MEG signature of social conformity: evidence from evoked and induced responses.

Zubarev I.1,2,5, Ossadtchi A.1,2, Klucharev V.2,4, and Shestakova A.1,2,3

1Department of Higher Nervous Activity and Psychophysiology, Saint-Petersburg State University, Russia
2National Research University, Higher School of Economics, Moscow, Russia
3MEG Centre, Moscow State University of Psychology and Education, Moscow, Russia
4Department of Psychology University of Basel, Switzerland
5Department of Biomedical Engineering and Computational Science, Aalto University, Espoo, Finland

Corresponding author: Ivan Zubarev, Centre for Cognition & Decision Making National Research University, Higher School of Economics, Moscow,
e-mail: ivan.zubarev@aalto.fi

Number of pages: 34
Number of figures, tables: Figures - 4, Tables - 2
Number of words: Abstract - 173, Introduction 454, and Discussion - 1374
Conflict of Interest: Authors report no conflict of interest

Acknowledgements: This work is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE). This work was partially supported by the Grant of Saint Petersburg State University 0.37.522.2013. The authors would also like to thank Dr. Vadim Nikulin (Charite Hospital, Berlin) for comments regarding the CSP methodology as well as the personnel of MEG center of Moscow City University of Psychology and Education (MSUPE) and personally prof. Tatiana Stroganova for valuable methodological comments and productive discussion.
Abstract

Humans often adjust their behavior to match the group norms. In this study, we used magnetoencephalographic (MEG) source imaging to investigate the electromagnetic responses to the perceived mismatch between individual and group opinions. After participants were exposed to group opinion that conflicted with their own, we observed an evoked response in the posterior medial prefrontal cortex (pMPFC) occurring around 200 ms, corresponding to the feedback-related negativity (FRN) – a component of the evoked response associated with processing negative feedback and reinforcement learning. This response was accompanied by an increase in power of theta oscillations (4-8 Hz) over a number of frontal sites (including OFC and pMPFC). The magnitude of both evoked and induced responses to the perceived conflict with social norms was stronger in participants who showed relatively low conformity. Overall, our results suggest that the activation of the pMPFC following conflicts with group opinion, as recoded by MEG, may reflect an enhanced control state – a process complimentary to the reinforcement learning signal in the ventral striatum reported in previous studies of social conformity.

Introduction

Defined as a tendency to align one’s attitudes, beliefs and behaviors to match the group norms, social conformity, is a well-documented phenomenon in social psychology (Cialdini and Goldstein, 2004). Yet only recently the neuroscience has focused on neurobiological mechanisms underlying conformity (see Izuma 2013, for a review). A number of fMRI studies demonstrated that being exposed to a group opinion conflicting with one’s individual opinion triggered the activations in the posterior medial prefrontal cortex (pMPFC) and ventral striatum (Klucharev et al., 2009; Berns et al., 2010; Campbell-Meiklejohn et al., 2010; Izuma and Adolphs, 2013). The pMPFC has been implicated in the generation of a so-called reward prediction error signal when the result of an action mismatches the expectation (Holroyd & Coles 2002; Nieuwenhuis et al. 2004; Rushworth et al. 2007; Cohen & Ranganath 2007; but also see
This signal presumably reflects a process of updating predictions of action values and thus guiding future action selection (Niv, 2009). This finding led to suggest that social conformity may be based on general action-monitoring and reinforcement learning mechanisms (Klucharev et al. 2009; Klucharev et al. 2011; Shestakova et al. 2013).

Several electroencephalographic (EEG) studies (Kim et al., 2012; Shestakova et al., 2013) demonstrated that a mismatch between individual opinion and the opinion of the group elicited the feedback-related negativity (FRN), an event-related brain potential (ERP) component associated with outcome evaluation and behavioral adaptation (see Walsh & Anderson 2012, for a review). Moreover, a growing body of literature links FRN to a modulation of ongoing theta-band (4-8Hz) oscillations over the anterior cingulate cortex (ACC), pMPFC and several other frontal sites (Cohen et al., 2007; Cavanagh et al., 2010, 2012; van de Vijver et al., 2011).

More specifically, we hypothesized that (1) the conflict between individual and group opinions would trigger evoked responses over the pMPFC, corresponding to the FRN, accompanied with the increase in power of theta oscillations in that region, and that (2) the dynamics of these responses would be predictive of individual differences in proneness to social conformity.

To test these hypotheses, we used a paradigm in which a person’s initial judgments, that is, perceived trustworthiness of faces, were open to the social influence of the opinion of a group. Participants rated the trustworthiness of faces and after each rating they were informed about an ‘average group rating’ of the face given by a large group of people. With this procedure, we introduced conflict between a person’s own judgment and the opinion of a group and compared MEG activity calculated over trials in which the group rating differed from the participant’s rating (conflict trials) with all no-conflict trials. Next, we probed whether such activity differed between participants who demonstrated high and low levels of conformity.

**Materials and Methods**

**Participants**
20 female volunteers took part in the experiment (mean age 24.2, range 18-28, right-handed (Oldfield, 1971), with normal or corrected to normal eyesight). All participants reported no history of neurological or psychiatric disease, drug abuse or head trauma. The data of one participant was discarded from the group analysis due to a large number of artifacts. The participants received monetary compensation of 500 rubles (equivalent of 16 US dollars) for participating in the experiment. The amount received covers typical one day food (grocery store) expenses for a single person in Moscow.

All participants were familiarized with experimental procedure and signed the informed consent form. The study was approved by the Ethics committee of Saint Petersburg State University. The participants’ personality traits were tested using the Eysenck Personality Inventory (Eysenck and Eysenck, 1994), Sensation Seeking scale (Aluja et al., 2010), a short version of Big Five questionnaire (Gosling et al., 2003), the Mehrabian Conformity Scale (Mehrabian, 1997), individual level of anxiety (Hajcak et al., 2003; Gu et al., 2010), the Locus of Control questionnaire (Rotter, 1966) and Spielberger’s State Trait Anxiety Inventory (Spielberger et al., 1970). We did not find any significant correlations between the behavioral results and the aforementioned personality traits (p > 0.2).

**Stimuli and procedure**

In present study we use the face judgment task (Campbell-Meiklejohn et al., 2010) where participants were instructed to rate the trustworthiness of faces. While their MEG was recorded (session 1), participants were exposed to a series of 222 photographs of female faces (stimuli duration=2s, inter-trial interval=2.5–3.0s, overall duration of the session=35 min). At the beginning of each trial (Fig.1A), the participants were exposed to a photograph of a female face for 2 seconds (with the face occupying approximately 60% of the picture). According to the instruction participants decided whether to trust a substantial sum of money (equivalent of 1500 US dollars) to the person presented on the screen or not. Participants rated each face using the
eight-point scale ranging from 1 (very untrustworthy) to 8 (very trustworthy). The participants indicated their choices by pressing the appropriate button. The participant’s rating (initial rating, blue rectangular frame) was visualized on the screen immediately after the button press. At the end of each trial, the participant was informed about the average rating of the same face given by a large group of students from the same Russian university (group rating, green rectangular frame). The difference between the participant’s and the group rating was additionally indicated by a score above the scale (0, +/-2 or +/-3 points). The frame and the number indicating the deviance from the group opinion appeared for both ‘conflict’ and ‘no-conflict’ trials. Actual group ratings were generated pseudorandomly as \( R_g = R_0 + M \), where \( R_g \) is the group rating, \( R_0 \) is initial rating given by the participant and \( M \) is a (pseudo) random modifier. Our sampling scheme used the adaptive algorithm described in Klucharev et al. (2009) ensuring that for 33% of the trials the “group rating” agreed with the participants’ initial rating (no-conflict trials, \( M = 0 \)), whereas in 67% of the trials the group rating were above or below the participants’ initial ratings by 2 or 3 points (conflict trials, \( M = [2,3,-2,-3] \)). Thus, the relative number of ‘more negative’, ‘more positive’ or ‘equal’ group ratings was approximately equal for every participant.

Participants were neither informed regarding the real purpose of the experiment nor were they informed about the way the group ratings were created.

After the first MEG session, the participants were offered a 30-min rest break outside the MEG room. Next, participants were instructed to rate the same set of faces again during the second session (subsequent rating, session 2). Before and after the first experimental session we also recorded two 10- min blocks of participants’ resting-state activity to estimate task independent brain noise covariance matrix. To probe the duration of the conformity-related effects, we asked participants to rate the trustworthiness of the same faces again three months after the MEG experiment (subsequent session 3 data was collected for 15 out of 20 participants).

**MEG acquisition and preprocessing**
We registered and processed MEG in accordance with the recent guidelines to good practices for conducting MEG studies (Gross et al., 2013): We used 306 channel Electa Neuromag System (102 magnetometers and 204 planar gradiometers, 1000 Hz sampling rate, the data was low-passed filtered with 333 Hz cut-off frequency). To control for cardiac and eye-motion related artifacts electrocardiographic (ECG) and electrooculographic (EOG) electrodes were mounted prior to MEG acquisition. Head movements were controlled using the continuous head position identification (cHPI) system. ECG electrodes were placed on the breastbone and on axillary furrow approximately near the 5th rib. Vertical EOG (vEOG) electrodes were placed above and below the center of left eye, horizontal EOG (hEOG) were placed on frontal processes of left and right zygomatic bones. ECG and EOG recordings were used as additional source of information for projecting out artifacts. Anatomical landmarks (NAS, LPA, RPA), cHPI-coil positions and 100 +/-5 additional head shape points were digitized using Polhemus Isotrak digital tracker system (Polhemus, Colchester, VT, USA). Participants were instructed to avoid movements and blink as little as possible during the experiment. The stimuli were presented on a semi-transparent display via a projector situated outside the room. The distance between participant’s head and the display was 1.5m. To assure equal distance between frontal and occipital sensors and participants’ head a special cushion was used whenever it was necessary. Magnetoencephalogram was preprocessed using the Neuromag Maxfilter software by the means of the temporally-extended signal space separation (tSSS) algorithm (Taulu and Hari, 2009) with a temporal autocorrelation threshold of 0.9 and segment length of 1 second. The MEG data was recalculated to compensate for head movements and to correspond to default head origin coordinates of (0, 0, 45)mm.

**Structural MRI acquisition and forward model**

Individual structural MRI T1-images were collected for each participant using the 1.5 T Philips Intera scanner. The reconstruction of cortical surfaces was performed with the Freesurfer image analysis suite (http://surfer.nmr.mgh.harvard.edu). Resulted cortical surface meshes were
imported into Brainstorm (Tadel et al., 2011) and down-sampled to 15,000 vertices for further processing. Forward modeling was performed using overlapping spheres method (Huang et al., 1999) as implemented in the Brainstorm Software. Due to unavailability of individual structural MRI the default MNI anatomy with 1mm resolution was used for 2 participants.

Analysis of behavioral data

To detect whether the conflict with the group rating had led to the subsequent change of trustworthiness ratings we compared participants’ initial and subsequent ratings assigned to the same face during the two experimental sessions (conformity effects). The trials were then grouped based on whether the group rating was more positive (by 2 or 3 points), more negative (by 2 or 3 points) or identical to the participant’s initial rating (total 5 groups). Mean size of the conforming face rating changes following conflict between the sessions was calculated. To ensure that actual conformity effects were probed and provide for comparability between subjects we also computed mean rating change for trials where group agreed with participant and subtracted it from each participant’s conformity scores data.

To account for limitations of repeated measurements inherent to experiments involving preference changes (Chen and Risen, 2010) and to isolate the conformity effects we used the following randomization approach. For each participant we performed 1000 randomizations of the behavioral data during which we permuted the “group ratings” while keeping the initial and subsequent ratings from the first and the second sessions constant. Such randomization scheme keeps all the properties of the “group rating” constant but destroys the possible causal connection of the subsequent rating with the “group ratings”. For each iteration, we calculated the mean “conforming rating change” (i.e. positive shift for trials where the group was more positive minus negative shift when the group was more negative about the face, similar to Kim et al. 2012) for the randomly generated surrogate data and built the sampling distribution of this quantity. Same analysis was repeated for the frequency of such conforming rating adjustments.
occurring for each subject (similar to Berns et al. 2010). Since the original group ratings were
generated based on random increments to subjects’ ratings (and thus are independent), the
permutability assumption is justified and we can use such a randomization scheme to recover the
distribution of 'conformal adjustments' under the null hypothesis of no causal connection
between group and individual ratings. Using this null hypotheses distribution we computed the z-
scores of the observed magnitude of conformity using the mean and the standard deviation of the
null-hypothesis distribution. The proposed scheme resolves the problems associated with
possible artificial correlations between the direction of participant's ratings changes and
randomly generated 'group opinion' induced by the scale finiteness and probes the actual
conformity effect.

The resulting two-modal distribution of z-scores suggested the existence of two readily
identifiable groups of participants. For the purposes of further analysis we assigned 6 of 19
participants falling into the higher mode (mean = 1.67, SD = 0.61) to 'conformists' group, while
other 13 participants forming the lower mode (mean = -0.39, SD = 0.43) were labeled as 'non-
conformists'.

Mean conforming rating changes were submitted to a 3-way ANOVA with the within-subject
factors conflict direction (group’s opinion is more positive vs. group’s opinion is more
negative), conflict size (small vs. large conflict) and the between-subject factor level of
conformity (conformists vs. non-conformists group). Additionally, we separately analyzed the
conformity effects following the intermediate initial ratings (4 and 5 points) to account for the
regression towards the mean and scale finiteness effects that may have given rise to artificial
correlations between group rating and subsequent rating changes. Based on the results of the
procedure described above we compared the MEG data between participants who demonstrated
stronger conformity effects and the participants who demonstrated conformity effects at chance
level.
MEG data analysis

MEG data analysis was performed using Brainstorm package (Tadel et al., 2011) and custom
written Matlab routines (The MathWorks, Inc.). Prior to the analysis, the recordings were down-
sampled to a 500 Hz sampling rate. Event-related magnetic fields (ERF) and time-frequency
maps were locked onto the presentation of the group rating for the conflict trials (i.e., when the
participant ’s ratings mismatched the group rating) and for the no-conflict trials (i.e., when the
participant ’s ratings matched the group rating) separately.

Sensor space evoked response field (ERF) analysis. For the ERF analysis, we extracted epochs
in the −200 ms to 800 ms time window locked to the stimulus onset (group opinion
presentation). To remove the DC (Direct Current) offset from each trial, a zero-order polynomial
detrend was applied based on the pre-stimulus interval (-200 ms, 0 ms). To identify time
windows of components of the evoked response that account for the differences in activation
between the conflict and no-conflict trials we computed Global Field Power (GFP) Root-Mean-
Squared (RMS) curves separately for all magnetometers and all gradiometers. We used
permutation tests followed by false discovery rate correction (FDR = 0.1, for 500 time points,) to
estimate time windows where the significant differential activation was observed. To increase
the reliability of the test we restricted analysis to time-windows that exceeded 20 ms in duration.

Source space evoked field analysis. To localize the cortical sources, we first computed the
averaged waveforms for conflict and no-conflict trials in sensor space. We then used the
Brainstorm implementation of cortically constrained Tikhonov-regularized normalized
minimum-norm estimate (MNE)(Hämäläinen and Ilmoniemi, 1994) with the fixed dipole
orientations and depth weighting coefficient 0.6 (Lin et al., 2006) to solve the inverse problem
and localize the sources of the evoked activity (Baillet et al., 2001). In the group analysis,
individual source-space ERF data was projected on default MNI brain with a 1mm resolution
using iterative closest point search algorithm as implemented in Brainstorm software (Tadel et al.
We then conducted a cluster mass-based permutation test on the cortical activation maps for conflict vs. no-conflict conditions (Maris and Oostenveld, 2007) using all participants data. The mass of the cluster was calculated as the sum of signed t-scores in the adjacent vertices and time points. The threshold for cluster inclusion was set to correspond to uncorrected $p$-value < 0.025 for a two-tailed t-test. For each cluster in the original data $p$-value was computed as a probability of observing a cluster of larger (positive or negative separately) mass over 2000 random permutations. We performed a soft version of maximum distribution-based correction for multiple comparisons using the FWER principle. Ten (and not one as in the standard procedure) largest clusters per each permutation were used to compute the $p$-value. To alleviate the multiple comparisons problem we limited the studied cortical surface to bilateral frontal lobes (including medial surface) delineated by postcentral gyrus posteriorly and superior temporal gyrus laterally. For the same reason we also limited the source analysis to three 100 ms time intervals (the earlier interval lasted from 150 to 250 ms, the intermediate from 350 to 450ms and the latter from 450 to 550 ms intervals) based on the results of GFP analysis. We could do so because the GFP does not provide any spatially specific information. The data for each cortical hemisphere was analyzed separately.

**Region of interest (ROI) analysis.** To elucidate the dynamics of MEG activation related to conformity, we analysed four significant ROIs (right MPFC, left OFC, left anterior MPFC and left posterior MPFC) identified at the earlier stage of analysis using the source space cluster-based permutation test described above. To estimate the temporal profile of ROI activation power we extracted the absolute value of the first mode PCA component loading of the difference waveform (conflict – no conflict), computed for all vertices comprising each ROI. We then performed permutation tests to probe whether such ROI activations differed between the two groups of participants by randomly assigning them to what we called “conformist” and “non-conformist” groups. Test results were corrected for multiple comparisons by identifying the
largest cluster (in the time domain) of significant t-scores over 2000 permutations.

**Sensor-Space time-frequency data analysis.** To analyze the induced oscillatory activity, we extracted epochs that included a 1-s pre-stimulus and 2-s post-stimulus locked to the presentation of the group opinion. DC offset was removed from each epoch by aligning the time series to the average amplitude of 1 sec pre-stimulus interval. In order to remove the phase-locked activity, we subtracted the averaged evoked response from each epoch before further analysis.

The data was filtered in the theta (4-8Hz) frequency band. We applied Hilbert transform to extract the amplitude envelope and calculated event-related (de)synchronization (ERS/ERD) coefficients by aligning the absolute Hilbert amplitude envelope values, averaged over 600 ms after stimulus presentation to the averaged baseline amplitude calculated based on the -400ms -0ms pre-stimulus interval for a cluster of 44 frontal gradiometers (Fig. 4,A).

The resulting ERS coefficients were analyzed using the two-way ANOVA to probe whether the increase in frontal theta amplitude differed between the conditions (conflict and no-conflict) and the two groups of participants ("conformists" and "non-conformists").

**Source space time-frequency data analysis**

In order to localize sources of theta oscillatory activity, we first computed a spatial filter that maximized the variance between the conflict and no-conflict conditions for each participant using the Common Spatial Pattern (CSP) approach (for details see (Blankertz et al., 2008)). We used only magnetometer data filtered in the theta frequency band. Tikhonov regularization with parameter lambda = 0.001 was applied to condition the covariance inverse matrix. We then applied MNE localization procedure to the CSP’s corresponding to the two largest eigenvalues of the generalized covariance matrix. Similarly to the ERF analysis, we projected individual MNE solutions obtained for the a.m. CSP components on the default anatomy for grand averaging.

**Results**
Behavioral results

Overall, the participants rated faces (session 1) as moderately trustworthy: mean rating = 4.3, SD = 0.67. When participants’ opinion conflicted with the group rating, they changed their ratings towards the group opinion in 46% of the trials (range from 33.5% to 61%, session 2). The randomization procedure performed on the behavioral data revealed two groups of participants: “conformists” (average conforming rating change = 0.66, std = 0.28) and “non-conformists” (average conforming rating change = 0.42, std = 0.14).

The 3-way ANOVA revealed the significant main effect of conflict direction (F(1.19) = 116.1, p = 0.00001) and significant interactions conflict direction x conflict size (F(1,19)=22.7, p = 0.00001) and conflict direction x level of conformity (F(1,19) = 6.31, p = 0.014). Therefore, the conformity effects were stronger after large conflicts with the group opinion than after small conflicts. Furthermore, the conformity effects were stronger for ‘conformists’ as compared to ‘non-conformists’. To account for possible artificial correlations between the group ratings and subsequent rating changes caused by repeated measurements we performed additional analysis using a subset of faces with intermediate initial ratings (4 and 5). The 3-way ANOVA also showed a significant main effect of conflict direction (F (1,19) = 12.54, p = 0.0007) and an interaction conflict direction x level of conformity (F(1,19) = 16.99, p =0.001).

To probe the duration of the conformity-related effects, we asked participants to rate the trustworthiness of the same faces again three months after the MEG experiment (session 3). The 3-way ANOVA of conformity effects revealed the significant main effect of conflict direction (F(1,14) = 101.2.2, p = 0.00001) and interaction conflict direction x level of conformity (F(1,14) = 14.57, p=0.0004). The conformity scores computed for the sessions 2 and 3 showed a significant positive correlation (r=0.68, p = 0.0051). Similarly to short-term conformity effects in session 2, we analyzed a subset of trials with the intermediate initial ratings (4 and 5). A 3-way ANOVA showed the significant main effect of conflict direction which occurred 3 months after
the MEG experiment (F (1, 14) = 5.88, p = 0.0188), this observation was also supported by the interaction \textit{conflict direction} \times \textit{conflict size} \times \textit{level of conformity} (F(1,14) = 6.29, p = 0.015) indicating that long-term conformity effects were observed only for large conflicts and only for ‘conformists’. Thus, in our study group opinion effectively modulated individuals’ judgments of trustworthiness.

**ERF results**

**GFP analysis of conflict-related activity.** To determine time-windows with the strongest conflict-related activity (conflict trials vs. no-conflict trials, duration greater than 20 ms) we analyzed evoked (GFP) activity for the magnetometers data using the permutation test (Fig. 2A). The test signified three time-windows: 204-240 ms, 414-440 ms and 496-518 ms after the onset of group ratings where the activity in conflict trials was stronger than the activity in no-conflict trials. Similar results were obtained with analysis of the gradiometer data (data not shown).

**Source-space analysis of conflict-related activity.** To identify the brain regions generating conflict-related activity we conducted a cluster-based permutation test in the MNE source space around the time-windows identified by the GFP analysis (150-250 ms, 350-450ms and 450-550 ms) and compared the activations between \textit{conflict} and \textit{no-conflict} trials. The test revealed four statistically significant clusters of activations (Fig 3A, Table 1) occurring only in the earliest time-window (150-250 ms): the bilateral pMPFC, left anterior medial prefrontal cortex (aMPFC), and left orbitofrontal cortex (left OFC). The right orbitofrontal cortex (right OFC) cluster approached significance (mass = 366.0, cluster p-value = 0.063). The four clusters that showed statistical significance were further used as ROI’s for the between-group analysis (‘conformists’ vs. ‘non-conformists’).

In the second (350-450ms), and third (450-550ms) time-windows no significant conflict-related clusters were identified, however three clusters approached statistical significance (left pMPFC,
left precentral gyrus and frontal superior sulcus, see Table 1 for details).

**ROI analysis of the conformity-related activity.** To identify whether the conflict-related activity differed between subjects with higher or lower tendency to conform to the group opinion we compared the amplitudes of difference activations (conflict minus no-conflict trial) between subjects that were assigned to the ‘conformists’ and ‘non-conformists’ within preselected ROIs. The permutation test indicated a larger conflict-related activity (cluster p-value = 0.031) for ‘non-conformists’ than for ‘conformists’ in the right pMPFC (Fig.3B).

**Time-frequency analysis of conflict- and conformity-related effects.** The ERS analysis of the induced oscillatory activity at the frontal sensors (Fig.4A.) revealed that both in conflict (mean magnitude increase = 17.3 percent, SD = 10.9) and no-conflict (mean magnitude increase = 7.2 percent, SD = 6.7) trials, the magnitude of the frontal theta activity (4-8Hz) increased relative to the prestimulus baseline (Fig. 4B). As expected, this increase was more pronounced for conflict trials, than for no-conflict trials (mean magnitude difference = 10, SD. = 7.7). This observation was supported by the two-way ANOVA performed for theta ERS coefficients with the conflict (conflict vs. no-conflict trials) as a within-group factor, and the level of conformity (conformists vs. non-conformist) as a between-group. The significant main effects of conflict (F(1,19) = 5.84, p = 0.022) and level of conformity (F(1,19) = 5.17, p = 0.03) indicate that ‘non-conformists’ showed a stronger increase in theta power than ‘conformists’. The interaction conflict x level of conformity was not significant (p = 0.16). Overall, we found a stronger synchronization of the theta band activity at the frontal sensors for ‘non-conformists’ than for ‘non-conformists’.

**Source analysis of conflict-related theta band effects.** To investigate brain regions generating the conflict-related induced theta activity we first computed the components accounting for most variance between conflict and no-conflict trials in the theta frequency band. Resulting CSP components explained on average 22.8% variance of the original data (SD = 6.25%, time-window = 0-600 ms). We then used the MNE inverse solution to localize the sources of these.
components. The localization of CSP components of the theta activity indicated multiple frontal
sources of activations including the bilateral MPFC (ventral and dorsal), OFC, bilateral temporal
poles, lingual gyrus and left dprsolateral prefrontal cortex (DLPFC) (Fig. 4C).

Discussion

The goal of this study was to elucidate temporal and spatial characteristics of electromagnetic
brain responses to the visual cues signaling that participants’ individual ratings of face
trustworthiness conflict with the group opinion. The GFP analysis revealed a stronger activation
around 204-240 ms in conflict trials where individual ratings differed from the group rating than
in no-conflict trials where individual ratings matched with the normative group ratings. The
timing of the activation in our study corresponds to the timing of the FRN, an evoked-response
component often associated with reward prediction error processing and adjustments of behavior
(Holroyd and Coles, 2002; Cohen et al., 2011a; Walsh and Anderson, 2012). The source analysis
revealed significant conflict-related activations in the bilateral pMPFC and left OFC.
Interestingly, the activation in the right pMFPC overlapped with the results of the previous fMRI
study (Klucharev et al., 2009) and occurred approximately within the same time-window as
conflict-related evoked responses reported in the previous EEG studies of social conformity (e.g.
Shestakova et al., 2012).

We also found that the conflict with the group opinion triggers an induced frontal theta activity
that has been implicated into performance monitoring and reinforcement learning (Cavanagh et
al., 2010, 2012; Cohen et al., 2011b; van de Vijver et al., 2011). Frontal theta dynamics observed
in our study was consistent with corresponding evoked responses: we observed a more
pronounced increase in the induced theta power in conflict as compared to no-conflict trials.
These results are in line with several recent studies suggesting that FRN can be a time-domain
manifestation of theta oscillatory perturbations in the MPFC (Cohen et al., 2008; Cavanagh et al.,
Source modeling of our results indicated that the conflict-related induced oscillatory activity in the theta band was generated in multiple regions including the MPFC, OFC, lateral PFC, supplementary motor area and the visual cortex, significantly overlapping with brain sources of the conflict-related evoked responses.

We further estimated individual proneness to social influence using a randomization approach and divided subjects in two groups: ‘conformists’ and ‘non-conformists’ based on their behavioral results. The ‘non-conformists’ group showed significantly smaller conformity (conformal changes of ratings), than the ‘conformists’ group. Furthermore, long-term conformity effects, measured 3 months after the MEG session, were significantly stronger in the ‘conformists’ group than in the ‘non-conformists’ group. We also found that the magnitude of both evoked and induced MEG responses differed significantly between ‘conformists’ and ‘non-conformists’. ROI analysis revealed that ‘non-conformists’ demonstrated a significantly stronger conflict-related evoked and theta synchronization responses than ‘conformists’.

Previous fMRI studies consistently reported that the magnitude of the conflict-related BOLD signals in the pMPFC correlated with individual differences in the proneness to conformity, with conformists demonstrating higher BOLD activations (see (Izuma and Adolphs, 2013), for a review). On a contrary, our results demonstrated a relative theta power increase and a larger evoked magnetic response in the pMPFC of ‘non-conformists’ than in the pMPFC of ‘conformists’. Several previous studies demonstrated a negative correlation between the theta power increase (ERS) localized in the MPFC and the amplitude of the BOLD in the same region (Meltzer et al. 2007; Hoffmann et al. 2013). It has been noted that ‘‘activations’’ of the MPFC as indicated by a BOLD signal increase may correspond to a deactivation of the so called “default mode network” (Raichle and Snyder, 2007) and vary with the degrees of engagement in the task (Rushworth et al., 2011). Thus, the frontal theta dynamics and its’ putative time-domain representations (FRN) may well co-localize with the decrease in the BOLD response (Winterer et al., 2007; Singh, 2012). Thus, our results add to the discussion of the relationship between the
Our MEG results support an important role of the pMPFC in the brain mechanisms of social influence. The activity of the ventral striatum also has been previously associated with social conformity (Klucharev et al., 2009; Berns et al., 2010; Campbell-Meiklejohn et al., 2010; Zaki et al., 2011; Izuma, 2013b). Since the ventral striatum was implicated in reinforcement learning (Delgado, 2007; O’Doherty et al., 2007; Niv, 2009), this region can contribute to the automatic learning mechanism underling conformity, while the activity of the pMPFC may underlie a distinct cognitively demanding process of a ‘control state’ evaluating whether to follow the group opinion or stick to the initial opinion. This hypothesis is consistent with recent accounts of the pMPFC role highlighting its function in cost computation, representing best alternative action and executive control functions (Folstein and Van Petten, 2008; Rushworth et al., 2012; Shenhav et al., 2013). Since the pMPFC is a part of the “default mode network”, the greater magnitude of biomagnetic conflict-related responses in the ‘non-conformist’ group may indicate the allocation of larger cognitive resources during the conflict with group norms as compared to the ‘conformist group’, where a more general reinforcement learning-like process engaging the ventral striatum could dominates over cortical activity of the pMPFC.

Our results suggest that a perceived conflict with a normative group opinion triggers the pMPFC activity similar to the FRN. To date, few studies have investigated MEG signatures of the FRN (Doñamayor et al., 2012; Thomas et al., 2013), closely-related error-related negativity (ERN) (Keil et al., 2010; Charles et al., 2013) and model-derived prediction error signal (Talmi et al., 2012). Although only partly consistent to each other, these studies registered the error and feedback-related activity at the ACC and MPFC (Thomas et al. 2013 Keil et al. 2010; Charles et al. 2013; Thomas et al. 2013) as well as at the posterior cingular cortex (Doñamayor et al. 2012). Differences in FRN localization in various studies can be due to the fact that this component may consist of two or more subcomponents. Previous EEG studies suggested that the FRN is produced by a superposition of two separate (negative and positive) subcomponents associated
with feedback processing (Holroyd et al., 2008; Baker and Holroyd, 2011; Krigolson et al., 2014). The negative subcomponent (‘N2’ component in EEG literature) is associated with processing of conflicting or unexpected events including unexpected negative and positive feedbacks. The N2 subcomponent has been also associated with high-order conflict processing and control state (Folstein and Van Petten, 2008) or with a state of enhanced control redirecting future behavior away from tasks and strategies that result in a conflict (Botvinick, 2007; Fritz and Dreisbach, 2013). While the second positive subcomponent (‘reward positivity’) is elicited specifically by unexpected positive events along with N2 and cancels out the negative subcomponent, resulting in a more positive waveform in EEG (Holroyd et al., 2008; Krigolson et al., 2014), but, presumably, not in MEG due to different source cancellation profiles (Irimia et al., 2012). Our results suggest that the observed MEG activity of the pMPFC may represent only the negative FRN subcomponent. Overall, we can speculate that the observed pMPFC activations may represent the processing of the conflict with descriptive social norms in terms of a self-control (a ‘lose-stay’, but not ‘lose-switch’ strategy), a process complimentary to reinforcement learning-related activity of the ventral striatum.

Alternatively, the weaker conflict-related activity of the pMPFC in ‘conformists’ group could be associated with a relatively weaker individual opinion about the trustworthiness of the faces resulting in a lower internal conflict (and smaller pMPFC activity) evoked by observed mismatch with group ratings. Consistent with this interpretation of our findings, Achtziger and colleagues demonstrated that people who tend to overweight new information display lower evoked activity upon receiving one that conflicts with prior information (Achtziger et al., 2014). Additional studies are clearly needed to clarify the function of the pMPFC activity during social conformity.

Our results also show the conflict-related activity in the OFC following the conflicts with the group opinion. The conflict-related activity in the OFC did not differ between conformists and non-conformists, probably reflecting a more general decision making (e.g. valuation) process. Additional analysis revealed a stronger activation of the OFC in conformists than non-
conformists, but only when the group agreed with the participants’ rating (196-226ms). Since the OFC has been robustly implicated in encoding stimulus value (e.g. Rushworth et al. 2007; Walton et al. 2011; Rushworth et al. 2012), the observed differential activity may also be related to a higher value of matching the group opinion for conformists.

Conclusions

Taken together our results suggest the involvement of the performance monitoring neural circuitry into the mechanisms underlying social influence. The evoked and induced activity of the pMPFC detected by MEG may underlie a control state following perceived conflicts with social norms. Additionally, the activity of the OFC to a perceived match with social norms can contribute to social conformity by assigning a value of being similar with the social group. Overall, our results further contribute to the growing literature investigating the neural mechanisms of social influence and support a profound role of the medial and orbital regions of the frontal cortex in conformity-related adjustments to social norms..

References


MEG oscillations. Hum Brain Mapp 28:805–816 Available at:


**Figure 1.** Behavioral task and behavioral results. A. The sequence of the events within a trial is shown. During the first session, subjects rated the trustworthiness 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 first session subjects rated again the same faces during the second session. During the third session three months later 15 out of 20 original rated the same set of faces one more time.
Figure 2. A. The GFP-RMS representation of conflict-related effects (magnetometers only).

Underlined areas represent time-windows, where statistically significant differences of conflict and no-conflict trials lasted for more than 20 ms.
**Figure 3.** Conflict-related effects (conflict trials vs. no-conflict trials). 

A. Results of the cluster level permutation test in MNE source space. aMPFC – anterior part of medial prefrontal cortex, pMFPC - posterior medial prefrontal cortex; OFC - orbitofrontal cortex. 

B. Time span of conflict-related activations (left). The time course of the right pMPFC activity for ‘conformist’ vs ‘non-conformist’ groups (right). 

Black blocks represent the duration of significant spatiotemporal clusters (see Table 1 for details). Red block overlaid over the back block indicate the time-window, of the significant conformity effect (‘conformists’ vs. ‘non-conformist’, see Table 2 for details).
Figure 4. Analysis of induced theta activity. A. Grand averaged Event-Related Synchronization (ERS) of theta activity in conflict and no-conflict trials (left) and in conflict trials between ‘conformist’ and ‘non-conformist’ groups (right) B. Source localization for the CSP components of induced theta conflict-related activity (conflict vs no-conflict trials) for all subjects.
Table 1. The localization of the conflict-related activity (conflict trials vs. no-conflict trials).

<table>
<thead>
<tr>
<th>GFP</th>
<th>Time-window</th>
<th>L/R</th>
<th>Structure</th>
<th>Cluster time window, ms</th>
<th>Number of vertices (unique)</th>
<th>Cluster mass, (t-score)</th>
<th>Cluster p-value, FWER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>150-250 ms</td>
<td>Right</td>
<td>MPFC</td>
<td>152-250</td>
<td>246</td>
<td>-491</td>
<td>0.042*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Right</td>
<td>OFC</td>
<td>150-194</td>
<td>194</td>
<td>366</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Left</td>
<td>OFC</td>
<td>158-196</td>
<td>140</td>
<td>-288,9</td>
<td>0.019*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Left</td>
<td>aMPFC</td>
<td>160-218</td>
<td>74</td>
<td>-179,4</td>
<td>0.048*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Left</td>
<td>pMPFC</td>
<td>186-248</td>
<td>157</td>
<td>-314,1</td>
<td>0.016*</td>
</tr>
<tr>
<td></td>
<td>350-500 ms</td>
<td>Left</td>
<td>pMPFC</td>
<td>414-442</td>
<td>99</td>
<td>212</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Left</td>
<td>Precentral Gyrus.</td>
<td>438-482</td>
<td>79</td>
<td>210</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>450-550 ms</td>
<td>Left</td>
<td>Frontal Superior Sulc.</td>
<td>482-532</td>
<td>94</td>
<td>291</td>
<td>0.053</td>
</tr>
</tbody>
</table>

Table 2 Time-windows of the conformity-related activity (within ROI determined during the localization of the conflict-related activity)
Conflicting-No-conflict (difference waveform)

<table>
<thead>
<tr>
<th>Side</th>
<th>Region</th>
<th>Time</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>OFC</td>
<td>210-224</td>
<td>-20.9</td>
<td>0.11</td>
</tr>
<tr>
<td>Right</td>
<td>pMPFC</td>
<td>216-240</td>
<td>-43</td>
<td>0.031*</td>
</tr>
<tr>
<td>Left</td>
<td>aMPFC</td>
<td>308-316</td>
<td>-19.25</td>
<td>0.14</td>
</tr>
</tbody>
</table>