Behavioral/Cognitive

The Brain's Temporal Dynamics from a Collective Decision to Individual Action

Caroline J. Charpentier,[1,2] **Christina Moutsiana,**[1] **Neil Garrett,**[1] **and Tali Sharot**[1]

[1]Affective Brain Lab, Department of Cognitive, Perceptual and Brain Sciences, Division of Psychology and Language Sciences, and [2]Institute of Cognitive Neuroscience, University College London (UCL), WC1N 3AR, London, United Kingdom

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 deviance; 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
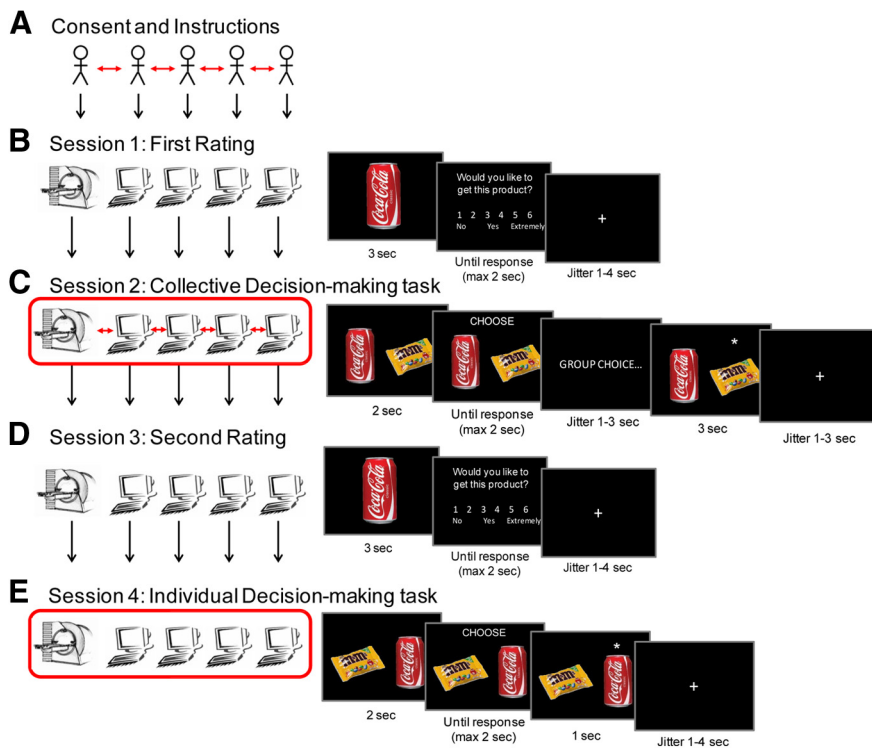
Figure 1. Experimental design. **A**, Participants came to the lab in groups of five and spent ~30 min together filling forms and learning about the task. Then one participant completed the task inside the MRI scanner while the other four completed it in separate computer booths. The experiment consisted 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 (**D**) and had the opportunity to make individual choices between all pairs of items (**E**).

scanner (10 females, 10 males, age range 21–33 years, mean 24.75) and the rest on separate desktop computers (data from one male subject was lost due to a computer crash). Participants were right-handed, had normal or corrected-to-normal vision, and no known history of previous psychological or neurological disorders. Because stimuli were food items, participants were required not to have any food allergies or restricted diet. To ensure familiarity with the items, participants were required to have lived in the UK for at least 6 months. Subjects gave written informed consent and were paid for their participation. The study was approved by the departmental ethics committee at the Birkbeck-UCL Centre for NeuroImaging.

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 (Fig. 1A). One participant of each group then performed the task inside the MRI scanner, whereas the remaining participants completed it in separate booths in front of computer screens. There were no confederates. Participants were told that all computers (including fMRI stimuli presentation) were synchronized via the university network to enable them to perform the task simultaneously as a group. The experimental paradigm consisted of four parts of 10–15 min each (Fig. 1B–E); (1) an initial rating task, (2) a collective decision-making task, (3) a second rating task (identical to initial rating task), and (4) an individual decision-making task. 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.

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.

Rating task. In sessions 1 and 3 (i.e., before and after the collective decision-making task) participants rated all 120 food/drink items (Fig.

1 B,D). 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).

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 did not align with the subjects' vote.

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 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 Stefl, 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 Stefl, 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: -19.22 ± 7.92, range: -40 to -10; high conformers' mean score: -0.63 ± 6.31, 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 ($x_i - \mu$) 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.

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 conformers/low conformers × 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 mm^3, 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 = 224 × 256, voxel size = 1 × 1 × 1 mm^3). Subjects spent $\sim$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 (MNI) EPI template, resampled into 2 × 2 × 2 mm^3 voxels, and smoothed using an 8 mm^3 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, $k \geq 20$ 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; Padoa-Schioppa and Assad, 2006; Rushworth et al., 2007; Elliott et al., 2010; Walton et al., 2011; Zaki et al., 2011; Levy and Glimcher, 2012). Using the Automated Anatomical Labeling (AAL) atlas in the SPM WfuPickAtlas 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 over all voxels in the anatomically defined subregions of the right OFC, as defined by the AAL atlas in the WfuPickAtlas toolbox (right inferior, superior, and medial OFC). The values were then entered into a two-way (high conformers/low conformers × 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 β 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 = **beta1.1** × trait conformity + ε.
(2) Relationship between mediator and outcome: choice reversal = **beta2.1** × OFC coupling + beta2.2 × trait conformity + ε.
(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., **beta1.1 × beta2.1**.
(4) Direct effect: relationship between initial predictor and outcome after controlling for the mediator ; i.e., equal to **beta2.2** from (2).
(5) Total effect: relationship between initial predictor and outcome without controlling for the mediator: choice reversal = **beta3.1** × trait conformity + ε. The total effect is also equal to the sum of the direct and indirect effect; i.e., **beta2.2 + beta1.1 × beta2.1**.

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 collective decision/vote inconsistent with collective decision). This revealed a main effect of the latter ($F_{(1,90)} = 4.27, p < 0.05$), such that when faced with a group decision that was inconsistent with their own initial choice, participants were more likely to change their subsequent choice to align with the groups' than when the collective decision was consistent with their initial vote. The effect interacted with trait conformity ($F_{(1,90)} = 5.01, p < 0.03$), such that it was observed only in high trait conformers ($t_{(45)} = 2.71, p < 0.01$), but not in low trait conformers ($t_{(45)} = 0.14, p > 0.8$).

Consistent with previous studies (Klucharev et al., 2009; Berns et al., 2010; Izuma and Adolphs, 2013), participants altered their subjective value ratings following collective decisions. Specifically, entering rating change toward the group choice into a two-way (trait conformity: high/low × consistency with group choice: initial subject preference consistent/
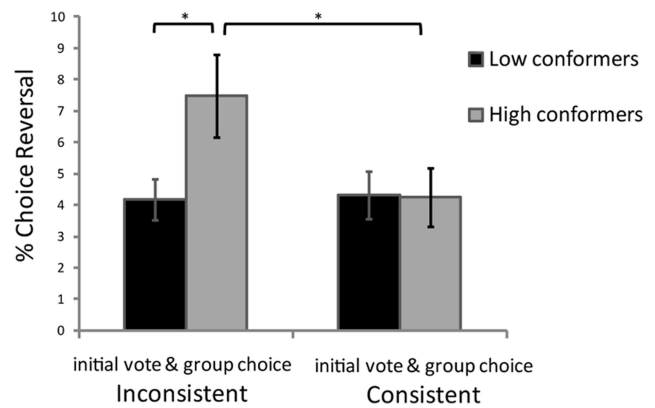

Figure 2. Collective decisions induce behavioral change. For high conformers (subjects scoring high on the trait conformity questionnaire), the likelihood of choice reversals following a collective decision that was inconsistent with their own vote was greater than when it was consistent and greater 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.

inconsistent with group choice) ANOVA revealed a main effect of group choice ($F_{(1,90)} = 28, p < 0.001$) such that participants were more likely to revise their ratings toward the group when the group choice was inconsistent with the individual's original preference than when it was already aligned with the subject's original preference. For inconsistent trials ratings both increased for items the group selected ($t_{(91)} = 4.85, p < 0.001$), and decreased for items the group rejected ($t_{(91)} = -6.23, p < 0.001$). There was no main effect of trait conformity ($F_{(1,90)} = 0.06, p > 0.8$) nor an interaction between consistency and trait conformity ($F_{(1,90)} = 2.5, p > 0.1$).

However, what was modulated by trait conformity was how well changes in ratings predicted choice reversals. Specifically, for each subject we conducted a linear regression across trials to test how well choice reversal was predicted by rating changes toward the group's choice. Entering the resulting β values into an ANOVA as above revealed a main effect of consistency with group choice ($F_{(1,90)} = 34.99, p < 0.001$). Specifically, rating changes positively predicted choice reversals toward the group when the subject's initial choice was inconsistent with the group ($t_{(91)} = 4.44, p < 0.001$), and did so significantly more than when it was already consistent. There was also a main effect of trait conformity ($F_{(1,90)} = 4.32, p < 0.05$) such that rating changes toward the group were more likely to predict choice reversals toward the group in high conformers than low conformers.

Together, our results suggest that decision values change under group influence and that this change is related to subsequent choice reversal to align with the group.

The proportion of choice reversals ranged from 0 to 20.8% in low conformers, and from 0 to 35% in high conformers. Because items were paired such that one item was originally rated low by the subject and the other rated high, a reversal indicates that following group choice, subjects select an item originally rated low by him/her over one rated high. Thus reversals, though not always frequent, suggest a large shift in preference.

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-

sion 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; $T = 4.06$; $p < 0.05$ FWE, SVC in anatomically defined OFC; threshold at $p < 0.001$ uncorrected, $k \geq 20$ 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 the reverse contrast. On its own, the right OFC response was not modulated by trait conformity, nor 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 β 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., 2012; Mumford et al., 2012). These are then averaged over the voxels that were identified above in the right OFC, resulting in 60 β values per subject (one for each collective choice outcome; see Materials and Methods for details). In the second step, these β 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
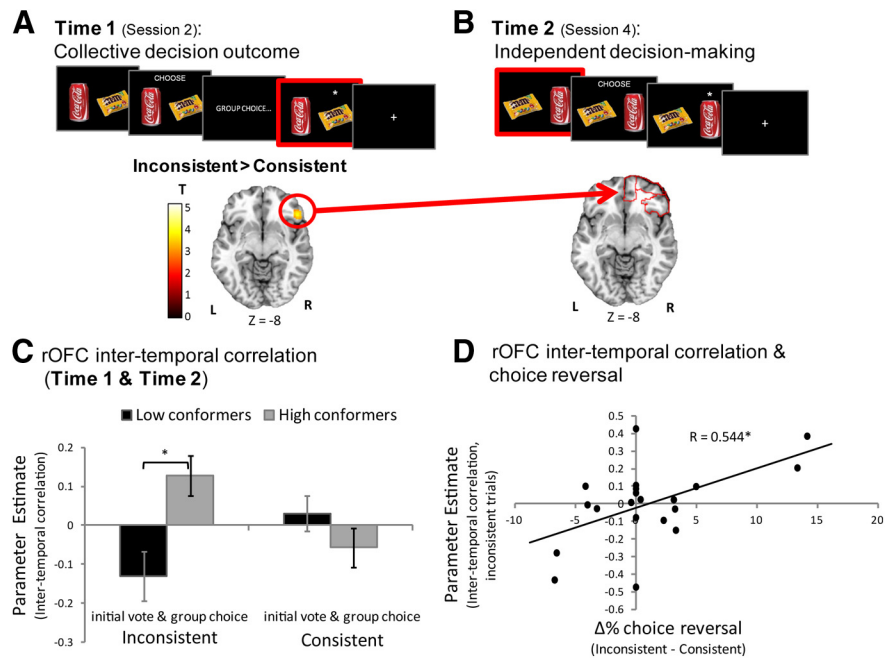

Figure 3. OFC response to a collective decision predicts subsequent response when making an individual choice. **A**, Enhanced BOLD response was observed in the right OFC when subjects learnt that the collective decision was inconsistent with their own vote, relative to when it was consistent ($p < 0.05$ FWE, anatomical SVC; displayed at $p < 0.005$, uncorrected). **B**, BOLD signal was extracted from this region for each specific trial and added as a parametric modulator predicting BOLD signal when an individual decision was to be made regarding the same pair at a later time. Parameter estimates were then averaged over the anatomically defined subregions of the right OFC as defined by the AAL atlas. For illustration purposes, the red frames in the trial design in **A** and **B** denotes the time of the trial when BOLD response was analyzed. **C**, A significant interaction emerged in the right superior OFC. Past OFC response to collective decision positively predicted subsequent OFC signal when making an individual selection between a pair of items for which subjects' initial vote was inconsistent with the collective decision in high conformers, a trend for a negative correlation emerged for low conformers, and no correlation emerged when the initial vote was consistent with group outcome. **D**, Across participants, the relationship between past and subsequent OFC response predicted behavioral change (i.e., choice reversals). Specifically, the greater the positive correlation between these signals on inconsistent trials, the more likely an individual was to reverse their choices to align with collective decisions that were initially inconsistent with their own. *$p < 0.05$, two-tailed t test; error bars indicate SEM.

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 × 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
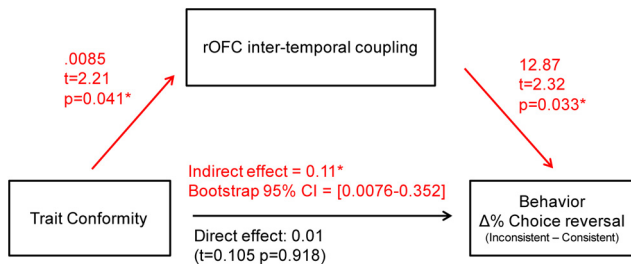
Figure 4. Mediation model: OFC temporal dynamics mediate the relationship between trait conformity and behavior. Values represent unstandardized coefficients for each effect of interest, with corresponding t statistics and p values or bootstrapping confidence interval (for indirect effect) to assess significance. Across subjects, the indirect effect of trait conformity on choice reversals, mediated by the strength of intertemporal coupling in the right OFC (red path) was significant.

positive intertemporal correlation in high conformers ($t_{(9)} = 2.4$, $p < 0.05$) and a trend for a negative intertemporal correlation in low conformers ($t_{(9)} = 2.0$, $p = 0.07$) 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, $t_{(18)} = 2.21$, $p < 0.04$), 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 illuminate 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., 2005). 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; Meshi et al., 2012), studies have also indicated a role for the mPFC (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

Baron RM, Kenny DA (1986) The moderator mediator variable distinction in social psychological-research- conceptual, strategic, and statistical considerations. J Pers Soc Psychol 51:1173–1182. CrossRef Medline

Beer JS, John OP, Scabini D, Knight RT (2006) Orbitofrontal cortex and social behavior: integrating self-monitoring and emotion-cognition interactions. J Cogn Neurosci 18:871–879. CrossRef Medline

Berns GS, Capra CM, Moore S, Noussair C (2010) Neural mechanisms of the influence of popularity on adolescent ratings of music. Neuroimage 49:2687–2696. CrossRef Medline

Black JM (1988) Preflight signaling in swans: a mechanism for group cohesion and flock formation. Ethology 79:143–157.

Bonnici HM, Chadwick MJ, Kumaran D, Hassabis D, Weiskopf N, Maguire EA (2012) Multi-voxel pattern analysis in human hippocampal subfields. Front Hum Neurosci 6:290. CrossRef Medline

Burke CJ, Tobler PN, Schultz W, Baddeley M (2010) Striatal BOLD response reflects the impact of herd information on financial decisions. Front Hum Neurosci 4:48. CrossRef Medline

Campbell-Meiklejohn DK, Bach DR, Roepstorff A, Dolan RJ, Frith CD (2010) How the opinion of others affects our valuation of objects. Curr Biol 20:1165–1170. CrossRef Medline

Campbell-Meiklejohn DK, Kanai R, Bahrami B, Bach DR, Dolan RJ, Roepstorff A, Frith CD (2012) Structure of orbitofrontal cortex predicts social influence. Curr Biol 22:R123–R124. CrossRef Medline

Chadwick MJ, Bonnici HM, Maguire EA (2012) Decoding information in the human hippocampus: a user's guide. Neuropsychologia 50:3107–3121. CrossRef Medline

Cikara M, Farnsworth RA, Harris LT, Fiske ST (2010) On the wrong side of the trolley track: neural correlates of relative social valuation. Soc Cogn Affect Neurosci 5:404–413. CrossRef Medline

Conradt L, Roper TJ (2003) Group decision-making in animals. Nature 421:155–158. CrossRef Medline

Couzin ID (2009) Collective cognition in animal groups. Trends Cogn Sci 13:36–43. CrossRef Medline

Danchin E, Giraldeau LA, Valone TJ, Wagner RH (2004) Public information: from nosy neighbors to cultural evolution. Science 305:487–491. CrossRef Medline

Edelson M, Sharot T, Dolan RJ, Dudai Y (2011) Following the crowd: brain substrates of long-term memory conformity. Science 333:108–111. CrossRef Medline

Elliott R, Agnew Z, Deakin JF (2010) Hedonic and informational functions of the human orbitofrontal cortex. Cereb Cortex 20:198–204. CrossRef Medline

Hayes AF (2009) Beyond Baron and Kenny: statistical mediation analysis in the new millennium. Commun Monogr 76:408–420. CrossRef

Hayes AF (2013) Introduction to mediation, moderation, and conditional process analysis: a regression-based approach. New York: Guilford.

Hornak J, Bramham J, Rolls ET, Morris RG, O'Doherty J, Bullock PR, Polkey CE (2003) Changes in emotion after circumscribed surgical lesions of the orbitofrontal and cingulate cortices. Brain 126:1691–1712. CrossRef Medline

Izuma K (2013) The neural basis of social influence and attitude change. Curr Opin Neurobiol 23:456–462. CrossRef Medline

Izuma K, Adolphs R (2013) Social manipulation of preference in the human brain. Neuron 78:563–573. CrossRef Medline

Kirman A (1993) Ants, rationality, and recruitment. Q J Econ 108:137–156. CrossRef

Klucharev V, Hytönen K, Rijpkema M, Smidts A, Fernández G (2009) Reinforcement learning signal predicts social conformity. Neuron 61:140–151. CrossRef Medline

Kringelbach ML (2005) The human orbitofrontal cortex: linking reward to hedonic experience. Nat Rev Neurosci 6:691–702. CrossRef Medline

Levy DJ, Glimcher PW (2012) The root of all value: a neural common currency for choice. Curr Opin Neurobiol 22:1027–1038. CrossRef Medline

Mehrabian A, Stefl CA (1995) Basic temperament components of loneliness, shyness, and conformity. Soc Behav Personal 23:253–263. CrossRef

Meshi D, Biele G, Korn CW, Heekeren HR (2012) How expert advice influences decision making. PLoS One 7:e49748. CrossRef Medline

Meyer-Lindenberg A, Hariri AR, Munoz KE, Mervis CB, Mattay VS, Morris CA, Berman KF (2005) Neural correlates of genetically abnormal social cognition in Williams syndrome. Nat Neurosci 8:991–993. CrossRef Medline

Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, Grafman J (2006) Human fronto-mesolimbic networks guide decisions about charitable donation. Proc Natl Acad Sci U S A 103:15623–15628. CrossRef Medline

Mumford JA, Turner BO, Ashby FG, Poldrack RA (2012) Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage 59:2636–2643. CrossRef Medline

Padoa-Schioppa C, Assad JA (2006) Neurons in the orbitofrontal cortex encode economic value. Nature 441:223–226. CrossRef Medline

Rilling JK, Goldsmith DR, Glenn AL, Jairam MR, Elfenbein HA, Dagenais JE, Murdock CD, Pagnoni G (2008) The neural correlates of the affective response to unreciprocated cooperation. Neuropsychologia 46:1256–1266. CrossRef Medline

Rissman J, Gazzaley A, D'Esposito M (2004) Measuring functional connectivity during distinct stages of a cognitive task. Neuroimage 23:752–763. CrossRef Medline

Rushworth MF, Behrens TE, Rudebeck PH, Walton ME (2007) Contrasting roles for cingulate and orbitofrontal cortex in decisions and social behaviour. Trends Cogn Sci 11:168–176. CrossRef Medline

Seeley TD, Buhrman SC (1999) Group decision making in swarms of honey bees. Behav Ecol Sociobiol 45:19–31. CrossRef

Seeley TD, Camazine S, Sneyd J (1991) Collective decision-making in honey-bees: how colonies choose among nectar sources. Behav Ecol Sociobiol 28:277–290. CrossRef

Sharot T, De Martino B, Dolan RJ (2009) How choice reveals and shapes expected hedonic outcome. J Neurosci 29:3760–3765. CrossRef Medline

Sharot T, Velasquez CM, Dolan RJ (2010) Do decisions shape preference? Evidence from blind choice. Psychol Sci 21:1231–1235. CrossRef Medline

Sobel ME (1982) Asymptotic confidence intervals for indirect effects in structural equation models. In: Sociological methodology (Leinhardt S, ed), pp 290–312. San Francisco: Jossey-Bass.

Spitzer M, Fischbacher U, Herrnberger B, Grön G, Fehr E (2007) The neural signature of social norm compliance. Neuron 56:185–196. CrossRef Medline

Stallen M, Smidts A, Sanfey AG (2013) Peer influence: neural mechanisms underlying in-group conformity. Front Hum Neurosci 7:50. CrossRef Medline

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15:273–289. CrossRef Medline

Walton ME, Behrens TE, Noonan MP, Rushworth MF (2011) Giving credit where credit is due: orbitofrontal cortex and valuation in an uncertain world. Ann N Y Acad Sci 1239:14–24. CrossRef Medline

Williams J, MacKinnon DP (2008) Resampling and distribution of the product methods for testing indirect effects in complex models. Struct Equ Modeling 15:23–51. CrossRef Medline

Zaki J, Schirmer J, Mitchell JP (2011) Social influence modulates the neural computation of value. Psychol Sci 22:894–900. CrossRef Medline