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

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

3 [1]Department of Higher Nervous Activity and Psychophysiology, Saint-Petersburg State

4 University, Russia

5 [2]National Research University, Higher School of Economics, Moscow, Russia

6 [3]MEG Centre, Moscow State University of Psychology and Education, Moscow, Russia

7 [4]Department of Psychology University of Basel, Switzerland

8 [5]Department of Biomedical Engineering and Computational Science, Aalto University, Espoo

9 Finland

10 Corresponding author: Ivan Zubarev, Centre for Cognition & Decision Making

11 National Research University, Higher School of Economics, Moscow,

12 e-mail: ivan.zubarev@aalto.fi

13 Number of pages: 34

14 Number of figures, tables: Figures - 4, Tables - 2

15 Number of words: Abstract - 173, Introduction 454, and Discussion - 1374

16 Conflict of Interest: Authors report no conflict of interest

24

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

Botvinick 2007). This signal presumably reflects a process of updating predictions of action values and thus guiding future action selection (Niv, 2009). This findings led to suggestthat 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

75   20 female volunteers took part in the experiment (mean age 24.2, range 18-28, right-handed

76   (Oldfield, 1971), with normal or corrected to normal eyesight). All participants reported no

77   history of neurological or psychiatric disease, drug abuse or head trauma. The data of one

78   participant was discarded from the group analysis due to a large number of artifacts. The

79   participants received monetary compensation of 500 rubles (equivalent of 16 US dollars) for

80   participating in the experiment. The amount received covers typical one day food (grocery store)

81   expenses for a single person in Moscow.

82   All participants were familiarized with experimental procedure and signed the informed consent

83   form. The study was approved by the Ethics committee of Saint Petersburg State University. The

84   participants' personality traits were tested using the Eysenck Personality Inventory (Eysenck and

85   Eysenck, 1994), Sensation Seeking scale (Aluja et al., 2010), a short version of Big Five

86   questionnaire (Gosling et al., 2003), the Mehrabian Conformity Scale (Mehrabian, 1997),

87   individual level of anxiety (Hajcak et al., 2003; Gu et al., 2010), the Locus of Control

88   questionnaire (Rotter, 1966) and Spielberger's State Trait Anxiety Inventory (Spielberger et al.,

89   1970). We did not find any significant correlations between the behavioral results and the

90   aforementioned personality traits (p > 0.2).

91   ***Stimuli and procedure***

92   In present study we use the face judgment task (Campbell-Meiklejohn et al., 2010) where

93   participants were instructed to rate the trustworthiness of faces. While their MEG was recorded

94   (session 1), participants were exposed to a series of 222 photographs of female faces (stimuli

95   duration=2s, inter-trial interval=2.5–3.0s, overall duration of the session=35min). At the

96   beginning of each trial (Fig.1A), the participants were exposed to a photograph of a female face

97   for 2 seconds (with the face occupying approximately 60% of the picture). According to the

98   instruction participants decided whether to trust a substantial sum of money (equivalent of 1500

99   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***

125   We registered and processed MEG in accordance with the recent guidelines to good practices for

126   conducting MEG studies (Gross et al., 2013): We used 306 channel Electa Neuromag System

127   (102 magnetometers and 204 planar gradiometers,1000 Hz sampling rate, the data was low-

128   passed filtered with 333 Hz cut-off frequency). To control for cardiac and eye-movement related

129   artifacts electrocardiographic (ECG) and electrooculographic (EOG) electrodes were mounted

130   prior to MEG acquisition. Head movements were controlled using the continuous head position

131   identification (cHPI) system. ECG electrodes were placed on the breastbone and on axillary

132   furrow approximately near the 5th rib. Vertical EOG (vEOG) electrodes were placed above and

133   below the center of left eye, horizontal EOG (hEOG) were placed on frontal processes of left and

134   right zygomatic bones. ECG and EOG recordings were used as additional source of information

135   for projecting out artifacts. Anatomical landmarks (NAS, LPA, RPA), cHPI-coil positions and

136   100 +/-5 additional head shape points were digitized using Polhemus Isotrak digital tracker

137   system (Polhemus, Colchester, VT, USA). Participants were instructed to avoid movements and

138   blink as little as possible during the experiment. The stimuli were presented on a semi-

139   transparent display via a projector situated outside the room. The distance between participant's

140   head and the display was 1.5m. To assure equal distance between frontal and occipital sensors

141   and participants' head a special cushion was used whenever it was necessary.

142   Magnetoencephaloram was preprocessed using the Neuromag Maxfilter software by the means

143   of the temporally-extended signal space separation (tSSS) algorithm (Taulu and Hari, 2009) with

144   a temporal autocorrelation threshold of 0.9 and segment length of 1 second. The MEG data was

145   recalculated to compensate for head movements and to correspond to default head origin

146   coordinates of (0, 0, 45)mm.

147   ***Structural MRI acquisition and forward model***

148   Individual structural MRI T1-images were collected for each participant using the 1.5 T Philips

149   Intera scanner. The reconstruction of cortical surfaces was performed with the Freesurfer image

150   analysis suite (http://surfer.nmr.mgh.harvard.edu). Resulted cortical surface meshes were

151 imported into Brainstorm (Tadel et al., 2011) and down-sampled to 15 000 vertices for further

152 processing. Forward modeling was performed using overlapping spheres method (Huang et al.,

153 1999) as implemented in the Brainstorm Software. Due to unavailability of individual structural

154 MRI the default MNI anatomy with 1mm resolution was used for 2 participants.

Analysis of behavioral data

156 To detect whether the conflict with the group rating had led to the subsequent change of

157 trustworthiness ratings we compared participants' initial and subsequent ratings assigned to the

158 same face during the two experimental sessions (conformity effects). The trials were then

159 grouped based on whether the group rating was more positive (by 2 or 3 points), more negative

160 (by 2 or 3 points) or identical to the participant's initial rating (total 5 groups). Mean size of the

161 conforming face rating changes following conflict between the sessions was calculated. To

162 ensure that actual conformity effects were probed and provide for comparability between

163 subjects we also computed mean rating change for trials where group agreed with participant and

164 subtracted it from each participant's conformity scores data.

165 To account for limitations of repeated measurements inherent to experiments involving

166 preference changes (Chen and Risen, 2010) and to isolate the conformity effects we used the

167 following randomization approach. For each participant we performed 1000 randomizations of

168 the behavioral data during which we permuted the "group ratings" while keeping the initial and

169 subsequent ratings from the first and the second sessions constant. Such randomization scheme

170 keeps all the properties of the "group rating" constant but destroys the possible causal connection

171 of the subsequent rating with the "group ratings". For each iteration, we calculated the mean

172 "conforming rating change" (i.e. positive shift for trials where the group was more positive

173 minus negative shift when the group was more negative about the face, similar to Kim et al.

174 2012) for the randomly generated surrogate data and built the sampling distribution of this

175 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.

226   2011).

227   We then conducted a cluster mass-based permutation test on the cortical activation maps for

228   *conflict* vs. *no-conflict* conditions (Maris and Oostenveld, 2007) using all participants data. The

229   mass of the cluster was calculated as the sum of signed t-scores in the adjacent vertices and time

230   points. The threshold for cluster inclusion was set to correspond to uncorrected *p*-value < 0.025

231   for a two-tailed t-test. For each cluster in the original data p-value was computed as a probability

232   of observing a cluster of larger (positive or negative separately) mass over 2000 random

233   permutations. We performed a soft version of maximum distribution-based correction for

234   multiple comparisons using the FWER principle. Ten (and not one as in the standard procedure)

235   largest clusters per each permutation were used to compute the p-value. To alleviate the multiple

236   comparisons problem we limited the studied cortical surface to bilateral frontal lobes (including

237   medial surface) delineated by postcentral gyrus posteriorly and superior temporal gyrus laterally.

238   For the same reason we also limited the source analysis to three 100 ms time intervals (the earlier

239   interval lasted from 150 to 250 ms, the intermediate from 350 to 450ms and the latter from 450

240   to 550 ms intervals) based on the results of GFP analysis. We could do so because the GFP does

241   not provide any spatially specific information. The data for each cortical hemisphere was

242   analyzed separately.

243   ***Region of interest (ROI) analysis***. To elucidate the dynamics of MEG activation related to

244   conformity, we analysed four significant ROIs (right MPFC, left OFC, left anterior MPFC and

245   left posterior MPFC) identified at the earlier stage of analysis using the source space cluster-

246   based permutation test described above. To estimate the temporal profile of ROI activation

247   power we extracted the absolute value of the first mode PCA component loading of the

248   difference waveform (conflict – no conflict), computed for all vertices comprising each ROI.

249   We then performed permutation tests to probe whether such ROI activations differed between

250   the two groups of participants by randomly assigning them to what we called "conformist" and

251   "non-conformist" groups. Test results were corrected for multiple comparisons by identifying the

252   largest cluster (in the time domain) of significant t-scores over 2000 permutations.

253   ***Sensor-Space time-frequency data analysis.*** To analyze the induced oscillatory activity, we

254   extracted epochs that included a 1-s pre-stimulus and 2-s post-stimulus locked to the presentation

255   of the group opinion. DC offset was removed from each epoch by aligning the time series to the

256   average amplitude of 1 sec pre-stimulus interval. In order to remove the phase-locked activity,

257   we subtracted the averaged evoked response from each epoch before further analysis.

258   The data was filtered in the theta (4-8Hz) frequency band. We applied Hilbert transform to

259   extract the amplitude envelope and calculated event-related (de)synchronization (ERS/ERD)

260   coefficients by aligning the absolute Hilbert amplitude envelope values, averaged over 600 ms

261   after stimulus presentation to the averaged baseline amplitude calculated based on the - 400ms -

262   0ms pre-stimulus interval for a cluster of 44 frontal gradiometers (Fig. 4,A).

263   The resulting ERS coefficients were analyzed using the two-way ANOVA to probe whether the

264   increase in frontal theta amplitude differed between the conditions (*conflict* and *no-conflict*) and

265   the two groups of participants *("conformists"* and *"non-conformists")*.

266   ***Source space time-frequency data analysis***

267   In order to localize sources of theta oscillatory activity, we first computed a spatial filter that

268   maximized the variance between the *conflict* and *no-conflict* conditions for each participant using

269   the Common Spatial Pattern (CSP) approach (for details see (Blankertz et al., 2008)).We used

270   only magnetometer data filtered in the theta frequency band. Tikhonov regularization with

271   parameter lambda = 0.001 was applied to condition the covariance inverse matrix. We then

272   applied MNE localization procedure to the CSP's corresponding to the two largest eigenvalues

273   of the generalized covariance matrix. Similarly to the ERF analysis, we projected individual

274   MNE solutions obtained for the a.m. CSP components on the default anatomy for grand

275   averaging.

276   **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

302 the MEG experiment (F (1, 14) = 5.88, p = 0.0188), this observation was also supported by the

303 interaction *conflict direction* x *conflict size* x *level of conformity* (F(1,14) = 6.29, p = 0.015)

304 indicating that long-term conformity effects were observed only for large conflicts and only for

305 'conformists'. Thus, in our study group opinion effectively modulated individuals' judgments of

306 trustworthiness.

307

308 **ERF results**

309 *GFP analysis of conflict-related activity.*   To determine time-windows with the strongest

310 conflict-related activity (conflict trials vs. no-conflict trials, duration greater than 20 ms) we

311 analyzed evoked (GFP) activity for the magnetometers data using the permutation test (Fig. 2A).

312 The test signified three time-windows: 204-240 ms, 414-440 ms and 496-518 ms after the onset

313 of group ratings where the activity in conflict trials was stronger than the activity in no-conflict

314 trials. Similar results were obtained with analysis of the gradiometer data (data not shown).

315 *Source-space analysis of conflict-related activity.* To identify the brain regions generating

316 conflict-related activity we conducted a cluster-based permutation test in the MNE source space

317 around the time-windows identified by the GFP analysis (150-250 ms, 350-450ms and 450-550

318 ms) and compared the activations between *conflict* and *no-conflict* trials. The test revealed four

319 statistically significant clusters of activations (Fig 3A, Table 1) occurring only in the earliest

320 time-window (150-250 ms): the bilateral pMPFC, left anterior medial prefrontal cortex

321 (aMPFC), and left orbitofrontal cortex (left OFC). The right orbitofrontal cortex (right OFC)

322 cluster approached significance (mass = 366.0, cluster p-value = 0.063). The four clusters that

323 showed statistical significance were further used as ROI's for the between-group analysis

324 ('conformists' vs. 'non-conformists').

325 In the second (350-450ms), and third (450-550ms) time-windows no significant conflict-related

326 clusters were identified, however three clusters approached statistical significance (left pMPFC,

327   left precentral gyrus and frontal superior sulcus, see Table 1 for details).

328   **ROI analysis of the conformity-related activity**. To identify whether the conflict-related

329   activity differed between subjects with higher or lower tendency to conform to the group opinion

330   we compared the amplitudes of difference activations (conflict minus no-conflict trial) between

331   subjects that were assigned to the 'conformists' and 'non-conformists' within preselected ROIs.

332   The permutation test indicated a larger conflict-related activity (cluster p-value = 0.031) for

333   'non-conformists' than for 'conformists' in the right pMPFC (Fig.3B).

334   **Time-frequency analysis of *conflict- and conformity-related effects*.** The ERS analysis of the

335   induced oscillatory activity at the frontal sensors (Fig.4A.) revealed that both in *conflict* (mean

336   magnitude increase = 17.3 percent, SD = 10.9) and *no-conflict* (mean magnitude increase = 7.2

337   percent, SD = 6.7) trials, the magnitude of the frontal theta activity (4-8Hz) increased relative to

338   the prestimulus baseline (Fig. 4B). As expected, this increase was more pronounced for *conflict*

339   trials, than for *no-conflict* trials (mean magnitude difference = 10, SD. = 7.7). This observation

340   was supported by the two-way ANOVA performed for theta ERS coefficients with the *conflict*

341   (conflict vs. no-conflict trials) as a within-group factor, and the *level of conformity* (conformists

342   vs. non-conformist) as a between-group. The significant main effects of *conflict* ($F(1,19) = 5.84$,

343    $p = 0.022$) and *level of conformity* ($F(1,19) = 5.17$, $p = 0.03$) indicate that 'non-conformists'

344   showed a stronger increase in theta power than 'conformists'. The interaction *conflict* x *level of*

345   *conformity* was not significant ($p = 0.16$). Overall, we found a stronger synchronization of the

346   theta band activity at the frontal sensors for 'non-conformists' than for 'non-conformists'.

347   ***Source analysis* of *conflict--related theta band effects***. To investigate brain regions generating

348   the conflict-related induced theta activity we first computed the components accounting for most

349   variance between conflict and no-conflict trials in the theta frequency band. Resulting CSP

350   components explained on average 22.8% variance of the original data (SD = 6.25%, time-

351   window = 0-600 ms). We then used the MNE inverse solution to localize the sources of these

352 components. The localization of CSP components of the theta activity indicated multiple frontal

353 sources of activations including the bilateral MPFC (ventral and dorsal), OFC, bilateral temporal

354 poles, lingual gyrus and left dprsolateral prefrontal cortex (DLPFC) (Fig. 4C).

355

356 **Discussion**

357 The goal of this study was to elucidate temporal and spatial characteristics of electromagnetic

358 brain responses to the visual cues signaling that participants' individual ratings of face

359 trustworthiness conflict with the group opinion. The GFP analysis revealed a stronger activation

360 around 204-240 ms in *conflict* trials where individual ratings differed from the group rating than

361 in *no-conflict* trials where individual ratings matched with the normative group ratings. The

362 timing of the activation in our study corresponds to the timing of the FRN, an evoked-response

363 component often associated with reward prediction error processing and adjustments of behavior

364 (Holroyd and Coles, 2002; Cohen et al., 2011a; Walsh and Anderson, 2012). The source analysis

365 revealed significant conflict-related activations in the bilateral pMPFC and left OFC.

366 Interestingly, the activation in the right pMFPC overlapped with the results of the previous fMRI

367 study (Klucharev et al., 2009) and occurred approximately within the same time-window as

368 conflict-related evoked responses reported in the previous EEG studies of social conformity (e.g.

369 Shestakova te al., 2012).

370 We also found that the conflict with the group opinion triggers an induced frontal theta activity

371 that has been implicated into performance monitoring and reinforcement learning (Cavanagh et

372 al., 2010, 2012; Cohen et al., 2011b; van de Vijver et al., 2011). Frontal theta dynamics observed

373 in our study was consistent with corresponding evoked responses: we observed a more

374 pronounced increase in the induced theta power in *conflict* as compared to *no-conflict* trials.

375 These results are in line with several recent studies suggesting that FRN can be a time-domain

376 manifestation of theta oscillatory perturbations in the MPFC (Cohen et al., 2008; Cavanagh et al.,

2010; van de Vijver et al., 2011). 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

403   BOLD and MEG signals (Winterer et al., 2007; Singh, 2012)

404   Our MEG results support an important role of the pMPFC in the brain mechanisms of social

405   influence. The activity of the ventral striatum also has been previously associated with social

406   conformity (Klucharev et al., 2009; Berns et al., 2010; Campbell-Meiklejohn et al., 2010; Zaki et

407   al., 2011; Izuma, 2013b). Since the ventral striatum was implicated in reinforcement learning

408   (Delgado, 2007; O'Doherty et al., 2007; Niv, 2009), this region can contribute to the automatic

409   learning mechanism underling conformity, while the activity of the pMPFC may underlie a

410   distinct cognitively demanding process of a 'control state' evaluating whether to follow the

411   group opinion or stick to the initial opinion. This hypothesis is consistent with recent accounts of

412   the pMPFC role highlighting its function in cost computation, representing best alternative

413   action and executive control functions (Folstein and Van Petten, 2008; Rushworth et al., 2012;

414   Shenhav et al., 2013). Since the pMPFC is a part of the "default mode network", the greater

415   magnitude of biomagnetic conflict-related responses in the 'non-conformist' group may indicate

416   the allocation of larger cognitive resources during the conflict with group norms as compared to

417   the 'conformist group', where a more general reinforcement learning-like process engaging the

418   ventral striatum could dominates over cortical activity of the pMPFC.

419   Our results suggest that a perceived conflict with a normative group opinion triggers the pMPFC

420   activity similar to the FRN. To date, few studies have investigated MEG signatures of the FRN

421   (Doñamayor et al., 2012; Thomas et al., 2013), closely-related error-related negativity (ERN)

422   (Keil et al., 2010; Charles et al., 2013) and model-derived prediction error signal (Talmi et al.,

423   2012). Although only partly consistent to each other, these studies registered the error and

424   feedback-related activity at the ACC and MPFC (Thomas et al. 2013 Keil et al. 2010; Charles et

425   al. 2013; Thomas et al. 2013) as well as at the posterior cingular cortex (Doñamayor et al. 2012).

426   Differences in FRN localization in various studies can be due to the fact that this component may

427   consist of two or more subcomponents. Previous EEG studies suggested that the FRN is

428   produced by a superposition of two separate (negative and positive) subcomponents associated

429   with feedback processing (Holroyd et al., 2008; Baker and Holroyd, 2011; Krigolson et al.,

430   2014). The negative subcomponent ('N2' component in EEG literature) is associated with

431   processing of conflicting or unexpected events including unexpected negative and positive

432   feedbacks. The N2 subcomponent has been also associated with high-order conflict processing

433   and control state (Folstein and Van Petten, 2008) or with a state of enhanced control redirecting

434   future behavior away from tasks and strategies that result in a conflict (Botvinick, 2007; Fritz

435   and Dreisbach, 2013). While the second positive subcomponent ('reward positivity') is elicited

436   specifically by unexpected positive events along with N2 and cancels out the negative

437   subcomponent, resulting in a more positive waveform in EEG (Holroyd et al., 2008; Krigolson et

438   al., 2014), but, presumably, not in MEG due to different source cancellation profiles (Irimia et

439   al., 2012). Our results suggest that the observed MEG activity of the pMPFC may represent only

440   the negative FRN subcomponent. Overall, we can speculate that the observed pMPFC activations

441   may represent the processing of the conflict with descriptive social norms in terms of a self-

442   control (a 'lose-stay', but not 'lose-switch' strategy), a process complimentary to reinforcement

443   learning-related activity of the ventral striatum.

444   Alternatively, the weaker conflict-related activity of the pMPFC in 'conformists' group could be

445   associated with a relatively weaker individual opinion about the trustworthiness of the faces

446   resulting in a lower internal conflict (and smaller pMPFC activity) evoked by observed mismatch

447   with group ratings. Consistent with this interpretation of our findings, Achtziger and colleagues

448   demonstrated that people who tend to overweight new information display lower evoked activity

449   upon receiving one that conflicts with prior information (Achtziger et al., 2014). Additional

450   studies are clearly needed to clarify the function of the pMPFC activity during social conformity.

451   Our results also show the conflict-related activity in the OFC following the conflicts with the

452   group opinion. The conflict-related activity in the OFC did not differ between conformists and

453   non-conformists, probably reflecting a more general decision making (e.g. valuation) process.

454   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

Achtziger A, Alós-Ferrer C, Hügelschäfer S, Steinhauser M (2014) The neural basis of belief updating and rational decision making. Soc Cogn Affect Neurosci 9:55–62 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22956673 [Accessed February 20, 2014].

Aluja A, Kuhlman M, Zuckerman M (2010) Development of the Zuckerman-Kuhlman-Aluja Personality Questionnaire (ZKA-PQ): a factor/facet version of the Zuckerman-Kuhlman Personality Questionnaire (ZKPQ). J Pers Assess 92:416–431.

Baker TE, Holroyd CB (2011) Dissociated roles of the anterior cingulate cortex in reward and conflict processing as revealed by the feedback error-related negativity and N200. Biol Psychol 87:25–34 Available at: http://www.ncbi.nlm.nih.gov/pubmed/21295109 [Accessed

479   June 1, 2013].

480   Berns GS, Capra CM, Moore S, Noussair C (2010) Neural mechanisms of the influence of

481   popularity on adolescent ratings of music. Neuroimage 49:2687–2696 Available at:

482   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2818406&tool=pmcentrez&ren

483   dertype=abstract [Accessed February 3, 2014].

484   Blankertz B, Tomioka R, Lemm S, Kawanabe M, Müller K (2008) Optimizing Spatial Filters for

485   Robust EEG Single-Trial Analysis. XX:1–12.

486   Botvinick MM (2007) Conflict monitoring and decision making: reconciling two perspectives on

487   anterior cingulate function. Cogn Affect Behav Neurosci 7:356–366 Available at:

488   http://www.ncbi.nlm.nih.gov/pubmed/18189009 [Accessed February 19, 2014].

489   Campbell-Meiklejohn DK, Bach DR, Roepstorff A, Dolan RJ, Frith CD (2010) How the opinion

490   of others affects our valuation of objects. Curr Biol 20:1165–1170 Available at:

491   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2908235&tool=pmcentrez&ren

492   dertype=abstract [Accessed January 23, 2014].

493   Cavanagh JF, Frank MJ, Klein TJ, Allen JJB (2010) Frontal theta links prediction errors to

494   behavioral adaptation in reinforcement learning. Neuroimage 49:3198–3209 Available at:

495   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2818688&tool=pmcentrez&ren

496   dertype=abstract [Accessed November 5, 2012].

497   Cavanagh JF, Zambrano-Vazquez L, Allen JJB (2012) Theta lingua franca: a common mid-

498   frontal substrate for action monitoring processes. Psychophysiology 49:220–238 Available

499   at:

500   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3262926&tool=pmcentrez&ren

501   dertype=abstract [Accessed March 9, 2014].

502   Charles L, Van Opstal F, Marti S, Dehaene S (2013) Distinct brain mechanisms for conscious

503   versus subliminal error detection. Neuroimage 73:80–94 Available at:

504   http://www.ncbi.nlm.nih.gov/pubmed/23380166 [Accessed January 21, 2014].

505   Chen MK, Risen JL (2010) How choice affects and reflects preferences: revisiting the free-

506   choice paradigm. J Pers Soc Psychol 99:573–594 Available at:

507   http://www.ncbi.nlm.nih.gov/pubmed/20658837 [Accessed February 24, 2014].

508   Cialdini RB, Goldstein NJ (2004) Social influence: compliance and conformity. Annu Rev

509   Psychol 55:591–621 Available at: http://www.ncbi.nlm.nih.gov/pubmed/14744228

510   [Accessed January 21, 2014].

511   Cohen MX, Elger CE, Ranganath C (2007) Reward expectation modulates feedback-related

512   negativity and EEG spectra. Neuroimage 35:968–978 Available at:

513   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1868547&tool=pmcentrez&ren

514   dertype=abstract [Accessed November 4, 2012].

515   Cohen MX, Ranganath C (2007) Reinforcement learning signals predict future decisions. J

516   Neurosci 27:371–378 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17215398

517   [Accessed May 29, 2013].

518   Cohen MX, Ridderinkhof KR, Haupt S, Elger CE, Fell J (2008) Medial frontal cortex and

519   response conflict: evidence from human intracranial EEG and medial frontal cortex lesion.

520   Brain Res 1238:127–142 Available at: http://www.ncbi.nlm.nih.gov/pubmed/18760262

521   [Accessed March 8, 2014].

522   Cohen MX, Wilmes K, Vijver I Van De (2011a) Cortical electrophysiological network dynamics

523   of feedback learning. Trends Cogn Sci 15:558–566 Available at:

524   http://www.ncbi.nlm.nih.gov/pubmed/22078930 [Accessed November 5, 2012].

525   Cohen MX, Wilmes K, Vijver I Van De (2011b) Cortical electrophysiological network dynamics

526   of feedback learning. Trends Cogn Sci 15:558–566 Available at:

527   http://www.ncbi.nlm.nih.gov/pubmed/22078930 [Accessed January 21, 2014].

528   Delgado MR (2007) Reward-related responses in the human striatum. Ann N Y Acad Sci

529   1104:70–88 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17344522 [Accessed July

530   11, 2014].

531   Doñamayor N, Schoenfeld MA, Münte TF (2012) Magneto- and electroencephalographic

532   manifestations of reward anticipation and delivery. Neuroimage 62:17–29 Available at:

533   http://www.ncbi.nlm.nih.gov/pubmed/22561022 [Accessed February 10, 2014].

534   Eysenck HJ, Eysenck SBG (1994) Manual for the Eysenck Personality Questionnaire. (EPQ-R

535   Adult):21 Available at:

536   http://books.google.com/books?id=dYQ8HAAACAAJ&dq=intitle:Manual+for+the+Eysenc

537   k+Personality+Inventory&hl=&cd=2&source=gbs_api\npapers3://publication/uuid/BEEE6

538   DC8-7B1A-47A6-8B66-3F028BFA6370.

539   Folstein JR, Van Petten C (2008) Influence of cognitive control and mismatch on the N2

540   component of the ERP: a review. Psychophysiology 45:152–170 Available at:

541   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2365910&tool=pmcentrez&ren

542   dertype=abstract [Accessed February 20, 2014].

543   Fritz J, Dreisbach G (2013) Conflicts as aversive signals: conflict priming increases negative

544   judgments for neutral stimuli. Cogn Affect Behav Neurosci 13:311–317 Available at:

545   http://www.ncbi.nlm.nih.gov/pubmed/23307475 [Accessed February 20, 2014].

546   Gosling SD, Rentfrow PJ, Swann WB (2003) A very brief measure of the Big-Five personality

547   domains. J Res Pers 37:504–528.

548 Gross J, Baillet S, Barnes GR, Henson RN, Hillebrand A, Jensen O, Jerbi K, Litvak V, Maess B,

549   Oostenveld R, Parkkonen L, Taylor JR, van Wassenhove V, Wibral M, Schoffelen J-M

550   (2013) Good practice for conducting and reporting MEG research. Neuroimage 65:349–363

551   Available at: http://www.ncbi.nlm.nih.gov/pubmed/23046981 [Accessed May 27, 2013].

552 Gu R, Huang YX, Luo YJ (2010) Anxiety and feedback negativity. Psychophysiology 47:961–

553   967.

554 Hajcak G, McDonald N, Simons RF (2003) Anxiety and error-related brain activity. Biol

555   Psychol 64:77–90.

556 Hämäläinen MS, Ilmoniemi RJ (1994) Interpreting magnetic fields of the brain: minimum norm

557   estimates. Med Biol Eng Comput 32:35–42 Available at:

558   http://www.ncbi.nlm.nih.gov/pubmed/8182960 [Accessed August 13, 2014].

559 Hoffmann S, Labrenz F, Themann M, Wascher E, Beste C (2013) Crosslinking EEG time-

560   frequency decomposition and fMRI in error monitoring. Brain Struct Funct Available at:

561   http://www.ncbi.nlm.nih.gov/pubmed/23443964 [Accessed January 23, 2014].

562 Holroyd CB, Coles MGH (2002) The neural basis of human error processing: Reinforcement

563   learning, dopamine, and the error-related negativity. Psychol Rev 109:679–709 Available

564   at: http://doi.apa.org/getdoi.cfm?doi=10.1037/0033-295X.109.4.679 [Accessed November

565   5, 2012].

566 Holroyd CB, Pakzad-Vaezi KL, Krigolson OE (2008) The feedback correct-related positivity:

567   sensitivity of the event-related brain potential to unexpected positive feedback.

568   Psychophysiology 45:688–697 Available at:

569   http://www.ncbi.nlm.nih.gov/pubmed/18513364 [Accessed January 25, 2014].

570 Huang MX, Mosher JC, Leahy RM (1999) A sensor-weighted overlapping-sphere head model

571   and exhaustive head model comparison for MEG. Phys Med Biol 44:423–440 Available at:

572   http://www.ncbi.nlm.nih.gov/pubmed/10070792.

573   Irimia A, Van Horn JD, Halgren E (2012) Source cancellation profiles of

574   electroencephalography and magnetoencephalography. Neuroimage 59:2464–2474

575   Available at:

576   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3254784&tool=pmcentrez&ren

577   dertype=abstract [Accessed January 23, 2014].

578   Izuma K (2013a) The neural basis of social influence and attitude change. Curr Opin Neurobiol

579   23:456–462 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23608704 [Accessed

580   September 18, 2013].

581   Izuma K (2013b) The neural basis of social influence and attitude change. Curr Opin Neurobiol

582   23:456–462 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23608704 [Accessed June

583   8, 2013].

584   Izuma K, Adolphs R (2013) Social manipulation of preference in the human brain. Neuron

585   78:563–573 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23664619.

586   Keil J, Weisz N, Paul-Jordanov I, Wienbruch C (2010) Localization of the magnetic equivalent

587   of the ERN and induced oscillatory brain activity. Neuroimage 51:404–411 Available at:

588   http://www.ncbi.nlm.nih.gov/pubmed/20149884 [Accessed November 12, 2012].

589   Kim B-R, Liss A, Rao M, Singer Z, Compton RJ (2012) Social deviance activates the brain's

590   error-monitoring system. Cogn Affect Behav Neurosci 12:65–73 Available at:

591   http://www.ncbi.nlm.nih.gov/pubmed/22038705 [Accessed June 1, 2013].

592   Klucharev V, Hytönen K, Rijpkema M, Smidts A, Fernández G (2009) Reinforcement learning

593   signal predicts social conformity. Neuron 61:140–151 Available at:

594 http://www.ncbi.nlm.nih.gov/pubmed/19146819 [Accessed September 21, 2013].

595 Klucharev V, Munneke M a M, Smidts A, Fernández G (2011) Downregulation of the posterior

596 medial frontal cortex prevents social conformity. J Neurosci 31:11934–11940 Available at:

597 http://www.ncbi.nlm.nih.gov/pubmed/21849554 [Accessed January 23, 2014].

598 Krigolson OE, Hassall CD, Handy TC (2014) How We Learn to Make Decisions : Rapid

599 Propagation of Reinforcement Learning Prediction Errors in Humans. :1–10.

600 Lin F-H, Witzel T, Ahlfors SP, Stufflebeam SM, Belliveau JW, Hämäläinen MS (2006)

601 Assessing and improving the spatial accuracy in MEG source localization by depth-

602 weighted minimum-norm estimates. Neuroimage 31:160–171 Available at:

603 http://www.ncbi.nlm.nih.gov/pubmed/16520063 [Accessed September 28, 2013].

604 Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J

605 Neurosci Methods 164:177–190 Available at:

606 http://www.ncbi.nlm.nih.gov/pubmed/17517438 [Accessed September 20, 2013].

607 Mehrabian A (1997) Analysis of affiliation-related traits in terms of the PAD Temperament

608 Model. J Psychol 131:101–117.

609 Meltzer J a, Negishi M, Mayes LC, Constable RT (2007) Individual differences in EEG theta and

610 alpha dynamics during working memory correlate with fMRI responses across subjects.

611 Clin Neurophysiol 118:2419–2436 Available at:

612 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2080790&tool=pmcentrez&ren

613 dertype=abstract [Accessed January 21, 2014].

614 Nieuwenhuis S, Holroyd CB, Mol N, Coles MGH (2004) Reinforcement-related brain potentials

615 from medial frontal cortex: origins and functional significance. Neurosci Biobehav Rev

28:441–448 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15289008 [Accessed November 5, 2012].

Niv Y (2009) Reinforcement learning in the brain. J Math Psychol 53:139–154 Available at: http://www.sciencedirect.com/science/article/pii/S0022249608001181 [Accessed July 19, 2014].

O'Doherty JP, Hampton A, Kim H (2007) Model-based fMRI and its application to reward learning and decision making. Ann N Y Acad Sci 1104:35–53 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17416921 [Accessed July 15, 2014].

Raichle ME, Snyder AZ (2007) A default mode of brain function: a brief history of an evolving idea. Neuroimage 37:1083–90; discussion 1097–9 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17719799 [Accessed January 20, 2014].

Rotter JB (1966) Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr 80:1–28.

Rushworth MFS, Behrens TEJ, Rudebeck PH, Walton ME (2007a) Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. Trends Cogn Sci 11:168–176 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17337237 [Accessed January 27, 2014].

Rushworth MFS, Buckley MJ, Behrens TEJ, Walton ME, Bannerman DM (2007b) Functional organization of the medial frontal cortex. Curr Opin Neurobiol 17:220–227 Available at: http://www.ncbi.nlm.nih.gov/pubmed/17350820 [Accessed January 21, 2014].

Rushworth MFS, Kolling N, Sallet J, Mars RB (2012) Valuation and decision-making in frontal cortex: one or many serial or parallel systems? Curr Opin Neurobiol 22:946–955 Available at: http://www.sciencedirect.com/science/article/pii/S0959438812000694 [Accessed July

639   10, 2014].

640   Rushworth MFS, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011) Frontal cortex and

641   reward-guided learning and decision-making. Neuron 70:1054–1069 Available at:

642   http://www.ncbi.nlm.nih.gov/pubmed/21689594 [Accessed February 20, 2014].

643   Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative

644   theory of anterior cingulate cortex function. Neuron 79:217–240 Available at:

645   http://www.sciencedirect.com/science/article/pii/S0896627313006077 [Accessed July 9,

646   2014].

647   Shestakova A, Rieskamp J, Tugin S, Ossadtchi A, Krutitskaya J, Klucharev V (2013)

648   Electrophysiological precursors of social conformity. Soc Cogn Affect Neurosci 8:756–763

649   Available at: http://www.ncbi.nlm.nih.gov/pubmed/22683703 [Accessed February 25,

650   2014].

651   Singh KD (2012) Which "neural activity" do you mean? fMRI, MEG, oscillations and

652   neurotransmitters. Neuroimage 62:1121–1130 Available at:

653   http://www.ncbi.nlm.nih.gov/pubmed/22248578 [Accessed May 22, 2013].

654   Spielberger CD, Gorsuch RL, Lushene RE (1970) The State-Trait Anxiety Inventory.

655   MANUAL:1–23.

656   Tadel F, Baillet S, Mosher JC, Pantazis D, Leahy RM (2011) Brainstorm: a user-friendly

657   application for MEG/EEG analysis. Comput Intell Neurosci 2011:879716 Available at:

658   http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3090754&tool=pmcentrez&ren

659   dertype=abstract [Accessed September 16, 2013].

660   Talmi D, Fuentemilla L, Litvak V, Duzel E, Dolan RJ (2012) An MEG signature corresponding

661   to an axiomatic model of reward prediction error. Neuroimage 59:635–645 Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3200436&tool=pmcentrez&ren

dertype=abstract [Accessed November 5, 2012].

Taulu S, Hari R (2009) Removal of magnetoencephalographic artifacts with temporal signal-

space separation: demonstration with single-trial auditory-evoked responses. Hum Brain

Mapp 30:1524–1534 Available at: http://www.ncbi.nlm.nih.gov/pubmed/18661502

[Accessed October 7, 2013].

Thomas J, Vanni-Mercier G, Dreher J-C (2013) Neural dynamics of reward probability coding: a

Magnetoencephalographic study in humans. Front Neurosci 7:214 Available at:

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3831091&tool=pmcentrez&ren

dertype=abstract [Accessed January 23, 2014].

Van de Vijver I, Ridderinkhof KR, Cohen MX (2011) Frontal oscillatory dynamics predict

feedback learning and action adjustment. J Cogn Neurosci 23:4106–4121 Available at:

http://www.ncbi.nlm.nih.gov/pubmed/21812570.

Walsh MM, Anderson JR (2012) Learning from experience: event-related potential correlates of

reward processing, neural adaptation, and behavioral choice. Neurosci Biobehav Rev

36:1870–1884 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22683741 [Accessed

November 4, 2012].

Walton ME, Behrens TEJ, Noonan MP, Rushworth MFS (2011) Giving credit where credit is

due: orbitofrontal cortex and valuation in an uncertain world. Ann N Y Acad Sci 1239:14–

24 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22145871 [Accessed February 3,

2014].

Winterer G, Carver FW, Musso F, Mattay V, Weinberger DR, Coppola R (2007) Complex

relationship between BOLD signal and synchronization/desynchronization of human brain

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

686   http://www.ncbi.nlm.nih.gov/pubmed/17133396 [Accessed January 23, 2014].

687 Zaki J, Schirmer J, Mitchell JP (2011) Social influence modulates the neural computation of

688   value. Psychol Sci 22:894–900 Available at:

689   http://www.ncbi.nlm.nih.gov/pubmed/21653908 [Accessed January 21, 2014].

690
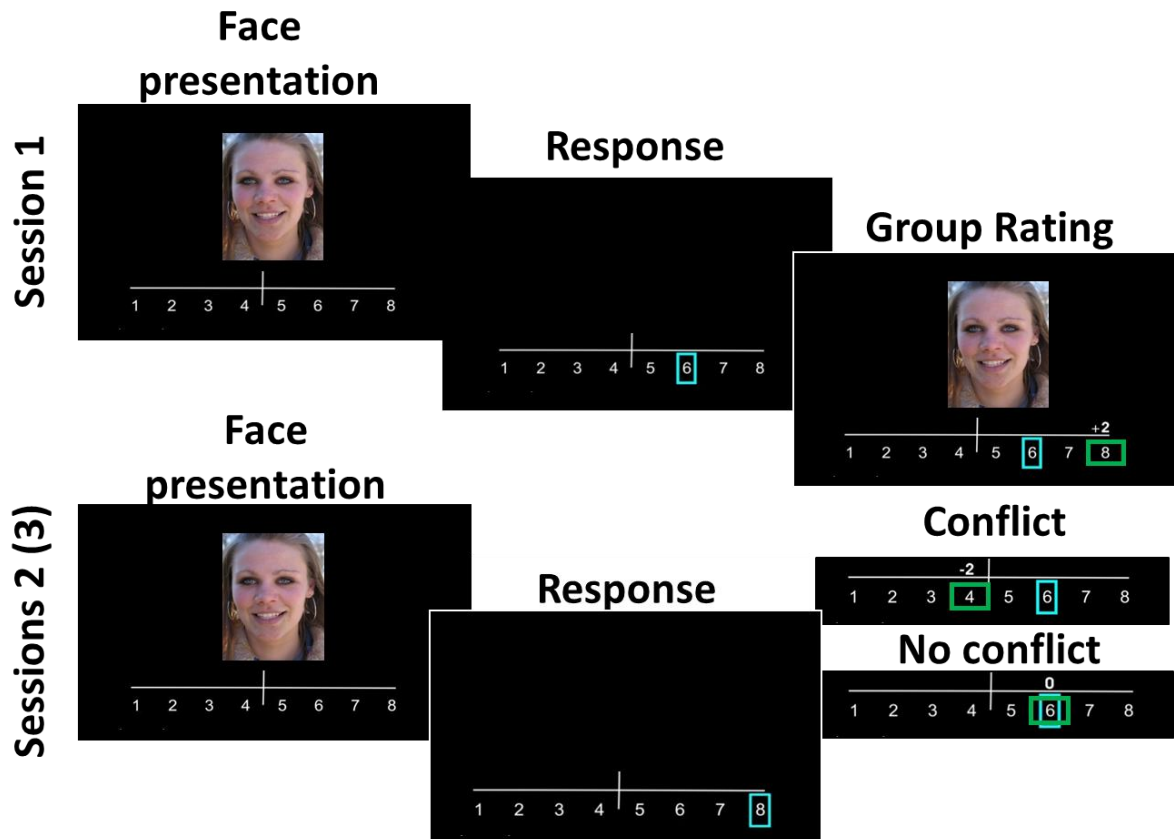
Figure 1. Behavioral task and behavioral results. **A.** The sequence of the events within a trial is shown. During the first session, subjects rated the trustworthiness of female faces and were subsequently presented with the group ratings that could be similar (no-conflict with group ratings), below 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.
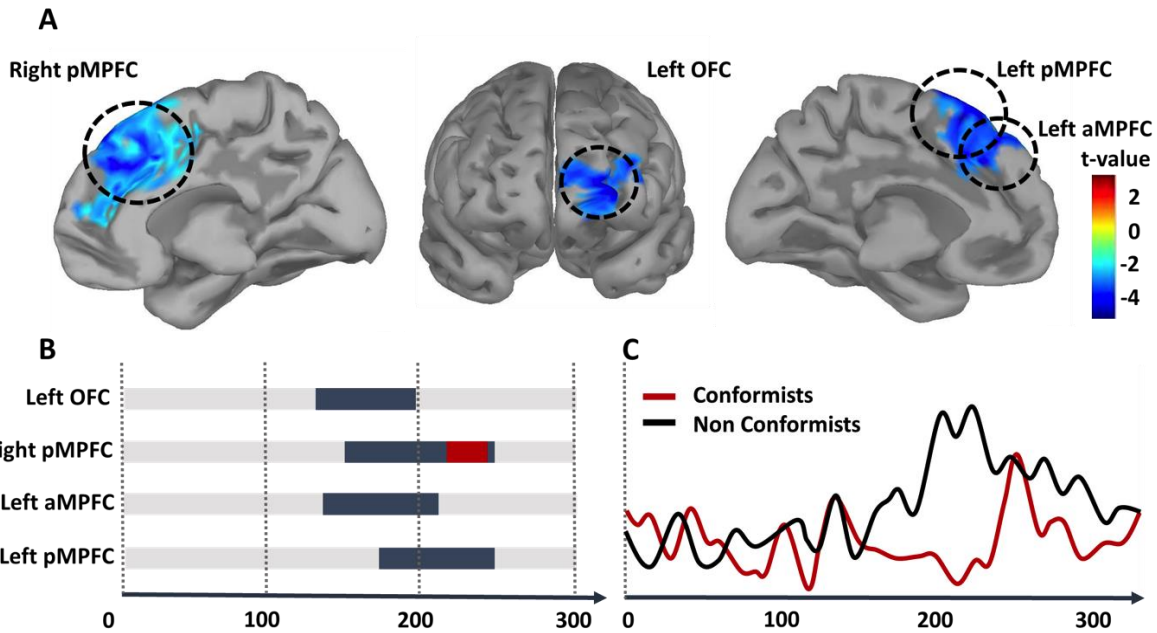
698

699
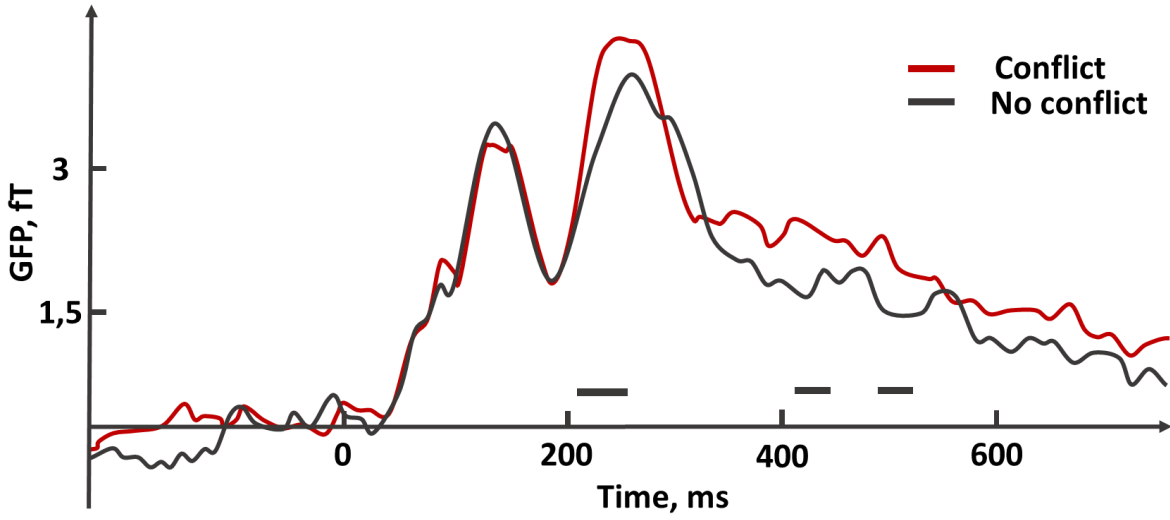
Figure 2. A. The GFP-RMS representation of conflict-related effects (magnetometers only). Underlined areas represent time-windows, where statistically significant differences of conflict and no-conflict trials lasted for more than 20 ms.

703

Figure 3. Conflict-related effects (conflict trials vs. no-conflict trails). **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).
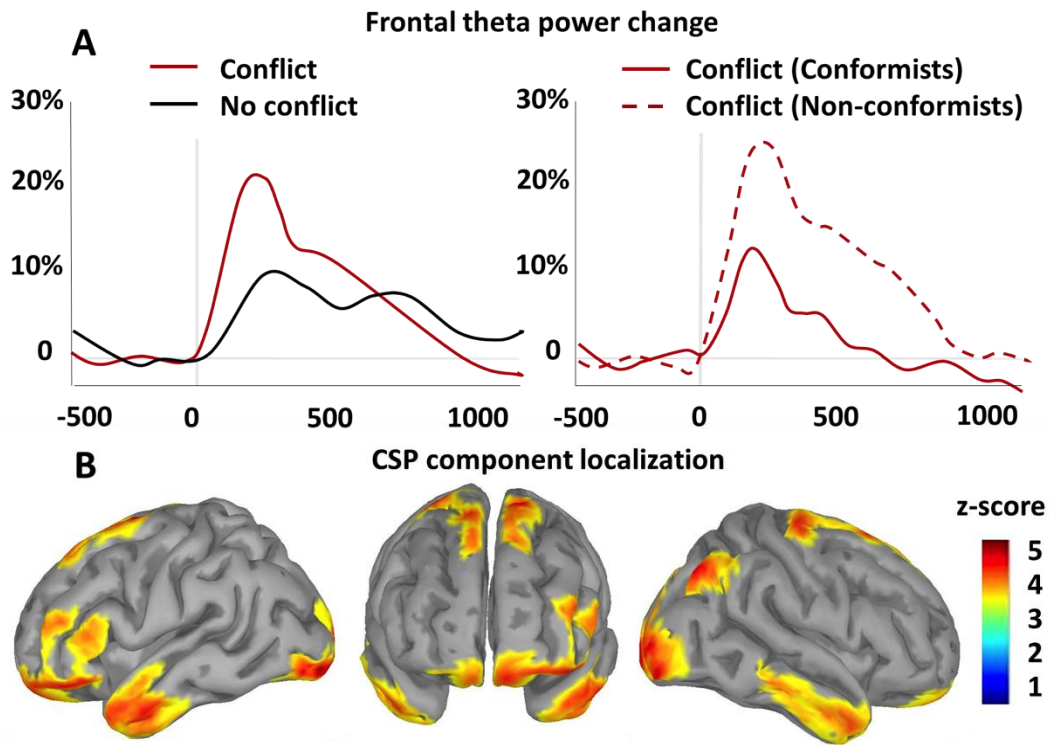
Figure 4. Analysis of induced theta activity. **A.** Grand averagedEvent-Related Synchronization (ERS) of theta activity in conflict and no-conflict trials (left) and in conflict trials between 'conformist' and 'non-conformist' groups (right) **B.** Source localization for the CSP components of induced theta conflict-related activity (conflict vs no-conflict trials) for all subjects.

719   Table1. The localization of the conflict-related activity (conflict trials vs. no-conflict trials) .

GFP Time - Window	L/R	Structure	Cluster time window, ms	Number of vertices (unique)	Cluster mass, (t-score)	Cluster p-value , FWER
150-250 ms	**Right**	**MPFC**	**152-250**	**246**	**-491**	**0.042***
	Right	OFC	150-194	194	366	0.063
	Left	**OFC**	**158-196**	**140**	**-288,9**	**0.019***
	Left	**aMPFC**	**160-218**	**74**	**-179,4**	**0.048***
	Left	**pMPFC**	**186-248**	**157**	**-314.1**	**0.016***
350-500 ms	Left	pMPFC	414-442	99	212	0.082
	Left	Precentral Gyrus.	438-482	79	210	0.083
450-550 ms	Left	Frontal Supperior Sulc.	482-532	94	291	0.053

720

721   Table 2 Time-windows of the conformity-related activity (within ROI determined during the

722   localization of the conflict-related activity)

L\R	ROI	Cluster time window, ms	Cluster mass (t-score)	Cluster p-value

				(FWER)
Conflict-No-conflict (difference waveform)				
Left	OFC	210-224	-20.9	0.11
Right	**pMPFC**	**216-240**	**-43**	**0.031***
Left	aMPFC	308-316	-19.25	0.14

723