RESEARCH ARTICLE

Private but not social information validity modulates social conformity bias

Li Li¹ | King King Li² | Jian Li¹

¹School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

²Shenzhen Audencia Business School, Shenzhen University, Shenzhen, China

Correspondence

Jian Li, School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China.
Email: leekin@gmail.com

Funding information

Ministry of Science and Technology of the People's Republic of China, Grant/Award Number: 2015CB559200; National Natural Science Foundation of China, Grant/Award Numbers: 31371019, 31421003, 31871140

Abstract

It is well established that human behaviors are susceptible to others' opinions. However, optimal decision theory mandates choices be made upon the estimated validities of different information sources and little is known about whether and how people could wean themselves off social conformity bias, especially when the social signals are uninformative. Here, we asked subjects to participate in a probabilistic urn guessing task based on their private information as well as observed choices from their partners. Specifically, we manipulated the information validity of these two sources such that only the private evidence was informative. Across trials, social conformity declined, manifested by the increased influence of the private evidence but steady effect of social information. Correspondingly, we found dorsomedial prefrontal cortex (dmPFC) was involved in detecting the conflict of private and social information and conforming to social signal whereas striatum was responsible for selectively updating the influence of private (but not social) evidence contingent on its inferred validity. Furthermore, functional coupling between striatum and dmPFC predicted the resistance toward the influence of social information. Together, these results may provide a mechanistic account of how the conformity bias toward uninformative social information can be remedied.

KEYWORDS

dACC, dmPFC, information validity, social conformity bias, social learning, striatum

1 | INTRODUCTION

Social conformity bias refers to the propensity to follow other people's decisions in the face of their own incongruent or even conflicting personal experience. This preference for social information has been demonstrated to permeate into a variety of cognitive processes such as economic behavior (Bikhchandani, Hirshleifer, & Welch, 1992; Burke, Tobler, Schultz, & Baddeley, 2010), social valuation and opinions heeding (Campbell-Meiklejohn, Bach, Roepstorff, Dolan, & Frith, 2010; Klucharev, Hytönen, Rijpkema, Smidts, & Fernández, 2009; Zaki, Schirmer, & Mitchell, 2011), belief formation (Anderson & Holt, 1997), memory (Edelson, 2011) as well as sensory perception (Asch, 1956). It has been generally believed that such bias is evolutionally adaptive and elicits social approval (Cialdini & Goldstein, 2004; Deutsch & Gerard, 1955; Montague & Lohrenz, 2007; Spitzer, Fischbacher, Herrnberger, Grön, & Fehr, 2007). Furthermore, according to statistical decision theory, taking into account of social signals might help to exploit the informational content and lead to more advantageous

choice selection, analogous to multisensory integration (Behrens, Hunt, Woolrich, & Rushworth, 2008; Burke, Tobler, Baddeley, & Schultz, 2010). Recently, an emerging body of neuroimaging research has started to reveal critical brain structures involved in conforming to social signals that included dorsomedial prefrontal cortex (dmPFC) (Behrens et al., 2008; Berns, Capra, Moore, & Noussair, 2010; Falk, Berkman, Mann, Harrison, & Lieberman, 2010; Klucharev, Munneke, & Smidts, 2011), dorsal anterior cingulate cortex (dACC), ventral striatum (VS), dorsolateral prefrontal cortex (dlPFC), and orbitofrontal cortex (Burke, Tobler, Baddeley, & Schultz, 2010; Campbell-Meiklejohn et al., 2010; Klucharev et al., 2009; Zaki et al., 2011).

On the other hand, the tendency to conform to social signals may backfire when they are not informative to guide choices, as has been vividly demonstrated in the herding behavior during the bubble burst periods of stock market as well as classic information cascade studies in the laboratory (Anderson & Holt, 1997; Banerjee, 1992; Camerer & Weigelt, 1991). However, the way in which social conformity bias can be dynamically updated remains unknown. Given the fact that humans

often possess both private and social evidence during decision making, it is therefore plausible that the relative reduction of social conformity bias might result from decreased effect of social evidence, elevated effect of private information, or both. Here, we provide participants with both private and social information and examine their choice behavior as well as neural activity measured by functional magnetic resonance imaging (fMRI) while the level of congruency between private and social signals is manipulated. It is likely that participants might update their reliance on private and social signals based on the feedback received.

Indeed, a vast array of previous literatures, notably in reinforcement learning, has uncovered the importance of feedback in adjusting action selection (Behrens, Woolrich, Walton, & Rushworth, 2007; Boorman, Behrens, Woolrich, & Rushworth, 2009; Dayan & Balleine, 2002). For example, in a counterfactual learning paradigm, monetary feedback conveys information that can be assigned to both chosen and unchosen options by updating either action values or policies (Boorman, Behrens, & Rushworth, 2011; Li & Daw, 2011; Lohrenz, McCabe, Camerer, & Montague, 2007). It has been shown that dopamine projected brain structures, such as striatum, might mediate such action value or policy update (Li & Daw, 2011; Lohrenz et al., 2007). We thus hypothesize that feedback may provide means to evaluate the validities of different sources of information and humans would dynamically adjust their behaviors according to the estimated validity. Specifically, in an environment where social evidence is completely uninformative, such behavior adjustment would nudge people to rely more on their own private evidence. Regarding social conformity, previous literatures consistently demonstrate the importance of dmPFC/dACC, which are heavily involved in conflicts monitoring and cognitive control (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Kragel et al., 2018). In social conformity setup where conflicts arise between private and social information, dmPFC shows greater activities and also predicts individual conformity behavior (De Martino, Bobadilla-Suarez, Nouguchi, Sharot, & Love, 2017; Klucharev et al., 2009). It is hypothesized that the neural signals in the dmPFC might reflect the deviation of private judgment from group opinions or social norms, which leads to the adjustment of individual behavior (Klucharev et al., 2009; Wu, Luo, & Feng, 2016). Furthermore, research using repetitive transcranial magnetic stimulation (rTMS) on the mPFC has established a causal link between the activity in this brain area and social conformity (Klucharev et al., 2011). Therefore, we assume that dmPFC activity might encode the discrepancy between private and social information and the tendency to conform to social information, under the influence of striatum activity, which conveys the information validity updating signal.

To test these hypotheses, we designed an urn guessing game where subjects had to decide which urn their randomly drawn bead came from, based on their personal experience and other players' predictions. After each trial, subjects received monetary feedback indicating whether their choice was correct or not. Importantly, we manipulated the validities of two sources of information such that decision accuracy based solely on private bead color information was two-thirds while the accuracy of other players' choices was 50% (see Section 2). Optimal decision theory predicts increased and decreased reliance on private and social evidence, respectively, as trial-and-error feedback started to better approximate the true underlying information validities.

However, it is also possible that susceptibility toward social information might convey evolutionary benefits and the mitigation of social conformity biases is solely manifested by the modulation of the decision weight on private evidence. Our regression analysis revealed that both private and social information affected subject's choices and social conformity bias decreased as trials proceeded. Importantly, such conformity bias mitigation was only due to the increasing decision weight on private information. In the meantime, we also examined participants' brain activity while they contemplated between choices, especially when their private information was in conflict with social signal. We found activities in the dmPFC (dACC) were involved in detecting the private versus social information conflict as well as conforming to social signals. Furthermore, we showed activities in the striatum that, in addition to encoding outcome valence, responded only to the validity of private (but not social) cues. Finally, we focused on brain areas of medial prefrontal cortex (mPFC) and striatum and their functional interactions since they were implicated in social conformity and policy/value update in previous research (Klucharev et al., 2009; Klucharev et al., 2011; Li & Daw, 2011).

2 | MATERIALS AND METHODS

2.1 | Participants

Thirty-five subjects (20 males) gave consent and took part in the fMRI study. All subjects were graduate or undergraduate students from Peking University. They were right-handed and had no history of neurological disorder with average age 22.1 ± 2.9 years. The experiment was conducted in accordance with the protocol approved by the Ethics Committee of School of Psychological and Cognitive Sciences, Peking University.

2.2 | Experimental design

We adopted a classic urn guessing game in which subjects had to guess which urn their randomly drawn bead came from (Anderson & Holt, 1997). There are two urns subjects could draw their bead from, one containing two red beads and one blue bead (i.e., *Red Urn*) and the other containing two blue beads and one red bead (*Blue Urn*). In each trial, one of the two urns was randomly selected by the computer. Each subject was told that he or she would play with up to two anonymous players in the game, and each player would in turn draw one bead from the selected urn, with replacement, and announce his or her prediction of the urn color (Figure 1a). When the subject assumes the role of Player 1, his or her decision would be solely based on the private information (bead color, whose color correctly predicts urn color with probability of 2/3). However, if he/she is the second or third player, his or her decision could be based on both the private information and the social signals, predictions made by preceding players (Figure 1b). Unknown to the subjects, however, we manipulated the validities of social information such that it was 50% correct throughout the task. Therefore, subject's optimal strategy is to choose based solely on the bead color of his or her private draw regardless of other players' predictions. The task has 192 trials covering all six

FIGURE 1 Task design. (a) The procedure of fMRI experiment. In half of the trials, social signals were presented first, whereas in the other half, private information was displayed first. (b) Example information screens for subjects assuming the roles of the first, second, or third player. Depending on subjects' roles and the congruency between private and social information, trials can be categorized into first player baseline (BAS), second player congruent (CON) and incongruent (INC), and third player congruent (CON), incongruent (INC) and mixed (MIX, only one piece of social information was congruent with the private signal) trials [Color figure can be viewed at wileyonlinelibrary.com]

possible conditions (Figure 1b) with 24 trials for the baseline condition, 36 trials each for two INCongruent conditions (second and third players), 24 trials each for two congruent conditions (second and third players), and 48 trials for the MIXed condition (third player). Due to random sampling noises, social information was 47 and 48% correct when subjects were Players 2 and 3, respectively. Also, social signals were not significantly correlated ($\phi = 0.093$, $\chi^2 = 0.926$, $p = 0.34$) when subjects took the role of Player 3. In half of the trials where subjects were Player 2 or 3, social signals were presented first, followed by private signal, whereas in the other half trials, the order was reversed to avoid potential sequence effects. All trials were equally divided into three sessions with a brief break in between and the sequence of the trials was randomized across subjects.

The experiment was implemented in Psychtoolbox 3 (<http://www.psychtoolbox.org/>). At the beginning of each trial, a fixation was displayed at the center of the computer screen for 0.5 s. Depending on subject's role, his or her private information and other players' decisions would appear sequentially on the screen. Once all the necessary information was on the screen, subjects had 4 s to decide.

After the decision, the selected choice was highlighted and remained on the screen until the decision time window (4 s) expired, after which was followed by the feedback (2 s). Subjects received ¥20 endowment before the experiment, and were rewarded ¥4 for each correct choice and punished ¥2 otherwise. The intertrial interval varied from 2 to 4 s (Figure 1a). At the end of the task, subjects' accumulated payoff was displayed and they were paid accordingly.

2.3 | Behavioral data analysis

All behavioral data analyses were conducted using R v3.2.2 (<https://www.r-project.org/>). The missing trials (on average $0.8 \pm 1.4\%$ of trials) were excluded before the data analysis. p-Values reported in the text were two-tailed.

2.4 | Mixed effects logistic regression model

We used the mixed effects logistic model to estimate the weights of private and social information on the behavioral choice via the R

FIGURE 2 The impact of social and private information on choice behavior. (a) The proportion of choices deviating from the private signal. (b) The impact of two sources of information on choice behavior across three sessions. The weights of private information gradually increased across sessions but the weights of social information remained unaltered. Socials 1 and 2 indicate the first and second social partner's decision, respectively. The error bars indicate SE. * $p < 0.05$. ** $p < 0.01$. *** $p < 0.001$ [Color figure can be viewed at wileyonlinelibrary.com]

package "lme4." In this model, the dependent variable was subject's choice of the red (1) or blue urn (0) for each trial. Independent variables included the private information (red bead coded 0.5, blue bead -0.5), first social prediction (red urn 0.5, blue urn -0.5), second social prediction (red urn 0.5, blue urn -0.5). In addition, we added the interaction term of two pieces of social information in the augmented model to test the synergistic effect as the social information accumulated. The coefficients of fixed effects and random effects across all subjects were estimated for all the regressors (Figure S1, Supporting Information).

To examine the dynamic effects of private and social information, we separately estimated their weights for each session. The regression coefficients of all the regressors and sessions were entered into an one-way repeated measure ANOVA for further analysis (Figure 2b).

To investigate the impact of information validity on the decision, we conducted another regression analysis that included the two sources of information at current trial (the private bead color and other players' urn predictions), accuracies of private and social information in the previous trial, and their interaction terms as the independent variables. Information accuracy was defined as whether the color of the bead or the urn prediction by other players was in accordance with the urn color in the previous trial. The coefficients of the interaction terms thus represented behavior weight change associated with the validities of previous information (Figure 4a and Table S5, Supporting Information). Since choice outcome (win/loss) and private information validity were correlated, we constructed a new model and confirmed the significance of the interaction term between private information and its previous validity by controlling the influence of the decision outcome (Figure S5, Supporting Information).

2.5 | fMRI data acquisition

The fMRI experiment was conducted in a Siemens Prisma 3T scanner. The blood oxygen level dependent (BOLD) signals were collected by the T2*-weighted echo-planar images (EPIs) sequence with repetition time (TR) of 2 s; echo time (TE) 25 ms; flip angle = 90°; matrix size = 64 × 64; and field-of-view = 224 mm². The in-plane resolution was

3.5 × 3.5 mm with a total of 33 slices and slice thickness = 3.5 mm, and no interslice gap. To reduce the fMRI signal dropout in the basal frontal and medial temporal regions, a tilted plane of acquisition method was used such that all slices were acquired 30° clockwise to the anterior commissure-posterior commissure axis and the phase encoding direction was set as from posterior to anterior. The experiment was split into three sessions with a brief break of ~60 s between sessions. Each session lasted for approximately 20 min, and the whole experiment lasted ~60 min. A T1-weighted MP-RAGE image was obtained with the following parameters: TR = 2,530 ms, TE = 2.98 ms, flip angle = 7°, matrix size = 256 × 256, field-of-view = 256 mm, 192 slices/stab, slice thickness = 1 mm, and no interslice gap. We also recorded the field map using a GRE sequence before obtaining the functional images and later used it to correct for magnetic field distortion.

2.6 | Image analysis

All image data analyses were implemented in SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm8/>). BOLD images were first slice timing corrected using the midpoint slice as the reference before FieldMap correction was applied for correcting the distortion induced by field inhomogeneities (Andersson, Hutton, Ashburner, Turner, & Friston, 2001). The EPI images were then realigned to the first volume to correct for head motion effects using a six-parameter rigid body transformation. Mean functional images were then coregistered to the T1 structural image. For the spatial normalization, functional images were coregistered to the T1 images which had been segmented into white matter, gray matter, and cerebrospinal fluid using SPM8 default tissue probability maps and normalized to the Montreal Neurological Institute (MNI) template and the final resolution of image is 3 × 3 × 3 mm³. Finally, functional images were smoothed with a Gaussian kernel of 6 mm full-width at half-maximum.

Different general linear models (GLMs) were then specified for each subject to examine the neural activities associated with corresponding cognitive processes. GLM 1 and 2 were constructed to test the neural activation patterns related to conflict monitoring and social conformity, respectively. More specifically, GLM 1 was established by

FIGURE 3 Neural activations at the time of decision. (a) Neural correlates of information conflict (red, incongruent vs. congruent) and behavioral choice (yellow, deviating from vs. following private information; $p < 0.05$ false discovery rate [FDR] correction). (b) Dorsomedial prefrontal cortex (dmPFC; peak $[-3\ 30\ 48]$) activity associated with deviating from versus following private information contrast was negatively correlated with the proportion of not following the private information across subjects [Color figure can be viewed at wileyonlinelibrary.com]

including two predictors of interest: the decision onsets of INCongruent and congruent conditions (trials where subjects played roles of Players 2 and 3) and we tested the difference of neural activations between these two conditions (Figure 3a). Other predictors of no interest were also included in the design matrix to account for the nuisance variance (the onsets of private and social information revelation, decision onsets of baseline condition, mixed condition and onsets of feedbacks). GLM 2 was established similarly to test differential neural activation related to the social conformity behavior (deviating from the private information vs. following the private information) in all trials. Nuisance variables such as onsets of private and social information revelation and outcome feedback onsets were also included as regressors of no interest. GLM3 was set up at the onsets of outcome feedback to explore differential neural activities between trials of valid (where private information prediction matched outcome) and invalid private information (Figure 4b). To further control for the potential contamination of feedback valence on striatum activity, we constructed GLM 4 by including both feedback valence (gain or loss) and private information validity as dummy regressors at the time of feedback presentation and focused on the neural activation specifically related to private information validity. The onsets of private and social information presentation and choice selection were also included in the GLM 3 and 4 as regressors of no interest. For each GLM, all the specified regressors as well as six head movement parameters were included in the final design matrix and convolved with a canonical hemodynamic response function (HRF).

The contrast images generated from each subject were then entered into the second level random-effect analysis with the significant voxels first survived an uncorrected threshold of $p < 0.0005$ ($z > 3.5$) and then $p < 0.05$ using the default false discovery rate (FDR) correction algorithm implemented in SPM8. In GLM 4, we specifically focused on the neural activities in the striatum and orthogonalized the private information validity regressor against feedback valence regressor. We thus reported our results at $p < 0.05$ FDR corrected after small volume correction (SVC) using a striatum template mask.

2.6.1 | Region of interest analyses

The striatum region of interest (ROI) was defined as brain activities that surpassed $p < 0.05$ FDR threshold after SVC correction from GLM4. In defining the small volume mask, we generated a full striatum template mask using the PickAtlas Toolbox (Maldjian, Laurienti, Kraft, & Burdette, 2003) and extracted BOLD time series of above threshold striatum ROI using MarsBaR (<http://www.mrc-cbu.cam.ac.uk/Imaging/marsbar.html>) and realigned the BOLD responses to the onset of outcome revelation. The dependent measure in time course plots was percentage change in hemodynamic signal averaged within the ROIs. It has been reported in previous literatures that the striatum BOLD responses typically reached their peaks around 2–6 s after the onset of outcome revelation (Knutson et al., 2004), we thus averaged the percentage signal change during the 2–6 s period after the outcome onset for each subject and conducted statistical tests between different conditions. For visual display of the time course for the striatum ROI after outcome revelation, we smoothed the time series using the method of interpolation implemented in the MATLAB spline function with a resampling resolution of 100 ms.

2.6.2 | Psychophysiological interaction analyses

We first extracted BOLD time series of the striatum ROI identified in Figure 4b and then deconvolved the BOLD signal to obtain the underlying neural activity. A new GLM was then constructed with the following regressors: (a) interaction between the deconvolved BOLD activity in the striatum ROI and a dummy indicator for accurate and inaccurate private information at outcome onsets, (b) the indicator function for accurate and inaccurate private information at the onsets of outcome revelation, and (c) the original BOLD time series in the striatum ROI. The first two regressors were also convolved with a canonical form of the HRF. We were interested in the neural correlates of the first regressor (Figure 5a). The threshold reported in the psychophysiological interaction (PPI) analyses are $z > 3.5$ at voxel level and FDR corrected ($p < 0.05$) at the cluster level for the whole brain.

Furthermore, in order to verify the modulation of the validity of private information on the activations of dmPFC at decision

FIGURE 4 Behavioral and neural correlates for choice behavior update. (a) The interaction of private information validity in the previous ($t-1$) trial and private signal content in the current (t) trial was significant ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$). (b) Significantly higher activity in the strium (red), vmPFC and PCC were identified in the accurate versus inaccurate private information contrast ($p < 0.05$ false discovery rate [FDR] correction). The activity of strium (yellow) remained significant after controlling for the confounding effect of feedback valence ($p < 0.05$, small volume correction [SVC] corrected). (c) Time courses of strium region of interest (ROI) after the onset of outcome revelation under conditions of accurate/inaccurate private information ($t(34) = 2.13$, $p = 0.04$). (d) Strium activities did not differ among different social information validities. ($F(2,68) = 2.23$, $p = 0.12$) [Color figure can be viewed at wileyonlinelibrary.com]

onset, we extracted the percentage signal change in the dmPFC (Figure 5a) and applied a two-way ANOVA to test the effects of the validity of private information (accurate vs. inaccurate) and the choice behavior (deviating from vs. following the private information, Figure 5b). Two subjects' data were excluded from the ANOVA due to the lack of deviating from private information trials when the previous private information was either accurate or inaccurate.

3 | RESULTS

3.1 | Impact of social and private information on choice behavior

In each trial, subjects had to decide which urn their bead came from. In theory, subjects' choice depends on both the color of the bead (private information) and other players' predictions (social information).

FIGURE 5 Validity of private information modulates dorsomedial prefrontal cortex (dmPFC) activity. (a) Stronger functional connectivity between strium and dmPFC when private information was accurate at feedback delivery ($p < 0.05$ FDR correction). (b) dmPFC at decision onset showed stronger activity when subjects deviated from private information and when the private information was accurate in the previous ($t-1$) trial.

$*p < 0.05$, $**p < 0.01$, $***p < 0.001$ [Color figure can be viewed at wileyonlinelibrary.com]

By design, the private information correctly predicts the urn color with probability of two-thirds while social information has no predictive power (50%) and these two sources of information are independent. Subjects were assigned randomly as the role of first, second, or third player and they were provided monetary feedback indicating whether their choice was correct by the end of each trial (Figure 1a).

We defined social conforming behavior as making the choice consistent with other players' decisions when their private draw was in conflict with other players' predictions. First, we calculated the proportion of the choices deviating from the private signal for each condition and compared the proportions across conditions. As we expected, the proportion of choices not following the private information was significantly higher under the incongruent conditions than the baseline when subject was the second or third player (Figure 2a, $F(5, 170) = 43.38$, $p < 0.001$; mean $\pm$ SE: baseline, $5 \pm 1.4\%$; second player INCongruent condition: $14 \pm 3\%$, $t(34) = 3.43$, $p = 0.002$; third player INCongruent condition: $53 \pm 6\%$, $t(34) = 7.66$, $p < 0.001$, compared to the baseline condition). However, the proportion of deviating from private information was not significantly different from baseline for both the second and third players when private and social information was congruent (second player congruent condition: $4 \pm 1.2\%$, $t(34) = 0.4$, $p = 0.66$; third player congruent condition: $6\% \pm 1.5\%$, $t(34) = -0.4$, $p = 0.72$; third player mixed condition: $6 \pm 1.5\%$, $t(34) = -0.8$, $p = 0.42$, compared to the baseline condition), indicating a potential floor effect.

We next tested whether both private and social information influenced subjects' choice using a mixed effects logistic regression model (see Section 2). The results showed that the coefficients of two sources of information were both significant (Figure S1, Supporting Information; the coefficients mean $\pm$ SE: private, 7.71 ± 0.69 , $z = 11.24$, $p < 0.001$; first social partner decision, Social 1, 3.14 ± 0.47 , $z = 6.72$, $p < 0.001$; second social partner decision, Social 2, 4.40 ± 0.77 , $z = 5.69$, $p < 0.001$). The interaction term of social information (first $\times$ second social partner, Social 1 $\times$ 2) showed a positive trend but did not reach statistical significance ($p = 0.10$).

To evaluate whether social conformity bias might change as a consequence of feedbacks, we first looked at the proportion of choices not following private signals for each condition across three sessions (Figure S2, Supporting Information). There was a clear tendency to break loose from social conformity bias across sessions (Figure S2, Supporting Information; second player INCongruent condition $F(2.68) = 7.56$, $p = 0.001$; third player, INCongruent condition $F(2.68) = 7.95$, $p < 0.001$; third player, MIXed condition $F(2.68) = 6.94$, $p = 0.002$). Interestingly, no significant cross session effect was observed when subjects were either making choice based on private information alone (first player, $F(2.68) = 3.01$, $p = 0.06$) or when their private and social signals were congruent (second player, $F(2.68) = 1.85$, $p = 0.16$; third player, $F(2.68) = 0.49$, $p = 0.62$). The above results could potentially be interpreted as a general disengagement effect: as trials proceeded, subjects were generally distracted or less engaged, resulting in taking less into account of social information. To test this possibility, we separately estimated the impact of private and social signals on subjects' choice behavior for three sessions (Figure 2b and Table S1, Supporting Information). Our results showed that the weights of private information increased across sessions (F

$(2.68) = 45.78$, $p < 0.001$), but the weights of social information were not affected by sessions (first social info, $F(2.68) = 1.46$, $p = 0.24$; second social info, $F(2.68) = 2.23$, $p = 0.12$). These results argued that the diminishment of social conformity bias was not a result from less impact of social signals per se (Figure 2b), but rather from increased reliance on private information, which indirectly decreased the relative weight of social signals.

3.2 | Neural correlates of information conflict and social conformity

To explore the neural correlates responsible for conflict monitoring and social conformity, we constructed GLMs to fit the BOLD data. We specifically focused on the contrasts between incongruent and congruent conditions of private and social information as well as deviating versus following private signals. Consistent with previous research, we found higher activities in the areas of the bilateral dlPFC, dmPFC, and angular gyrus in the incongruent–congruent contrast (red areas in Figure 3a and Table S2 (Supporting Information), $p < 0.05$ FDR correction). In addition, similar areas of dlPFC, dmPFC, and angular gyrus showed higher activation when subjects followed social signals versus private information (yellow areas in Figure 3a and Table S3 (Supporting Information), $p < 0.05$ FDR corrected). We then performed a conjunction analysis of these two contrasts (Figure 3a and Table S4 (Supporting Information), $p < 0.05$ FDR corrected). Together, these results suggested that dmPFC, dlPFC, and angular gyrus are involved in both information conflict detection and conforming to social signals.

Previous literature suggested that activities in dmPFC might play a critical role in mediating the effect of social information on subjects' choice behavior (De Martino et al., 2017; Klucharev et al., 2009; Klucharev et al., 2011). To test this hypothesis, we performed an ROI analysis on dmPFC, identified in the above conjunction analysis (Figure 3a). We found a significant negative correlation between activities of deviating versus following private information contrast and the degrees to which each subject's choice deviated from her private information (percentage of choices not following private information) (Figure 3b; $r = -0.45$, $p = 0.007$). Similar correlations were found when we focused on the INCongruent conditions alone: the activities of dmPFC conforming to social information (vs. not conforming) negatively correlated with individual degree of social conformity bias (Figure S3a, Supporting Information; $r = -0.70$, $p < 0.001$), indicating that when conforming to social information greater dmPFC activation was associated with people who were less susceptible to social influence. This ROI analysis result was further corroborated by the significant negative correlation between reaction time associated with social conforming behavior and social conformity index (Figure S3b, Supporting Information; $r = -0.53$, $p = 0.001$). Similar ROI analyses were also performed for dlPFC and angular gyrus, the other brain areas identified in the above conjunction analysis. However, their activities were not correlated with individual differences of social conformity bias (Figure S4, Supporting Information). Together, these data implied that subjects who were characterized as less susceptible to the social information had to exert more cognitive control in order to

conform to other people's opinions (Botvinick et al., 2001; De Martino et al., 2017; Klucharev et al., 2011).

3.3 | Updating the influence of private information on choice behavior

In our experiment, choices solely based on private information were correct two-thirds of the time. However, social signals were uninformative in terms of improving subjects' choice accuracy. To learn the information-outcome mapping successfully, participants should update the weights of private and social signals separately based on their corresponding validities. Since in each trial, monetary outcome indicated which urn was the correct choice, we defined bead color consistent with correct urn color as accurate private information and vice versa. Accuracies of social information were defined similarly. We constructed a mixed effects logistic regression model to test whether the identities of private and social information at the current trial, as well as the information accuracies at the previous trial and their interaction items would influence subjects' choice in the current trial (see Section 2 for detail). As expected, the regression results showed that both private and social information at current trial significantly influenced choice behavior (Figure 4a, $p_s < 0.001$). Importantly, interaction term between private information (current trial) and accuracy of private information (previous trial) was also significant (Figure 4a; $z = 4.48$, $p < 0.001$), indicating subjects increased reliance on private information when previous private signal was correct. Interestingly, the interaction terms for social signals were not significant (Table S5, Supporting Information; first social decision $z = -0.29$, $p = 0.78$; second social decision $z = -0.33$, $p = 0.74$), consistent with our previous finding that the impact of social signal did not change across sessions (Figure 2b).

The above results could potentially be confounded by the fact that private information accuracy tends to correlate with monetary feedback. To exclude the possibility that greater reliance on private information was due to positive feedback in the previous trial, we constructed an augmented model by adding previous outcome and the interaction term between previous outcome and current private information as covariates of no interest. Again, we found that the interaction of previous private information validity and current private signal was significant ($z = 3.14$, $p = 0.002$), but the interaction of previous outcome and current private signal was not (Figure S5, Supporting Information; $z = 0.35$, $p = 0.73$). These data further confirmed that only private information validity modulated subjects' social conformity.

3.4 | Neural correlates of validity inference and reliance update

Since the accuracies of private and social information can only be inferred after outcome feedback was presented, we also examined brain responses associated with feedback revelation. We hypothesized that inferred information validities might modulate activities in brain regions associated with value and belief updating, such as VS, dlPFC, and ventromedial prefrontal cortex (vmPFC) (Burke, Tobler, Schultz, & Baddeley, 2010; Schultz, Dayan, & Montague, 1997). As

expected, a simple accurate versus inaccurate contrast of private information at the outcome onset revealed heightened activation at striatum, posterior cingulate cortex and vmPFC (red areas in Figure 4b and Table S6 (Supporting Information), $p < 0.05$ FDR). To remove the potential confound of outcome valence (gain and loss), we constructed a new GLM with both outcome valence and private information validity as dummy regressors at outcome onset. Again, accurate private information was associated with higher activity in the striatum (Figure 4b yellow areas, $p < 0.05$ FDR after SVC using a striatum template mask; Figure S6, Supporting Information). Our results demonstrated that higher striatal activity was associated with accurate private information (Figure 4c; $t(34) = 2.13$, $p = 0.04$). However, no striatum activity difference was observed between accurate and inaccurate social information contrast (Figure 4d; $F(2,68) = 2.23$, $p = 0.12$). Further ROI analyses were implemented to compare the BOLD responses related to information validity under two different feedback outcomes (win/loss). Private information validity effect was significant when outcome feedback was negative (loss; Figure S7a, Supporting Information; $t(33) = 2.88$, $p = 0.007$), but not when outcome was positive (win; $t(33) = -0.19$, $p = 0.85$). In addition, the BOLD responses were not significantly different to different validities of social information under either the win or the loss outcome conditions (Figure S7b, Supporting Information; $p_s > 0.13$). These results further confirmed that part of the striatum seemed to specifically respond to the accuracy of private information.

3.5 | Functional connectivity between striatum and dmPFC

Given that behavioral results showed significant correlation between private signal validity and subjects' subsequent adjustment of behavioral strategy, we sought to test the hypothesis that the brain areas responding to private information validity might be functionally coupled with brain regions associated with conflict detection and social conforming behavior. Thus, we performed a functional connectivity analysis using seed region of striatum identified in Figure 4b and then explored brain areas that showed stronger functional connectivity when private information was accurate (vs. inaccurate).

A PPI analysis revealed that the connectivity between striatum and brain regions such as dmPFC and bilateral dlPFC (Figure 5a; $p < 0.05$ FDR correction) was significantly stronger when the private information was accurate at the time of outcome feedback. We also examined the effect of private information validity at the time of decision. Subjects tended to conform less to social information if private signal was accurate in the previous trial (Figure 4a). We hypothesize that the valid previous private information would increase dmPFC activity at the time of decision and prompt subjects to be more behaviorally aligned with the prediction of private information. To test this, we ran a two-way ANOVA with previous private signal validity and whether subjects deviated from private information as separate factors. Consistent with our prediction, we found both main effects of private information validity and private information deviation ($F(1,32) = 6.83$, $p = 0.01$; $F(1, 32) = 27.96$, $p < 0.001$, respectively) as well as the interaction effect ($F(1,32) = 13.78$, $p < 0.001$) were significant (Figure 5b). Similar results were obtained with bilateral dlPFC

(Figure S8, Supporting Information). These results, both at the time of outcome feedback and decision onset, suggest that the private information validity influence subjects' susceptibility to social information via modulating the functional connectivity between striatum and dmPFC, whose activity further dictates the degree of social conformity bias (Figures 3 and 5b).

4 | DISCUSSION

How human choice behavior is influenced by the confluence of private and social information has been an active research topic in the fields of psychology and social neuroscience. Aided by formal decision theory and neuroimaging techniques, it has been shown that brain regions such as cingulate cortex (posterior cingulate cortex and dACC) and dmPFC play a critical role in conforming to social influences (Burke, Tobler, Schultz, & Baddeley, 2010; De Martino et al., 2017; Huber, Klucharev, & Rieskamp, 2015; Klucharev et al., 2009; Klucharev et al., 2011; Park, Goïame, O'Connor, & Dreher, 2017; Wu et al., 2016; Zaki et al., 2011). However, whether and how human can dynamically adjust their behavior based on the informational content remains largely unknown. Using an urn guessing game together with fMRI, we showed that subjects' choices were consistently influenced by social signals, in spite of their (in)validities, thus confirming the existence of social conformity bias in a laboratory setup. We further showed that brain areas such as dmPFC (dACC), dlPFC, and angular gyrus were involved in detecting private and social information conflict and more activated when subjects' decisions deviated from private information. This observation accorded well with previous literatures indicating the involvement of brain regions such as dACC, dmPFC, and dlPFC in conflict monitoring, mentalization, metacognition, and cognitive control in social conformity tasks (De Martino et al., 2017; Klucharev et al., 2009; Kragel et al., 2018; Shenhav et al., 2017; Shenhav, Botvinick, & Cohen, 2013; Wu et al., 2016) as well as the nondefault option value tracking in foraging tasks (Kolling, Behrens, Mars, & Rushworth, 2012). Interestingly, however, we also found that dmPFC activity was negatively correlated with subjects' propensity to deviate from the default option of following private information (Figure 3b). Such an activation pattern and individual difference correlation result were more aligned with the theory that the dmPFC activity might represent subjective task difficulty, and the amount of cognitive control based on the expected value of such control rather than foraging value (De Martino et al., 2017; Hampton, Bossaerts, & O'Doherty, 2008; Hill et al., 2017; Kolling et al., 2016; Kragel et al., 2018; Shenhav et al., 2013; Shenhav et al., 2017; Shenhav, Straccia, Cohen, & Botvinick, 2014). It should be noted that, in our study, we found brain areas that encompass both traditionally defined dACC and dmPFC responded to both social and private information conflict as well as social information conformity (Figure 3a). It is thus possible that there exists a finer functional segregation in the mPFC with dACC more involved in private and social information conflict monitoring and dmPFC implicated in social conformity itself, based on the degree of cognitive control a certain decision context might demand (Kragel et al., 2018; Ruff & Fehr, 2014; Schlund et al., 2016; Shenhav et al., 2013; Shenhav et al., 2014; Wisniewski, Reverberi, Tusche, &

Haynes, 2015). Through this lens, the reduced conformal adjustment from the downregulation of the medial frontal cortex in a previous transcranial magnetic stimulation study could be interpreted as either reducing the error signal between subject's own judgment and group opinion or simply impairing participant's tendency to follow group opinion (Klucharev et al., 2009; Klucharev et al., 2011). Interestingly, a recent study reported that the activities in the dmPFC seemed to track participants' resistance to social information (De Martino et al., 2017), suggesting its role might be more consistent with social conformity. Indeed, in our experiment, PPI analysis at the revelation of outcome feedback showed that the functional connectivity between striatum and the more anterior dmPFC was strengthened when private information was valid and subjects' subsequent choices were less vulnerable to the social influence (Figure 5a). Thus, our individual variance data provide evidence to support the task difficulty account of dmPFC function in conforming to social information. Interestingly, one recent study on how group opinions influenced individual's game App recommendation found that the activities of temporoparietal junction positively correlated with individual susceptibility to peer recommendation opinions (Cascio, O'Donnell, Bayer, Tinney, & Falk, 2015), suggesting a pull-and-push balance between brain regions in the process of conforming to social signals.

The ubiquity of social conformity bias can be adaptive from an evolutionary perspective since it allows efficient information integration from different sources to improve decision efficacy. Indeed, formal decision theory suggests that agents should take into account of the uncertainty of each information source and weigh them according to their precisions (Pouget, Deneve, & Duhamel, 2002). However, social conforming behavior is only beneficial when the social signals are informative (Behrens et al., 2008; Bikhchandani et al., 1992; Burke, Tobler, Baddeley, & Schultz, 2010; Huber et al., 2015). Otherwise, such tendency might backfire as it has been clearly evidenced in the investors' herding behavior during the boom and bust cycles of the stock markets (Anderson & Holt, 1997; Camerer & Weigelt, 1991). While much research has focused on the neural mechanisms of social conforming behavior itself (Bikhchandani, Hirshleifer, & Welch, 1998; Burke, Tobler, Schultz, & Baddeley, 2010; Klucharev et al., 2009; Park et al., 2017; Zaki et al., 2011), we instead focused on whether and how the social conformity bias may be alleviated. One convenient and intuitive hypothesis is that as subjects gradually learned the trustworthiness of both private and social signals, they would increase and decrease the weights assigned to private and social information, respectively (Dunlap, Nielsen, Dornhaus, & Papaj, 2016). Recent research in social reinforcement learning suggested that people do update their expectations of social partner based on the feedback received or the variance within social information (Campbell-Meiklejohn, Simonsen, Frith, & Daw, 2017; De Martino et al., 2017; Jones et al., 2011; King-Casas et al., 2005; Park et al., 2017; Tamir & Mitchell, 2010). Another possibility is that due to the evolutionary constraint, innate tendency to follow social information may be hard to overrule (Asch, 1956), especially in an uncertain environment; people instead only update the weight of private information and indirectly adjust the relative weights of these two information channels. Such an approach would prepare people to flexibly update their behavioral strategy while still remain sensitive to

social information in an uncertain environment, analogous to the spontaneous recovery of fear after extinction (Clem & Schiller, 2016). Our results tend to accord with the latter interpretation. In our experiment, we provided actual feedback to the subjects such that they were able to infer the validities of both private and social signals and dynamically adjusted their behaviors accordingly; whereas extant social conformity studies mainly focused on the general tendency to conform to social signals. As a consequence, private and social signal validities could not be inferred (thus no updating of the influences of different sources). In our task, as subjects learned through trial-and-error, they were ostensibly less influenced by the social information. However, they adjusted their behavior such that the impact of private information increased while the influence of social information remained unchanged. It is the relative weight adjustment, instead of decreasing the impact of social information directly, that gradually rescued subjects from falling prey to the social conformity bias. We found that activities in the striatum, in addition to encoding outcome valence, also responded to the accuracy of private information. Updating the reliance of private, but not social information was associated with BOLD signal changes in the striatum and the functional coupling between striatum and dmPFC was strengthened when private information was valid. This activation pattern, together with the behavioral sensitivity to previous private signal validity suggests an intriguing possibility that striatum activity not only encoded feedback valence signal, but also carried information about the inferred private information accuracy (Bromberg-Martin, Matsumoto, & Hikosaka, 2010; Zink, Pagnoni, Martin-Skurski, Chappelow, & Berns, 2004). Furthermore, functional connectivity between striatum and dmPFC was strengthened when private information was valid. Such a functional coupling also correlated with elevated dmPFC activity at decision onset in the next trial, making subjects less likely to commit social conformity bias. Future work will be needed to investigate whether the lack of updating the weight on social information is specifically tied to the fact that social signal is simply uninformative in our task. As in previous research, our subjects differ considerably in their degrees of committing social conformity bias; it thus remains to be tested how subjects across social conformity bias spectrum would update their reliance on private and social information (Cook, Den Ouden, Heyes, & Cools, 2014; De Martino et al., 2017; Toelch, Bruce, Newson, Richerson, & Reader, 2013). Nevertheless, the present work illustrates and highlights intimate interactions between striatum and dmPFC in modulating people's social conformity propensity.

In conclusion, our results corroborate previous results and highlight the importance of dmPFC (dACC) in detecting information conflict and conforming to social influences. They also echo well with individual difference data supporting the hypothesis that dmPFC is related to the encoding of task difficulty and allocation of cognitive control (Kolling et al., 2016; Kragel et al., 2018; Shenhav et al., 2014; Shenhav et al., 2017). More importantly, the alleviation of social conformity bias was mediated by the private information weight updating in the striatum, which subsequently modulated its functional coupling with dmPFC and increased activity of dmPFC when subjects faced information conflict again. These observations should shed light to the question of whether and how subjects might overcome the social conformity bias based on feedback and may provide a general test bed

for studying the learning and estimation of information validities from different channels.

ACKNOWLEDGMENTS

This work is supported by the Ministry of Science and Technology of China (2015CB559200) and National Science foundation of China (projects 31421003, 31871140, and 31371019).

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ORCID

Li Li <https://orcid.org/0000-0002-2294-8287>

Jian Li <https://orcid.org/0000-0002-3941-2622>

REFERENCES

- Anderson, L. R., & Holt, C. A. (1997). Information cascades in the laboratory. *The American Economic Review*, 87, 847–862.
- Andersson, J. L., Hutton, C., Ashburner, J., Turner, R., & Friston, K. (2001). Modeling geometric deformations in EPI time series. *NeuroImage*, 13, 903–919.
- Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, 70, 1–70.
- Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, 107, 797–817.
- Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature*, 456, 245–249.
- Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, 10, 1214–1221.
- Berns, G. S., Capra, C. M., Moore, S., & Noussair, C. (2010). Neural mechanisms of the influence of popularity on adolescent ratings of music. *NeuroImage*, 49, 2687–2696.
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100, 992–1026.
- Bikhchandani, S., Hirshleifer, D., & Welch, I. (1998). Learning from the behavior of others: Conformity, fads, and informational cascades. *The Journal of Economic Perspectives*, 12, 151–170.
- Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual choice and learning in a neural network centered on human lateral frontopolar cortex. *PLoS Biology*, 9, 1–13.
- Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron*, 62, 733–743.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108, 624–652.
- Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in motivational: Rewarding, aversive, and alerting. *Neuron*, 68, 815–834.
- Burke, C. J., Tobler, P. N., Schultz, W., & Baddeley, M. (2010). Striatal BOLD response reflects the impact of herd information on financial decisions. *Frontiers in Human Neuroscience*, 4, 1–11.
- Burke, C. J., Tobler, P. N., Baddeley, M., & Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 14431–14436.
- Camerer, C., & Weigelt, K. (1991). Information mirages in experimental asset markets. *Journal of Business*, 64, 463–493.

- Campbell-Meiklejohn, D. K., Bach, D. R., Roepstorff, A., Dolan, R. J., & Frith, C. D. (2010). How the opinion of others affects our valuation of objects. *Current Biology*, 20, 1165–1170.
- Campbell-Meiklejohn, D. K., Simonsen, A., Frith, C. D., & Daw, N. D. (2017). Independent neural computation of value from other people's confidence. *The Journal of Neuroscience*, 37, 673–684.
- Cascio, C. N., O'Donnell, M. B., Bayer, J., Tinney, F. J., & Falk, E. B. (2015). Neural correlates of susceptibility to group opinions in online word-of-mouth recommendations. *Journal of Marketing Research*, 52, 559–575.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591–621.
- Clem, R. L., & Schiller, D. (2016). New learning and unlearning: Strangers or accomplices in threat memory attenuation? *Trends in Neurosciences*, 39, 340–351.
- Cook, J. L., Den Ouden, H. E. M., Heyes, C. M., & Cools, R. (2014). The social dominance paradox. *Current Biology*, 24, 2812–2816.
- Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, 36, 285–298.
- Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal Psychology*, 51, 629–636.
- Dunlap, A. S., Nielsen, M. E., Dornhaus, A., & Papaj, D. R. (2016). Foraging bumble bees weigh the reliability of personal and social information. *Current Biology*, 26, 1–5.
- Edelson, M. (2011). Following the crowd : Brain substrates of long-term memory conformity. *Science*, 333, 108–111.
- Falk, E. B., Berkman, E. T., Mann, T., Harrison, B., & Lieberman, M. D. (2010). Predicting persuasion-induced behavior change from the brain. *The Journal of Neuroscience*, 30, 8421–8424.
- Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 6741–6746.
- Hill, C. A., Suzuki, S., Polania, R., Moisa, M., O'doherty, J. P., & Ruff, C. C. (2017). A causal account of the brain network computations underlying strategic social behavior. *Nature Neuroscience*, 20, 1142–1149.
- Huber, R. E., Klucharev, V., & Rieskamp, J. (2015). Neural correlates of informational cascades: Brain mechanisms of social influence on belief updating. *Social Cognitive and Affective Neuroscience*, 10, 589–597.
- Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Libby, V., Glover, G., ... Casey, B. J. (2011). Behavioral and neural properties of social reinforcement learning. *The Journal of Neuroscience*, 31, 13039–13045.
- King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, 308, 78–83.
- Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement learning signal predicts social conformity. *Neuron*, 61, 140–151.
- Klucharev, V., Munneke, M. A. M., & Smidts, A. (2011). Downregulation of the posterior medial frontal cortex prevents social conformity. *The Journal of Neuroscience*, 31, 11934–11940.
- Knutson, B., Bjork, J. M., Fong, G. W., Hommer, D., Mattay, V. S., & Weinberger, D. R. (2004). Amphetamine modulates human incentive processing. *Neuron*, 43, 261–269.
- Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural mechanisms of foraging. *Science*, 336, 95–98.
- Kolling, N., Wittmann, M. K., Behrens, T. E. J., Boorman, E. D., Mars, R. B., & Rushworth, M. F. S. (2016). Value, search, persistence and model updating in anterior cingulate cortex. *Nature Neuroscience*, 19, 1280–1285.
- Kragel, P. A., Kano, M., van Oudenhove, L., Ly, H. G., Dupont, P., Rubio, A., ... Wager, T. D. (2018). Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. *Nature Neuroscience*, 21, 283–289.
- Li, J., & Daw, N. D. (2011). Signals in human striatum are appropriate for policy update rather than value prediction. *The Journal of Neuroscience*, 31, 5504–5511.
- Lohrenz, T., McCabe, K., Camerer, C. F., & Montague, P. R. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 9493–9498.
- Maldjian, J. A., Laurienti, P. J., Kraft, R. A., & Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *NeuroImage*, 19, 1233–1239.
- De Martino, B., Bobadilla-Suarez, S., Nouguchi, T., Sharot, T., & Love, B. C. (2017). Social information is integrated into value and confidence judgments according to its reliability. *The Journal of Neuroscience*, 37, 6066–6074.
- Montague, P. R., & Lohrenz, T. (2007). To detect and correct: Norm violations and their enforcement. *Neuron*, 56, 14–18.
- Park, S. A., Goïame, S., O'Connor, D. A., & Dreher, J.-C. (2017). Integration of individual and social information for decision-making in groups of different sizes. *PLoS Biology*, 15, e2001958.
- Pouget, A., Deneve, S., & Duhamel, J. R. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews. Neuroscience*, 3, 741–747.
- Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews. Neuroscience*, 15, 549–562.
- Schlund, M. W., Brewer, A. T., Magee, S. K., Richman, D. M., Solomon, S., Ludlum, M., & Dymond, S. (2016). The tipping point: Value differences and parallel dorsal-ventral frontal circuits gating human approach – Avoidance behavior. *NeuroImage*, 136, 94–105.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, 275, 1593–1599.
- Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, 79, 217–240.
- Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, 40, 99–124.
- Shenhav, A., Straccia, M. A., Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience*, 17, 1249–1254.
- Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, 56, 185–196.
- Tamir, D. I., & Mitchell, J. P. (2010). Neural correlates of anchoring-and-adjustment during mentalizing. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 10827–10832.
- Toelch, U., Bruce, M. J., Newson, L., Richerson, P. J., & Reader, S. M. (2013). Individual consistency and flexibility in human social information use. *Proceedings of the Royal Society B: Biological Sciences*, 281, 20132864–20132864.
- Wisniewski, D., Reverberi, C., Tusche, A., & Haynes, J.-D. (2015). The neural representation of voluntary task-set selection in dynamic environments. *Cerebral Cortex*, 25, 4715–4726.
- Wu, H., Luo, Y., & Feng, C. (2016). Neural signatures of social conformity: A coordinate-based activation likelihood estimation meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, 71, 101–111.
- Zaki, J., Schirmer, J., & Mitchell, J. P. (2011). Social influence modulates the neural computation of value. *Psychological Science*, 22, 894–900.
- Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., & Berns, G. S. (2004). Human striatal responses to monetary reward depend on saliency. *Neuron*, 42, 509–517.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Li L, Li KK, Li J. Private but not social information validity modulates social conformity bias. *Hum Brain Mapp*. 2019;40:2464–2474. <https://doi.org/10.1002/hbm.24536>