doi:10.1093/scan/nsr064 SCAN (2012) 7, 937–950

Inconsistencies in spontaneous and intentional trait inferences

Ning Ma,[1] Marie Vandekerckhove,[1] Kris Baetens,[1] Frank Van Overwalle,[1] Ruth Seurinck,[2] and Wim Fias[2,3]
[1]Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Bryussel, Belgium, [2]Ghent University, Dunantlaan 2, B-9000 Gent, Belgium, and
[3]GIfMI (Ghent Institute for Functional and Metabolic Imaging), Universitair Ziekenhuis Gent, De Pintelaan 185, B-9000 Gent, Belgium

This study explores the fMRI correlates of observers making trait inferences about other people under conflicting social cues. Participants were presented with several behavioral descriptions involving an agent that implied a particular trait. The last behavior was either consistent or inconsistent with the previously implied trait. This was done under instructions that elicited either spontaneous trait inferences ('read carefully') or intentional trait inferences ('infer a trait'). The results revealed that when the behavioral descriptions violated earlier trait implications, regardless of instruction, the medial prefrontal cortex (mPFC) was more strongly recruited as well as the domain-general conflict network including the posterior medial frontal cortex (pmFC) and the right prefrontal cortex (rPFC). These latter two areas were more strongly activated under intentional than spontaneous instructions. These findings suggest that when trait-relevant behavioral information is inconsistent, not only is activity increased in the mentalizing network responsible for trait processing, but control is also passed to a higher level conflict monitoring network in order to detect and resolve the contradiction.

Keywords: spontaneous trait inferences; person impression; person attribution; inconsistencies

INTRODUCTION

When meeting a novel person, we usually make quick impressions about the type of person he or she is, that is, about the traits or other stable characteristics of this person. Sometimes, we make relatively accurate evaluations on the basis of these rapid observations. However, information may sometimes reveal that we are incorrect, and that we have to revise our initial impressions. What happens when we detect judgment errors and have to correct our initial trait impression of others? Which brain areas are involved in this process? Most previous neuroscientific research explored the brain activity involved in trait inference about others when behavioral information was either consistent with previous trait expectations (Mitchell *et al.*, 2006; Ma *et al.*, 2011) or when no information was given (Moran *et al.*, 2009). To our knowledge, there are no fMRI studies that investigated neurological activity when behavioral information violated earlier trait expectations. This process is the topic of the present article. We also explore whether this process is alike for trait violations that are detected during spontaneous or intentional trait inferences.

Recent neuroimaging studies using functional magnetic resonance imaging (fMRI) suggest that social mentalizing, that is, mind reading about others, involves a brain network consisting of two core areas: the medial prefrontal cortex (mPFC) and the temporo-parietal junction (TPJ; see for a review, Van Overwalle, 2009). The TPJ is essential for processing temporary goals, intentions and beliefs of others (Saxe and Wexler, 2005; Saxe and Powell, 2006; for reviews, Saxe, 2006; Van Overwalle and Baetens, 2009). Several fMRI studies indicate that the TPJ or a nearby area—the posterior superior temporal sulcus (pSTS)—is also involved in trait attributions based on behavioral descriptions (Mitchell *et al.*, 2004, 2006; Harris *et al.*, 2005). The most essential area involved in all types of trait inference, based either on prior knowledge of familiar figures or on novel behavioral information, is the mPFC (Harris *et al.*, 2005; Mitchell *et al.*, 2005; Todorov *et al.*, 2007; Van Duynslaeger *et al.*, 2007; Ma *et al.*, 2011). The involvement of the mPFC is especially clear in trait inferences and is naturally supported by the retrieval of memory about the self and others (Vandekerckhove *et al.*, 2005).

Conflict detection and resolution

Given that the mPFC area is preferentially involved in the attribution of enduring traits, it seems very plausible to assume that this area is also recruited when a conflict is detected between previous trait impressions and novel behavioral information. But is this enough? Perhaps such a conflict would quickly surpass the computational capacities of the mPFC. In that case, control must be passed over to another higher level brain area to weight and decide among the conflicting social cues.

One likely candidate for higher level control in the case of contradictory cues is the domain-general conflict monitoring network (Cohen *et al.*, 2000), which involves the posterior

Received 29 March 2011; Accepted 13 September 2011
Advance Access publication 17 October 2011
This research was performed at GIfMI (Ghent Institute for Functional and Metabolic Imaging). An OZR-G.0650.10 Grant of the Vrije Universiteit Brussel (to F.V.O.).
Correspondence should be addressed to Frank Van Overwalle, Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, B - 1050 Brussel, Belgium. E-mail: frank.vanoverwalle@vub.ac.be

medial frontal cortex (pmFC) including the dorsal part of anterior cingulate cortex (dACC) and the lateral prefrontal cortex (lateral PFC). Roughly, the pmFC is located posterior to the 30 mm *y*-coordinate, while the mPFC is located anterior to it. The conflict monitoring network has been studied mainly with the aid of cognitive conflict tasks, including stroop and flanker tasks. The findings converge on the suggestion that the pmFC detects the conflict and engages the lateral PFC to resolve it (Cohen *et al.*, 2000; MacDonald *et al.*, 2000; Botvinick *et al.*, 2004; Kerns *et al.*, 2004; van Veen and Carter, 2006).

Is this domain-general conflict network also involved in social cognitive conflict? Mohanty *et al.* (2007) conducted a study to differentiate the cognitive and emotional role of the ACC and found that the dorsal ACC (i.e. part of the pmFC) is engaged more during cognitive conflicts in a standard Stroop task, while the mPFC is more involved in a Stroop task based on emotional words. Exploring conflicts in a more social context, Zaki *et al.* (2010) showed participants silent video clips portraying the facial expression of a speaker together with a short description of the topic of conversation. This topic either matched the valence of the facial expression or not (e.g. a smiling male face while speaking about the death of his dog). Results revealed that inconsistencies between these verbal and non-verbal social cues activated the pmFC and lateral PFC associated with domain-general conflict monitoring processes, and further increased activity in the mPFC associated with domain-specific mentalizing. Perhaps of more interest, Schiller and coworkers (Schiller *et al.*, 2009) provided mixed (positive and negative) behavioral descriptions and then requested participants to give an overall evaluation of the actor. Results showed that encoding of this information activated the pmFC and lateral PFC (domain-general conflict monitoring) and the mPFC (domain-specific mentalizing), together with a number of other areas. As far as we are aware, no prior fMRI study investigated conflicting behavioral information implicating personality traits. However, based on these prior findings, it seems plausible to suggest that domain-general conflict areas such as the pmFC and lateral PFC are involved when encoding inconsistent trait relevant information, together with the mPFC as core area responsible for mentalizing about traits.

Spontaneous and intentional trait inference and conflict

According to dual-process models in the social cognition literature, social information processing is based on two different systems: either automatic processing or controlled reasoning (Uleman, 1999; Smith and DeCoster, 2000; Satpute and Lieberman, 2006; Keysers and Gazzola, 2007). The automatic system is assumed to operate fast, and to automatically rely on prior knowledge and beliefs via associative links. In contrast, the controlled system would be slow and heavily demanding of people's computational resources (De Neys and Glumicic, 2008). With respect to social inference, spontaneous trait inferences (STI) are relatively automatic in the sense that they require little mental effort, and are difficult to suppress or modify, while intentional trait inferences (ITI) involve a deliberate attempt to make a relevant social judgment, which requires more mental effort (Uleman *et al.*, 2005).

To explore the role of a spontaneous *vs* intentional processing mode on trait inferences, Mitchell *et al.* (2006) conducted an fMRI study in which ITI instructions requested the participants 'to form an impression of the target individual' (p. 50) described in behavioral statements, while STI instructions asked participants 'to encode the order in which statements were paired with a particular individual' (p. 50). This study found that the mPFC was only marginally stronger activated under ITI than STI for trait-diagnostic descriptions. To avoid contamination of the intentional instruction on the spontaneous trials, Ma *et al.* (2011) conducted a between-participants study in which half of the participants were given spontaneous ('read carefully') instructions while the other half were given intentional ('infer the person's trait') instructions. The results indicated that the same core social mentalizing areas including mPFC and TPJ were recruited under spontaneous and intentional instructions.

While none of these fMRI studies explored trait inferences given conflicting social information, an event-related potential (ERP) study by Van Duynslaeger *et al.* (2007) investigated conflicting trait information under spontaneous and intentional instructions. The results from a LORETA source analysis on the ERP data indicated that during the stage when traits were inferred (at ~600 ms), the TPJ was strongly activated under spontaneous instructions regardless of either trait-consistent or trait-inconsistent information. More importantly, under intentional instructions, the mPFC was strongly recruited given trait-consistent information, while a large area in the medial paracentral frontal cortex (extending to the pmFC) was involved given trait-inconsistent information. This seems to suggest that trait conflicts may activate a domain-general conflict monitoring network only under intentional instructions to infer a trait, but not under spontaneous and occasional trait processing. However, given the rougher spatial resolution of LORETA, an fMRI study is needed to confirm the brain activation given conflicting trait information.

Present research and hypotheses

To investigate the brain areas involved in trait violations under spontaneous and intentional instructions, we applied fMRI to acquire a high spatial resolution. We used a modified version of the paradigm used in the ERP study by Van Duynslaeger *et al.* (2007) and the fMRI study by Ma *et al.* (2011). Participants were given behavioral descriptions that were either consistent or inconsistent with a previously inferred trait. Half of them were requested to make trait inferences about each target person (ITI), whereas the

other half was instructed to read the stimulus material carefully without mentioning anything about impression formation or trait inference (STI) to avoid contamination of the intentional instruction on spontaneous processing.

Based on the previous neural imaging research (Van Duynslaeger *et al.*, 2007; Ma *et al.*, 2011; for a review, Van Overwalle, 2009), we expect that the mentalizing areas including the TPJ and mPFC are recruited during both STI and ITI. However, their role remains unclear under trait violations. Several possibilities are open. Either the mentalizing network recruits more activity to resolve the inconsistency, the domain-general conflict network takes over control, or both processes take place interactively. As indicated by the ERP study of Van Duynslaeger *et al.* (2007), it seems most likely that the mentalizing network continues to operate under spontaneous processing, while the conflict network takes over under intentional processing.

Two memory measures were offered immediately after the presentation of all stimulus material, to validate the occurrence of trait inferences. These were (i) trait-cued recall in which participants have to remember the sentences read during scanning with a trait as cue to aid their memory (Winter and Uleman, 1984) and (ii) sentence completion in which participants have to complete the sentences with a critical word that strongly implies the trait (Bartholow *et al.*, 2001, 2003). Enhanced recall on these measures indicates that trait inferences were made while reading the behavioral sentences and were integrated with these sentences. For both memory measures, we predict better recall for diagnostic (i.e. consistent and inconsistent) trait-implying information as opposed to trait-irrelevant information, and even more so for trait-inconsistent behavior because resolving inconsistencies leads to more elaborate processing and hence deeper encoding in memory (Srull and Wyer, 1989; Stangor and McMillan, 1992).

METHODS

Participants

Participants were all right-handed, 16 women and 14 men, with ages varying between 18 and 43 years. In exchange for their participation, they were paid €10. Participants reported no abnormal neurological history and had normal or corrected-to-normal vision. Half of the participants (10 women and 5 men) received a spontaneous trait instruction (STI), while the other half (6 women and 9 men) received an intentional trait instruction (ITI). Informed consent was obtained in a manner approved by the Medical Ethics Committee at the Hospital of University of Ghent (where the study was conducted) and the Free University Brussels (of the principal investigator, F.V.O.).

Procedure and stimulus material

The design and stimulus material were borrowed from earlier studies on trait inference using fMRI (Ma *et al.*, 2011) and ERP (Van Duynslaeger *et al.*, 2007). Participants read a large number of events that described the behavior of a fictitious agent and from which a strong trait could be inferred. The events involved positive and negative moral traits. To avoid associations with familiar and/or existing names, fictitious 'Trek'-like names were used. There were 48 agents in this experiment. For each agent, a series of four or three behavioral sentences was presented (for 30 and 18 agents, respectively). Each sentence consisted of six words and was presented at once in the middle of the screen for a duration of 5.5 s. To optimize estimation of the event-related fMRI response, each sentence was separated by a variable interstimulus interval of 2.5–4.5 s randomly drawn from a uniform distribution, during which participants passively viewed a fixation crosshair. All agents were randomly presented, while all sentences involving the same agent were presented in a fixed order. To induce a trait expectation, all sentences except the last, implied the same trait by the description of a positive or negative moral behavior (e.g. 'Tolvan gave her brother a *compliment*' to induce the trait friendly). The last, critical sentence determined the degree of consistency with the previously inferred trait. This was determined by the last word in each sentence: trait consistent (TC), trait inconsistent (TI) or irrelevant (IRR). TC sentences described behaviors that were consistent with the inferred trait (e.g. 'Tolvan gave her sister a *hug*' is consistent with the trait friendly). TI sentences (e.g. 'Tolvan gave her mother a *slap*') are opposite to the inferred trait with respect to valence. IRR sentences described neutral behaviors (e.g. 'Tolvan gave her mother a *bottle*').

To make sure that the participants were attending to the task and instructions, they were informed that control questions would be asked during scanning. This was done after all behavioral sentences on an agent were presented, to avoid disruption of the trait inference process. Under STI, for about one-third of the agents, participants had to respond a control question asking whether the agent was a female or not by pressing a response key. Given that gender is automatically induced from pronouns during sentence reading, this control question interferes minimally with the spontaneous processes under study. Under ITI, the participants had to respond after the last trait-implying sentence whether the implied trait was correct or not, and after the last irrelevant sentence whether the agent was female or not. Note that the gender questions might draw attention to and possibly insinuate an importance of the gender of the agents when making trait inferences. Although these effects are probably minimal, we cannot exclude them entirely, which might be reason for some concern given that in the present sample women dominated the spontaneous condition (10 women) while men dominated the intentional trait instruction (9 men).

Immediately after leaving the scanner, the participants were given the cued recall and the sentence completion task in the same order for all participants. In the cued recall task, participants had to write as many behavioral

sentences as possible with the aid of words that consisted of the implied traits. In the sentence completion task, participants had to complete the last word of randomly selected incomplete sentences they had read during scanning.

Imaging procedure

Images were collected with a 3 T Magnetom Trio MRI scanner system (Siemens Medical Systems, Erlangen, Germany), using an 8-channel radiofrequency head coil. Stimuli were projected onto a screen at the end of the magnet bore that participants viewed by way of a mirror mounted on the head coil. Stimulus presentation was controlled by E-Prime 2.0 (www.pstnet.com/eprime; Psychology Software Tools) under Windows XP. Immediately prior to the experiment, participants completed a brief practice session. Foam cushions were placed within the head coil to minimize head movements. We first collect a high-resolution T1-weighted structural scan (MP-RAGE) followed by one functional run of 922 volume acquisitions (30-axial slices; 4-mm thick; 1 mm skip). Functional scanning used a gradient-echo echoplanar pulse sequence (TR = 2 s; TE = 33 ms; 3.5 × 3.5 × 4.0-mm in-plane resolution).

Image processing

The fMRI data were preprocessed and analyzed using SPM5 (Wellcome Department of Cognitive Neurology, London, UK). For each functional run, data were preprocessed to remove sources of noise and artifact. Functional data were corrected for differences in acquisition time between slices for each whole-brain volume, realigned within and across runs to correct for head movement, and coregistered with each participant's anatomical data. Functional data were then transformed into a standard anatomical space (2-mm isotropic voxels) based on the ICBM 152 brain template (Montreal Neurological Institute). Normalized data were then spatially smoothed [6-mm full-width-at-half-maximum (FWHM)] using a Gaussian kernel.

Statistical analysis

Whole-brain and ROI analysis

Statistical analyses were performed using the general linear model of SPM5 of which the event-related design is modeled with one regressor for each consistency condition using a canonical hemodynamic response function and its temporal derivative (with event duration set to the default of zero for all conditions) and six movement artifact regressors. The conditions were defined by the participants' individual onset times of the first two introductory TC sentences ($n = 24$), the last TI sentence ($n = 24$) and all the IRR sentences ($n = 36$). We choose the first two introductory TC sentences rather than the last TC sentence, to control for novelty. Indeed, the last TI sentence introduces an opposite trait that is moderately novel (i.e. same trait concept but difference valence). By taking the first two TC sentences, we also obtain moderate novelty for the TC condition on average (i.e. new trait concept for the first sentence, repetition of trait concept for second sentence). Note that an analysis with all TC sentences yielded approximately the same results. Comparisons of interest were implemented as linear contrasts using a random-effects model. To detect all relevant areas not only under intentional trait inference but also under the shallower, spontaneous processing of traits, a voxel-based statistical threshold of $P \leq 0.005$ (uncorrected) was used for all comparisons with a minimum cluster extent of 10 voxels (Lieberman and Cunningham, 2009; Ma *et al.*, 2011). Statistical comparisons between conditions were conducted using analysis of variance (ANOVA) procedures on the parameter estimates associated with each trial type.

Regions of interest (ROI) analyses were broadly determined, not only by the hypothesized regions but also by earlier findings (Ma *et al.*, 2011), which indicated that similar material containing concrete action verb (e.g. 'gives a slap') may also activate the mirror system concurrently with the mentalizing system. These ROI were performed with the small volume correction in SPM5, and were taken from the meta-analysis by Van Overwalle (2009) and Van Overwalle and Baetens (2009) as being involved in mentalizing: 0 −60 40 [precuneus (PC)], ±50 −55 10 (pSTS), ±50 −55 25 (TPJ), 0 50 5 [ventral mPFC (vmPFC)], 0 50 35 [dorsal mPFC (dmPFC)], ±45 5 −30 [temporal pole (TP)]; action understanding via mirror areas: ±40 −40 45 [anterior intraparietal sulcus (aIPS)], ±40 5 40 [premotor cortex (PMC)] and conflict monitoring: 0 20 45 (pmFC), ±40 25 20 (lateral PFC). The ROI involved a sphere of 15 mm radius around the centers (in MRI coordinates) of the areas mentioned above, except the pmFC and lateral PFC where a 20 mm radius was applied because these two regions are larger (Van Overwalle, 2009, 2011; Baumgartner *et al.*, 2011; Radke *et al.*, 2011). Significant ROI were identified using a threshold of $P < 0.05$, FWE corrected. In addition, the mean percentage signal change in each ROI was extracted using the MarsBar toolbox (http://marsbar.sourceforge.net) and analyzed using ANOVA and *t*-tests with a threshold of $P < 0.05$.

PPI analysis

Functional connectivity between activation in ROIs and other regions was assessed using a PPI analysis. This analysis tests the hypothesis that activity in one brain area can be explained by an interaction between the presence of a cognitive process and activity in another part of the brain. We conducted PPI analyses to estimate the functional connectivity with the conflict system (pmFC) and mentalizing system (dmPFC). Each of these regions was selected as seed region, which denotes activity within that region as the physiological regressor in the PPI analysis. Inconsistent trait processing (inconsistent > consistent) was the psychological regressor. A third regressor in the analysis represented the interaction between the first and second regressors.

The seed region was defined in the following manner (see also Burnett and Blakemore, 2009). First, as explained above, we created a ROI for each seed region based on a sphere with a radius of 20 mm (pmFC) and 15 mm (dmPFC) around the center. Next, for each participant, we set a threshold of $P < 0.05$ uncorrected (minimum voxel extent 4) *t*-contrast map for the inconsistent > consistent contrast. We then determined an individually tailored ROI by searching the nearest local maximum of each participant to the center of the seed region, and then created around it an individual ROI as a sphere with an 8 mm radius. For the pmFC as seed region, one data set did not contain a peak surviving at $P < 0.05$ (uncorrected) and was excluded from the spontaneous group for the subsequent PPI analysis. For the dmPFC as seed region, we excluded three participants from the intentional group and four participants from the spontaneous group. Finally, we collected the results from all single-participant inconsistent > consistent contrasts to conduct a group-level analysis. A threshold of $P < 0.05$ (minimum voxel extent 4) with FWE correction was applied for the group-level analysis (Penny *et al.*, 2003; Burnett and Blakemore, 2009). To make sure that somewhat less strong connectivity would not go unnoticed, we also applied the same threshed with a less stringent cluster-wise correction.

Relationship with behavioral measures

To verify the relationship between brain activation during inconsistent trait processing and behavior, we conducted two additional analyses. First, we tested the relationship between brain activation and participants' agreement with the implied trait given mixed information, as measured during intentional instructions (trait ratings). To do so, in the inconsistent condition, we compared brain activity when participants agreed with the trait implied in the final (inconsistent) behavioral descriptions *vs* the trait implied in the initial (consistent) description. Second, we tested the relationship between memory for behavioral information (i.e. sentence completion) and brain activation during inconsistent trait inferences under both intentional and spontaneous conditions. We compared brain activation when the last, inconsistent behavioral item was remembered correctly *vs* incorrectly. For these two analyses, we used the same thresholds as for the whole-brain and ROI analysis mentioned above.

RESULTS

Behavioral data

Trait ratings

Under intentional instructions, participants were asked whether they agreed or not with the implied trait. The results of the inconsistency condition showed that 62% of the trait ratings were in agreement with the earlier (consistent) information, while only 25% were in agreement with the final (inconsistent) information. In the other 12%, no response was given. An ANOVA with consistency as within-participants factor (consistent *vs* inconsistent) revealed that this difference was significant, $F(1,14) = 24.95$, $P < 0.001$.

Memory measures

In order to make sure that trait inferences were made not only under intentional instructions, but also under spontaneous instructions, we analyzed the memory measures (Table 1). Our prediction was that if traits are inferred during reading of the sentences, then these traits would be stored in memory together with the sentences and so facilitate (i) recall of the sentences by the aid of a trait cue and (ii) recall of the critical words in the sentences that induce a trait. We conducted an ANOVA with instruction (spontaneous *vs* intentional) as between-participants factor and consistency (TC, TI and IRR) as within-participants factor. For trait-cued recall, the analysis revealed a significant main effect of consistency, $F(2,56) = 13.23$, $P < 0.001$ and instruction, $F(1,28) = 7.58$, $P < 0.01$, as well as a significant interaction, $F(2,56) = 8.13$, $P < 0.001$. A post hoc LSD test revealed the predicted differences between all types of consistency under intentional instructions (TI > TC > IRR; see Table 1) but revealed no differences not under spontaneous instructions. For sentence completion, the analysis revealed a significant main effect of consistency, $F(2,56) = 18.04$, $P < 0.001$, and instruction, $F(1,28) = 9.29$, $P < 0.005$ and again their interaction, $F(2,56) = 4.26$, $P < 0.05$. A post hoc LSD test showed better memory for TI sentences as opposed to IRR sentences under both instructions, and also for TC sentences under ITI (Table 1). The results of these memory measures suggest that, consistent with expectations, intentional instructions show that both consistent and inconsistent traits were inferred. However, the STIs, although showing the predicted trend, were unreliable for consistent trait-implying behaviors, and reached significance only for inconsistent trait implications. Presumably, the presence of inconsistent trait-implying behaviors focused participants predominantly to the potential conflict, and as such they tended to neglect the smaller differences between trait-implying and trait-irrelevant descriptions under spontaneous processing.

Table 1 Memory (per cent correct) as a function of instruction and trait consistency

	Spontaneous			Intentional		
	Consistent	Inconsistent	Irrelevant	Consistent	Inconsistent	Irrelevant
Cued recall	5_c	7_c	4_c	9_b	23_a	2_c
Sentence completion	7_c	10_b	1_c	24_a	32_a	3_c

Means in a row sharing the same subscript do not differ significantly from each other according to a Fisher's LSD test, $P < 0.05$.

Imaging data

We concentrate the analysis on our main question, that is, the role of inconsistent trait inferences by contrasting the inconsistent > consistent conditions. In the interest of brevity, we report only on those areas that reached a conventional level of significance of $P < 0.05$ (after FWE correction) either in the whole-brain, random-effects analysis or in the ROI analysis focusing on the mentalizing, mirror and conflict monitoring networks.

Under ITI, the inconsistent > consistent contrast revealed significant activation in a large number of brain regions, including the left pSTS, dorsal mPFC and precuneus related to the mentalizing network, as well as the right aIPS and right PMC related to the mirror network. In addition, areas involved in general conflict monitoring were also recruited, including the pmFC and right PFC. Under STI, basically the same networks were activated, although generally to a somewhat lesser degree, except for the larger activation of the dorsal and ventral mPFC (Figure 1 & Table 2). The activation included parts of the mentalizing network (right TPJ extending to right pSTS and mPFC), the mirror network (right aIPS and right PMC), as well as conflict monitoring (pmFC and right PFC).

To explore the joint activation between instructions, we ran a conjunction analysis of the inconsistent > consistent contrast under ITI and STI (also with a whole-brain threshold of $P < 0.005$). Joint activation was found significant in the dorsal mPFC, but failed to reach conventional levels of significance in the pmFC and right PFC ($P = 0.17$, FWE corrected). This suggests that areas responsible for trait inference (dorsal mPFC) share common brain activation under the two instructions, but that areas involved in conflict monitoring (pmFC and PFC) share an overlap that is relatively small and unreliable.

Next, we analyzed the critical difference between spontaneous and intentional instructions for the

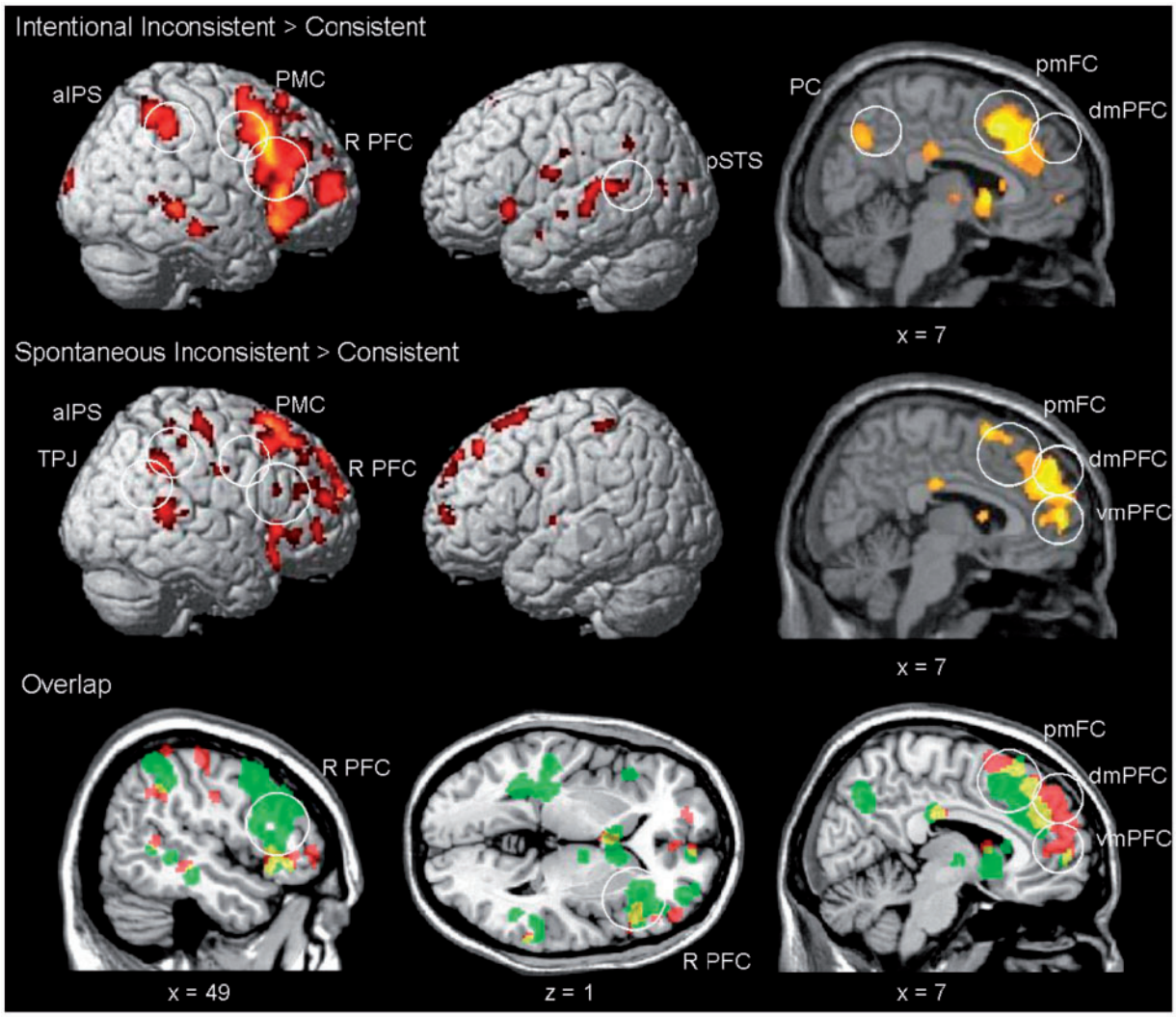

Fig. 1 The inconsistent > consistent contrast under spontaneous and intentional instructions, and their overlap. Whole-brain activation thresholded at $P < 0.005$ (uncorrected) with at least 10 voxels. Circles indicate ROIs with significant activation. The overlap was created using MRIcroN, showing selected areas under intentional (red) or spontaneous (green) instructions at the same whole-brain threshold $P < 0.005$, and their overlap (yellow).

Table 2 Peak voxel, number of voxels and *t*-value of the inconsistent > consistent contrasts from the ROI analysis and additional regions of the whole-brain analysis (other regions)

Anatomical label	*x*	*y*	*z*	Voxels	Max *t*
Intentional: inconsistent > consistent					
ROI					
Precuneus	10	−62	38	268	4.10**
R pSTS	46	−48	4	14	3.12
L pSTS	−36	−50	8	243	3.95**
L TPJ	−38	−50	34	36	3.34
R aIPS	48	−40	44	333	3.92**
L aIPS	−38	−50	36	22	3.23
R PMC	46	10	46	671	4.32**
Dorsal mPFC	4	38	34	157	3.94**
Ventral mPFC	10	56	0	30	3.19
pmFC	8	18	46	1065	5.25**
R PFC (inferior frontal gyrus)	34	24	2	1398	5.99**
Other regions					
R occipital	18	−100	12	52	4.20
L pyramis	−12	−72	−38	122	3.79
L occipital	−28	−70	14	2741	4.56
L supramaginal	−38	−50	34	76	3.34
R mid-temporal	52	−38	−4	173	3.39
R sub-gyral	50	−26	−14	139	3.72
R cingulate	4	−22	32	74	3.67
R caudate	8	12	0	482	4.95*
L inferior frontal	−44	20	−2	81	4.10
R frontal gyrus	34	52	4	447	4.07
Spontaneous: inconsistent > consistent					
ROI					
R pSTS	52	−46	8	59	3.19
R TPJ	52	−46	38	72	3.80**
R aIPS	52	−44	38	21	3.82**
L aIPS	−44	−32	58	11	2.96
R PMC	34	18	46	27	3.83**
Dorsal mPFC	4	46	28	799	5.64**
Ventral mPFC	6	56	10	392	4.67**
pmFC	12	16	60	329	5.44**
R PFC (frontal gyrus)	30	36	22	74	3.72*
Other regions					
L posterior cingulate	−2	−44	24	64	3.32
R cingulate	2	−18	30	148	4.32
R postcentral	50	−18	56	110	3.59
R ventricle	−2	6	6	69	3.45
R inferior frontal gyrus	36	20	−14	252	3.61
R inferior frontal gyrus	46	48	0	69	3.58
Conjunction analysis: inconsistent > consistent					
ROI					
Dorsal mPFC	4	42	32	120	3.77**
pmFC	10	26	56	55	3.44
R PFC (inferior frontal gyrus)	42	24	2	22	3.45
Intentional > spontaneous: inconsistent > consistent					
ROI					
L pSTS	−46	−48	10	224	3.37
L TPJ	−48	−50	12	29	3.30
R PMC	44	2	40	13	3.14
pmFC	8	14	48	216	4.16**
R PFC (extending to insula)	32	24	2	65	4.62**
Spontaneous > intentional: inconsistent > consistent					
ROI					
R aIPS	40	−36	56	16	2.95
Dorsal mPFC	−12	46	42	15	2.99

Coordinates refer to the MNI (Montreal Neurological Institute) stereotaxic space. RIOs are spheres with 15 mm radius around 0 −60 40 (PC), ±50 −55 10 (pSTS), ±50 −55 25 (TPJ), ±45 5 −30 (TP), ±40 −40 45 (aIPS), ±40 5 40 (PMC), 0 50 5 (vmPFC), 0 50 35 (dmPFC) and 20 mm radius around 0 20 45 (pmFC) and ±40 25 20 (lateral PFC). All regions thresholded at $P < 0.005$.

R, right; L, left.

*$P < 0.10$, **$P < 0.05$, FWE corrected. For other regions corrected after whole-brain analysis and for ROIs corrected after small volume analysis.

inconsistent > consistent contrast. An ANOVA with Consistency (inconsistent *vs* consistent) and Instruction (STI *vs* ITI) as factors revealed that the pmFC and right PFC were significantly more strongly recruited under ITI than STI, indicating that inconsistencies recruit more brain activity in the monitoring network especially under intentional instructions.

In summary, the whole-brain and ROI analyses indicate that basically the same neural networks were recruited by ITI and STI. Of most importance, we found that the areas responsible for conflict monitoring (pmFC and right PFC) were strongly recruited when there was a violation of a previously implied trait. Contrary to our hypothesis, this activation occurred not only under ITI, but also under STI, although to a somewhat lesser extent.

Signal change estimations

We also extracted percentage signal change estimations from each ROI in order to explore the potential differences between instructions in more depth (Figure 2). Note that while the prior ROI analysis considers only the active part of the ROI above threshold, this signal change analysis takes the whole ROI into consideration. We conducted an ANOVA with inconsistency (inconsistent *vs* consistent), instruction (STI *vs* ITI) and region as factors. An effect of instruction on the processing of trait inconsistencies should be revealed by an interaction between instruction and inconsistency. The analysis revealed a main effect of instruction, $F(1,28) = 10.54$, $P < 0.01$ and of inconsistency, $F(1,28) = 4.22$, $P < 0.05$. In addition, there was also the predicted interaction between instruction and inconsistency, $F(1,28) = 9.48$, $P < 0.005$, which was further modulated by a three-way interaction with region, $F(15,420) = 5.36$, $P < 0.001$. As can be seen in Figure 2, under ITI, trait inconsistencies increased activation significantly ($P < 0.05$) in the right hemisphere in most relevant areas of the mentalizing, mirror and conflict networks (pSTS, TPJ, aIPS, PMC and PFC), as well as in the pmFC, while none of these differences reached significance under STI (although some of these areas did reach significance in the ROI analysis reported above). Surprisingly, this pattern was reversed for the activation of the (ventral and dorsal) mPFC, as this area was more strongly activated under STI than ITI given trait inconsistencies. The other areas were not modulated by the interaction between instruction and inconsistency.

Taken together, the signal change data appear to be more sensitive to differences between instructions as they point to more difference than revealed earlier by the standard ROI analysis. It confirms the strong role in dealing with trait violations of mentalizing and mirror areas in the right hemisphere, as well as the role of the conflict areas, especially under intentional instructions. In contrast, the core area involved in mentalizing about traits (mPFC) appears to more activated under spontaneous processing.

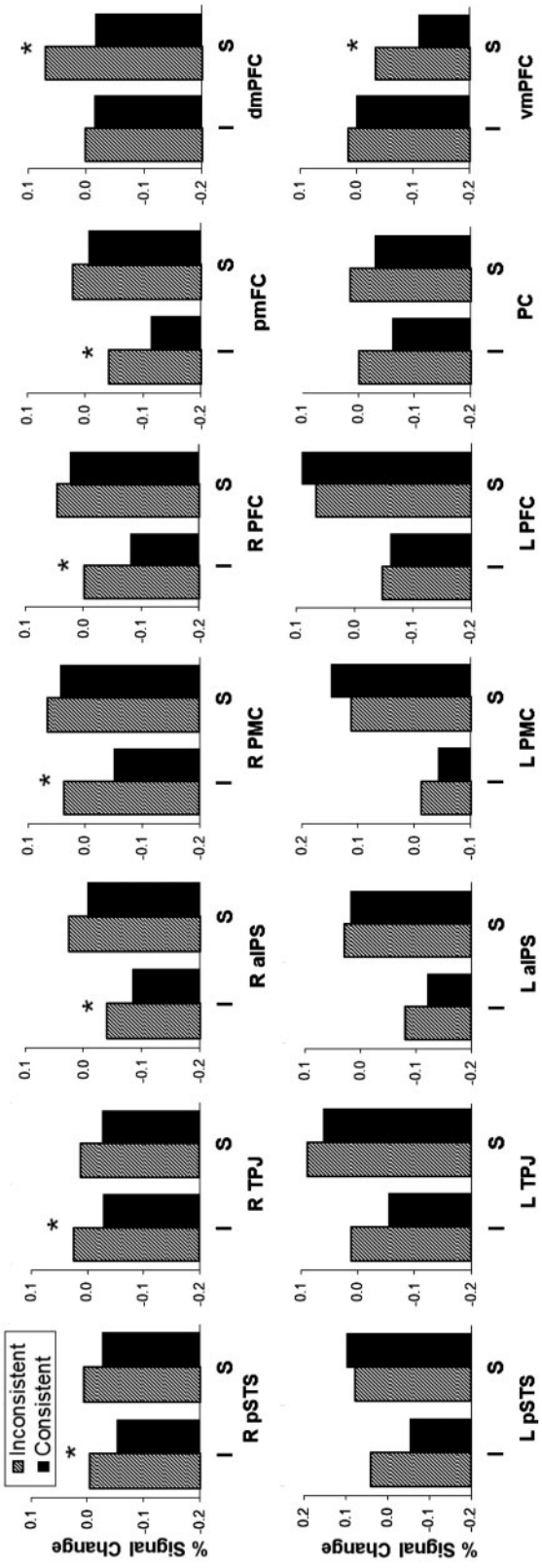

Fig. 2 Percentage signal change based on ROIs with 15 or 20 mm sphere created by Marsbar in function of instruction and consistency. S, spontaneous; I, intentional. Asterisks denote significant differences between consistent and inconsistent conditions by an LSD test, $P < 0.05$, for ROI that revealed a significant interaction between consistency and instruction.

Table 3 Results of PPI analysis with the seed regions from pmFC and dmPFC on spontaneous inconsistent trait processing (inconsistent > consistent) from the ROI analysis and additional regions of the whole-brain analysis (other regions)

	x	*y*	*z*	Voxels	Max *t*
Seed region in pmFC: inconsistent > consistent					
ROI					
L pSTS	−52	−48	12	165	5.35**$_a$
L TPJ	−52	−50	12	94	5.09*$_a$
L PMC	−40	14	30	167	5.77**$_a$
pmFC	−6	10	62	249	4.59**
L PFC	−46	18	24	792	5.99**$_a$
Other regions					
R occipital	12	−82	12	2506	5.5**
Seed region in dmPFC: inconsistent > consistent					
ROI					
L TPJ/ L pSTS	−46	−52	16	90	3.7*
L PMC	−48	2	50	133	4.07**
L PFC	−46	30	10	195	3.93**
Other regions					
L occipital	−16	−78	4	4046	6.62**
L fusiform	−30	−64	−24	285	4.81*
L inferior frontal	−48	32	−12	358	4.38**

ROI are spheres with 15 mm radius around ±50 −55 10 (pSTS), ±50 −55 25 (TPJ), ±40 5 40 (PMC), 0 50 5 (vmPFC), 0 50 35 (dmPFC) and 20 mm radius around 0 20 45 (pmFC) and ±40 25 20 (lateral PFC). All regions thresholded at $P < 0.01$.R, right; L, left.
*$P < 0.10$, **$P < 0.05$, cluster-wise corrected; $_aP < 0.10$, FWE corrected. For other regions corrected after whole-brain analysis and for ROIs corrected after small volume analysis.

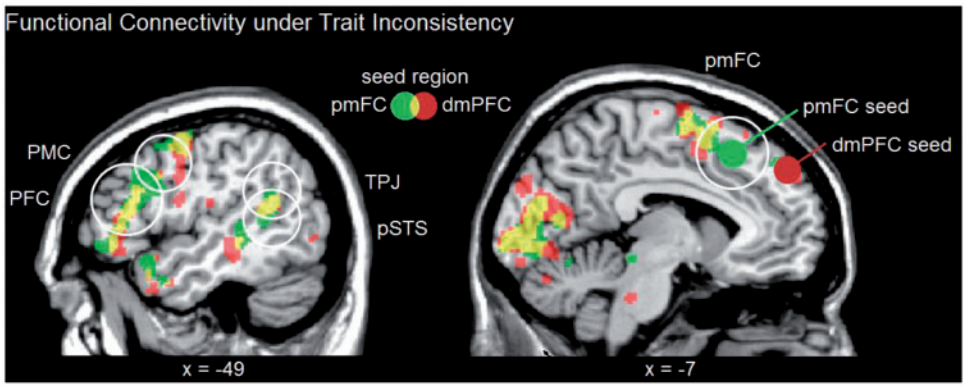

Fig. 3 Shared connectivity during spontaneous trait inconsistency from pmFC and dmPFC. Increases in functional connectivity during trait inconsistencies compared with trait consistencies (inconsistent > consistent) from the pmFC (green voxels) and dmPFC (red voxels). Yellow voxels denote overlapping connectivity among these two seed regions. Note that the location of the seed regions is approximate, as it was individually tailored to the highest activation of each individual within the ROI.

Functional connectivity

To explore how mentalizing and conflict monitoring networks interact with each other and with other brain areas under inconsistent trait processing, we explored their functional connectivity by selecting one area that has the highest level of control or abstractness in each network (Botvinick *et al.* 2004; Van Overwalle, 2011), that is, the pmFC (conflict monitoring) and the dmPFC (trait mentalizing). Specifically, we ran a PPI analysis, taking the pmFC and dmPFC as seed region under the inconsistent > consistent contrast. Under the intentional instructions, the PPI analysis did not reveal any significant interaction with other brain areas. In contrast, under spontaneous processing, the PPI analysis revealed a significant interaction of the seed regions (pmFC and dmPFC) with each other, as well as with other social brain areas, resulting in an extensive shared overlap including the left TPJ (extending to the left pSTS), left PMC and left PFC. Occipital areas were also involved (Table 3 and Figure 3).

Relationship between fMRI and behavioral data

To get more insight in the processes underlying mixed information processing and inconsistency resolution, we explored whether the experienced conflict lead to downstream behavioral decisions. We analyzed the relationship between fMRI activation and trait ratings (intentional

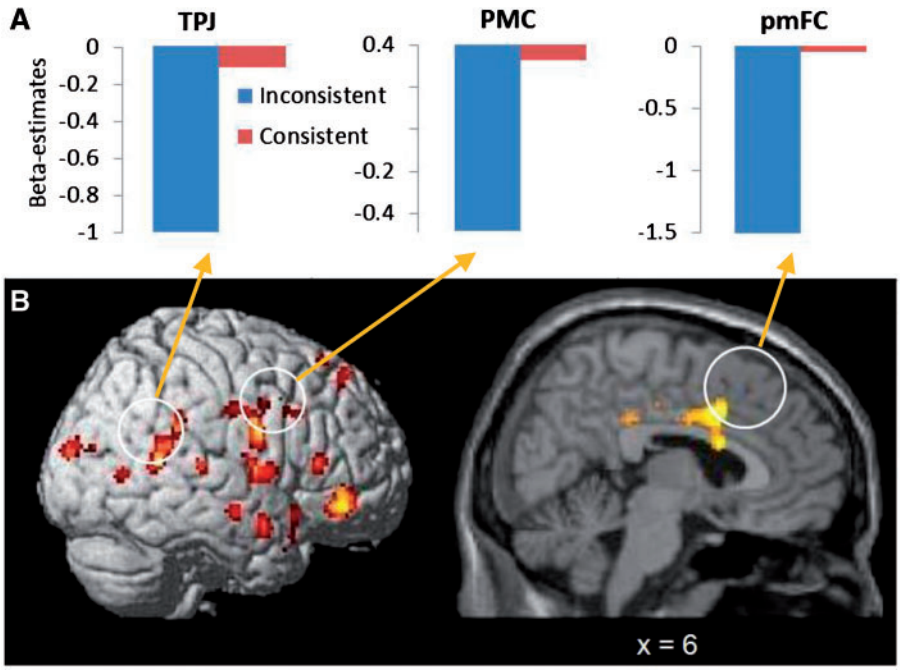

Fig. 4 Relationship between behavioral descriptions and trait ratings under intentional instructions. (**A**) β-estimates at peak coordinates for significant ROIs in function of trait ratings in the inconsistency condition that were in agreement with initial (consistent) > final (inconsistent) behavioral descriptions (no ROIs were significant in the opposite contrast). (**B**) View of whole-brain activation thresholded at $P < 0.005$ (uncorrected) with at least 10 voxels. Circles indicate ROIs showing significant activation in the right TPJ (MRI peak coordinates: 46 −40 26; $P < 0.10$, FWE corrected), right PMC (36 2 34, $P < 0.05$) and pmFC (6 8 34; $P < 0.05$). Note that under spontaneous instructions, ratings were not requested.

instructions only) as well as with memory for behavioral information (i.e. sentence completion; both instructions).

Relationship with trait ratings

To analyze how behavioral information contributes to an trait impression about the target, we contrasted brain activity for trait ratings that agreed with the trait implied by the initial (consistent) *vs* the final (inconsistent) descriptions, a procedure analogous to prior research (Schiller *et al.*, 2009; Zaki *et al.*, 2010). Higher activation in the consistent > inconsistent contrast was revealed in the right TPJ, right PMC and pmFC, consistent with the role of the mentalizing, mirror and conflict networks in building a coherent trait impression of the actor (Figure 4). This suggests that inconsistencies were resolved mainly by focusing on and using consistent descriptions, consistent with the behavioral data showing that 62% of the trait ratings were in agreement with the consistent information. No ROIs were significant in the opposite contrast, also in line with the behavioral data that only 25% of the ratings were based on the inconsistent information.

Relationship with memory

Inconsistent behavioral descriptions that were followed by correct *vs* incorrect memory in the sentence completion task were compared. There was higher activation in the dmPFC and vmPFC given correct > incorrect memory given spontaneous instructions (Figure 5), while there was no significant difference under intentional instructions. This is consistent with the fMRI analysis above, where we found the strongest activation in the same parts of the mPFC under spontaneous instructions. The present analyses add to this finding that this increased activation leads to higher recall of the crucial information implicating the actor's trait.

DISCUSSION

Perceivers integrate complex and sometimes conflicting social information into a coherent view of what others are like, and what traits they possess. Although the neural bases of cognitive conflict resolution have been extensively investigated, the mechanisms underlying the resolution of social conflicting information still remain unclear. The present study addressed this issue by exploring perceivers' STI and ITI about a social agent based on inconsistent social behavioral descriptions. Previous work demonstrated that verbal behavioral descriptions preferentially engage areas of the mentalizing network (Van Overwalle, 2009), while cognitive conflicts (e.g. Stroop task) engage the conflict monitoring network (Botvinick *et al.*, 2004). Based on this evidence, we hypothesized that social cognitive conflict would engage brain areas responsive to domain-specific mentalizing as well as domain-general cognitive control.

In line with our hypotheses, inconsistent trait-implying behavioral information increased the activation of the

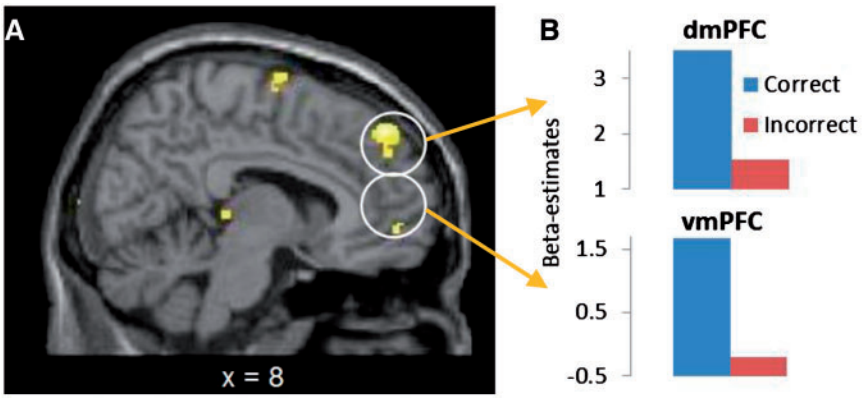

Fig. 5 Relationship between inconsistent behavioral descriptions and memory (sentence completion) under spontaneous instructions. (**A**) Midline view of whole-brain activation thresholded at $P < 0.005$ (uncorrected) with at least 10 voxels. Circles indicate ROIs showing significant activation in the dmPFC (MRI peak coordinates: 8 48 46) and vmPFC (8 54 −6; P's < 0.05, FWE corrected). There were no significant ROIs under intentional instructions. (**B**) β-estimates at peak coordinates for the significant ROIs in function of correct *vs* incorrect memory (spontaneous instructions).

mPFC, a core mentalizing area responsible for trait inference (Van Overwalle, 2009). This occurred regardless of a spontaneous or intentional instruction, although it was stronger under spontaneous processing. Crucially, the domain-general conflict monitoring network consisting of the pmFC and lateral PFC was also recruited (Botvinick *et al.*, 2004). The engagement of the conflict monitoring network did occur not only under ITI as hypothesized (see also Van Duynslaeger *et al.*, 2007), but also under STI, although to a somewhat weaker degree. A functional connectivity analysis provided further insight in how trait inconsistencies were resolved. Under spontaneous instructions, there was an increased interaction between mentalizing and conflict monitoring networks, showing an extensive set of overlapping areas between the pmFC and dmPFC seed regions (assumed to be at the highest level of control, see Botvinick *et al.*, 2004; Van Overwalle, 2011) and additional areas of conflict monitoring (left PFC), mentalizing (left TPJ) and mirroring (left PMC). This demonstrates that the connectivity is largely shared and reciprocal. This suggests that the failure of the mentalizing system to provide a coherent trait impression even after increased activity automatically triggers a domain-general network involved in conflict resolution. In addition, this conflict network directs behavior by modulating the engagement of the mentalizing network (Zaki *et al.*, 2010). This results in a cycle of forward triggering of the conflict network (when mentalizing fails) and backward biasing by the conflict network of the mentalizing circuit.

However, we found this reciprocal collaboration only under spontaneous instructions, but not under intentional instructions. There was no brain activity related to any of the two seed regions. That this collaboration existed under spontaneous processing suggests that the inconsistent information was salient enough to interrupt spontaneous trait processing, leading to more deliberate conflict resolution. In contrast, the lack of such collaboration under intentional processing, together with the almost negligible activation of the mPFC, suggests that control was quickly passed to the domain-general conflict network. Stated differently, the intentional instruction may have made the inconsistency very salient (increasing pmFC activity) and so may have prevented participants to make robust trait inferences (reducing mPFC activity). A related explanation is that the elaborate and deeper processing of the discrepancy under intentional instructions might have resulted in many individually tailored solutions to resolve the discrepancy, precluding a convergence in the connectivity pattern between brain areas.

To our knowledge, the present research is the first fMRI study that explores conflict detection and resolution while making trait inferences, under both spontaneous and intentional processing. Our results demonstrate that the mPFC is a core area not only of trait inference (as revealed in earlier research), but also of trait conflict resolution. Presumably, it increases its activation, perhaps because of enhanced attempts at understanding the trait implications of the behavioral inconsistency. More importantly, our results show that contradictory trait-relevant information also recruits a domain-general conflict network (Botvinick *et al.*, 2004), presumably because the mentalizing system is unable to resolve the inconsistency on its own.

Convergent neuroscientific evidence suggests that the pmFC and lateral PFC are involved in conflict monitoring and resolution (Van Veen and Carter, 2002, 2006; Botvinick *et al.*, 2004; Kerns *et al.*, 2004; Carter and Van Veen, 2007). Recently, Zaki *et al.* (2010) and Van Duynslaeger *et al.* (2007) extended the study of conflict processing into social cognition. Zaki and his coworkers (2010) studied emotional conflict as a consequence of contradictory facial expressions and emotion-eliciting events. They found that the TPJ and mPFC were more strongly engaged when participants relied on the contextual event rather than the face during decisions about incongruent emotional cues. Van Duynslaeger *et al.* (2007) conducted an ERP study to explore the neural

correlates of trait inferences given inconsistent behavioral information. Their results revealed an engagement of a large area that extended to the pmFC when behavioral information contradicted traits implied by previous behavioral descriptions.

There is also growing evidence indicating that the pmFC and lateral PFC are activated not only during explicit cognitive conflict processing (Kerns *et al.*, 2004; Mulert *et al.*, 2005) but also under spontaneous monitoring (Ursu *et al.*, 2009). This is in line with behavioral research indicating that the conflict monitoring process operates flawlessly under automatic and controlled reasoning (De Neys and Glumicic, 2008). Combined with the current results, this suggests that conflicting information in a social context is often detected and resolved in an automatic (spontaneous) and controlled (intentional) manner alike, even though the conflict network is less actively recruited under spontaneous processing. However, the results do not necessarily suggest that perceivers are unaware of the conflict. Although the inconsistency might initially go unnoticed, the activation of the lateral PFC suggests that some level of awareness is present, as this area is crucial for deliberating processing (Lieberman, 2007; Mason and Morris, 2010). Hence, automatic conflict detection may be interrupted by awareness of the conflict and controlled reasoning (De Neys *et al.*, 2006, 2008; Evans, 2007, 2008). The present finding that STI were better remembered than an irrelevant baseline only after inconsistent information (but not after consistent information) seems to support this suggestion.

Additional analyses seeking evidence for relationships between fMRI and behavioral data shed some light on how mixed information was processed and guided judgments. Under intentional instructions, we found that mixed descriptions were most often resolved by agreeing with the trait implications of the consistent descriptions, although about one-fourth of the discrepancies were resolved by agreeing with the inconsistent information. The relationship analysis further revealed that when participants relied more on inconsistent than consistent information in making their trait ratings, this did not reveal any brain activation, in line with the limited use of this information. However, when participants relied more on consistent than inconsistent information in making trait judgments, this recruited two areas related to goal identification (TPJ and PMC) and conflict detection (pmFC). No functional association was found with the mPFC, but this is not surprising since that area was relatively small in the intentional condition. Our results are consistent with Zaki *et al.* (2010) who reported that when perceivers relied more on behavioral as opposed to facial cues in making judgments; they increasingly recruited the mPFC and TPJ. None of our areas were documented by Schiller *et al.* (2009), most likely because their analysis was focused on the encoding of consistent information in the face of inconsistencies (i.e. response to the information that lead to the liking or disliking of a person).

Under intentional reasoning, we found no correlation between brain activation during encoding the last, inconsistent behavioral descriptions and memory after scanning. A possible reason is that an explicit instruction to make a trait inference confronts the participants with the inconsistencies and necessity to resolve them, and hence forces them to reconsider all relevant information including prior information (so that memory for the last item weakens). In contrast, under spontaneous instructions, increased activation of the mPFC was associated with better memory for key words from the descriptions. This may perhaps suggest that trait inferences that recruit this area, relied more on inconsistent information. This unexpected finding is reported for the first time, because prior research was silent on the issue of inconsistency resolution during spontaneous social processing. How might this spontaneous inconsistency process evolve? When inconsistencies are detected, perhaps, observers might search in memory for similar schemata that may fit the inconsistency, or they may simply favor the last, inconsistent descriptions over earlier information. According to the model proposed by Botvinick *et al.* (2004), the detection of the conflict is related to the enhanced activation of the pmFC, while the increased load in working memory explains the enhanced activation of the lateral PFC. Because activation in the lateral PFC was actually lower during spontaneous instructions, it is unlikely that a controlled search process was triggered under spontaneous processing. Consequently, it seems more likely that inferences were made by favoring the last, inconsistent information. Given that memory after the experiment is not a very reliable measure to understand how perceivers reconcile conflicting information, future research should confirm the present findings with on-line memory paradigms (i.e. during scanning) to measure the content of spontaneous inferences.

Some differences between spontaneous and intentional processing in other brain areas besides the conflict network were revealed after trait violations. The strongest differences were found given the signal change analysis. The findings suggest, in line with the interpretation offered earlier, that under spontaneous processing the mentalizing system (mPFC) attempts at resolving the inconsistency, but when that attempt fails, control is passed to more deliberate reasoning, which involves besides the conflict areas, most relevant mentalizing and mirror areas at the right hemisphere. Interestingly, the present results point to a novel role of the mirror network. This network is typically activated when images are presented on human body movements (Van Overwalle and Baetens, 2009), but also when such information is verbally presented in a very vivid and concrete manner (e.g. 'slapped' his father) as in the present material (Tettamanti *et al.*, 2005). The present results suggest that the mirror network is an active process that may be elicited also by conflict monitoring (Botvinick *et al.*, 2004), perhaps in order to resolve the inconsistency by activating more vivid images of the action.

Implications and future research

The present study is consistent with a small but growing number of neuroimaging studies characterizing a domain-specific network for social mentalizing and a domain-general network for cognitive control. It is interesting to note a strong parallel with related research on cognitive reasoning (e.g. induction and deduction, analogical and transitive inferences) that typically recruits the pmFC and lateral PFC. It has been found that when this material involves social agents that invite thinking about their mental capacities and preferences, the mPFC is also recruited (see Van Overwalle, 2011 for a meta-analysis). Thus, here also is a distinction between a specific network for social mentalizing and a domain-general network for comparative or 'conflict' processing. To illustrate with a recent fMRI study by Raposo *et al.* (2010), participants indicated for a large variety of words how pleasant or unpleasant they were (i) for themselves, (ii) for a friend or (iii) they had to indicate the difference between their own pleasantness judgment and that of their friend. Consistent with the presumed role of mentalizing, the authors found that the mPFC was recruited more strongly for other ratings in contrast to self ratings. More critically, when the role of mentalizing was cancelled out by subtracting self ratings from relational self-friend ratings, only the lateral PFC was activated, revealing the role of this area in comparative reasoning between self and other.

The present evidence also shows rough parallels with research on the cognitive regulation of emotional responses. Cognitive reappraisal of emotion is a complex process that requires generating, maintaining and implementing a different cognitive reframe (e.g. distancing), as well as tracking changes in one's emotional states. According to Ochsner and colleagues (Ochsner and Gross, 2008; Ochsner *et al.*, 2009), cognitive reappraisal recruits lateral portions of the PFC implicated in working memory and selective attention, the pmFC implicated in monitoring control processes, and the mPFC implicated in reflecting upon one's own or someone else's affective states. All these areas were also involved in the present study. Thus, exerting control over conflict in emotional and social domains seems to recruit the same domain-specific mentalizing network and domain-general conflict monitoring network.

Future studies could further explore these biasing mechanisms underlying social cognitive conflict resolution. Apart from processing modes as in the present study, future work could manipulate stimulus factors to explore how conflicts are resolved in a particular direction or bias, such as, for example, in the direction of the actual behavior (by increasing its vividness) or in the direction of an enabling context (by providing context information). This biasing does not necessarily involve simply ignoring one cue in favor of another. Instead, quite often participants must decide how to weight and integrate contradictory pieces of information to judge the traits of an agent (Schiller *et al.*, 2009). For example, in the face of an extreme behavior (e.g. slaps his mother), behavioral cues may be discounted on the basis of contextual evidence (e.g. humiliated by his mother) and conflict may be low. By manipulating the extremity of the behavior, the strength of the trait expectancy and the relevance of the enabling context, one can systematically measure the relative impact of social cognitive conflict signals.

Taken together, our data sheds some new light on the process of social cognitive conflict resolution where domain-specific neural networks specialized for mentalizing about others increases processing, while additional control is exerted by domain-general cognitive monitoring mechanisms.

Conflict of Interest

None declared.

REFERENCES

Bartholow, B.D., Fabiani, M., Gratton, G., Bettencourt, B.A. (2001). A psychophysiological examination of cognitive processing of and affective responses to social expectancy violations. *Psychological Science*, *12*, 197–204.

Bartholow, B.D., Pearson, M.A., Gratton, G., Fabiani, M. (2003). Effects of alcohol on person perception: a social cognitive neuroscience approach. *Journal of Personality and Social Psychology*, *85*, 627–38.

Baumgartner, T., Gotte, L., Gugler, R., Fehr, E. (2011). The mentalizing network orchestrates the impact of parochial altruism on social norm enforcement. *Human Brain Mapping*. (13 May 2011 Epub ahead of print; doi: 10.1002/hbm.21298).

Botvinick, M.M., Cohen, J.D., Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8*, 539–46.

Burnett, S., Blakemore, S.-J. (2009). Functional connectivity during a social emotion task in adolescents and in adults. *European Journal of Neuroscience*, *29*, 1294–301.

Carter, C.S., Van Veen, V. (2007). Anterior cingulate cortex and conflict detection: an update of theory and data. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 367–79.

Cohen, J.D., Botvinick, M., Carter, C.S. (2000). Anterior cingulate and prefrontal cortex: who's in control? *Nature Neuroscience*, *3*, 421–3.

De Neys, W. (2006). Dual processing in reasoning: two systems but one reasoner. *Psychological Science*, *17*, 428–33.

De Neys, W., Clumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*, 1248–99.

De Neys, W., Vartanian, O., Goel, V. (2008). Smarter than we think: when our brains detect that we are biased. *Psychological Science*, *19*, 483–9.

Evans, J.St.B.T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking and Reasoning*, *13*, 321–39.

Evans, J.St.B.T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–78.

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.

Kerns, J.G., Cohen, J.D., MacDonald, A.W.III, Cho, R.Y., Stenger, V.A., Carter, C.S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–6.

Keysers, C., Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in Cognitive Sciences*, *11*, 194–6.

Lieberman, M.D. (2007). Social cognitive neuroscience: a review of core processes. *Annual Review of Psychology*, *58*, 259–89.

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.

Ma, N., Vandekerckhove, M., Van Overwalle, F., Seurinck, R., Fias, W. (2011). Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: spontaneous inferences activate only its core areas. *Social Neuroscience*, *6*, 123–38.

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

Mason, M.F., Morris, M.W. (2010). Culture, attribution and automaticity: a social cognitive neuroscience view. *Social Cognitive and Affective Neuroscience*, *5*, 292–306.

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

Mitchell, J.P., Cloutier, J., Banaji, M.R., Macrae, C.N. (2006). Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *Social Cognitive and Affective Neuroscience*, *1*, 49–55.

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

Mohanty, A., Engels, A.S., Herrington, J.D., et al. (2007). Differential engagement of anterior cingulate cortex subdivisions for cognitive and emotional function. *Psychophysiology*, *44*, 343–51.

Moran, J.M., Heatherton, T.F., Kelley, W.M. (2009). Modulation of cortical midline structures by implicit and explicit self-relevance evaluation. *Social Neuroscience*, *4*, 197–211.

Mulert, C., Menzinger, E., Leicht, G., Pogarell, O., Hegerl, U. (2005). Evidence for a close relationship between conscious effort and anterior cingulate cortex activity. *International Journal of Psychophysiology*, *56*, 65–80.

Ochsner, K.N., Gross, J. (2008). Cognitive emotion regulation: insights from social cognitive and affective neuroscience. *Current Directions in Psychological Science*, *17*, 153–8.

Ochsner, K.N., Hughes, B., Robertson, E.R., Cooper, J.C., Gabrieli, J.D. (2009). Neural systems supporting the control of affective and cognitive conflicts. *Journal of Cognitive Neuroscience*, *21*, 1842–55.

Penny, W., Harrison, L., Stephan, K. (2003). Attention to visual motion fMRI data – GLM and PPI analyses. ftp://ftp.fil.iom.ucl.ac.uk/spm/data/attention/README_GLM_PPI.txt (date last accessed October 29, 2011).

Radke, S., de Lange, F.P., Ullsperger, M. (2011). Mistakes that affect others: an fMRI study on processing of own errors in a social context. *Experimental Brain Research*, *211*, 405–13.

Raposo, A., Vicens, L., Clithero, J.A., Dobbins, I.G., Huettel, S.A. (in press). Contributions of frontopolar cortex to judgments about self, others, and relations. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nsq033.

Satpute, A.B., Lieberman, M.D. (2006). Integrating automatic and controlled processes into neurocognitive models of social cognition. *Brain Research*, *1079*, 86–97.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*, 235–9.

Saxe, R., Powell, L.J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychological Science*, *17*, 692–9.

Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, *43*, 1391–9.

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.

Smith, E.R., DeCoster, J. (2000). Associative and rulebased processing: a connectionist interpretation of dual-process models. In: Chaiken, S., Trope, Y., editors. *Dual-Process Theories in Social Psychology*. London: Guilford, pp. 323–38.

Srull, T.K., Wyer, R.S. (1989). Person memory and judgment. *Psychological Review*, *96*, 58–83.

Stangor, C., McMillan, D. (1992). Memory for expectancy-congruent and expectancy-incongruent information: a review of the social and social developmental literatures. *Psychological Bulletin*, *111*, 42–61.

Tettamanti, M., Buccino, G., Saccuman, M.C., et al. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–81.

Todorov, A., Gobbini, M.I., Evans, K.K., Haxby, J.V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, *45*, 163–73.

Uleman, J.S. (1999). Spontaneous versus intentional inferences in impression formation. In: Chaiken, S., Trope, Y., editors. *Dual-Process Theories in Social Psychology*. New York: Guilford Press, pp. 141–60.

Uleman, J.S., Blader, S.L., Todorov, A. (2005). Implicit impressions. In: Hassin, R.R., Uleman, J.S., Bargh, J.A., editors. *The New Unconscious*. New York: Oxford University Press, pp. 362–92.

Ursu, S., Clark, K.A., Aizenstein, H.J., Stenger, V.A., Carter, C.S. (2009). Conflict-related activity in the caudal anterior cingulate cortex in the absence of awareness. *Biological Psychology*, *80*, 279–86.

Vandekerckhove, M.M.P., Markowitsch, H.J., Mertens, M., Woermann, F. (2005). Bi-hemispheric engagement in the retrieval of autobiographical episodes. *Behavioral Neurology*, *16*, 203–10.

Van Duynslaeger, M., Van Overwalle, F., Verstraeten, E. (2007). Electrophysiological time course and brain areas of spontaneous and intentional trait inferences. *Social Cognitive and Affective NeuroScience*, *2*, 174–88.

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

Van Overwalle, F. (2011). A dissociation between social mentalizing and general reasoning. *Neuroimage*, *54*, 1589–99.

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

Van Veen, V., Carter, C.S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior*, *77*, 477–82.

Van Veen, V., Carter, C.S. (2006). Conflict and cognitive control in the brain. *Current Directions in Psychological Science*, *15*, 237–40.

Winter, L., Uleman, J.S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*, 237–52.

Zaki, J., Hennigan, K., Weber, J., Ochsner, K.N. (2010). Social cognitive conflict resolution: contributions of domain-general and domain-specific neural systems. *The Journal of Neuroscience*, *30*, 8481–8.