The brain's temporal dynamics from a collective decision to individual action
Charpentier, C.J., Moutsiana, C., Garrett, N. and Sharot, T.

The article was originally published by the Society for Neuroscience, in Journal of Neuroscience, 34 (17), pp. 5816-5823, 2014 and is available at:

https://dx.doi.org/10.1523/JNEUROSCI.4107-13.2014

Copyright © 2014 Charpentier et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: (http://westminsterresearch.wmin.ac.uk/).

In case of abuse or copyright appearing without permission e-mail repository@westminster.ac.uk
Social animals constantly make decisions together. What determines if individuals will subsequently adjust their behavior to align with collective choices? Here, using functional magnetic resonance imaging in humans, we characterize a novel temporal model of brain response from the time a collective decision is made to the time an individual action is required. We reveal that whether a behavioral modification will occur is determined not necessarily by the brain’s response to the initial social influence, but by how that response (specifically in the orbitofrontal cortex; OFC) is mirrored at a later time when the individual selects their own action. This result suggests that the OFC may reconstitute an initial state of collective influence when individual action is subsequently needed. Importantly, these dynamics vary across individuals as a function of trait conformity and mediate the relationship between this personality characteristic and behavioral adjustment toward the group.

**Key words:** collective behavior; decision-making; fMRI; OFC; orbitofrontal cortex; social influence

**Introduction**

Our society depends heavily on collective decision-making to function. From choosing a business strategy to selecting a family dinner, decisions are frequently made together. Other social animals also exhibit collective decision-making, including bees selecting a nest site, birds choosing a travel route, and primates deciding when to move after a rest period (Black, 1988; Seeley et al., 1991; Kirman, 1993; Seeley and Buhrman, 1999; Conradt and Roper, 2003). Group decisions inevitably produce conflict-of-interest when the preference of an individual is different from that of the majority. Yet, members may follow the group’s choice rather than exhibit defiance; foraging at an unpreferred site or deciding to move with the herd despite requiring additional rest. Such social influence can be long lasting, evident at a later time when the individual is faced with the same choice again in a nongroup setting (e.g., deciding alone on a dinner option or business strategy).

What determines whether an individual will subsequently follow the group’s decision? The propensity to change one’s action and follow the group is likely determined by neural responses at two points in time: when a collective decision is revealed (Phase 1), and later when an individual selects an independent action (Phase 2). Yet, how these responses combine to predict behavior is unknown.

Studies examining the brain’s response to social influence at Phase 1 have revealed activation in a large network of regions related to subsequent alteration in subjective preferences (Klucharev et al., 2009; Berns et al., 2010; Campbell-Meiklejohn et al., 2010; Izuma and Adolphs, 2013; for review, see Izuma, 2013), perceptual decisions (Stallen et al., 2013), economic decisions (Burke et al., 2010), and memories (Edelson et al., 2011). These include the insula, ventral striatum, anterior and posterior cingulate cortices, prefrontal and inferior frontal cortices, temporoparietal junction, amygdala, and hippocampus. Less is known regarding the brain’s response during Phase 2, when an individual selects their own action following social influence. In a study examining socially induced changes in preferences for faces, the ventral striatum and orbitofrontal cortex (OFC) were found to mediate the effect of social influence at this stage (Zaki et al., 2011). The role of the OFC, which is involved both in value learning and social cognition (Kringelbach, 2005; Padoa-Schioppa and Assad, 2006; Rushworth et al., 2007; Elliott et al., 2010; Walton et al., 2011; Meshi et al., 2012), was further strengthened by a study associating gray matter volume in the lateral OFC with the tendency to follow the opinion of others (Campbell-Meiklejohn et al., 2012).

To date it is unknown how neural response to learning about the preference of others is related to neural response when an individual needs to select their own action, and how the combination of these responses determines an individual’s choice. Here, we provide a novel mechanistic account that considers the temporal dynamics of the decision-making process.

**Materials and Methods**

**Participants**

Ninety-eight subjects (51 females, 47 males, age range 18–56 years, mean 26.14) were recruited through an advertisement on a University College London (UCL) website. Twenty of the subjects performed the task in an MRI.
Procedure

Overview. Participants arrived at the lab in groups of five (two groups were run with only three participants). They spent approximately half an hour together filling in forms and learning about the task. Then one participant completed the task inside the MRI scanner while the other four completed it in separate computer booths. The experiment consisted of four sessions. In Session 1 (rating session) subjects were presented with 120 items and asked how much they would like to receive the item at the end of the study. These ratings were used to pair stimuli in the decision-making task. Participants then completed a collective decision-making task in which 60 pairs of items were presented and each subject voted for their preferred item. The “majority vote” was then revealed with an asterisk (*). Subsequently, subjects rated all items again and had the opportunity to make individual choices between all pairs of items.

Stimuli. Stimuli consisted of 120 pictures of various food and drink items. All pictures were resized to a resolution of 300 × 300 pixels and were displayed on a black background. Ratings from Session 1 were used to pair items in Session 2. Specifically, item pairing was determined by a Matlab script that maximized rating difference between the two items of each pair. Participants were thus presented (in Session 2) with as many pairs as possible of one item they rated high and one item they rated low (in Session 1). On average, 79% (±18% SD) of pairs had a rating difference of two or more points between the two items, therefore making most choices relatively “easy.”

This was done to maximize conflict of interest induced when the collective decision-making task except that subjects made decisions in an independent context, without a group choice. A star appeared next to the participant’s initial vote and half were consistent with the majority such that half were inconsistent with the participant’s initial vote and half were consistent. A fixation cross was then presented for a jittered duration (1–3 s).

Individual decision-making task. In Session 4, participants were presented with all 60 pairs again in a randomized order and asked to make choices between pairs once more (Fig. 1E). The design was similar to the collective decision-making task except that subjects made decisions in an independent context, without a group choice. A star appeared next to the product chosen by the participant for 1 s after their response, followed by a fixation cross (1–4 s jittered). The critical question was whether at this stage they reversed their initial choices to go along with the group.

Outcome and incentive structure. Participants were instructed that at the end of the experiment, one trial would be chosen by a MATLAB function, and this “incentive trial” would determine the item they would receive as a reward. Before each decision block, they received additional instructions regarding how their choices in that block would influence this reward if the incentive trial were to be selected from that block. In Session 2, the collective decision-making trials were divided into two blocks of 30 choices (block order counterbalanced across subjects) that differed only in the instructions about how choices in each block would impact on the item given as a reward at the end. Specifically, before one collective decision block participants were told they would receive the item selected by the majority vote should the incentive trial be picked from that block. In the other collective decision block and in the individual decision block they were told they would receive the item they voted for.

Figure 1. Experimental design. A, Participants came to the lab in groups of five and spent ~30 min together filling in forms and learning about the task. Then one participant completed the task inside the MRI scanner while the other four completed it in separate computer booths. The experiment consisted of four sessions. B, In Session 1 (rating session) subjects were presented with 120 items and asked how much they would like to receive the item at the end of the study. These ratings were used to pair stimuli in the decision-making task. C, Participants then completed a collective decision-making task in which 60 pairs of items were presented and each subject voted for their preferred item. The “majority vote” was then revealed with an asterisk (*). Subsequently, subjects rated all items again and had the opportunity to make individual choices between all pairs of items.

D, Session 3: Second Rating. Participants received instructions regarding each part of the study at the beginning of that session and did not know in advance what would be required in subsequent sessions.

E, Session 4: Individual Decision-making task. On each trial, a picture of a food/drink item was presented for 3 s. Then participants indicated how much they would like to receive each item at the end of the experiment on a scale from (1) “Not at all” to (6) “Extremely.” Participants were encouraged to give accurate ratings and were told their ratings would influence the item they would receive at the end of the study (for more details, see Outcome and incentive structure). They had up to 2 s to respond using a keyboard or MRI-compatible response box. To prevent lateralized motor activations, half of the scanned participants gave ratings 1–3 with their left hand and ratings 4–6 with their right hand, and the other half did the opposite. Finally, a fixation cross appeared for a jittered duration (1–4 s).

Collective decision-making task. In Session 2, participants made collective decisions. On each trial, they were presented with pairs of products from the first session for 2 s. When the word “CHOOSE” appeared, they had 2 s to indicate which of the two items they would rather consume at the end of the experiment (Fig. 1C). Participants were told that the other members of the group were making their own choice simultaneously. After a jittered delay of 1–3 s, during which they were led to believe that the computer was counting the votes of the five participants in the group, a star appeared for 3 s next to the item supposedly receiving the majority vote. Unbeknownst to all the participants in the group, these collective outcomes were in fact manipulated by a MATLAB program such that half were inconsistent with the participant’s initial vote and half were consistent. A fixation cross was then presented for a jittered duration (1–3 s).
for should the incentive trial be picked from that block. Results were not affected by this difference and were therefore collapsed across both blocks in Session 2. Before the rating sessions, participants were told their rating would influence their reward such that they were more likely to receive items they rated higher. At the end of the study, our MATLAB program selected the incentive trial from one of the choice blocks consistent with these instructions.

Conformity scale. Participants completed the Mehrabian Conformity Scale (Mehrabian and Stell, 1995). This scale measures people’s tendency to identify with and follow other people’s ideas, preferences, and behaviors rather than stand against them (Mehrabian and Stell, 1995). Statements include “I often rely on, and act upon, the advice of others” or “Generally, I’d rather give in and go along for the sake of peace than struggle to have my way.” These are rated from −4 (very strong disagreement) to +4 (very strong agreement). Five subjects failed to complete this scale. Participants were divided into two groups (each N = 46) using a median split (low conformers’ mean score: 8.20, high conformers’ mean score: 8.63, range: 9 to 16). Of the 20 fMRI subjects, 10 fell into each group.

Behavioral data analysis
Trials were classified into ones where the collective decision was consistent with the subject’s vote and ones where it was inconsistent. Trials in which participants’ vote contradicted their initial ratings (i.e., the item voted for by the subject was rated lower by him/her than the item rejected) were removed from all analyses, as these trials could not be reliably classified to either condition (mean number of trials excluded = 5.88 of 60).

Behavioral change following collective decisions was calculated as the proportion of trials in which participants reversed their choice in the last session following the collective decision. For each subject one score was calculated for trials in which the collective decision was consistent with the individual’s vote and one score for trials in which the collective decision was inconsistent.

To investigate whether participants altered their ratings following collective decisions, rating changes were calculated for each item and participant by subtracting the mean-corrected rating of Session 1 from the mean-corrected rating of Session 2 (i.e., difference scores). The mean-corrected score is the distance of a particular stimulus’ rating from the average rating for that participant and session (xi − μ) and indicates the value of the stimulus relative to all other stimuli in that session. Mean-correction is commonly used because ratings often decrease across sessions, due to the subject becoming bored and tired (Sharot et al., 2009, 2010). Thus, differences in mean-corrected scores across sessions indicate the change in preference of that stimulus relative to all other stimuli. For each pair, the rating change toward the group choice was then calculated using the equation below.

\[
\text{Rating change toward the group} = (\text{item selected by the group: } 2^{nd} - 1^{st} \text{ rating}) - (\text{item rejected by the group: } 2^{nd} - 1^{st} \text{ rating})
\]

To assess whether rating changes predicted choice reversals, we conducted a linear regression for each subject relating rating change (computed as explained above) to choice reversal (coded as a binary variable, 1 if choice was changed, 0 if not) across trials. Betas were then averaged across participants.

Choice reversal scores, rating change scores, and betas relating rating change and choice reversals were analyzed in a two-way (high formmers/low formmers × collective decision inconsistent/consistent with individual vote) ANOVA, followed by planned t tests.

MRI data acquisition
A Siemens Avanto 1.5T MRI scanner equipped with a 32-channel Siemens head coil was used. To correct for inhomogeneities of the static magnetic field we acquired field maps to be used in the unwarping stage of data preprocessing. Then five functional scanning sessions were acquired using a prescan normalized gradient echo-planar sequence with the following parameters: volume repetition time = 2.975 s, echo time = 50 ms, flip angle = 90°, matrix = 64 × 64, voxel size = 3 × 3 × 3 mm³, 35 axial slices sampled for whole brain coverage, tilt = −30°. A T1-weighted MPRAGE structural scan was acquired at the end of the experiment (176 sagittal slices, repetition time = 2.73 s, echo time = 3.57 ms, flip angle = 7°, matrix = 256 × 256, voxel size = 1 × 1 × 1 mm³). Subjects spent ~1 h in the scanner.

MRI data analysis
Preprocessing. MRI data analysis was performed using SPM8 software (Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm). The first six dummy volumes of each functional session were discarded from the analyses to allow for T1 equilibration effect. A field map was then created for each functional session using the FieldMap toolbox. Using this field map file for phase correction, images were realigned to the first functional volume of each session and unwarped. They were then spatially normalized to the standard Montreal Neurological Institute (MNII) EPI template, resampled into 2 × 2 × 2 mm³ voxels, and smoothed using an 8 mm³ full-width half-maximum Gaussian kernel.

Decision-making task: model and analysis. For each participant, blood oxygen level-dependent (BOLD) signal was modeled using a General Linear Model (GLM) including the following onsets: motor response onset, time of pair presentation (modeled as a 2 s window for consistent and inconsistent trials separately). In addition, in the collective decision-making task the time of collective decision presentation (i.e., when the star appears to indicate collective outcome) was modeled as a 3 s window for consistent and inconsistent trials separately (30 trials per condition). Six movement parameters were added as regressors of no interest.

ROI definition. We identified voxels where BOLD response during the collective decision outcome presentation was greater for outcomes that were inconsistent with the individual’s choice relative to outcomes that were consistent [p < 0.05, family wise error (FWE) corrected; threshold at p < 0.001 uncorrected, k ≥ 20 contiguous voxels]. Small volume correction (SVC) was applied to a priori anatomically defined regions of the OFC, previously shown to be involved in value learning, decision-making and/or social cognition (Kringelbach, 2005; Fadiga-Schioppa and Assad, 2006; Rushworth et al., 2007; Elliott et al., 2010; Walton et al., 2011; Zaki et al., 2011; Levy and Glűmchen, 2012). Using the Automated Anatomical Labeling (AAL) atlas in the SPM WufuPickAtlas toolbox (Tzourio-Mazoyer et al., 2002), six regions were defined: left and right medial OFC, left and right inferior lateral OFC (the most lateral subregion), and left and right superior lateral OFC that lies between the two. The three subdivisions of the right OFC are shown in Figure 3B.

Intertemporal analysis. This analysis investigated whether BOLD response to collective outcomes in regions identified in the collective decision-making task (Phase 1) was related to subsequent BOLD response when subjects made autonomous choices in the last session (Phase 2). For each participant we first extracted the average BOLD signal over all voxels in the emerging region above (i.e., cluster in the right inferior lateral OFC; see Fig. 3A) for each trial using the “spm_summarise.m” function. This was done using a new GLM model in which each trial was modeled separately, resulting in 60 β values per subject. These β values were then entered as a parametric modulator of the time subjects were presented with the pairs in Session 4 (individual choice), separately for consistent and inconsistent trials. The parameter estimates for this analysis were then extracted and averaged across all voxels in the anatomically defined subregions of the right OFC, as defined by the AAL atlas in the WufuPickAtlas toolbox (right inferior, superior, and medial OFC). The values were then entered into a two-way (high formmers/low formmers × consistent trials/inconsistent trials) ANOVA in SPSS. Using Bonferroni correction for the presence of three regions of interest (ROIs), resulting p values were threshold at p < 0.016. Across participants parameter estimates were correlated with the difference in behavioral change scores between inconsistent and consistent trials in the region where the above interaction was significant using SPSS.

Mediation analysis
We used linear regression models implemented in the PROCESS for SPSS macro (Model 4; Hayes, 2013) to investigate whether the effect of
trait conformity on choice reversals was mediated by the strength of coupling between right inferior OFC activity at time 1 and right superior OFC activity at time 2 (see Fig. 4). For each subject, three values were entered in the model: their trait conformity score on the Mehrabian Conformity Scale, their $\beta$ value representing the intertemporal coupling of OFC activity (mean parameter estimate extracted from intertemporal analysis, see paragraph above), and their total proportion of choice reversals for inconsistent versus consistent trials. The following effects were calculated by the model:

1. Relationship between initial predictor and mediator: OFC coupling = $\beta_{1.1} \times$ trait conformity + $\epsilon$.
2. Relationship between mediator and outcome: choice reversal = $\beta_{2.1} \times$ OFC coupling + $\beta_{2.2} \times$ trait conformity + $\epsilon$.
3. Indirect effect of interest (effect of trait conformity on choice reversal via OFC coupling strength); this indirect effect is defined as the product of the parameter estimates from (1) and (2); i.e., $\beta_{1.1} \times \beta_{2.1}$.
4. Direct effect: relationship between initial predictor and outcome after controlling for the mediator; i.e., equal to $\beta_{2.2}$ from (2).
5. Total effect: relationship between initial predictor and outcome without controlling for the mediator: choice reversal = $\beta_{3.1} \times$ trait conformity + $\epsilon$. The total effect is also equal to the sum of the direct and indirect effect; i.e., $\beta_{2.2} + \beta_{1.1} \times \beta_{2.1}$.

The significance of the indirect effect was estimated using the bias-corrected bootstrapping method of the PROCESS macro with the following parameters: $k = 10,000$ bootstrap repetitions, $\alpha = 95\%$ confidence interval (Hayes, 2009, 2013). This approach has been described as one of the more general and powerful methods for testing intervening variable effects, in that it makes no assumption about the shape of the sampling distribution or the significance of the other effects and requires no estimation of a standard error (Williams and MacKinnon, 2008; Hayes, 2009). The significance of all the other effects (relationship between initial predictor and mediator, between mediator and outcome, direct effect, and total effect) was estimated with classical linear regression models implemented in the PROCESS macro. A $t$ statistic and $p$ value were therefore obtained for each of these effects, whereas the significance of the indirect effect was determined by whether or not zero was included in the bootstrap confidence interval.

**Results**

**Collective decisions lead to behavioral modifications**

Findings indicated that collective decisions resulted in changes to individual choice behavior, the extent of which was strongly modulated by trait conformity (Fig. 2). The percentages of choice reversals following collective decisions (between Sessions 2 and 4) were entered into a two-way ANOVA with trait conformity as a between-subject factor (median split based on Mehrabian Conformity Scale scores: high trait conformity/low trait conformity) and agreement between the individual’s initial vote and the group’s decision as a within-subject factor (vote consistent with group choice/vote inconsistent with group choice). A $t$ statistic and $p$ value were therefore obtained for each of these effects, whereas the significance of the indirect effect was determined by whether or not zero was included in the bootstrap confidence interval.

For high conformers (subjects scoring high on the trait conformity questionnaire), the likelihood of choice reversals following a collective decision that was inconsistent with their own vote was greater than when it was consistent and greater than for low conformers. Choice reversals were trials in which subjects selected an item during Session 4 (individual decision-making task) that was different from the item they voted for during Session 2 (collective decision-making task). The two-way (high/low conformers × consistent/inconsistent collective choices) interaction was significant ($p < 0.05$, two-tailed $t$ test; error bars indicate SEM.

**Figure 2.** Collective decisions induce behavioral change. For high conformers (subjects scoring high on the trait conformity questionnaire), the likelihood of choice reversals following a collective decision that was inconsistent with their own vote was greater than when it was consistent and greater than for low conformers. Choice reversals were trials in which subjects selected an item during Session 4 (individual decision-making task) that was different from the item they voted for during Session 2 (collective decision-making task). The two-way (high/low conformers × consistent/inconsistent collective choices) interaction was significant ($p < 0.05$, two-tailed $t$ test; error bars indicate SEM.

Brain response to collective outcome (Phase 1)

How do collective decisions lead to such rapid change in choice behavior? The starting point in this process is likely to be the individual’s biological response to learning that a collective deci-
tion does not align with their own vote. Thus, we first searched for brain regions where BOLD response during collective outcome presentation was greater when that outcome was inconsistent with a participant’s vote relative to when it was consistent. A significant effect was observed in the right inferior lateral OFC (peak voxel MNI coordinates: 50, 30, −10; \( T = 4.06; p < 0.05 \text{ FWE, SVC in anatomically defined OFC; threshold at } p < 0.001 \text{ corrected, } k \geq 20 \text{ contiguous voxels; Fig. 3A). Note that the effect expanded into the adjacent inferior frontal gyrus (IFG), though these voxels did not survive small-volume correction in anatomically defined IFG. No other region exhibited a significant effect and neither did it predict subsequent choice reversal. Rather, as we evidence below, the extent to which this signal was related to subsequent OFC activity when participants had the opportunity to reverse their choices was both modulated by trait conformity and predicted choice reversal.

**Intertemporal relationship between OFC response to collective outcome (Phase 1) and individual choice (Phase 2) predicted behavioral modification**

To examine the relationship between the OFC’s response across the two time points, to collective decisions and to individual choice, we used an intertemporal correlation analysis. This analysis asks whether the response of a brain region to a specific stimulus/event can predict subsequent brain activity to the identical stimulus/event, but under different circumstances. In this case, the question is whether OFC response to learning that the group selected a banana over an apple (for example) predicts subsequent OFC response when the subject is faced with an autonomous choice between an apple and a banana, and importantly, how this intertemporal relationship differs across conditions and individuals.

In the first step of the analysis, parameter estimates are extracted for each unique pair of items (e.g., apple–banana) during the time collective outcomes are revealed (such \( \beta \) extraction has been used in different contexts including multivariate studies, functional connectivity, and mediation studies; for previous examples, see Rissman et al., 2004; Bonnici et al., 2012; Chadwick et al., 2011; Mumford et al., 2012). These are then averaged over the voxels that were identified above in the right OFC, resulting in 60 \( \beta \) values per subject (one for each collective choice outcome; see Materials and Methods for details). In the second step, these \( \beta \) values are entered as a parametric modulator of BOLD activity during the time participants have to make the same choice again in the last session (e.g., selecting individually between an apple and a banana). The resulting parameter estimates thus reflect the intertemporal correlation between BOLD response at time 1 (Session 2) and time 2 (Session 4). We can then examine how this intertemporal association varies across individuals and conditions.

Because subregions of the OFC have been suggested to have distinct roles in the different stages of decision-making (Kringelbach, 2005; Elliott et al., 2010; Walton et al., 2011), with more lateral regions important for learning from the consequence of a choice, and more medial parts necessary for selecting the significant information to make a choice (for review, see Walton et al., 2011), an ROI analysis was conducted in each of the three anatomical subregions of the right OFC. The AAL atlas (Tzourio-Mazoyer et al., 2002) defines the inferior lateral OFC, medial OFC, and superior OFC that lies between the two (Fig. 3B; see Materials and Methods for more details). The intertemporal parameter estimates resulting from the analysis mentioned above were extracted and averaged across all voxels in each of these three anatomical ROIs separately for consistent and inconsistent trials. Intertemporal parameter estimates were then entered into a two-way (high trait conformity/low trait conformity \( \times \) collective outcomes consistent with initial vote/collective outcomes inconsistent with initial vote) ANOVA in SPSS. A significant interaction was observed in the anatomical right superior OFC only \( F(1,18) = 8.62, p < 0.01; \) surviving Bonferroni correction for three subregions; Fig. 3C). The interaction was characterized by an absence of relationship between past and present OFC BOLD response in both groups on trials in which the collective outcome was consistent with the individual’s vote, but a strong
positive intertemporal correlation in high conformers ($t_{19} = 2.4, p < 0.05$) and a trend for a negative intertemporal correlation in low conformers ($t_{19} = 2.0, p = 0.07$) when the outcome was inconsistent.

Across all participants the exact nature of this coupling predicted behavioral change (Fig. 3D). The greater the positive correlation between (1) past response to inconsistent collective outcomes in the right (inferior lateral) OFC, and (2) subsequent right superior OFC response when selecting an individual action, the more likely participants were to reverse their choices to align with the group's ($r = 0.54, p < 0.02$; Fig. 3D).

The initial response to inconsistent collective outcomes across participants in the right OFC may represent encoding or learning of this information. What predicts behavioral change is not this signal itself, but the relation of this signal to subsequent activity in the OFC when making an individual decision, which may represent subsequent use of this information or reconstitution of the initial state of collective influence. Importantly, this intertemporal relationship between past and subsequent OFC response was observed only in the inconsistent choice condition and not in the consistent choice condition. Thus, it cannot be interpreted as a simple response to the stimuli pair itself as in both conditions the stimuli pairs were presented, but only in the former condition did we find the effect.

**Mediation analysis: OFC temporal dynamics mediate the relationship between trait conformity and choice reversal**

We have shown that right OFC response to inconsistent collective outcomes predicted subsequent activity in an adjacent right OFC region during individual decision-making in high, but not low, conformers, and that across all subjects the strength of this intertemporal coupling was positively correlated with choice reversals. To formally determine whether the effect of trait conformity on choice reversal was mediated by this right OFC intertemporal coupling, we conducted a mediation analysis across subjects (Hayes, 2009, 2013), in which (1) conformity scores and (2) betas representing the coupling between OFC signal at time 1 and OFC signal at time 2 were included as predictors of choice reversals (Fig. 4). The effects were estimated using linear regression models (see Materials and Methods). The mediation model first confirmed that as reported above a significant relationship exists between conformity scores and intertemporal coupling strength of right OFC activity (unstandardized coefficient = 0.008, $t_{19} = 2.21, p < 0.04$), as well as between intertemporal coupling strength and choice reversals when controlling for conformity scores (unstandardized coefficient = 12.87, $t_{17} = 2.32, p < 0.03$). The mediation analysis then revealed that the direct effect of conformity scores on choice reversals, when controlling for right OFC intertemporal coupling strength, was not significant (unstandardized coefficient = 0.01, $t_{17} = 0.10, p > 0.9$). The indirect effect was estimated using a bias-corrected bootstrapping method. One of the major strengths of this approach, compared with more classical methods such as the causal steps approach (Baron and Kenny, 1986) or the Sobel test (Sobel, 1982), is that it allows testing intervening variable effects without making any assumption about the significance of the other effects or about the sampling distribution of the indirect effect, thus making it a more powerful, valid, and general method for testing indirect effects (Williams and MacKinnon, 2008; Hayes, 2009). Critically, the indirect effect of interest was significant (unstandardized effect = 0.11, bootstrap 10,000 repetitions, 95% confidence interval = 0.0076–0.352). This result suggests that the effect of trait conformity on behavior is mediated by the strength of temporal coupling in the right OFC.

Together, our findings demonstrate that behavioral changes induced by inconsistent collective decisions can be predicted by the temporal dynamics of OFC activity between the time of initial social influence and the time when individual action is needed. Individual differences in this process are tightly related to people's trait conformity.

**Discussion**

Decisions made in a group have significant consequences on the individual behavior of its members. The current study provides a temporal model of brain activity that describes when collective decisions lead to individuals changing their actions.

We first showed that a disagreement between a collective decision and an individual's preference engaged the right inferior OFC. Importantly, our findings suggest a dynamic interaction between the brain's response to such disagreement, and the subsequent tendency to change one's action accordingly. Specifically, we found that right inferior OFC response during the time of initial exposure to conflicting collective choice was coupled with right superior OFC response when an individual needs to select their own action. This was true only when the group's choice conflicted with the individual's preference, not when the two were compatible, and thus cannot be interpreted as a simple response to the stimuli set per-se. The greater this temporal coupling, the more likely an individual was to reverse their choice to align with the group. This suggests that an individual's initial response to a collective decision does not necessarily determine whether they will subsequently follow the group. Rather, the influence of a collective decision on individual action is determined by how this neural response is mirrored at a later time, when the individual needs to select their own action. This result suggests that the OFC may reconstitute an initial state of collective influence when individual action is subsequently needed, possibly reflecting how social information is used.

Critically, we found that this process varies across individuals as a function of trait conformity. People who scored high on a personality measure of conformity were more likely to reverse their choices to align with the group. Our mediation analysis revealed that this association between personality and behavior was mediated by the temporal coupling of the OFC. Once the OFC intertemporal coupling was accounted for, the direct relationship between trait conformity and choice reversal was no longer significant. These findings illustrate the process by which group choices are, or fail to be, instrumental in a person's
decision-making, and highlight the underlying source of individual differences in this domain.

The OFC has been implicated in valuation and decision-making, especially in instances where behavior must be adjusted rapidly in response to changes in the environment (Kringelbach, 2005; Elliott et al., 2010; Walton et al., 2011; Levy and Glimcher, 2012). Interestingly, the OFC has also been shown to play an important role in social cognition (Moll et al., 2006; Spitzer et al., 2007; Rilling et al., 2008; Cikara et al., 2010), with clinical studies showing that lesions to the OFC induce a disruption of appropriate social behavior (Hornak et al., 2003; Beer et al., 2006). Our findings tie the two sets of studies together, suggesting that the OFC plays a critical role in learning from social information and adjusting behavior in response. The extent of adjustment is predicted by a person's trait conformity. It has been shown that the regulating role of the OFC on behavior could be genetically mediated (Meyer-Lindenberg et al., 2004). An intriguing question is whether these genetic variations are related to the individual differences reported here.

Responses in the OFC to the different stages of the process were observed in different parts of the OFC. The response to inconsistent collective outcomes was observed in the right inferior lateral OFC, whereas choice reversal was associated with the right superior OFC (in a more medial part of lateral OFC). The spatial resolution of our data does not allow us to make claims regarding the specific roles of these subregions. However, we note that such distinctions are consistent with the notion that subregions of the OFC may have different roles in the decision-making process (Elliott et al., 2010; Walton et al., 2011). For example, it has been suggested that more lateral regions play a central role in learning from choice outcomes, and more medial parts play a role in selecting the relevant information to make a choice (Walton et al., 2011). Although the OFC has previously been reported to be involved in social influence (Zaki et al., 2011; Campbell-Meiklejohn et al., 2012; Meski et al., 2012), studies have also indicated a role for the pPFC (Klucharev et al., 2009; Berns et al., 2010; Izuma and Adolphs, 2013; for a review see Izuma, 2013). An open question is whether these regions mediate different functions in the process of social influence, such as mediating social influence for different types of stimuli and/or decisions.

Collective decision-making appears a conserved strategy for determining group action across social animals (Black, 1988; Seeley et al., 1991; Kirman, 1993; Seeley and Buhrman, 1999; Conradt and Roper, 2003). The prevalence of this solution, practiced by different cultures and species, may rely on individuals aligning themselves with the group, even when the group’s preference is not in accord with their own. This can be adaptive as it allows individuals to live and work smoothly in groups. Such benefits may have predisposed humans and other social animals to alter their behavior to align with collective decisions (Conradt and Roper, 2003; Danchin et al., 2004; Couzin, 2009). We show that these behavioral modulations, the neurobiological basis of which we describe here vary, however, across individuals. Thus, groups will include members that at times demonstrate alternative autonomous behavior, perhaps securing an adaptive social balance between group cohesion and progressive societal change.

References