doi:10.1093/scan/nst019

Neural mechanisms underlying the integration of situational information into attribution outcomes

Tobias Brosch,[1,2] Daniela Schiller,[3] Rachel Mojdehbakhsh,[2] James S. Uleman,[2] and Elizabeth A. Phelps[2,4]

[1]Department of Psychology, University of Geneva, 1205 Geneva, Switzerland, [2]Department of Psychology, New York University, New York, NY 10003, USA, [3]Departments of Psychiatry and Neuroscience, and Friedman Brain Institute, Mount Sinai School of Medicine, New York, NY 10029, USA, and [4]Nathan Kline Institute, Orangeburg, New York, NY 10962, USA

When forming impressions and trying to figure out why other people behave the way they do, we should take into account not only dispositional factors (i.e. personality traits) but also situational constraints as potential causes for a behavior. However, in their attributions, people often ignore the importance of situational factors. To investigate the neural mechanisms underlying the integration of situational information into attributions, we decomposed the attribution process by separately presenting information about behaviors and about the situational circumstances in which they occur. After reading the information, participants judged whether dispositional or situational causes explained the behavior (*attribution*), and how much they liked the person described in the scenario (*affective evaluation*). The dorsolateral prefrontal cortex showed increased blood oxygenation-level-dependent activation during the encoding of situational information when the resulting attribution was situational, relative to when the attribution was dispositional, potentially reflecting a controlled process that integrates situational information into attributions. Interestingly, attributions were strongly linked to subsequent affective evaluations, with the dorsomedial prefrontal cortex emerging as potential substrate of the integration of attributions and affective evaluations. Our findings demonstrate how top-down control processes regulate impression formation when situational information is taken into account to understand others.

Keywords: attribution; evaluation; person perception; disposition; situation; fundamental attribution error

INTRODUCTION

We constantly try to explain other peoples' behavior in order to understand and negotiate social situations. The enduring dispositions of a person can explain what causes that person's behavior and also the situational context in which the behavior unfolds. Mike may smile at Tom (behavior), for example, because he is a friendly person (disposition) or because he is currently trying to sell Tom a used car (situation). People should always consider both dispositional and situational factors as potential causes for a given behavior. Very often, however, people attribute behaviors to dispositional causes, even though the behavior could be entirely explained by the situation in which it occurred (Jones and Harris, 1967). This cognitive bias, referred to as *Fundamental Attribution Error* (Ross, 1977) or *correspondence bias* (Gilbert and Malone, 1995), has become a textbook example of flawed human reasoning.

Several competing theories have been put forward to explain why people often discount situational information (see Gawronski, 2004; Sabini *et al*., 2001, for reviews), some of which emphasizing a single factor, and others proposing a dual process. Single factor theories focus on one principal explanation such as the higher perceptual salience of behavior compared with the situation ('behavior engulfs the field;' Taylor and Fiske, 1978), or layperson theories in which stable dispositions are the main determinant of behavior (Dweck and Leggett, 1988). Such mechanisms would lead to an immediate discounting of the situational information as a potential explanation for an observed behavior. Dual-process theories posit that people may be generally aware that situational circumstances can affect behavior, but may under certain circumstances not take this into account during their attributions (Gawronski, 2004). In this context, it has been suggested that although dispositional inferences are drawn automatically, the integration of situational information requires a more controlled, top-down process. Consistent with this idea, it has been shown that when people perceive another person's behavior, they spontaneously generate the disposition implied by the action. Reading that 'Alice solved the mystery halfway through the book', for example, automatically activates the disposition 'clever' (Uleman *et al*., 1996). This dispositional inference, however, can be corrected by taking into account the influence of any relevant situational information (e.g. 'the book was written for pre-teens') in a controlled and cognitively more demanding process. However, if this process fails, the situational information will not be taken into account.

As a first step toward understanding the neural mechanisms underlying the Fundamental Attribution Error, here we investigated the brain systems underlying the encoding and integration of situational information during the attribution process. Dual-process theories of attribution would predict the recruitment of additional neurocognitive mechanisms during the processing of situational information only if this information is taken into account for the attribution, but not when it is ignored or discarded. In contrast, these mechanisms are not expected to play a role during the encoding of behavioral information only, where dispositional inferences are expected to occur automatically.

The dorsolateral prefrontal cortex (DLPFC) has been linked to top-down cognitive control, detection of appropriate behavior among competing responses and inhibition of inappropriate automatic reactions (MacDonald *et al*., 2000) and is thus a potential neural substrate of the process for the integration of situational information as predicted by dual-process theories of attribution (see also Lieberman *et al*., 2002). Previous work investigating the neural basis of social inferences has revealed a network consisting of medial prefrontal cortex (MPFC), temporo-parietal junction (TPJ) and precuneus underlying our capacity to ascribe intentions, beliefs and traits to others (Saxe and Kanwisher, 2003; Mitchell *et al*., 2004, 2006; Harris

Received 28 February 2012; Accepted 12 February 2013

Advance Access publication 27 February 2013

This study was funded by a Seaver Foundation grant to the Center for Brain Imaging, a grant from the National Institutes of Health (5R01MH062104 to E.A.P.) and a grant from the Swiss National Science Foundation (PA00P1_131435 to T.B.).

Correspondence should be addressed to Tobias Brosch, Department of Psychology, University of Geneva, 40, Boulevard du Pont d'Arve, CH-1205 Geneva, Switzerland. E-mail: tobias.brosch@unige.ch

et al., 2005; Van Overwalle, 2009; Van Overwalle and Baetens, 2009). The dorsal part of the MPFC (DMPFC) in particular is involved in actively forming impressions of others (Mitchell *et al.*, 2005a, 2005b, 2006; Mitchell *et al.*, 2005c). Precuneus/posterior cingulate cortex and amygdala have been suggested to be involved in rapid person evaluations based on descriptions of behaviors (Schiller *et al.*, 2009). Using functional magnetic resonance imaging (fMRI), we investigated how these brain regions interact when we attribute causes to other people's behaviors and use the attribution outcome to inform our evaluations.

To investigate the neural mechanisms underlying the encoding and integration of situational information specifically, we needed to isolate the different components of the attribution process. We therefore presented, in separate segments, information about a certain behavior ('Mike smiled at Tom') and about the situational circumstances in which it occurred ('Mike is selling a car'). The behaviors were either positive or negative and appeared either before or after the situational information (to avoid that effects of interest were confounded with temporal effects, Figure 1). This experimental design allowed us to examine blood oxygenation-level-dependent (BOLD) signal that is specific to situational information and compare cases where it was integrated into the attribution *vs* not. After reading the information, 19 participants judged to what extent dispositional or situational causes explain the behavior (*attribution*), and how much they liked the person described in the scenario (*affective evaluation*).

We then extracted participants' BOLD signal during the presentation of each type of information from independently defined regions of interest (ROIs). To test our main hypothesis of a central role of DLPFC in the integration of situational information into the attribution outcome, we retrospectively sorted the BOLD data based on the attribution ratings to compare brain activation during the encoding of information that was either taken into account for the subsequent attribution or was discounted or ignored. That is, we compared the encoding of situational information in trials where participants subsequently made a situational attribution to trials where they subsequently made a dispositional attribution (i.e. discounted the situational information). Similarly, we compared the encoding of behavioral information in trials where participants subsequently made a dispositional attribution to trials where they subsequently made a situational attribution.

We were furthermore interested in the link between attribution and affective evaluation. The way we explain someone's behavior should have a strong impact on how we feel about that person. Tom, for instance, could attribute Mike's smiling behavior to dispositional ('Mike is a really nice guy') or to situational ('Mike is a slick salesman') causes. In each case, Tom's subsequent interactions with Mike would be dramatically different. The repeated association of a person with positive behaviors results in more positive evaluations, and the repeated association with negative behaviors in more negative evaluations (Kerpelman and Himmelfarb, 1971). Here, we propose that this link should be especially strong for behaviors that are attributed to dispositional factors, as dispositions allow us to predict whether a person is likely to behave in a friendly or exploitative manner in the future. We predict that subjects would evaluate a person more positively when they attribute a positive behavior to dispositional but not situational factors. By the same token, they would evaluate a person more negatively when they attribute a negative behavior to dispositional but not situational factors. Finally, we wanted to identify how the neural processing of behavioral and situational information interacts with subsequent affective evaluations. To this end, we compared

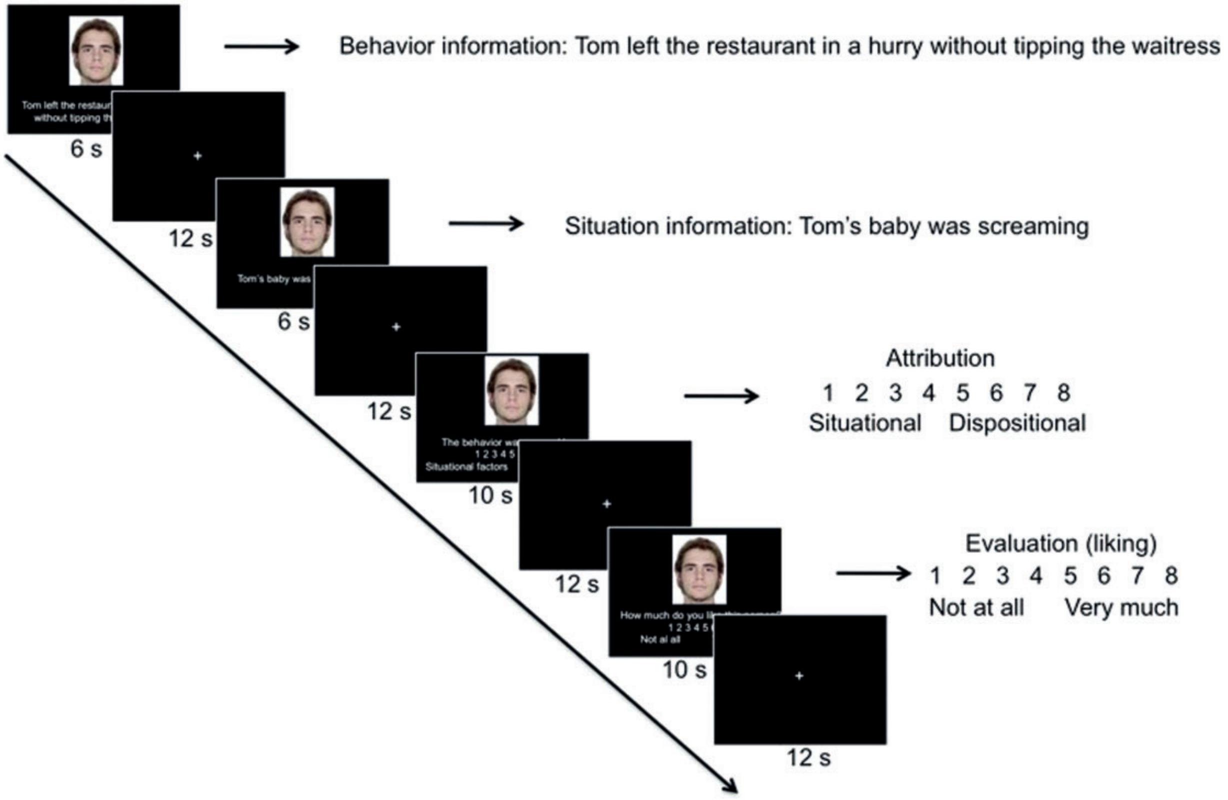

Fig. 1 Experimental sequence. We used 32 scenarios consisting of separately presented information about a behavior and about the situational circumstances in which it occurred. Half of the scenarios described a positive behavior, the other half a negative behavior. The order of the presentation of behavioral and situational information, respectively, was counterbalanced across trials. After reading the information, participants were asked to what extent the behavior was attributable to dispositional or situational causes (*attribution*), and how much they liked the person described in the scenario (*evaluation*).

BOLD signal in brain regions known to play a role in mentalizing and evaluation (MFPC, precuneus, TPJ, amygdala) in trials where the valence of the evaluation was incongruent with the behavior (implying an integration of the situational information into the evaluation) *vs* trials where it was congruent (implying that the situational information was discarded for the evaluation).

METHODS

Participants

We recruited 19 right-handed normal volunteers (six males) between 18 and 41 years of age (mean = 22.8, s.d. = 5.48). The study was approved by the New York University Committee on Activities Involving Human Subjects. All participants gave informed consent and were paid for their participation.

Stimuli

We began by constructing 100 scenarios describing the behavior of a person in a given situation. Each scenario included information about a behavior (e.g. 'Tom left the restaurant in a hurry without tipping the waitress', 'Jenny called her grandmother and aunt to catch up', 'Jim belched loudly during a theater performance') and information about the situational background in which the behavior took place (e.g. 'Tom's baby was screaming', 'Jenny was sitting in traffic', 'Jim had had indigestion all day'). Half of the scenarios described a positive behavior, the other half a negative behavior. Behavioral and situational information were presented separately, with the order of presentation counterbalanced across scenarios. We paired each profile with a photo of a face of neutral expression. The scenarios were pretested ($n = 30$) to select the 32 scenarios with the largest inter-individual variability in the attributions. The goal of the selection was to identify scenarios where the behavior was judged to be due to the disposition of the person by some participants, and due to the influence of the situation by other participants, to ensure that attributions were driven by participants' interpretations of the scenarios rather than by scenario-specific effects.

Procedure

During the fMRI task, each of 32 scenarios started with the presentation of behavioral or situational information for 6 s. After a 12-s inter-stimulus interval, the other type of information (situational or behavioral) was presented. Subsequently, participants judged whether the behavior was caused mainly by situational or by dispositional factors (*attribution*) on a Likert scale from 1 to 8. Participants also indicated how much they liked the person (*affective evaluation*, on a Likert scale from 1 = 'not at all' to 8 = 'very much'). After the fMRI session, participants completed a memory task. Finally, participants completed the Need for Cognition questionnaire (Cacioppo and Petty, 1982), an individual difference measure of the extent to which people engage in and enjoy effortful cognitive activities.

Behavioral results

To ensure that attributions were driven by participants' idiosyncratic interpretations rather than by general stimulus-specific effects, we first quantified the inter-individual variability in the attributions that different participants gave for the same scenario by computing the mean range and the mean standard deviation for the dispositionality ratings of each scenario. There were large inter-individual differences in the attribution ratings [mean range of ratings per scenario = 6.09 (on a scale from 1 to 8), mean minimum rating = 1.34, mean maximum rating = 7.44, and mean s.d. = 1.95], indicating that identical scenario information led to dispositional attributions for some participants and to situational attributions for others. We also probed the scenarios for overall differences in the number of dispositional *vs* situational attributions. Mean attribution scores across all scenarios were not significantly different from the mean of the attribution scale (4.3, $t(31) = 1.45$, $P = 0.16$), indicating that overall participants did not take either behavioral or dispositional information more into account for their attributions. This was confirmed by the absence of differences in the number of participants who made dispositional and situational attributions for a given scenario when analyzing responses by stimulus rather than participant (two-tailed *t*-tests comparing the mean proportion of dispositional evaluations to 0.5, $t(31) = 0.87$, $P = 0.39$). This procedure may seem somewhat counterintuitive: in our experiment, participants did not consistently discard situational information, even though it is thought to be a pervasive phenomenon, implying that in real-life situations people will make dispositional attributions most of the time. However, in order to experimentally investigate this effect and the underlying neural mechanisms, we selected our stimuli (based on pilot data) so that participants would make dispositional attributions in ~50% of the cases, in order to compare equal numbers of trials where the situational information was integrated into the attribution outcome *vs* not.

Memory analysis

To confirm that any observed behavioral or neural differences on the basis of subsequent attributions and evaluations were not merely reflecting differences in episodic memory encoding, we assessed performance on a subsequent memory-recognition task for the information presented during the task. In each trial, participants were presented with one sentence they encountered during one of the scenarios, consisting of either behavioral or situational information, paired with four similar distractor sentences that differed in small details. Participants had to select which of the five sentences had been presented during the fMRI task. We examined the memory performance for type of information bias (i.e. behavioral information remembered better than situational or vice versa), valence bias (i.e. negative sentences remembered better than positive or vice versa), attribution-relevance bias (i.e. information leading to dispositional ratings remembered better than information leading to situational ratings or vice versa) and evaluation-relevance bias (i.e. information guiding the evaluation remembered better than information not guiding the evaluation or vice versa). To test for these memory biases, we compared mean recognition accuracies using paired two-tailed *t*-tests. There was no difference in subsequent memory for behavioral *vs* situational information, $t(18) = 1.7$, ns, and no difference for positive *vs* negative information, $t(18) = 0.89$, ns. Furthermore, there was no difference for situational information that later led to high dispositionality ratings compared with low dispositionality ratings, $t(18) = 1.42$, ns; similarly, no difference for behavioral information that later led to high dispositionality ratings compared with low dispositionality ratings, $t(18) = 0.27$, ns. No differences in subsequent memory was observed for evaluation-relevant *vs* evaluation-irrelevant information, $t(18) = 0.17$, ns. Together, these results eliminate differential memory encoding as an alternative explanation of the findings in our procedure. Thus, attributions and impressions were not driven by episodic memory for specific item information.

fMRI acquisition

A 3 T Siemens Allegra head-only scanner and Siemens standard head coil were used for data acquisition. Anatomical images were acquired using a T1-weighted protocol (256 × 256 matrix, 176 1-mm sagittal slices). Functional images were acquired using a single-shot gradient echo EPI sequence (repetition time, 2.0 s; echo time, 25 ms; field of view, 192 cm, flip angle = 75°). We obtained 39 contiguous oblique-axial slices (3 × 3 × 3-mm voxels) parallel to the anterior commissure–posterior commissure line.

fMRI analysis

Functional images were analyzed using the general linear model (GLM) for event-related designs in SPM8 (Wellcome Department of Imaging Neuroscience, London, UK; http://www.fil.ion.ucl.ac.uk/spm). All images were first realigned, corrected for slice timing, normalized to an EPI template (resampled voxel size of 3 mm), spatially smoothed (8 mm FWHM Gaussian kernel) and high pass-filtered (cutoff 120 s). Statistical analyses were performed on a voxel-wise basis across the whole brain. Individual events were modeled using a boxcar function convolved by the canonical double-gamma hemodynamic response function (HRF). Five event types were defined, including presentation of behavior information, presentation of situational information, attribution screen, evaluation screen and fixation baseline. In subsequent analyses, behavioral and situational information were further classified as attribution-relevant/attribution-irrelevant and evaluation-relevant/evaluation-irrelevant, based on the individual dispositionality and liking ratings.

Attribution relevance

Based on the dispositionality ratings, we sorted the data to compare brain activation during the encoding of information that was taken into account for the attribution *vs* information that was discounted or ignored. Behavioral information that led to dispositional attributions [high dispositionality ratings (5–8)] was classified as attribution-relevant, behavioral information that led to situational attributions [low dispositionality ratings (1–4)] as attribution-irrelevant. Similarly, situational information that led to a situational attribution was classified as attribution-relevant, situational information that led to a dispositional attribution was classified as attribution-irrelevant.

Evaluation relevance

Based on the liking ratings, we sorted the data to compare brain activation during the encoding of information that was taken into account for the evaluation *vs* information that was discounted or ignored. Behavioral information leading to evaluations that were congruent with the valence of the behavior (e.g. a positive behavior leading to a positive evaluation) was classified as evaluation-relevant, behavioral information that led to a valence-incongruent evaluation was classified as evaluation-irrelevant. Similarly, situational information leading to evaluations that were incongruent with the valence of the behavior (suggesting a correction of the evaluation for the situational circumstances) was classified as integrated situational information, and situational information in scenarios where the evaluation was consistent with the behavior as discounted situational information.

To account for residual movement artifacts after realignment, movement parameters derived from realignment corrections (three translations, three rotations) were entered as additional covariates of no interest. The GLM was then used to generate parameter estimates of activity at each voxel, for each condition and each participant. Statistical parametric maps were generated from linear contrasts between the HRF parameter estimates for the different conditions. We performed random-effect group analyses on the contrast images from the individual analyses, using one-sample *t*-tests. To define our ROIs for the encoding of information independently from our specific hypotheses, we contrasted activation during the encoding of all scenario information (behavioral and situational) to resting baseline, whole-brain FDR corrected at $P < 0.05$, minimum cluster size = 20 voxel). As expected, this analysis revealed a network of brain regions previously implicated in impression formation, evaluation and cognitive control (DLPFC, DMPFC, TPJ, precuneus and amygdala, Table 1, as well as low- and high-level visual regions). We extracted the BOLD response from each of the ROIs (mean betas for a sphere of 8-mm centered at the peak coordinate for cortical regions, 4 mm for subcortical regions) and compared the mean activations during presentation of the different kinds of information (two-tailed *t*-tests). We complemented this analysis using whole-brain contrasts thresholded by a combined criterion of $P < 0.005$, and minimum cluster size = 20 contiguous voxels (Lieberman and Cunningham, 2009), in order to validate the primary ROI-based analysis and to explore whether additional brain regions were involved.

Table 1 ROIs defined by increased BOLD response during the processing of behavioral and situational information relative to resting baseline

Region	BA	Side	Coordinates		
			x	*y*	*z*
DMPFC[a,b]	9/10	M	0	47	40
DLPFC[a]	45	L	−51	20	22
DLPFC	45	R	60	23	25
TPJ	39	L	−45	−49	28
TPJ[c]	39	R	48	−58	22
Precuneus[c]	7	M	−6	−58	40
Amygdala	–	L	−30	−1	−20
Amygdala	–	R	27	−4	−20

[a] Regions that show increased activation during the encoding of situational information when subsequent attributions were situational. No regions showed differential activation during the encoding of behavioral information as a function of subsequent attributions.

[b] Regions that show increased activation during the encoding of situational information when the evaluation was situationally corrected.

[c] Regions that show increased activation during the encoding of behavioral information when affective evaluations were based on the behavior.

BA, Brodmann area; L, left; M, middle; R, right.

RESULTS AND DISCUSSION

We first compared the encoding of situational information on trials where participants subsequently made a situational attribution to trials where they subsequently made a dispositional attribution (i.e. discounted the situational information). The left DLPFC showed greater activation during the encoding of situational information that was taken into account for the attribution compared with situational information that was later discarded, $t(18) = 2.37$, $P = 0.03$. We then performed the same analysis for behavioral information, comparing activation during the encoding of behavioral information when participants later made dispositional attributions to activation when participants later made situational attributions. No other ROI showed differential activation as a function of subsequent attributions. A statistically significant interaction of *information type* (behavioral/situational) × *subsequent attribution* (dispositional/situational), $F(1, 18) = 5.2$, $P = 0.04$, confirmed that the BOLD difference in left DLPFC as a function of the subsequent attribution was specific to situational information (and not behavioral information), consistent with the predictions of dual-process theories of attribution.

To validate the primary ROI-based analysis and to explore whether additional brain regions are involved, we performed a second analysis using a whole-brain contrast comparing the presentation of situational information when it influenced the later attribution *vs* when it was discarded. Confirming the ROI analysis, this independent analysis again revealed an increase in BOLD response in left DLPFC (peak coordinates $x = -51$, $y = 32$, $z = 31$, Figure 2A and B) during situational information that was taken into account for the attribution. This difference was larger in participants with high Need for

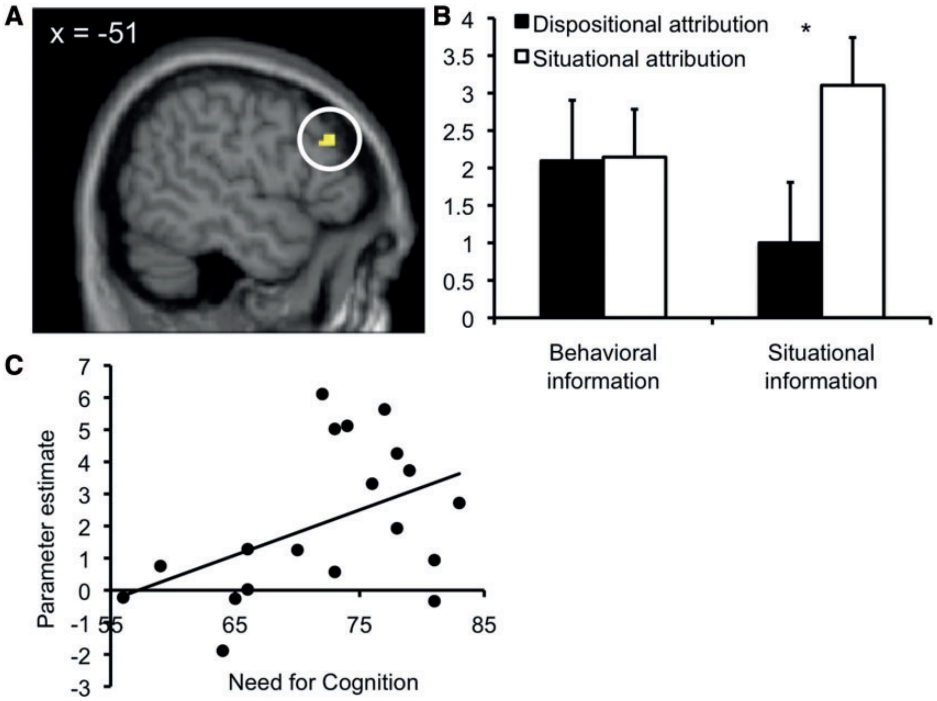

Fig. 2 Left DLPFC showed higher activation for situational information during trials that led to subsequent situational attributions compared with dispositional attributions. (**A**) Activation in left DLPFC [contrast (attribution-relevant situational information > attribution-irrelevant situational information), peak coordinates $x = -51$, $y = 32$, $z = 31$], (**B**) Parameter estimates (arbitrary units), (**C**) Correlation of parameter estimates in left DLPFC for contrast (attribution-relevant situational information > attribution-irrelevant situational information) with Need for Cognition.

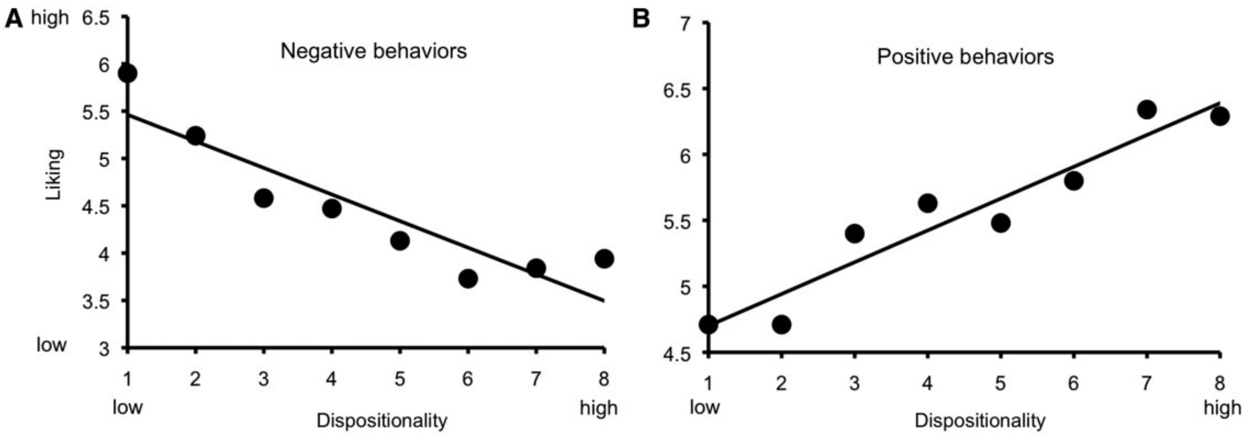

Fig. 3 Correlations between attribution and evaluation (figure shows the mean ratings for each level of attribution). (**A**) Negative behavior attributed to dispositional causes was associated with decreased liking, $r = -0.38$; (**B**) Positive behavior attributed to dispositional causes was associated with increased liking, $r = 0.31$.

Cognition, $r(18) = 0.46$, $P < 0.05$ (Figure 2C). The whole-brain analysis additionally revealed increased BOLD responses in a region of the DMPFC (peak coordinates $x = -6$, $y = 44$, $z = 40$, more lateral and anterior than the ROI reported in Table 1), right caudate (peak coordinates $x = 15$, $y = 20$, $z = 10$) and left middle frontal gyrus (peak coordinates $x = -21$, $y = 11$, $z = 43$) during situational information that was taken into account for the attribution. Again, no differences were found for a whole-brain contrast comparing the BOLD responses during the encoding of behavioral information as a function of subsequent attribution results.

Taken together, these findings support suggestions by cognitive dual-process theories of attribution (Trope, 1986; Gilbert *et al.*, 1988). Left DLPFC, a region linked to controlled, effortful processing and the inhibition of inappropriate automatic reactions (MacDonald *et al.*, 2000), showed increased BOLD response during the processing of situational information when participants subsequently included this information in their attributions, compared with when participants ignored or discarded it and made a dispositional attribution. This may reflect a controlled process subserving the integration of situational information into the attribution outcome.

THE LINK BETWEEN ATTRIBUTION AND AFFECTIVE EVALUATION

We predicted that the way we explain someone's behavior should have a strong impact on how we feel about that person. Using Pearson correlations, we found that dispositionality ratings and liking were positively correlated for positive behaviors, $r(303) = 0.31$, $P < 0.001$,

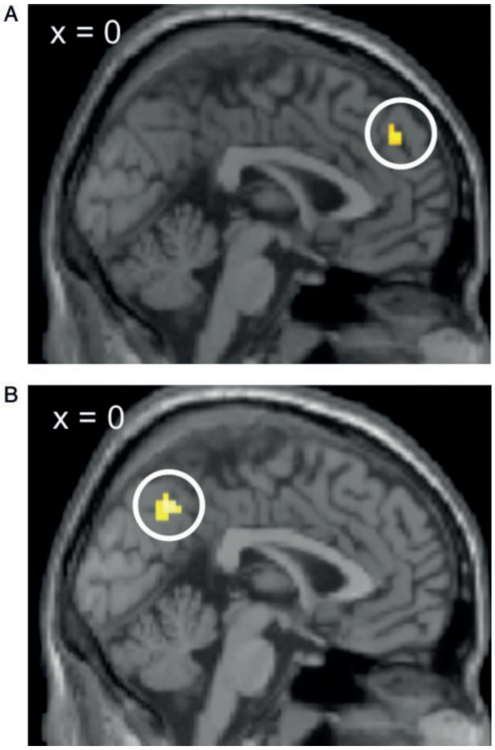

Fig. 4 Brain regions processing evaluation-relevant information. (**A**) Situational information, contrast (integrated situational information > discounted situational information): DMPFC showed increased activation during the processing of situational information that was integrated into the evaluation compared with situational information that was not. (**B**) Behavioral information, contrast (evaluation-relevant behavioral information > evaluation-irrelevant behavioral information): Precuneus, together with amygdala, TPJ and MTL (not shown), showed increased activation during the processing of evaluation-relevant behavioral information.

and negatively for negative behaviors, $r(303) = -0.38$, $P < 0.001$ (Figure 3). That is, positive behaviors that were attributed to dispositional factors were associated with higher liking than positive behaviors attributed to situational factors, whereas negative behaviors attributed to dispositional factors were associated with less liking than negative behaviors attributed to situational factors.

To identify how the neural processing of behavioral and situational information interacts with subsequent affective evaluations, we re-sorted the BOLD data to examine differential activation during the encoding of information that was consistent *vs* inconsistent with subsequent evaluations. We compared BOLD signal in trials where the valence of the evaluation was incongruent with the behavior (implying an integration of the situational information into the evaluation) *vs* trials where it was congruent (implying that the situational information was discarded for the evaluation). If subjects, for instance, disliked Mike the salesman (situation) although he smiled at Tom (positive behavior), they likely took into consideration the situation instead of assuming Mike is simply a nice guy (which may nevertheless be the case).

We first examined how situational information was processed when it had an effect on subsequent affective evaluations *vs* trials where it was ignored. In this analysis, only DMPFC showed increased BOLD signal to integrated situational information ($P = 0.03$). This finding was confirmed by a whole-brain analysis, but only at a lower threshold (peak coordinates $x = 0$, $y = 47$, $z = 37$, $P < 0.01$, minimal cluster size = 20 contiguous voxels, Figure 4A).

Thus, the behavioral ratings obtained during the experiment revealed a strong link between attributions and subsequent affective evaluations. At the neural level, DLPFC and DMPFC reflected the attribution process, with BOLD responses differing depending on whether situational information was included in the resulting attribution, or not. In addition, DMPFC reflected whether situational information was integrated into the evaluation, or not. Together, this pattern of results suggests a neural mechanism underlying the observed link between attribution outcomes and evaluation outcomes. DLPFC has been linked to controlled, effortful processing and the inhibition of inappropriate automatic reactions (MacDonald *et al.*, 2000). DLPFC activation may thus reflect a controlled integration mechanism that operates during the processing of situational information, determining whether situational information is taken into account during impression formation via modulations of DMPFC. Consistent with this proposed mechanism, a recent meta-analysis reported increased connectivity between DLPFC and DMPFC during mentalizing tasks (Gilbert *et al.*, 2010).

We also examined the BOLD responses to the encoding of behavioral information when it had an effect on subsequent evaluations *vs* trials when it was ignored. Here, we found that the right TPJ ($P = 0.03$) and precuneus ($P = 0.01$) showed BOLD increases for evaluation-congruent compared with incongruent behavioral information. In a previous study, we identified precuneus/posterior cingulate cortex and left amygdala as regions involved in the formation of rapid evaluations based only on behavioral information (Schiller *et al.*, 2009). Based on these findings, we performed an additional ROI analysis for the left amygdala using the coordinates from the previous study (mean betas for a sphere of 4-mm centered at the peak coordinates $x = -24$, $y = -7$, $z = -14$) and confirmed increased BOLD signal for evaluation-relevant behavioral information ($P = 0.04$). We complemented this ROI-based analysis with a whole-brain contrast comparing evaluation-relevant and evaluation-irrelevant behavioral information (Figure 4B). This analysis confirmed increased BOLD signal during evaluation-relevant behavioral information in TPJ (peak coordinates $x = 51$, $y = -58$, $z = 28$) and precuneus (peak coordinates $x = 0$, $y = -61$, $z = 43$) and additionally revealed increased BOLD signal in medial temporal gyrus (peak coordinates $x = 54$, $y = -13$, $z = -20$).

CONCLUSIONS

In this study, we provide new evidence for the manner in which information is encoded when deciding whether dispositions or situations are the causes of others' behaviors, and when using this information to evaluate others. Taken together, our data support suggestions by cognitive dual-process theories of attribution (Trope, 1986; Gilbert *et al.*, 1988), describing a controlled process subserving the integration of situational information into the attribution outcome. Our results suggest that the neural substrate underlying this process might be the DLPFC, as it showed increased activation only when the situational information was indeed integrated by the subjects (see also Lieberman *et al.*, 2002).

The absence or failure of this process, accompanied by less DLPFC activation, may play a role in the occurrence of the *Fundamental Attribution Error* (Ross, 1977) or *correspondence bias* (Gilbert and Malone, 1995), the pervasive tendency to make dispositional attributions when trying to explain the causes of other peoples' behavior. The fact that no brain regions showed relatively greater activation during the encoding of behavioral information when the attribution outcome was dispositional is consistent with an automatic dispositional inference that occurs during the encoding of behavioral information, regardless of whether the attribution is corrected for the situational information or not. Note that for experimental reasons, we selected our stimuli so that participants would make dispositional attributions in 50% of the cases. This is at odds with the behavior of people in real-life situation, where dispositional attributions are expected most of the time, but is necessary in order to compare equal numbers of trials where the situational information is taken into account *vs* trials where this is not the case.

Our findings demonstrate how top-down control processes regulate impression formation when situational information is taken into account to understand others.

REFERENCES

Cacioppo, J.T., Petty, R.E. (1982). The need for cognition. *Journal of Personality and Social Psychology*, *42*, 116–31.

Dweck, C.S., Leggett, E.L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–73.

Gawronski, B. (2004). Theory-based bias correction in dispositional inference: the fundamental attribution error is dead, long live the correspondence bias. *European Review of Social Psychology*, *15*, 183–217.

Gilbert, S.J., Gonen-Yaacovi, G., Benoit, R.G., Volle, E., Burgess, P.W. (2010). Distinct functional connectivity associated with lateral versus medial rostral prefrontal cortex: a meta-analysis. *Neuroimage*, *53*, 1359–67.

Gilbert, D.T., Malone, P.S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.

Gilbert, D.T., Pelham, B.W., Krull, D.S. (1988). On cognitive business: when person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*, 733–40.

Harris, L.T., Todorov, A., Fiske, S.T. (2005). Attributions on the brain: neuro-imaging dispositional inferences, beyond theory of mind. *Neuroimage*, *28*, 763–9.

Jones, E.E., Harris, V.A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, *3*, 1–24.

Kerpelman, J., Himmelfarb, S. (1971). Partial reinforcement effects in attitude acquisition and counterconditioning. *Journal of Personality and Social Psychology*, *19*, 301–5.

Lieberman, M.D., Cunningham, W.A. (2009). Type I and Type II error concerns in fMRI research: re-balancing the scale. *Social Cognitive and Affective Neuroscience*, *4*, 423–8.

Lieberman, M.D., Gaunt, R., Gilbert, D.T., Trope, Y. (2002). Reflexion and reflection: a social cognitive neuroscience approach to attributional inference. *Advances in Experimental Social Psychology*, *34*, 199–249.

MacDonald, A.W.3rd, Cohen, J.D., Stenger, V.A., Carter, C.S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, *288*, 1835–8.

Mitchell, J.P., Banaji, M.R., Neil Macrae, C. (2005a). General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage*, *28*, 757–62.

Mitchell, J.P., Banaji, M.R., Neil Macrae, C. (2005b). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience*, *17*, 1306–15.

Mitchell, J.P., Neil Macrae, C., Banaji, M.R. (2004). Encoding-specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience*, *24*, 4912–7.

Mitchell, J.P., Neil Macrae, C., Banaji, M.R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron*, *50*, 655–63.

Mitchell, J.P., Neil Macrae, C., Banaji, M.R. (2005c). Forming impressions of people versus inanimate objects: social-cognitive processing in the medial prefrontal cortex. *Neuroimage*, *26*, 251–7.

Ross, L. (1977). The intuitive psychologist and his shortcomings: distortions in the attribution process. In: Berkowitz, L., editor. *Advances in Experimental Social Psychology*, Vol. 10, New York: Academic Press, pp. 173–220.

Sabini, J., Siepmann, M., Stein, J. (2001). The really fundamental attribution error in social psychological research. *Psychological Inquiry*, *12*, 1–15.

Saxe, R., Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage*, *19*, 1835–42.

Schiller, D., Freeman, J.B., Mitchell, J.P., Uleman, J.S., Phelps, E.A. (2009). A neural mechanism of first impressions. *Nature Neuroscience*, *12*, 508–14.

Taylor, S.E., Fiske, S.T. (1978). Salience, attention and attribution: top of the head phenomena. In: Berkowitz, L., editor. *Advances in Experimental Social Psychology*, Vol. 11, Academic Press: New York, pp. 249–88.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*, 239–57.

Uleman, J.S., Newman, L.S., Moskowitz, G.B. (1996). People as flexible interpreters: evidence and issues from spontaneous trait inference. In: Zanna, M.P., editor. *Advances in Experimental Social Psychology*, Vol. 28, San Diego: Academic Press, pp. 211–79.

Van Overwalle, F. (2009). Social cognition and the brain: a meta-analysis. *Human Brain Mapping*, *30*, 829–58.

Van Overwalle, F., Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*, 564–84.