The results from a multilevel linear mixed model revealed a non-significant two-way interaction between group (ingroup observation *vs* outgroup observation) and feedback type (loss *vs* gain), $b = 0.69$, $SE = 0.46$, $t(23) = 1.49$, $P = 0.15$, indicating that the differentiation between losses and gains in the P300 was statistically comparable during both ingroup observation and outgroup observation. The effect of feedback monitoring, therefore appears limited to the early, unconscious detection of feedback as reflected in differences existing solely in the FRN signal.

point between this maximum and 350 ms following feedback onset. If there was no negative-going deflection, the FRN was scored as a zero. Cz was selected a priori because of previous work showing the feedback- and performance monitoring ERPs are maximal at fronto-central sites (e.g. Inzlicht and Gutsell, 2007; Hirsh and Inzlicht, 2008; Luck and Kappenman, 2012; Nash *et al.*, 2014).

Results

All data and analyses code can be found on Open Science Framework at osf.io/mvgr4. First looking at participants' reported attitudes toward ingroup and outgroup members, a one-way repeated measures analysis of variance, with group member (ingroup *vs* outgroup) as the repeated factor and feelings thermometer ratings at the dependent variable, revealed a significant effect of group member rating, $F(1,24) = 4.57$, $P = 0.04$, $\eta_p^2 = 0.16$, such that participants reported feeling more warm toward their minimal ingroup members ($M = 5.56$, s.d. = 1.23) relative to minimal outgroup members ($M = 5.08$, s.d. = 1.58).

FRN and ΔFRN results

First, we modeled the effects of group member observation (contrast codes: 0.5 = ingroup, −0.5 = outgroup) and feedback type (contrast codes: 0.5 = loss, −0.5 = gain) on feedback monitoring for the raw FRN waveform components. A two-level multilevel model was used to account for FRN activity nested within participants by estimating a random intercept and random slope for each participant. We used an unstructured covariance matrix and the between-within method of estimating degrees of freedom. Effect sizes were estimated with semi-partial R^2 (Edwards *et al.*, 2008).

As expected, the model output showed a significant effect of feedback type, $b = 0.97$, SE = 0.26, $t(24) = 3.66$, $P = 0.001$, semi-partial $R^2 = 0.32$, indicating that across member observation conditions, participants elicited a larger (i.e. more negative) FRN in response to losing feedback ($M = -2.41\,\mu V$) than; s.d. = 1.61) compared to winning feedback ($M = -1.76$, s.d. = 1.57). Importantly, the output also revealed a significant interaction between group member observation and feedback type, $b = -0.62$, SE = 0.27, $t(24) = -2.19$, $P = 0.039$, semi-partial $R^2 = 0.17$. Figure 2 illustrates the ERPs and graph figures. Parsing apart the interaction and looking first across feedback type (within observation conditions), simple effects tests showed that the FRN amplitude was significantly larger for losses (losses ($M = -2.65\,\mu V$, s.d. = 1.70) than wins ($M = -1.68\,\mu V$, s.d. = 1.66), but only when participants were being observed by an ingroup member, $t(24) = 3.66$, $P = 0.001$, $d = 0.58$. This differentiation in feedback type did not hold when participants were being observed by an outgroup member, $t(24) = 1.65$, $P = 0.11$, $d = 0.28$, with more similar feedback activity in response to losses ($M = -2.17$, s.d. = 1.78) and wins ($M = -1.82$, s.d. = 1.63). In addition to our main test, it was recommended by reviewers that we also look across conditions of group member observation (as an exploratory test in addition to our main hypotheses; see Luck and Gaspelin, in press). The simple effects revealed that there was no difference in the FRN for win trials $t(24) = 0.74$, $P = 0.47$, $d = 0.09$, but a marginally significant difference in the FRN for loss trials $t(24) = 1.78$, $P = 0.088$, $d = 0.28$.

Next, in line with standard ERP practice, we modeled the data using a difference-wave score (ΔFRN = lossFRN – gainFRN) to represent feedback-monitoring activity. The difference-wave approach is helpful in isolating and drawing statistical inferences from waveform components because they have lower signal to noise ratio than those of the raw ERPs (Luck, 2005). Here, ΔFRN activity was modeled as a function of group membership of the observer (i.e. ingroup *vs* outgroup). Mirroring the above analyses with the raw ERP scores, the model indicated a main effect of group member observation, $b = -0.62$, SE = 0.27, $t(24) = -2.19$, $P = 0.039$, semi-partial $R^2 = 0.17$. See Table 1 for descriptive means.

Here, we have evidence that the typical, and robust, FRN patterning (i.e. differentiation between losses and gains and larger amplitude for losses) is reduced when a minimally categorized outgroup member is observing a person perform. This lack of differentiation in the amplitudes of gains and losses is inconsistent with the standard patterning of the FRN (for a recent ERPs meta-analysis see Gillan and Robbins, 2014). This atypical pattern in FRN amplitudes also mirrors the psychiatric cases of obsessive-compulsive disorder, a condition characterized by dysfunctional performance/feedback-monitoring systems of the brain (e.g. O'Toole *et al.*, 2012; Gillan and Robbins, 2014). Additionally, when we compare feedback type across condition, it appears the lack of differentiation during outgroup observation arose from the FRN signal in response to negative feedback stimuli as opposed to positive feedback stimuli. This mirrors previous work showing that the FRN, as a proxy for signaling performance adaptation and agentic control, is most sensitive to errors and negative feedback (e.g. Bellebaum *et al.*, 2010).

The moderating role of reported group bias

Next, we wanted to test whether participants' FRN activity was moderated by self-reported attitudes toward ingroup and outgroup members. Prior to analyses, we created an intergroup bias score from the two feelings thermometer attitude ratings (ingroup and outgroup ratings). We subtracted the outgroup attitude ratings from the ingroup attitude ratings, with larger positive values indicating an ingroup favoring bias ($M_{\text{overall.bias}} = 0.48$, $\text{s.d.}_{\text{overall.bias}} = 1.13$). Each bias score was then grand mean centered. The model's two-way interaction between group membership observation and feedback type remained significant, $b = -0.62$, SE = 0.26, $t(23) = -2.36$, $P = 0.027$. Importantly, there was also a significant three-way interaction with intergroup bias score, $b = -0.05$, SE = 0.02, $t(23) = -2.20$, $P = 0.038$. For interpretability and graphing purposes, we modeled the interaction with ΔFRN. Figure 3 illustrates the two-way interaction for the effect on ΔFRN. As can be seen, participants' intergroup bias score significantly moderated the effects of group membership of observer on the ΔFRN amplitude.

A series of simple slopes analyses were conducted, once with participants who showed the presence of an intergroup bias ($M = 2.0_{\text{bias.present}}$, $\text{s.d.}_{\text{bias.present}} = 1.55$) and again with participants who showed no bias (i.e. equivalent attitudes ratings to both ingroup and outgroup members). These analyses revealed that when intergroup bias was present, group membership of observer predicted ΔFRN activity, $t(23) = 3.23$, $P = 0.004$; but this was not the case at when intergroup bias was absent, $t(23) = 0.09$, $P = 0.92$. Here we see that participants who reported no group bias showed no difference in feedback monitoring when in the presence of either group member. However, participants who reported group bias revealed differentiated patterns in their feedback monitoring, showing increased levels of ΔFRN activity while being observed by an ingroup member and reduced levels of ΔFRN activity while being observed by an outgroup member. Tying together the initial FRN findings with these ones, we provide further support that group categorization leads to a failure in feedback monitoring when in the

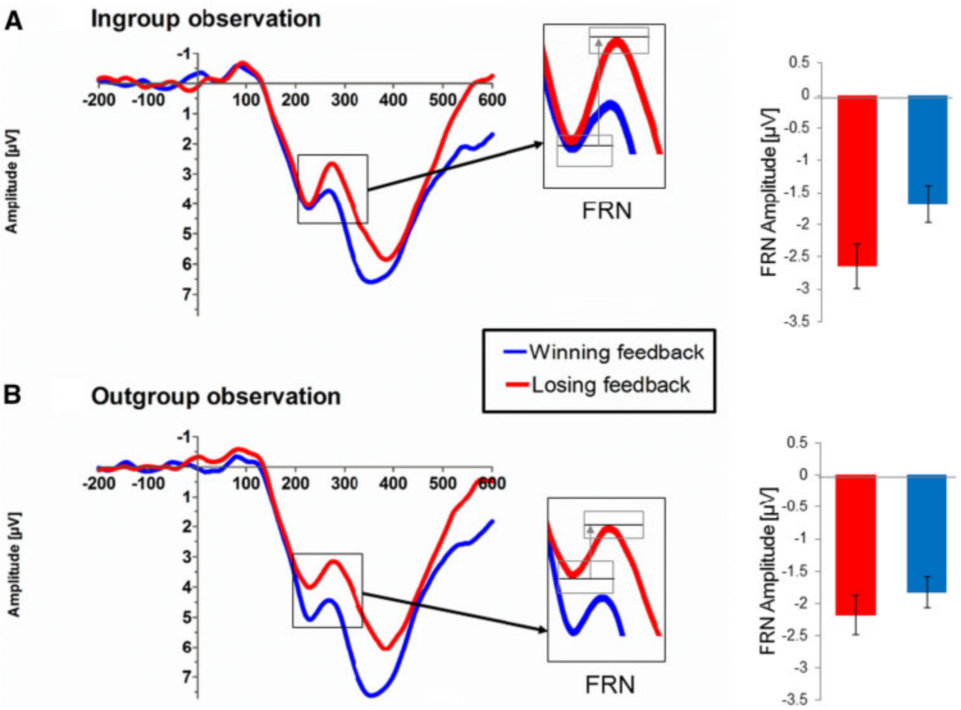

Fig. 2. Base-to-peak calculations of the stimulus-locked FRN at Cz electrode following wins and losses on the time-estimation task during (A) ingroup member observation and (B) outgroup member observation. In both cases, the FRN peaks at 280–300 ms. The bar graphs illustrate the interaction between feedback monitoring, showing relative increased differentiation between feedback type during ingroup observation compared to outgroup observation. More negative values indicate greater FRN activity.

Table 1. Means (s.d.) for electroencephalographic (EEG) measures of FRN activity in response to punishing and rewarding feedback

Observer block	lossFRN (punishing feedback)	gainFRN (rewarding feedback)	ΔFRN (lossFRN-gainFRN)
Ingroup observer	-2.65_a (1.70)	-1.68_b (1.66)	−0.97 (1.32)
Outgroup observer	-2.17_a (1.78)	-1.82_a (1.63)	−0.35 (1.06)

Note. Means across rows with different subscripts differ significantly at $P < 0.05$ (two tailed).

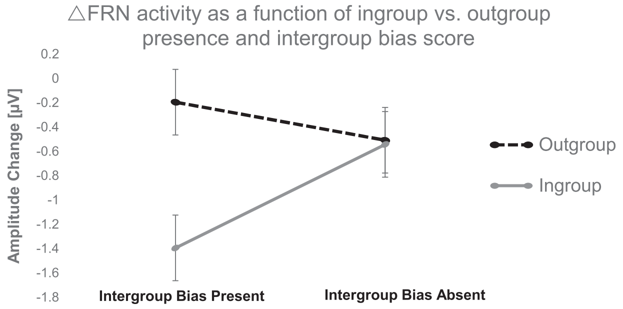

Fig. 3. Predicted difference-wave FRN (ΔFRN at Cz electrode) in the presence of an ingroup member and outgroup member as a function of the level of reported bias. Modeled at the levels of intergroup bias present and intergroup bias absent. More negative values indicate greater differentiation (i.e. difference-wave score) between wins and losses.

presence of an outgroup member, where it appears that the weakened FRN signal is predicted by polarized intergroup attitudes. That is, only biased individuals show the greatest disruption in feedback monitoring when around an outgroup member. We take this as evidence that group bias, and the motivation to identity with one group over another, plays an important role in the pattern of feedback monitoring that is observed (see Supplementary Materials for a Study 2 where we run a non-social/baseline control condition).

Discussion

This study set out to test whether a social intergroup context is capable of influencing neural feedback monitoring. Using online brain measures and leveraging classic social identity literature, the goal was to explore how the underlying mechanisms that drive brain-based reinforcement learning are affected by even the most minimal social categorization and group biases. A particular strength of this study is that we directly manipulated group membership in the lab using confederates, which ensured the physical presence of other people and allowed us create an artificial intergroup dynamic while simultaneously recording changes in brain activity.

On the basis of previous theory suggesting that people are less likely to regulate their behaviors around the outgroup (e.g. Brewer, 1979; Mackie *et al.*, 1990; van Knippenberg and Wilke, 1992; van Knippenberg, 1999), we predicted that the mere presence of an outgroup member would lead to the dampening of the basic feedback-monitoring signal, though we expected it to be maintained in the presence of an ingroup member. In addition, we predicted that these divergent patterns would be associated with individuals' biased group attitudes.

Confirming our hypotheses, we provide direct causal evidence that a simple intergroup interaction—the mere presence of a novel ingroup and outgroup member with no prior history—influences feedback monitoring. By creating an artificial intergroup dynamic in the lab, arbitrarily placing people into minimal groups, we found specifically that participants elicited a diminished FRN in the presence of an outgroup member, with the FRN effect remaining in the presence of an ingroup member (similar to the FRN in typical, non-social contexts). This idea is supported by models of the FRN as a neural marker of reinforcement learning, in which impaired FRN amplitudes are tied to the inability to adequately process feedback for the purpose of signaling remedial action during performance (e.g. Gehring and Willoughby, 2002; Holroyd and Coles, 2002; Stahl, 2010). Importantly, given that the current analyses were exploratory, pre-registered replication studies with confirmatory analytical designs should be done in future research.

Broadly, we add a layer to our understanding of how social preferences shape brain functioning. Although the presence of others should not necessarily impinge on a person's processing of feedback, previous work has shown that simply knowing that a person is being observed amplifies the brain's response to feedback (Simon *et al*., 2014). In this experiment, we wanted to extend this to show that it matters which person is doing the observing and whether they belong to the ingroup or not. Indeed, our findings offer evidence that the motivational features stimulated by an intergroup context—even a minimal, laboratory-based one—are enough to affect the basic neural reinforcement learning signal.

In addition, the findings hint at a proximal mechanism, which aligns with longstanding knowledge in the social identity literature: People are less motivated to care about, and attend to, outgroup members and the associated interactions. The current findings suggest that a muted neural feedback signal is associated with the failure to process information simply as a result of a person receiving that information in the presence of an outgroup member. Disregarding the outgroup member in this case might actually translate into a person not being as invested in their own goal-directed behaviors. That said, it is important to consider the fact that intergroup interactions are highly context dependent. Certain situations may actually cause a person to be more motivated to pay attention to an outgroup member, like when the interaction is hostile or negative.

The findings illustrate the personal cost of bias in intergroup interactions, where social categorization may act as a double-edged sword. Even though group biases motivate a person to monitor their performance when around an ingroup member, it appears these biases can also lead a person to be less invested in their own performance when in the presence of an outgroup member (i.e. dampened FRN). This reduction in feedback monitoring means the closing off of a system that allows for the constant updating of goal-directed action, ultimately hindering one's own learning and performance (e.g. Morris *et al*., 2008; Holroyd and Yeung, 2011). This would be particularly costly in social contexts where mixed group interactions are encouraged or expected. In short, group biases, although rooted in the motivation to boost one's identity, may actually end up being detrimental to the self.

Conclusion

This experiment demonstrates that intergroup contexts are capable of altering a person's neural feedback monitoring. We found that when a person is placed in even the simplest intergroup situation and then observed by others, they elicit a diminished feedback signal in the presence of an outgroup member. In particular, the biases that arose from this intergroup context were associated with the divergent effects in feedback monitoring.

These results have implications for how a person monitors their ongoing behavior and how they learn from the social environment. Even seemingly innocuous social contexts, like the presence of people we dislike, can constrain neural feedback monitoring. In sum, although social identity can bolster a person's sense of self, the resulting group biases may actually end up backfiring, impairing a person's fundamental ability to learn from important feedback.

Funding

Funding for this research was provided by grants to Nicholas Hobson and Michael Inzlicht from the Social Sciences and Humanities Research Council.

Supplementary data

Supplementary data are available at *SCAN* online.

Conflict of interest. None declared.

References

Amodio, D.M. (2014). The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, **15**(10), 670–82.

Anderson, S.M., Chen, S. (2002). The relational self: an interpersonal social-cognitive theory. *Psychological Review*, **4**, 619–45.

Anderson, S.M., Glassman, N.S. (1996). Responding to significant others when they are not there: effects on interpersonal inference, motivation, and affect. In: Sorrentino, R.M., Higgins, E.T., editors. *Handbook of Motivation and Cognition*, Vol. **3**, pp. 262–321. New York: Guilford Press.

Baker, T.E., Holroyd, C.B. (2011). Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. *Biological Psychology*, **87**, 25–34.

Baron, R.S. (1986). Distraction-conflict theory: progress and problems. *Advances in Experimental Social Psychology*, **19**, 1–39.

Bellebaum, C., Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience*, **27**, 1823–35.

Bellebaum, C., Kobza, S., Thiele, S., Daum, I. (2010). It was not MY fault: event-related brain potentials in active and observational learning from feedback. *Cerebral Cortex*, **20**, 2874–83.

Brewer, M.B. (1979). Ingroup bias in the minimal group situation: a cognitive-motivational analysis. *Psychological Bulletin*, **86**, 307–24.

Brewer, M.B. (1999). The psychology of prejudice: ingroup love or outgroup hate? *Journal of Social Issues*, **55**, 429–44.

Carlson, J., Foti, D., Harmon-Jones, E., Mujica-Parodi, L., Hajcak, G. (2011). The neural processing of rewards in the human striatum: correlation of fMRI and event-related potentials. *NeuroImage*, **57**, 1608–16.

Cikara, M., Fiske, S.T. (2011). Bounded empathy: neural responses to out-group targets' (mis)fortunes. *Journal of Cognitive Neuroscience*, **23**, 3791–803.

Cikara, M., Fiske, S.T. (2012). Stereotypes and Schadenfreude: behavioral and physiological markers of pleasure at others' misfortunes. *Social Psychological & Personality Science*, **3**, 63–71.

Cikara, M., Van Bavel, J.J. (2014). The neuroscience of intergroup relations: an integrative review. *Perspectives on Psychological Science*, **9**, 245–74.

Cohen, M.X., Elger, C.E., Ranganath, C. (2007). Reward expectation modulates feedback- related negativity and EEG spectra. *Neuroimage*, **35**, 968–78.

Cohen, M.X., Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience*, **27**, 371–80.

Coull, A., Yzerbyt, V.Y., Castano, E., Paladino, M., Leemans, V. (2001). Protecting the ingroup: motivated allocation of cognitive resources in the presence of threatening ingroup members. *Group Processes and Intergroup Relations*, **4**, 327–39.

De Bruijn, E.R.A., Mars, R.B., Bekkering, H., Coles, M.G. (2012). Your mistake is my mistake ... or is it? Behavioural adjustments following own and observed actions in cooperative and competitive contexts. *Quarterly Journal of Experimental Psychology*, **65**, 317–25.

De Bruijn, E.R.A., Miedl, S.F., Bekkering, H. (2008). Fast responders have blinders on: ERP correlates of response inhibition in competition. *Cortex*, **44**, 580–6.

De Bruijn, E.R.A., Miedl, S.F., Bekkering, H. (2011). How a coactor's task affects monitoring of own errors: evidence from a social event-related potential study. *Experimental Brain Research*, **211**, 397–404.

DeWall, N.C., Baumeister, R.F., Gailliot, M.T., Maner, J.K. (2008). Depletion makes the heart grow less helpful as a function of self-regulatory energy and genetic relatedness. *Personality and Social Psychology Bulletin*, **34**, 1653–62.

Edwards, L.J., Muller, K.E., Wolfinger, R.D., Qaqish, B.F., Schabenberger, O. (2008). An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine*, **27**, 6137–57.

Esses, V.M., Haddock, G., Zanna, M.P. (1993). Values, stereotypes, and emotions as determinants of intergroup attitudes. In: Mackie, D.M., Hamilton, D.L., editors. *Affect, Cognition and Stereotyping: Interactive Processes in Group Perception*, pp.137–166. New York: Academic Press.

Faul, F., Erdfelder, E., Lang, A.G., Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, program, and biomedical sciences. *Behavioral Research Methods*, **39**, 175–91.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Gehring, W. J., Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, **295**, 2279–82.

Gehring, W.J., Willoughby, A.R. (2004). Are all medial frontal negativities created equal? Toward a richer empirical basis for theories of action monitoring. In: Falkenstein, M.U.M., editor. *Errors, Conflicts, and the Brain. Current Opinions on Performance Monitoring*, pp. 14–20. Leipzig, Germany: Max Planck Institute of Cognitive Neuroscience.

Gillan, C.M., Robbins, T.W. (2014). Goal-directed learning and obsessive-compulsive disorder. *Philosophical Transactions of the Royal Society B*, **369**, 1–11.

Golby, A.J., Gabrieli, J.D.E., Chiao, J.Y., Eberhardt, J.L. (2001). Differential fusiform responses to same-race and other-race faces. *Nature Neuroscience*, **4**, 845–50.

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

Hajcak, G., Moser, J.S., Holroyd, C.B., Simons, R.F. (2005). The feedback-related negativity reflects the binary evaluation of good versus bad outcomes. *Biological Psychology*, **71**, 148–54.

Hirsh, J.B., Inzlicht, M. (2008). The devil you know: neuroticism predicts neural response to uncertainty. *Psychological Science*, **19**, 962–7.

Holroyd, C.B., Coles, M.G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, **109**, 679–709.

Holroyd, C.B., Krigolson, O.E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, **44**, 913–7.

Holroyd, C.B., Yeung, N. (2011). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, **16**, 122–8.

Huguet, P., Galvaing, M.P., Monteil, J.M., Dumas, F. (1999). Social presence effects in the Stroop Task: further evidence for an attentional view of social facilitation. *Journal of Personality and Social Psychology*, **77**, 1011–25.

Inzlicht, M., Gutsell, J.N. (2007). Running on empty: neural signals for self-control failure. *Psychological Science*, **18**, 933–7.

Karoly, P. (1999). A goal systems-self-regulatory perspective on personality, psychopathology, and change. *Review of General Psychology*, **3**, 264–91.

Koban, L., Corradi-Dell'Acqua, C., Vuilleumeir, P. (2013). Integration of error agency and representation of others' pain in the anterior insula. *Journal of Cognitive Neuroscience*, **25**, 258–72.

Koban, L., Pourtois, G., Bediou, B., Vuilleumier, P. (2012). Effects of social context and predictive relevance on action outcome monitoring. *Cognitive Affective and Behavioral Neuroscience*, **12**, 460–78.

Koban, L., Pourtois, G., Vocat, R., Vuilleumier, P. (2010). When your errors make me lose or win: event-related potentials to observed errors of cooperators and competitors. *Social Neuroscience*, **5**, 360–74.

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

Luck, S.J., Gaspelin, N. (in press). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*.

Luck, S.J., Kappenman, E.S. (2012). *The Oxford Handbook of Event-Related Potential Components*. New York: Oxford University Press.

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., Poulsen, C. (2003). Electrophysiological responses to errors and feedback in the process of action regulation. *Psychological Science*, **14**, 47–53.

Ma, Q., Shen, Q., Zu, Q., Li, D., Shu, L., Weber, B. (2011). Empathic responses to others' gains and losses: an electrophysiological investigation. *NeuroImage*, **54**, 2472–80.

Mackie, D.M., Worth, L.T., Asuncion, A.G. (1990). Processing of persuasive in-group messages. *Journal of Personality and Social Psychology*, **58**, 812–22.

Markus, H., Wurf, E. (1987). The dynamic self-concept: a social psychological perspective. *Annual Review of Psychology*, **38**, 299–337.

Miltner, W.H., Braun, C.H., Coles, M.G.H. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: evidence for a generic neural system for error detection. *Journal of Cognitive Neuroscience*, **9**, 788–98.

Molenberghs, P. (2013). The neuroscience of in-group bias. *Neuroscience & Biobehavioral Reviews*, **37**, 1530–6.

Molenberghs, P., Halász, V., Mattingley, J.B., Vanman, E.J., Cunnington, R. (2012). Seeing is believing: neural mechanisms of action-perception are biased by team membership. *Human Brain Mapping*, **34**, 2055–68.

Molenberghs, P., Morrison, S. (2014). The role of the medial prefrontal cortex in social categorization. *Social Cognitive and Affective Neuroscience*, **9**, 292–6.

Morris, S.E., Heerey, E.A., Gold, J.M., Holroyd, C.B. (2008). Learning-related changes in brain activity following errors and performance feedback in schizophrenia. *Schizophrenia Research*, **99**, 274–85.

Morrison, S., Decety, J., Molenberghs, P. (2012). The neuroscience of group membership. *Neuropsycholgia*, **50**, 2114–20.

Nash, K.N., Prentice, M., Hirsh, J.B., McGregor, I.D., Inzlicht, M. (2014). Muted neural response to distress among securely attached people. *Social Cognitive Affective Neuroscience*, **9**, 1239–45.

Nieuwenhuis, S., Holroyd, C.B., Mol, N., Coles, M.G. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience Biobehavioral Review*, **28**, 441–8.

Olvet, D.M., Hajcak, G. (2009). The stability of error-related brain activity with increasing trials. *Psychophysiology*, **46**, 957–61.

O'Toole, S.A., Weinborn, M., Fox, A.M. (2012). Performance monitoring among non- patients with obsessive-compulsive symptoms: ERP evidence of aberrant feedback monitoring. *Biological Psychology*, **91**, 221–8.

Paulus, P.B. (1983). Group influence on individual task performance. In: Paulus, P.B., editor. *Basic Group Processes*, pp. 14–20. New York: Springer.

Radke, S., de Lange, F. P., Ullsperger, M., de Bruijn, E.R.A. (2011). Mistakes that affect others: an fMRI study on processing of own errors in a social context. *Experimental Brain Research*, **211**, 405–13.

Ratner, K.G., Amodio, D.M. (2012). Seeing "us vs. them": minimal group effects on the neural encoding of faces. *Journal of Experimental Social Psychology*, **49**, 298–301.

Richard, F.D., Bond, C.F.J., Stokes-Zoota, J.J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, **7**, 331–65.

Segalowitz, S.J., Santesso, D.L., Murphy, T.I., Homan, D., Chantziantoniou, D.K., Khan, S. (2010). Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology*, **47**, 260–70.

Simon, D., Becker, M.P.I., Mothes-Lasch, M., Miltner, W.H.R., Straube, T. (2014). Effects of social context on feedback-related activity in the human ventral striatum. *NeuroImage*, **99**, 1–6.

Stahl, J. (2010). Error detection and the use of internal and external error indicators: an investigation of the first-indicator hypothesis. *International Journal of Psychophysiology*, **77**, 43–52.

Tajfel, H. (1974). Social identity and intergroup behavior. *Social Science Information*, **13**, 65–93.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Review of Psychology*, **33**, 1–59.

Tajfel, H., Billig, M., Bundy, R.P., Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, **1**, 149–78.

Tajfel, H., Turner, J.C. (1979). An integrative theory of inter- group conflict. In: Austin W.G., Worschel S., editors. *The Social Psychology of Intergroup Relations*, pp. 33–47. Pacific Grove, CA: Brooks/Cole Publishing.

Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., Wetherell, M.S. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Oxford: Blackwell.

Van Bavel, J.J., Packer, D.J., Cunningham, W.A. (2008). The neural substrates of in-group bias: a functional magnetic resonance imaging investigation. *Psychological Science*, **19**, 1131–9.

Van Bavel, J.J., Packer, D.J., Cunningham, W.A. (2011). Modulation of the fusiform face area following minimal exposure to motivationally relevant faces: evidence of in-group enhancement (not out-group disregard). *Journal of Cognitive Neuroscience*, **23**, 3343–54.

van Knippenberg, D. (1999). Social identity and persuasion: reconsidering the role of group membership. In: Abrams, D., Hogg, M.A., editors. *Social Identity and Social Cognition*, pp. 315–331. Oxford, UK: Blackwell.

van Knippenberg, D., Wilke, H. (1992). Prototypicality of arguments and conformity to in-group norms. *European Journal of Social Psychology*, **22**, 141–55.

Van Meel, C.S., Van Heijningen, C.A.A. (2010). The effect of interpersonal competition on monitoring internal and external error feedback. *Psychophysiology*, **47**, 213–22.

Yasuda, A., Sato, A., Miyawaki, K., Kumano, H., Kuboki, T. (2004). Error-related negativity reflects detection of negative reward prediction error. *Neuroreport*, **15**, 2561–5.

Yeung, N., Sanfey, A.G. (2004). Independent coding of reward magnitude and valence in the human brain. *The Journal of Neuroscience*, **24**, 6258–64.

Zajonc, R.B. (1965). Social facilitation. *Science*, **149**, 269–74.