Reinforcement Learning Signal Predicts Social Conformity

Vasily Klucharev,[1,2,*] Kaisa Hytönen,[1,2] Mark Rijpkema,[1] Ale Smidts,[2] and Guillén Fernández[1,3]
[1]Donders Institute for Brain, Cognition, and Behaviour, Radboud University Nijmegen, 6500 HB Nijmegen, The Netherlands
[2]Rotterdam School of Management, Erasmus University, 3062 PA Rotterdam, The Netherlands
[3]Department of Neurology, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands
*Correspondence: v.klucharev@fcdonders.ru.nl
DOI 10.1016/j.neuron.2008.11.027

SUMMARY

We often change our decisions and judgments to conform with normative group behavior. However, the neural mechanisms of social conformity remain unclear. Here we show, using functional magnetic resonance imaging, that conformity is based on mechanisms that comply with principles of reinforcement learning. We found that individual judgments of facial attractiveness are adjusted in line with group opinion. Conflict with group opinion triggered a neuronal response in the rostral cingulate zone and the ventral striatum similar to the "prediction error" signal suggested by neuroscientific models of reinforcement learning. The amplitude of the conflict-related signal predicted subsequent conforming behavioral adjustments. Furthermore, the individual amplitude of the conflict-related signal in the ventral striatum correlated with differences in conforming behavior across subjects. These findings provide evidence that social group norms evoke conformity via learning mechanisms reflected in the activity of the rostral cingulate zone and ventral striatum.

INTRODUCTION

Human behavior is guided not only by subjective values or attitudes, but also by the perceived behavior of others, in particular by social norms (Cialdini and Goldstein, 2004). Despite a growing body of literature describing neural mechanisms of decision making (Montague et al., 2006; Spitzer et al., 2007), we know little about how and why such social influence on human decisions occurs.

Conformity refers to the act of changing one's behavior to match the responses of others (Cialdini and Goldstein, 2004). The behavior and judgment of other people provides information on the normal and expected behavior in these circumstances and what is typically approved or disapproved. The effect of group opinion on individual judgments and decisions have been robustly replicated (Cialdini and Goldstein, 2004) since Solomon Asch's pioneering work on the line-judgment conformity experiments in which a third of the participants conformed to the erroneous majority opinion of the confederates, even when the majority claimed that two lines different in length by several inches were the same length (Asch, 1951). Conformity has been extensively studied in social psychology, and three central motivations for conforming behavior are suggested: a desire to be accurate by properly interpreting reality and behaving correctly, to obtain social approval from others, and to maintain a favorable self-concept (Cialdini and Goldstein, 2004). Whereas psychological studies emphasize the rewarding value of social approval or affiliation with others (Cialdini and Goldstein, 2004), behavioral economics focuses more on the effects of punishment for violation of the norm (Fehr and Fischbacher, 2004). In fact, both approaches may suggest that conformity is underlined by reinforcement learning, i.e., social norms selectively reinforce certain behaviors. Here, we utilize the cognitive neuroscience approach (Phelps and LeDoux, 2005) to provide a useful framework for studying reinforcement learning mechanisms of conformity.

Recent neuroscientific and computational models assume that goal-directed behavior requires continuous performance monitoring (Montague et al., 2006). Successful behavioral patterns are reinforced while errors call for adjustments of behavior. Many reinforcement learning models include a "prediction error"—a difference between the expected and obtained outcome (Schultz, 2006). Reward prediction error guides decision making by signaling the need for adjustment of behavior. Importantly, a conflict with social norms is not a usual behavioral error, i.e., it is not a typical behavioral mistake but rather any action that deviates from the behavior of the majority. Conformity with social norms requires neural signals related to deviations from it (Montague and Lohrenz, 2007). Here, we hypothesize that a perceived deviation from group norms triggers a neural response that is similar to prediction error in reinforcement learning—indicating a need to change individuals' future behavior in line with group norms. Event-related brain potential and functional magnetic resonance imaging (fMRI) studies suggest that the rostral cingulate zone (RCZ) has a specific role in reinforcement learning and generation of feedback- and error-related responses (Gehring et al., 1993). The RCZ is the region on the border of Brodman areas 6, 8, 24, and 32 (Picard and Strick, 1996). Cognitive neuroscience provides strong evidence to imply that activity of the RCZ, the region in the posterior medial frontal cortex, indicates the need for adjustments both when the action goal was not achieved and when the likelihood of failure is high (Cohen and Ranganath, 2007; di Pellegrino et al., 2007; Ridderinkhof et al., 2004). The magnitude of the RCZ activity has also been shown to predict the strength of subsequent behavioral adjustments during simple choice decisions (Cohen and Ranganath, 2007; Kerns et al., 2004). The reinforcement learning theory of performance monitoring suggests that the RCZ activity is modulated by a midbrain dopaminergic signal which indicates whether an action outcome is worse or better than expected, regardless of the primary cause of the deviation from the prediction (Holroyd and Coles, 2002). The RCZ is not alone in monitoring behavioral outcomes. In fact, a growing body of research has identified a distributed neural network involved in this process which includes the ventral striatum, i.e., the nucleus accumbens (NAc). Indeed, unpredictable reward modulates the activity of the human NAc (Berns et al., 2001; McClure et al., 2003; O'Doherty, 2004). The NAc has also been implicated in social learning (Rilling et al., 2002). Overall, previous studies have demonstrated that the NAc is involved into gain prediction in response to reward cues (Knutson and Wimmer, 2007). Importantly, the cell bodies of the majority of dopamine neurons that show an actual prediction error signal are located in the midbrain (substantia nigra and ventral tegmental area [Schultz, 2006]). These midbrain neurons project heavily to the NAc and the RCZ. Thus, assuming that the BOLD signal may primarily reflect inputs (and local computation), it is possible that with human fMRI such a full prediction error signal would show up primarily in the NAc and the RCZ rather than in the midbrain where it originates.

In the current study, we hypothesized that, if conformity is based on reinforcement learning, (1) a conflict with group opinion triggers a "prediction error" response manifested in activity of the RCZ and the NAc and (2) this activity predicts the subsequent adjustment of the behavior, i.e., social conformity. To test our hypothesis, we designed a paradigm in which the subject's initial judgments of facial attractiveness were open to influence by group opinion. Facial attractiveness is a highly important social characteristic (Langlois et al., 2000) and an everyday target of normative influence, for example by fashion magazines and cosmetics commercials. During fMRI (Experiment N1), female subjects rated the attractiveness of female faces, and after each rating they were informed of an "average European rating" of the face—group rating. Actual group ratings were systematically manipulated during the experiment. We assumed that group opinion (group ratings) signaled the normative opinion (a "descriptive norm" representing typical behavior [Cialdini and Goldstein, 2004]) about the attractiveness of each individual face. Thus, with our procedure, we introduced a conflict between the subject's own judgment and the normative group opinion. To identify subsequent conformity with the group, subjects rated the same set of faces again after the fMRI session.

To identify the neural activity related to "social (normative) conflict" we first compared the brain responses in all trials in which the group rating differed from the subject's rating (conflict trials) with all no-conflict trials. To model subsequent conformity effects, we then calculated a contrast within conflict trials: conflicts with group ratings followed by conformity (i.e., where perceived facial attractiveness subsequently changed in line with group ratings) versus conflicts with group ratings not followed by conformity (where perceived facial attractiveness did not change).

We found that the perceived difference of individual ratings from group ratings triggered long-term conforming behavioral adjustments, i.e., subjects changed their attractiveness ratings to align them with group ratings. As we expected, a conflict with group opinion activated RCZ and deactivated the NAc, which implies that conflict with normative group opinion triggers neuronal signals similar to the prediction error signal of reinforcement learning. Subsequent conformity was predicted by the larger conflict-related responses. Conjunction analysis (testing the conjunction null hypothesis [Nichols et al., 2005]) revealed a spatial overlap between the conflict-related activity and activity which predicted subsequent conformity. Furthermore, the individual strength of the conformity-related activity in the ventral striatum correlated with differences in conforming behavior across subjects. Finally, we conducted an fMRI control experiment (Experiment N2) to examine the social relevance of our results using a nonsocial version of the task in which group normative opinion was replaced with computer-generated ratings. We found that conforming behavior and related effects in the RCZ and the NAc were particularly strong in the social condition (Experiment N1). Overall, this data provides evidence that social conformity is based on mechanisms similar to reinforcement learning: a conflict with group opinion triggers a prediction error signal, indicating a need for adjustment of judgments, i.e., social conformity.

RESULTS

Experiment N1: Behavioral Results

Twenty-four female participants (three subjects were later excluded from the analysis, see Experimental Procedures) were invited to participate in a study investigating brain mechanisms of facial attractiveness. During fMRI participants rated the attractiveness of 222 female faces and after each rating they were informed about a group rating ("average European rating") of the face (Figure 1). To test whether group opinion affected perceived facial attractiveness, subjects were unexpectedly asked to rate the faces again during a behavioral session approximately 30 min after scanning. This time faces were again presented in a random order, but without group ratings. In agreement with our expectations, participants changed their ratings of attractiveness, aligning themselves with group ratings (Figure 2): on average, participants decreased their attractiveness ratings when group ratings had been more negative than their own initial rating, whereas more positive group ratings were associated with more positive re-evaluation of faces. Participants did not change their ratings significantly if group ratings matched their initial ratings (no-conflict trials). One-way ANOVA analysis (three-level factor of *group ratings*) revealed a significant main effect of the factor *group ratings* on changes in attractiveness ratings ($F(2,20) = 31.1$ $p = 0.0001$). Therefore, group opinion effectively modulated judgments of individuals even when the group was not physically present and so could not directly affect participants. The conformity effect was especially strong for highly ambiguous faces: for faces whose initial ratings varied most across participants (standard deviation ≥ 1.621, see Supplemental Data, Figure S1 available online for details).

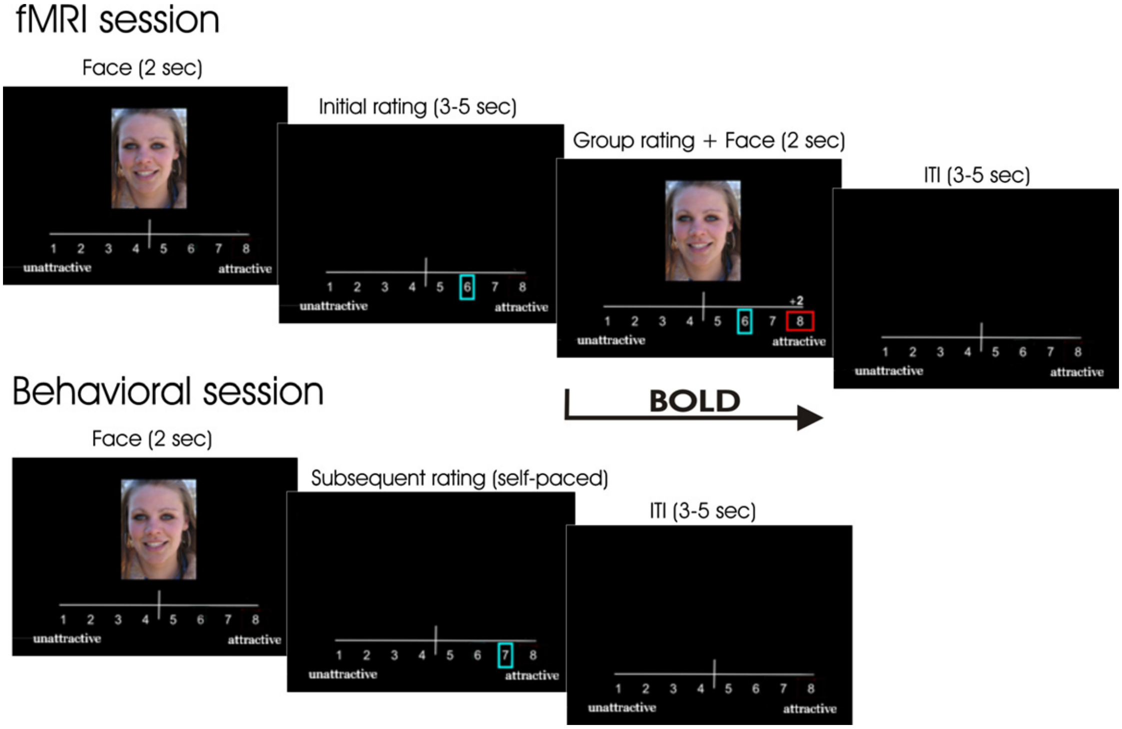

Figure 1. The Task (fMRI Session) Evoking a Conflict with Group Ratings Followed by the Behavioral Session

The sequence of the events within a trial is shown. During the fMRI session (Experiment N1), subjects rated the attractiveness of female faces and were subsequently presented with the group ratings that could be similar (no conflict with group ratings), below or above (as is shown in the figure) subjects' rating (conflict with group ratings). Thirty minutes after the fMRI session subjects rated again the same faces during the Behavioral session in order to identify the subsequent conformity effects. The control experiment (Experiment N2) had the same trial structure, but a different cover story.

Given the fact that group ratings were often ''more extreme'' than participants' initial ratings, one may argue that the behavioral effect of conformity is simply caused by an increase in variance of the scale used, i.e., variation in ratings of faces is greater in the subsequent behavioral session than in the initial fMRI session. To exclude this simple ''range'' effect, we compared variances of ratings for the first (fMRI) session and the second (behavioral) session (see Figure S2). In contrast to the expectations of the ''range'' hypothesis, the variance slightly decreased in the second session (from 2.96 to 2.7, t(1,20) = 1.85, p = 0.08). Thus our behavioral finding can not be explained by a simple increase in response variance, but entails a true adjustment to group feedback (see Supplemental Data for detailed analyses).

To establish an even closer relationship between group ratings and individual behavior, we performed a correlation analysis between the magnitude of the conflict (i.e., the difference value between subjects' own and group ratings during the fMRI session) and the subsequent change in the perceived facial attractiveness separately for each participant. We found a significant correlation among all participants (mean values: r = 0.21, n = 222, p = 0.005, SD = 0.06, min value: r = 0.13, max value: r = 0.33), except for one subject who showed a correlation that just failed to reach statistical significance (r = 0.126, p = 0.07). The larger the conflict with group opinion, the more pronounced the conformity effect was, even at the level of individual participants. We later used the individual correlation coefficients as conformity scores (i.e., a measure of the individual tendency to conform, thereby distinguishing conformists from nonconformists), and correlated them with individual fMRI conformity effects.

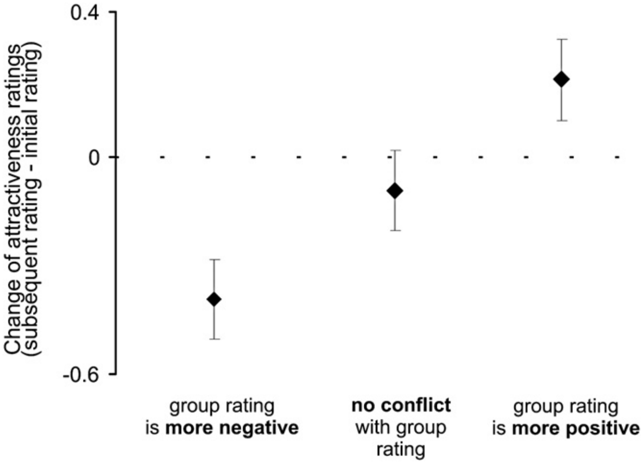

Figure 2. Mean Behavioral Conformity Effects

On average the attractiveness ratings changed in line with the group ratings. The picture illustrates the change of the faces' attractiveness measured during the behavioral session as compared to the initial ratings during the fMRI session. Bars indicate the standard error of the mean.

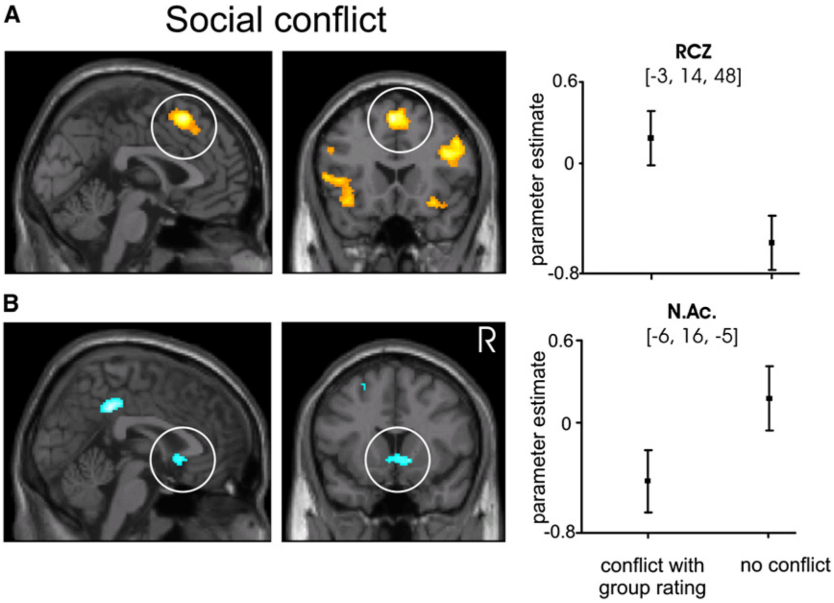

Figure 3. Social *Conflict* Effects: Neural Response to Group Ratings in Conflict versus No-Conflict Trials

Left: z-maps of activations (A) and deactivations (B) induced by a conflict with group ratings. Right: the signal change of the hemodynamic response for conflict and no conflict trials. RCZ, rostral cingulate zone. N.Ac., nucleus accumbens. R, right hemisphere. All maps are thresholded at $p < 0.001$; the clusters are significant at $p < 0.05$ (FDR corrected). Bars indicate standard error of the mean.

Our study therefore revealed that conformity leads to the transmission of facial preferences from the group to the individual. Overall the behavioral results indicate that the manipulation of social normative influence was successful in inducing conformity effects in the judgment of facial attractiveness.

Experiment N1: fMRI Results

To study brain activity associated with the perception of social conflict, we compared neural activity occurring during all trials in which the group rating conflicted with the subject's rating with all trials in which the group rating did not conflict with the subject's rating—the *conflict contrast*. As expected, the conflict with group opinion activated the RCZ (Figure 3A). The location of the cluster maximum (x = −3, y = 14, z = 48) matched closely the results of a previous meta-analysis on error monitoring (x = 1, y = 15, z = 43; for details see Ridderinkhof et al., 2004). In addition, conflict trials activated more strongly than no-conflict trials (Table 1) the insular cortex, the precuneus, the cerebellar tonsil, and the middle frontal gyrus, all areas known to be engaged in general error processing (Diedrichsen et al., 2005; Ridderinkhof et al., 2004). Furthermore, the conflict deactivated (i.e., more activity for no-conflict than conflict trials) the ventral striatum (NAc) and the posterior cingulate cortex, brain areas that are known to be involved in reward processing and error detection (McCoy and Platt, 2005; Schultz, 2006). Our results thus indicate that a mismatch with group opinion triggers

Table 1. Significant Activation Clusters for Social Conflict Contrast

Brain Region	HEM	x	y	z	No. of Voxels	Z
Activations						
Rostral cingulate zone (RCZ): medial/superior frontal gyrus, cingulate gyrus BA 6/8/24/32	L/R	−3	14	48	591	5.26
Precuneus, cuneus, BA 7/19	L	−20	−69	37	233	3.94
Precuneus, BA7/19	R	12	−75	45	989	4.97
Middle frontal gyrus, BA9	L	−36	−3	37	666	4.61
Middle frontal gyrus, BA9	R	36	14	23	844	4.87
Cerebellum	L	−34	−58	−28	357	4.30
Insula, BA13	L	−41	18	4	276	4.22
Insula, BA13	R	27	16	13	149	3.92
Middle frontal gyrus, precentral gyrus, BA 6	R	29	−3	51	149	4.19
Midbrain	R	10	−21	−14	52	3.66*
Midbrain	L	−3	−15	−3	27	3.55*
Midbrain	L/R	3	−27	−3	32	3.55*
Deactivations						
Posterior cingulate gyrus, BA 31	L/R	0	−38	40	240	4.32
Middle/superior frontal gyrus, BA 6/8	L	−24	18	38	206	4.24
Ventral striatum (nucleus accumbens, caudate)	L/R	−6	16	−5	198	4.06

Local maxima within these clusters are reported together with the number of voxels (No. of Voxels); BA, Brodmann area; HEM, hemisphere; L, left; R, right; x, y, z are Talairach coordinates of the local maximum; * with small volume correction.

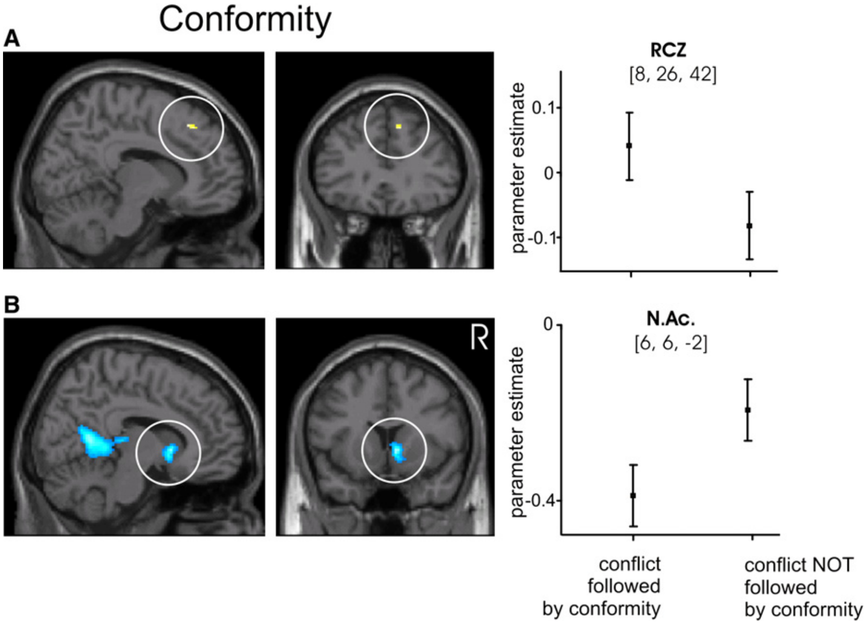

Figure 4. *Conformity* Effects: The Social Conflicts Followed by the Subsequent Change of Facial Attractiveness in Line with Group Ratings (i.e., Conformity) versus the Normative Conflicts that Were Not Followed by Changes in Attractiveness Ratings (i.e., No Conformity)

Left: z-maps of activations (A) and deactivations (B) predicting the conformity with group ratings. Right: the signal change of the hemodynamic response for trials followed by conformity and by no conformity. RCZ, rostral cingulate zone. N.Ac., nucleus accumbens. All maps are thresholded at $p < 0.001$. Bars indicate standard error of the mean.

a neural response in the RCZ and the NAc that is similar to prediction error signal.

The posterior cingulate cortex has been implicated into the “default” network (Buckner et al., 2008)—a specific, anatomically defined brain system preferentially active when individuals are not focused on the external environment. The deactivation of the cingulate cortex in the current study could therefore indicate an additional cognitive demand triggered by the conflict with the group opinion. Interestingly, a recent study showed that the posterior cingulate cortex is affected by dopamine depletion (Nagano-Saito et al., 2008). Furthermore, animal studies have demonstrated that neurons of the posterior cingulate monitor the omission of expected reward, suggestive of a prediction error-like signal (see McCoy and Platt, 2005, for a review).

Prediction error signals are intimately associated with dopamine neurons in the midbrain (Schultz, 2006). We therefore conducted an ROI analysis in the midbrain dopaminergic region covering the entire area, including substantia nigra, ventral tegmental area (VTA) and other structures. The spherical ROI had a radius of 15 mm and was centered at the coordinate −1, −18, −9 (x, y, z) (Aron et al., 2004). We found significant clusters of activity in the midbrain triggered by conflict with the group opinion (see Table 1 and Figure S4) and no significant deactivations. The activity of the midbrain, the RCZ and the NAc could reflect a degree of the social conflict with normative group opinion or a degree of reward subjects experienced when their ratings matched the normative ratings in no-conflict trials. However, given the fact that no-conflict trials were not followed by behavioral changes we focused our further analysis on conflict trials that triggered conformity.

Next, we hypothesized that the social conflict response in the RCZ and the NAc is predictive of changes in participants’ opinions on facial attractiveness. The activation of the RCZ and deactivation of the NAc should therefore be particularly strong during those conflict trials that effectively changed subjects’ opinion, i.e., were followed by conformity. To test this hypothesis, we compared brain activity during those conflict trials that were followed by changes in perceived attractiveness of faces in line with group ratings with conflict trials where there were no such changes—the *conformity contrast*. Indeed, the activation of the RCZ region of interest predicted subsequent conformity: the activity in the RCZ elicited by the conflicts with group opinion that were followed by conformity was stronger than that elicited by conflicts that were not followed by conformity (Figure 4A). Furthermore, the deactivation of the NAc region of interest during the perceived conflict with group opinion also predicted conformity (Figure 4B). In addition, we conducted a whole-brain analysis of conformity effects and found that the conformity-related suppression of activity in the NAc was significant, even without small volume correction. In the global search we found that conformity was also predicted by a deactivation of extrastriate visual cortex (BA 18,19) and parahippocampal cortices (Figure 4B; Table 2). We also checked conformity effects in the fusiform gyrus, a region implicated in face and attractiveness processing (Iaria et al., 2008). We did not find statistically significant effects in the selected ROIs (for fusiform gyrus: spheres of radius 10 mm, x, y, z: 34,−54,−21 and −32,−42,−25, based on a previous study [Iaria et al., 2008]). These null-findings might indicate that observed

Table 2. Significant Activation Clusters for the Social Conformity Contrast

Brain Region	HEM	x	y	z	No. of Voxels	Z
Activations						
Rostral cingulate zone: cingulate gyrus, BA 24/32	R	8	26	42	12	4.22*
Deactivations						
Lingual gyrus, posterior cingulate, parahippocampal gyrus, BA 18/29/30	L/R	10	−58	4	1588	5.61
Ventral striatum (nucleus accumbens), caudate head.	R	6	6	−2	169	5.60

*With small volume correction.

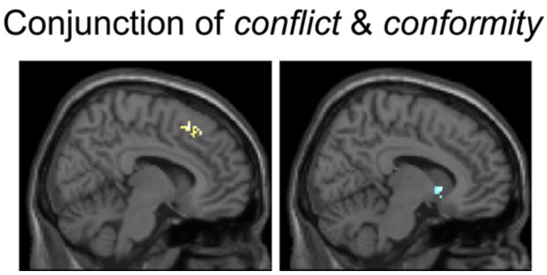

Figure 5. Results of the Conjunction Analysis of Social *Conflict* and *Conformity* Effects

Both the conflict with group ratings and the subsequent conformity activated RCZ (left part of the figure: local maxima at x = 6, y = 16, z = 46) and deactivated the NAc (right part of the figure: local maxima at x = 6, y = 6, z = −2). Maps are thresholded at $p < 0.001$; clusters are significant at $p < 0.05$ (FDR corrected).

conformity effects are not triggered by an immediate perceptive re-evaluation of facial attractiveness. ROI analysis of conformity effects in the midbrain also did not reveal effects reaching the level of statistical significance. Thus, the midbrain shows a nonspecific conflict-related signal in contrast to the neural signal at the RCZ and the NAc that is predictive of conformity effects.

To control the specificity of conformity effects in the RCZ and the NAc for conformal behavioral changes we conducted an additional analysis by calculating subsequent ‘‘anticonformity’’ effects—contrasting conflict trials followed by changes against the group versus conflict trials followed by unchanged ratings. However, we did not find any significant effect (thresholded at $p < 0.001$), even using an ROI analysis centered in the RCZ and the NAc. Furthermore, a direct contrast of conflict trials followed by changes in line with the group versus conflict trials followed by changes against the group showed significant activation of the RCZ (x, y, z: 8,5,40) and deactivation of the NAc (x, y, z: 1,4,−5) ROIs. These results indicate that observed conformity effects are specific for conformal adjustments and not related generally to changes in behavior.

To support more directly the hypothesis that social conformity is indeed triggered by social conflict-related neural activity in the RCZ and the NAc, we conducted a conjunction analysis (testing the conjunction null hypothesis, see Nichols et al. [2005] for details), aiming to identify those brain regions that are activated in both the *conflict* and the *conformity* contrast. The conjunction analysis revealed the activation of the RCZ and the deactivation of the NAc in both contrasts (Figure 5). Thus, the very same brain regions in the medial prefrontal cortex and the ventral striatum are sensitive for social conflict and predict conformity with group opinion.

To link individual performance differences to individual differences in brain activity, we compared neural responses of conformists (i.e., people conforming easily to group opinion) with nonconformists (see Behavioral Results for details). We split subjects in two groups using a median split on conformity scores: conformists (mean r = 0.26, n = 11) and nonconformists (mean r = 0.16, n = 10). We hypothesized that individual differences in levels of conformity are based on variability in response to social conflict, e.g., conformists generally show a greater degree of conflict-related activity than nonconformists, and for that reason the conflict-signal of the conformists reaches more easily the hypothetical threshold that triggers conformity. The current view on the functional role of the neural prediction error signal seems to suggest a threshold for error-related activity (similar to perceptual and motor decision making models, (e.g., Schall et al., 2002) that triggers the adjustment of future behavior (Schultz, 2006). Only an activity that crosses such a threshold evokes a change of behavior.

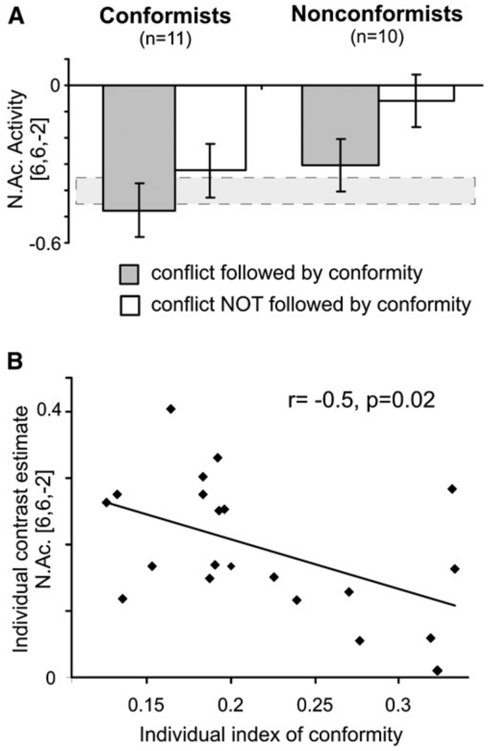

Figure 6. Nucleus Accumbens (NAc) Recruitment during a Conflict with Group Opinion Predicts Individual Differences in Conformity

(A) Conformists (subjects easily conforming to group ratings) showed the stronger conflict-related deactivation of the nucleus accumbens. Error bars indicate standard error of the mean. Grey rectangular area indicates a putative threshold of conformity.

(B) Significant correlation of the neural *conformity* effect with the individual level of conformity. Due to a higher probability of any conflict to trigger conformity, conformists showed a smaller difference (*conformity* effect) between neural responses to the conflicts with group ratings followed by conformity and those that were not followed by conformity.

This mechanism of conformity predicts that (1) the neural conflict-related signal is stronger in conformists than nonconformists and (2) the difference in conflict-related signal in trials that did and did not follow conformity (conformity effects) has to be weaker in conformists due to a higher chance of any conflict-related response crossing the hypothetical threshold, assuming that the threshold is similar across subjects. Figure 6A shows that the conflict-related response in the NAc was stronger for conformists than for nonconformists (prediction 1). This observation is supported by a MANOVA

Table 3. Comparison of Social (Experiment N1) and Nonsocial (Control Experiment N2) Experiments

Brain Region	HEM	x	y	z	No. of Voxels	Z
Significant Conflict × Social Task Interaction						
RCZ	R/L	3	16	43	49	4.42*
NAc	R	12	16	−2	3	3.35*
Midbrain		3	−35	−3	39	3.66*
Midbrain		−1	−27	−17	3	3.5*
Significant Conformity × Social Task Interaction						
RCZ	R	10	22	43	3	3.40*
NAc	R	8	6	−3	3	3.21*

*With small volume correction.

(*conformists/nonconformists* as a between-subject factor, subsequent *conformity* as a within-subjects factor): we found a significant effect of the *conformists/nonconformists* factor (F(1,20) = 19.9, p = 0.0003). Furthermore, the neural conformity effect was weaker for conformists than for nonconformists (prediction 2). We found a significant interaction between *conformists/nonconformists* and *conformity* factors (F(1,20) = 6.1, p = 0.023), due to the smaller difference in the conflict-related signal in trials that did and did not follow conformity for conformists in comparison to nonconformists (see Figure 6A). Moreover, Figure 6B illustrates the significant negative correlation of the neural *conformity* effect (*conformity* contrast) with the individual level of conformity (r = −0.5, n = 21, p = 0.021). The NAc has been previously linked to individual differences (Cohen, 2007; Schonberg et al., 2007; Tobler et al., 2007) in reinforcement learning and thus could also mediate individual differences in conforming behavior.

Experiment N2: Assessment of the Social Relevance of the Results

First, we conducted a behavioral control study to demonstrate the social nature of the experimental task. The control experiment demonstrated that social relatedness between the subjects, and the "group" is correlated with the degree of conformity in our task (see behavioral control study in Supplementary materials for details). Then, to assess the social relevance of our fMRI findings, we employed a nonsocial version of the experimental paradigm in which the normative group ratings were replaced with computer ratings—a method commonly used in social cognitive neuroscience (e.g., Spitzer et al., 2007; Zink et al., 2008). All other aspects of the paradigm and experimental setup were identical to the original fMRI design (task design and analysis). Twenty-two healthy females (aged 19–29 years, mean 22.1 years) participated in the nonsocial control study. The average age of subjects was not significantly different from those in experiment N1 (t(1,20) = 1.6, p = 0.1). One participant was rejected from the study due to large head motions exceeding 3 mm.

We made statistical comparisons of data from both the original and control fMRI experiment. We found an interaction between the *conflict* factor (within group factor: conflict versus nonconflict) and the *social task* factor (between group factor: social versus computer feedback) at the RCZ, NAc and midbrain region (see Table 3 and Figure 7A). Thus, the activity of the RCZ, NAc and midbrain was significantly more strongly affected by a conflict with social group opinion than by a conflict with a computer. The primary analysis of the control experiment showed that the mismatch with the computer activated the right insula, precuneus and precentral gyrus, in a similar way to the conflict with social group opinion (Table S5). We found conflict-related effects in the RCZ (x, y, z: 3,12,44) and the NAc (x, y, z: 18,14,−6 and −10,12,−7), only using a looser statistical threshold for the SPM analysis (p < 0.006). The conflict-related effects were thus strongly attenuated in the nonsocial experiment.

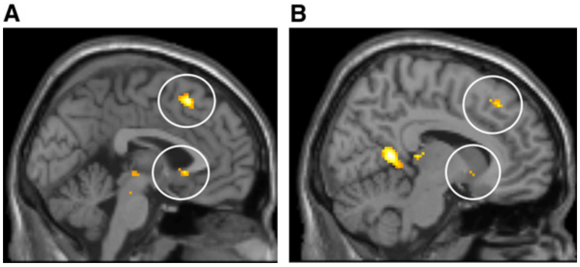

Figure 7. Comparison of Social and Nonsocial (Control) fMRI Studies

(A) *Conflict* × *social task* interaction. White circles indicate RCZ (local maxima at x = 3, y = 16, z = 43) and NAc (local maxima at x = 12, y = 16, z = −2).

(B) *Conformity* × *social task* interaction. White circles indicate RCZ (local maxima at x = 10, y = 22, z = 43) and NAc (local maxima at x = 8, y = 6, z = −3). The maps are thresholded at p < 0.001.

To explore further these results we studied the *conformity* × *social task* (between-group: social versus computer feedback) interaction. We found a significant *conformity* × *social task* interaction in the RCZ and the NAc (see Table 3 and Figure 7B). Our results indicate that conformity-related neural effects in the RCZ and the NAc are particularly strong for the social version of the task. Overall, the behavioral and fMRI results confirm that the observed effects in the RCZ and the NAc (sites receiving substantial dopamine inputs) are related to social conformity and are modulated by social factors.

Next, we studied the main effect of congruent behavioral adjustments in the control study by comparing neural responses for all conflict trials that were followed either by congruent behavioral changes (i.e., facial attractiveness subsequently changed in accordance with the computer rating) or by no behavioral changes (facial attractiveness ratings not changed). We found activation predicting adjustments in accordance with computer ratings (RCZ – x, y, z: 1,4,49 and NAc – x, y, z: 8,3,−7 and −10,8,−5) only with a decreased threshold (p < 0.003). Thus, the reinforcement mechanisms in both experiments were rather similar but the effects were strongly modulated by the social context. By and large, social descriptive norms of facial attractiveness were stronger and more effective reinforcers than computer-generated "norms."

In addition to distinct neural results, the social and nonsocial conditions were also dissociable behaviorally (see Figure S6). Overall, subjects changed their opinion more after a conflict with a social group than after a conflict with a computer (MANOVA, F(3,38) = 5.5, p = 0.004), both when group opinion was more negative and more positive than subjects' opinion (t(1,20) = 2.23, p = 0.03 and t(1,20) = 2.46, p = 0.01). To establish an even closer relationship between computer ratings and individual behavior, we performed a correlation analysis between the magnitude of the conflict and the subsequent change in the perceived facial attractiveness separately for each participant. We found a weak correlation (mean values: r = 0.15, n = 222, p = 0.05, SD = 0.02; min value: r = −0.01, max value: r = 0.28). Importantly, 12 out of 21 participants did not show a significant correlation. Moreover, the correlation was significantly weaker in the computer condition than in the social one (t(1,20) = 3.8, p = 0.001). Thus, the results demonstrate the social nature of the experimental paradigm (see Supplemental Data for additional details).

Overall, the results of all studies support the hypothesis that social conformity is based on neural mechanisms similar to those implemented in reinforcement learning. A conflict with social normative opinion triggers a conflict-related response at the RCZ and the NAc that is similar to prediction error in reinforcement learning; if the conflict-related signal exceeds a "learning" threshold then social conformity is triggered. Furthermore, the NAc activity shows a correlation with individual levels of conformity that indicates a close link of observed neural effects with actual behavior. The observed effects were particularly strong in the social context.

DISCUSSION

We found a robust behavioral effect of group opinion on perceived facial attractiveness. A conflict with a normative opinion triggered a long-term conforming adjustment of subjects' own rating. This result is in line with a recent study that demonstrated the social influence of others on an individual's face preferences (Jones et al., 2007). Furthermore, our results could explain the finding that there is considerably greater agreement in attractiveness ratings between individuals who share a close relationship (Bronstad and Russell, 2007): the ratings are homogenized within groups due to the strong conformity that is known to exist within social groups.

Social norms prescribe behaviors that a member of a group can enact, and norms are thought to exist "if any departure of real behavior from the norm is followed by some punishment" (Homans, 1950). Indeed, social norms reward or punish people (Bendor and Swistak, 2001) and can be seen as positive or negative reinforcers for socially appropriate or inappropriate behaviors. In other words, a conflict with social norms indicates an error that is similar to a reinforcement learning signal calling for an adjustment of the behavior. In the present study, we examined neural activity during a conflict with group opinion to test the hypothesis that the reinforcement learning signal guides conforming changes in social judgments. Our results were consistent with the reinforcement learning hypothesis of social conformity.

We found that a conflict with group opinion activates the RCZ and deactivates the NAc, both of which are known to be involved in the computation of the prediction error. Human neuroimaging studies consistently implicate the RCZ in monitoring response conflicts and errors and in differential processing of unfavorable outcomes such as monetary losses, abstract performance feedback, primary negative reinforcers (see Ridderinkhof et al. [2004] for an extensive review). Overall, the RCZ is engaged when the need for adjustments of the behavior becomes evident. It has been shown that the RCZ is activated by an unfair offer in an ultimatum game (Sanfey et al., 2003), by social exclusion (Eisenberger et al., 2003), and by the incorrect prediction of social rejection (or acceptance) by others (Somerville et al., 2006). RCZ activity is also modulated by the moral character of the partner in the trust game (Delgado et al., 2005) and by moral judgments (Greene et al., 2004). Furthermore, a recent study (Pochon et al., 2008) indicated a role of the RCZ in situations of choice difficulty: greater RCZ activation was found when participants had to choose between alternatives of similar desirability (indicating a high decision conflict) than when they made easier (low decision conflict) decisions. Our findings suggest a new interpretation of the role for the RCZ in social cognition: the RCZ is monitoring the incongruence of our judgments with social descriptive norms that are normally negatively reinforced by social rejection, exclusion, and moral or even physical punishment.

Activity of the NAc represents the value of the expected reward (Knutson et al., 2005; Knutson and Wimmer, 2007) and thus decreases for aversive stimuli (Besson and Louilot, 1995; Singer et al., 2006). In line with the previously reported inhibitory response to aversive stimuli, we found that the NAc activity during a conflict with group opinion is deactivated relative to a no-conflict situation. We investigated the social relevance of conflict-related effects in the RCZ and the NAc using a nonsocial control experiment. These effects were modulated by the social context, suggesting a social nature of the conflict. By and large, our findings indicate that the NAc, together with the RCZ, participates in the generation of the neural response indicating a conflict with group descriptive norms.

Recent learning theories have revealed the role of error monitoring in subsequent performance adjustments: errors indicate a need for behavioral changes (Schultz, 2006). The present study shows that the amplitude of the conflict-related responses in the RCZ and the NAc predicts the subsequent conforming change in behavior. We demonstrated that the conflict-related activity in RCZ in the trials that were followed by conformity was stronger than in trials that were not followed by conformity. The pattern was reversed in the NAc. Our finding indicates that when the conflict-related signals are strong enough, the performance is adjusted and subjects conform to the group normative opinion. These results establish a link between conflicts with descriptive norms and conformity. In addition, conjunction analysis revealed the clear spatial overlap of the neural activity underlying the conflict-related signal and conformity. Importantly, the effects predicting behavioral changes were strongest for the social version of the experiment. In accordance with our hypothesis, the conflict and conformity effects found may be enhanced by social situations rather than representing a specific social mechanism.

A previous study demonstrated that the magnitude of feedback-related negativity (FRN), whose neural generators are located in the RCZ (Herrmann et al., 2004), predicted whether subjects would change decision behavior on the subsequent trial of a simple computer strategic game (Cohen and Ranganath, 2007). Other studies have linked the magnitude of FRN to overall learning or decision making (Frank et al., 2005; Yeung and Sanfey, 2004) and to changes in reaction time on the subsequent trial (Gehring et al., 1993). Cingulate lesions in monkeys impair their ability to use previous reinforcements to guide decision-making behavior (Kennerley et al., 2006). The role of the RCZ in behavioral adjustment is also consistent with the ''conflict-monitoring hypothesis'' (Botvinick et al., 1999). This hypothesis suggests that the cingulate cortex is activated by the occurrence of response conflict during the so-called Stroop or Simon tasks. The monitoring of response conflict by the RCZ serves as a signal that aims to minimize the amount of conflict on subsequent performance. Indeed, the RCZ activity during response-conflict tasks predicted the adjustment of behavior (Kerns et al., 2004). Importantly, in our study the behavioral task did not evoke a response conflict, because the subjects responded before the conflicting group ratings were presented. Therefore, in the present study the RCZ activity did not indicate a response conflict but a neural signal similar to prediction error calculated as a perceived difference of own judgments from group opinion.

Our results extend the functional role of the NAc to social learning underlying conformity. We found that deactivation of the NAc during conflict with group opinion robustly predicts subsequent conformity and correlates with individual differences in conforming behavior. The NAc is often viewed as an integrator of memory, motivation, and goal-directed behaviors (Carelli, 2002). Thus, the individual variability of conformity could also be based on individual differences in the amplitude of the NAc conflict-related responses evoked by conflict with group opinion. The error signal at the ventral striatum, of which the NAc is part, has been previously correlated with individual differences (e.g., Tobler et al., 2007). A recent study reported that individual behavioral differences predicted the variability of the prediction error activity, particularly in the ventral striatum (Cohen, 2007). We found that conformists demonstrated stronger deactivation of the NAc during conflict with group opinion, indicating a stronger prediction error. We also found that differences in conflict-related responses in trials that did and did not follow conformity (conformity effects) were weaker in conformists, indicating a higher probability of conformity after any social conflict. Previous studies demonstrated that the social context modulated the activity of the NAc, for example, the perceived fairness of a person seen in pain affected the activity of ventral striatum (Singer et al., 2006) or social comparison modulated the activity of ventral striatum during the processing of rewards (Fliessbach et al., 2007). We suggest that the social comparison of the obtained reward could also be based on a prediction error mechanism that is similar to that reported in the current study.

Our findings expand the knowledge of the neuronal mechanisms of social norms. Previous studies probing the neural mechanism of conformity or social norms have focused on the differences in neural responses to normative feedback delivered by a social group versus a computer (Berns et al., 2005), on the pathology of norms (King-Casas et al., 2008), and on a modulation of neural activity related to decision making by the possibility of punishment for violation of the norm (Spitzer et al., 2007). These studies uncovered the effects of the normative context (a prior group opinion [Berns et al., 2005] or the possibility of punishment [Spitzer et al., 2007]) on decision making but did not investigate closely the mechanism predicting conforming behavioral adjustments on trial-by-trial basis. The current study has for the first time revealed that the same regions are activated when there is a conflict with group opinion and predict subsequent adjustments of judgments. Our result provides evidence that behavioral conformity to descriptive group norms is triggered by the social conflict monitoring mechanism that is similar to the reinforcement learning signal (Holroyd and Coles, 2002). It is important to note here that there can be different mechanisms underlying conformity (Cialdini and Goldstein, 2004). Informational conformity (in contrast to normative conformity) serves an informational function in helping to be accurate, especially if normative information is provided before the actual decision (e.g., study [Berns et al., 2005]). From a neuroscience perspective, informational conformity assumes an attention-related neural mechanism, i.e., an activation of sensory cortices by normative information. In contrast, we found that neural activity predicting conformity to group norm was similar to a reinforcing learning signal. Therefore, conformity investigated in the current study is most probably normative and based on reinforcing social approval. In other words, group opinion works as a reinforcer for the individual's behavior. Both reward for being aligned with the group and aversion to being non-aligned may have acted as reinforcers. Further studies will help to generalize the observed mechanisms to the male population and other social situations (including injunctive or moral norms) leading to conformity.

In summary, the present study shows that group opinion affects our judgments of facial attractiveness, which play a critical role in human social interaction (see Langlois et al. [2000] for a review). Our results support the view in social psychology and economics that conformity is based on reinforcing social feedback, and we go on to propose a neural mechanism of conformity that agrees with the concept of reinforcement learning from animal learning theory. The fMRI results indicate that social conformity is based on mechanisms that comply with reinforcement learning. This process starts when a deviation from group opinion is detected by neural activity in the paracingulate region and ventral striatum. These regions then produce a neural signal similar to the prediction error signal in reinforcement learning that indicates a need for social conformity: a strong conflict-related signal in the RCZ and the NAc triggers adjustment of judgments in line with group opinion. Both the NAc and the RCZ receive midbrain dopaminergic innervations (Schultz, 2006). Moreover, animal studies robustly demonstrated that prediction error signal is dopamine mediated (Schultz, 2006). Our results suggest that a phasic change in presumably dopamine-related activity occurs when individual judgments differ from normative group opinion. Dopamine-dependent synaptic plasticity is thus a potential cellular mechanism for long-term conforming adjustments of judgments (Schultz, 2006). Overall, our results suggest that social conformity is underlined by the neural error-monitoring activity which signals probably the most fundamental social mistake—that of being ''too different'' from others.

EXPERIMENTAL PROCEDURES

Participants

In addition to the participants investigated in the behavioral study (see Supplemental Data for details), a total of 46 young right-handed women participated in the social (Experiment N1) and nonsocial control (Experiment N2) neuroimaging experiments with two experimental sessions: an fMRI session and a behavioral session, separated by approximately 30 min. None of the subjects reported a history of drug abuse, head trauma or neurological or psychiatric illness. Twenty-four healthy students (aged 19–27 years, mean 21.8 years) participated in the social version of the experiment (Experiment N1). Two participants were rejected from the study due to large head motions exceeding 3 mm, one subject was excluded due to her reported suspicion about the cover story of the experiment. Twenty-two healthy students (aged 19–29 years, mean 22.1 years) participated in the nonsocial control study described in the Results section (Experiment N2).

Stimuli

A set of 222 digital photos of European females (aged 18–35 years, from free internet sources) were used as stimuli. Color portraits of moderately attractive (mean 4.2, SD = 1.2 of the 8 point scale) females and moderate smile (rated AU6A/C+AU12B/C in accordance with the facial action coding system [FACS] by a certified FACS coder [Ekman et al., 1978]) were selected from a set of 1000 stimuli, all made with a highly similar photographic style and appearance. Attractiveness is a socially important facial feature (Langlois et al., 2000); judgments of facial attractiveness are fast, effortless, and consistent across subjects (Willis and Todorov, 2006). Therefore, a mismatch of individual judgments of facial attractiveness with group opinion should create a strong normative conflict. Social standards of female facial attractiveness are also constantly influenced by social norms, e.g., via fashion magazines and cosmetics commercials. Previous studies showed that individuals adjust their judgments of attractiveness in various situations (Geiselman et al., 1984; Kenrick and Guttierres, 1980). Ratings of facial attractiveness are modulated by social environment (Jones et al., 2007; Little et al., 2008) and thus it makes them an optimal and important model for studying social conformity.

Only female portraits and female subjects were selected. Crossgender rating of attractiveness is related to mate selection that has very specific neural mechanisms (Cloutier et al., 2008). In contrast, within-gender ratings of attractiveness can be generalized to other types of conforming behavior. One subject was excluded from the analysis due to reported homosexual orientation and motion artifacts.

Experiment

Subjects were informed that they were participating in a pan-European project "Seeing Beauty" to study human perception of attractiveness. They were told that the project team was conducting the same studies in France (Paris), Italy (Milan), and Netherlands (Nijmegen). The logos of European "collaborators" (Milan School of Design, French Institute of Beauty, and Dutch Royal Academy of Art) were included at the bottom of the written instructions. During the fMRI session subjects were exposed to a series of 222 photographs of female faces (stimuli duration = 2 s, intertrial interval [ITI] = 3–5 s, see Figure 1). Subjects were instructed to rate the face on an 8 point scale, ranging from very unattractive (1) to very attractive (8). Subjects indicated their rating by pressing the appropriate button. Eight buttons were used, four for each hand. The subject's rating (initial rating, green rectangle frame) was visualized on screen immediately after the face stimulus. Three to five seconds later, at the end of each trial, the subject was informed (by red rectangle frame) of the rating of the same face given by an "average European female participant from Milan and Paris" (group rating). The difference between the subject's and the group rating was also indicated by a score shown above the scale (0, ±2, or ±3 points). Importantly, the frame and the number indicating the conflict with group opinion were present during both "conflict" and "no-conflict" trails. Actual group ratings were programmed using the following criteria: in 33% of trials, group ratings agreed with subject's ratings, whereas in 67% of trials group ratings were pseudorandomly above or below subject's rating by ±2 or 3 points, i.e., using an adaptive algorithm that kept the overall ratio of "more negative" or "more positive" group ratings approximately equal during the experiment. Subjects were told that group ratings which matched with their own rating to within ±1 points produced the frame of the group rating visually overlapping with the frame of the subject's own rating. Subjects were not informed about the real purpose of the experiment and the manipulation of the group ratings. All photographs were randomized across subjects and conditions. Importantly, the sign of the difference between individual and group ratings does not play a role similar to positive and negative prediction error; in fact the prediction error (a deviation from the group) was always negative (see Supplemental Data for an additional discussion). Thirty minutes after the fMRI session in the unexpected (unannounced) subsequent behavioral session subjects were instructed to rate again—at their own pace—the attractiveness of the same faces presented in a new randomized order without the normative ratings (subsequent rating, Figure 1). At the end of the experiment, subjects were questioned using the self-monitoring scale on interpersonal influence (Snyder and Gangestad, 1986) (see Supplemental Data).

Our setup imitates social psychological studies investigating persuasion, where subjects are informed of a dominant behavior in a group (Cialdini and Goldstein, 2004). Social psychology suggests two types of social norms (Cialdini and Goldstein, 2004): (1) injunctive norms have a moral tone and characterize what people should do, whereas (2) descriptive norms represent typical behavior or what most people actually do, regardless of its appropriateness. In the current study, we investigated descriptive social norms that send out the message, "If a lot of people are doing this, it's probably a wise thing to do." It is also important to note that in our study subjects were not involved in a standard reinforcement task, i.e., they could not learn correct answers or a correct evaluation criteria because there was no correct answer, the normative feedback was pseudorandom.

Data Acquisition

Functional MRI was performed with ascending slice acquisition, using a T2*-weighted echo-planar imaging sequence (Sonata 1.5 T, Siemens, Munich; 33 axial slices; volume repetition time [TR], 2.28 s; echo time [TE], 35 ms; 90° flip angle; slice matrix, 64 × 64; slice thickness, 3.5 mm; slice gap, 0.5 mm; field of view, 224 mm). For structural MRI, we acquired a T1- weighted MP-RAGE sequence (176 sagittal slices; volume TR, 2.25 s; TE, 3.93 ms; 15° flip angle; slice matrix, 256 × 256; slice thickness, 1.0 mm; no gap; field of view, 256 mm).

MRI Data Analysis

Image analysis was performed with SPM5 (Wellcome Department of Imaging Neuroscience, London, UK). The first three EPI volumes were discarded to allow for T1 equilibration, and the remaining images were realigned to the first volume. Images were then corrected for differences in slice acquisition time, spatially normalized to the Montreal Neurological Institute (MNI) T1 template, resampled into $3 \times 3 \times 3$ mm^3 voxels, and spatially smoothed with a Gaussian kernel of 8 mm full-width at half-maximum. Data were high-pass filtered (cutoff at 1/128 Hz).

Statistical analysis was performed within the framework of the general linear model (Friston et al., 1995). Conflict and no-conflict trials were modeled separately, as were no-conflict trials (mean number of trials 73, SD = 0.8), conflict trials (mean number of trials 148, SD = 1.5), conflict trials followed by conformity (as tested during the behavioral session, mean number of trials 61, SD = 7.8), conflict trials that were not followed by conformity (mean number of trials 52, SD = 8.7). The regressors were convolved with the canonical hemodynamic response function of SPM5. In addition, the realignment parameters were included to model potential movement artifacts. In a whole-brain analysis, statistical tests were familywise error rate corrected for multiple comparisons across the entire brain. For the regions of interest, a small volume correction was used for the analysis of the conformity effects to correct for multiple comparisons across the search volume. For the RCZ and the NAc the search volumes were defined as a sphere with 15 mm radius around the center ($x = 4$, $y = 15$, $z = 43$ and $x = \pm 11$, $y = 11$, $z = -2$, respectively) based on the results of a previous study (Knutson et al., 2005) and meta-analysis (Ridderinkhof et al., 2004).

Hypothesis Testing

The individual contrasts were submitted to group-level random effects analysis. The main effect of social conflict was estimated by contrasting the group ratings in conflict and no-conflict trials. The main effect of conformity was investigated by comparing neural responses for all conflicting group ratings followed by conformity (i.e., facial attractiveness subsequently changed in accordance with the group rating) and all conflicting group ratings that were not followed by conformity (facial attractiveness not changed). In addition, a conjunction analysis was performed to confirm the regional overlap between the main effects of social conflict and conformity by testing the conjunction null hypothesis using the minimum T-statistic as implemented within SPM5 (Nichols et al., 2005). To assess the relationship between neural activity and individual level of conformity across subjects, individual contrast estimates within the RCZ and the NAc local maxima were extracted and entered in correlation analyses (two-tailed).

Image preprocessing and data analysis of Experiment N2 was identical to that of Experiment N1. Differences of neural responses in Experiments N1 and N2 were investigated by two-sample t test.

SUPPLEMENTAL DATA

The Supplemental Data include six figures, five tables, and supplemental text and can be found with this article online at http://www.neuron.org/supplemental/S0896-6273(08)01020-9.

ACKNOWLEDGMENTS

We thank Paul Gaalman for assistance in fMRI experiments and Stijn van Osselaer, Stefano Puntoni, and Theo Vosse for the incisive discussion of the results. We gratefully acknowledge the financial support of the Erasmus Research Institute of Management (ERIM).

Accepted: November 12, 2008
Published: January 14, 2009

REFERENCES

Aron, A.R., Shohamy, D., Clark, J., Myers, C., Gluck, M.A., and Poldrack, R.A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. J. Neurophysiol. *92*, 1144–1152.

Asch, S. (1951). Effects of group pressure upon the modification and distortion of judgments. In Groups, Leadership and Men Research in Human Relations, H. Guetzkow, ed. (Pittsburgh: Carnegie Press), pp. 177–190.

Bendor, J., and Swistak, P. (2001). The evolution of norms. Am. J. Sociol. *106*, 1493–1545.

Berns, G.S., McClure, S.M., Pagnoni, G., and Montague, P.R. (2001). Predictability modulates human brain response to reward. J. Neurosci. *21*, 2793–2798.

Berns, G.S., Chappelow, J., Zink, C.F., Pagnoni, G., Martin-Skurski, M.E., and Richards, J. (2005). Neurobiological correlates of social conformity and independence during mental rotation. Biol. Psychiatry *58*, 245–253.

Besson, C., and Louilot, A. (1995). Asymmetrical involvement of mesolimbic dopaminergic neurons in affective perception. Neuroscience *68*, 963–968.

Botvinick, M., Nystrom, L.E., Fissell, K., Carter, C.S., and Cohen, J.D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. Nature *402*, 179–181.

Bronstad, P.M., and Russell, R. (2007). Beauty is in the 'we' of the beholder: greater agreement on facial attractiveness among close relations. Perception *36*, 1674–1681.

Buckner, R.L., Andrews-Hanna, J.R., and Schacter, D.L. (2008). The brain's default network: anatomy, function, and relevance to disease. Ann. N Y Acad. Sci. *1124*, 1–38.

Carelli, R.M. (2002). The nucleus accumbens and reward: neurophysiological investigations in behaving animals. Behav. Cogn. Neurosci. Rev. *1*, 281–296.

Cialdini, R.B., and Goldstein, N.J. (2004). Social influence: compliance and conformity. Annu. Rev. Psychol. *55*, 591–621.

Cloutier, J., Heatherton, T.F., Whalen, P.J., and Kelley, W.M. (2008). Are attractive people rewarding? Sex differences in the neural substrates of facial attractiveness. J. Cogn. Neurosci. *20*, 941–951.

Cohen, M.X. (2007). Individual differences and the neural representations of reward expectation and reward prediction error. Soc. Cogn. Affect Neurosci. *2*, 20–30.

Cohen, M.X., and Ranganath, C. (2007). Reinforcement learning signals predict future decisions. J. Neurosci. *27*, 371–378.

Delgado, M.R., Frank, R.H., and Phelps, E.A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. Nat. Neurosci. *8*, 1611–1618.

di Pellegrino, G., Ciaramelli, E., and Ladavas, E. (2007). The regulation of cognitive control following rostral anterior cingulate cortex lesion in humans. J. Cogn. Neurosci. *19*, 275–286.

Diedrichsen, J., Hashambhoy, Y., Rane, T., and Shadmehr, R. (2005). Neural correlates of reach errors. J. Neurosci. *25*, 9919–9931.

Eisenberger, N.I., Lieberman, M.D., and Williams, K.D. (2003). Does rejection hurt? An FMRI study of social exclusion. Science *302*, 290–292.

Ekman, P., Friesen, W., and Hager, J. (1978). Facial Action Coding System (Palo Alto, CA: Consulting Psychologists Press).

Fehr, E., and Fischbacher, U. (2004). Third-party punishment and social norms. Evol. Hum. Behav. *25*, 63–87.

Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elger, C.E., and Falk, A. (2007). Social comparison affects reward-related brain activity in the human ventral striatum. Science *318*, 1305–1308.

Frank, M.J., Woroch, B.S., and Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. Neuron *47*, 495–501.

Friston, K.J., Frith, C.D., Turner, R., and Frackowiak, R.S. (1995). Characterizing evoked hemodynamics with fMRI. Neuroimage *2*, 157–165.

Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., and Donchin, E. (1993). A neural system for error detection and compensation. Psychol. Sci. *4*, 385–390.

Geiselman, R.E., Haight, N.A., and Kimata, L.G. (1984). Context effects on the perceived physical attractiveness of faces. J. Exp. Soc. Psychol. *20*, 409–424.

Greene, J.D., Nystrom, L.E., Engell, A.D., Darley, J.M., and Cohen, J.D. (2004). The neural bases of cognitive conflict and control in moral judgment. Neuron *44*, 389–400.

Herrmann, M.J., Rommler, J., Ehlis, A.C., Heidrich, A., and Fallgatter, A.J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). Brain Res. Cogn. Brain Res. *20*, 294–299.

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

Homans, G.C. (1950). The Human Group (New York: Harcourt).

Iaria, G., Fox, C.J., Waite, C.T., Aharon, I., and Barton, J.J. (2008). The contribution of the fusiform gyrus and superior temporal sulcus in processing facial attractiveness: Neuropsychological and neuroimaging evidence. Neuroscience *155*, 409–422.

Jones, B.C., DeBruine, L.M., Little, A.C., Burriss, R.P., and Feinberg, D.R. (2007). Social transmission of face preferences among humans. Proc. Biol. Sci. *274*, 899–903.

Kennerley, S.W., Walton, M.E., Behrens, T.E., Buckley, M.J., and Rushworth, M.F. (2006). Optimal decision making and the anterior cingulate cortex. Nat. Neurosci. *9*, 940–947.

Kenrick, D.T., and Guttierres, S.E. (1980). Contrast effect and judgements of physical attractiveness. J. Appl. Soc. Psychol. *38*, 131–140.

Kerns, J.G., Cohen, J.D., MacDonald, A.W., 3rd, Cho, R.Y., Stenger, V.A., and Carter, C.S. (2004). Anterior cingulate conflict monitoring and adjustments in control. Science *303*, 1023–1026.

King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., and Montague, P.R. (2008). The rupture and repair of cooperation in borderline personality disorder. Science *321*, 806–810.

Knutson, B., and Wimmer, G.E. (2007). Splitting the difference: how does the brain code reward episodes? Ann. N Y Acad. Sci. *1104*, 54–69.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., and Glover, G. (2005). Distributed neural representation of expected value. J. Neurosci. *25*, 4806–4812.

Langlois, J.H., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M., and Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. Psychol. Bull. *126*, 390–423.

Little, A., Burriss, R., Jones, B., DeBruine, L., and Caldwell, C. (2008). Social influence in human face preference: men and women are influenced more for long-term than short-term attractiveness decisions. Evol. Hum. Behav. *29*, 140–146.

McClure, S.M., Berns, G.S., and Montague, P.R. (2003). Temporal prediction errors in a passive learning task activate human striatum. Neuron *38*, 339–346.

McCoy, A.N., and Platt, M.L. (2005). Expectations and outcomes: decision-making in the primate brain. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. *191*, 201–211.

Montague, P.R., and Lohrenz, T. (2007). To detect and correct: norm violations and their enforcement. Neuron *56*, 14–18.

Montague, P.R., King-Casas, B., and Cohen, J.D. (2006). Imaging valuation models in human choice. Annu. Rev. Neurosci. *29*, 417–448.

Nagano-Saito, A., Leyton, M., Monchi, O., Goldberg, Y.K., He, Y., and Dagher, A. (2008). Dopamine depletion impairs frontostriatal functional connectivity during a set-shifting task. J. Neurosci. *28*, 3697–3706.

Nichols, T., Brett, M., Andersson, J., Wager, T., and Poline, J.B. (2005). Valid conjunction inference with the minimum statistic. Neuroimage *25*, 653–660.

O'Doherty, J.P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. Curr. Opin. Neurobiol. *14*, 769–776.

Phelps, E.A., and LeDoux, J.E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. Neuron *48*, 175–187.

Picard, N., and Strick, P.L. (1996). Motor areas of the medial wall: a review of their location and functional activation. Cereb. Cortex *6*, 342–353.

Pochon, J.B., Riis, J., Sanfey, A.G., Nystrom, L.E., and Cohen, J.D. (2008). Functional imaging of decision conflict. J. Neurosci. *28*, 3468–3473.

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. Science *306*, 443–447.

Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. (2002). A neural basis for social cooperation. Neuron *35*, 395–405.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E., and Cohen, J.D. (2003). The neural basis of economic decision-making in the Ultimatum Game. Science *300*, 1755–1758.

Schall, J.D., Stuphorn, V., and Brown, J.W. (2002). Monitoring and control of action by the frontal lobes. Neuron *36*, 309–322.

Schonberg, T., Daw, N.D., Joel, D., and O'Doherty, J.P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. J. Neurosci. *27*, 12860–12867.

Schultz, W. (2006). Behavioral theories and the neurophysiology of reward. Annu. Rev. Psychol. *57*, 87–115.

Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J., and Frith, C.D. (2006). Empathic neural responses are modulated by the perceived fairness of others. Nature *439*, 466–469.

Snyder, M., and Gangestad, S. (1986). On the nature of self-monitoring: matters of assessment, matters of validity. J. Pers. Soc. Psychol. *51*, 125–139.

Somerville, L.H., Heatherton, T.F., and Kelley, W.M. (2006). Anterior cingulate cortex responds differentially to expectancy violation and social rejection. Nat. Neurosci. *9*, 1007–1008.

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., and Fehr, E. (2007). The neural signature of social norm compliance. Neuron *56*, 185–196.

Tobler, P.N., Fletcher, P.C., Bullmore, E.T., and Schultz, W. (2007). Learning-related human brain activations reflecting individual finances. Neuron *54*, 167–175.

Willis, J., and Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. Psychol. Sci. *17*, 592–598.

Yeung, N., and Sanfey, A.G. (2004). Independent coding of reward magnitude and valence in the human brain. J. Neurosci. *24*, 6258–6264.

Zink, C.F., Tong, Y., Chen, Q., Bassett, D.S., Stein, J.L., and Meyer-Lindenberg, A. (2008). Know your place: Neural processing of social hierarchy in humans. Neuron *58*, 273–283.