Situation and person attributions under spontaneous and intentional instructions: an fMRI study

Jenny Kestemont, Marie Vandekerckhove, Ning Ma, Nicole Van Hoeck, and Frank Van Overwalle Vrije Universiteit Brussel, Department of Psychology, Brussels, Belgium

This functional magnetic resonance imaging (fMRI) research explores how observers make causal beliefs about an event in terms of the person or situation. Thirty-four participants read various short descriptions of social events that implied either the person or the situation as the cause. Half of them were explicitly instructed to judge whether the event was caused by something about the person or the situation (intentional inferences), whereas the other half was instructed simply to read the material carefully (spontaneous inferences). The results showed common activation in areas related to mentalizing, across all types of causes or instructions (posterior superior temporal sulcus, temporo-parietal junction, precuneus). However, the medial prefrontal cortex was activated only under spontaneous instructions, but not under intentional instruction. This suggests a bias toward person attributions (e.g. fundamental attribution bias). Complementary to this, intentional situation attributions activated a stronger and more extended network compared to intentional person attributions, suggesting that situation attributions require more controlled, extended and broader processing of the information.

Keywords: situation attribution; person attribution; spontaneous inferences

INTRODUCTION

To interact with other persons and the social world, we need to understand and predict other persons' mental states and behaviors. The knowledge that we have on others' mental belief states allows us to predict their further actions and to engage in joint cooperation. The majority of recent social neuroscientific research has focused on this capacity to mentalize (or theory of mind), that is, how we, as observers, understand the internal mental states of other persons (such as goals, thoughts, desires, intentions, beliefs and feelings; Frith and Frith, 2001; Saxe et al., 2004a). However, much less research effort has gone into understanding the beliefs entertained by the observers themselves. How do they perceive causes of behaviors in the first place? Do they see behaviors as caused by another person (either temporary states or enduring dispositions) or by situational constraints (either occasional circumstances or stable features)? This distinction has been at the forefront in earlier behavioral research on attribution in social psychology. The lack of neuroscientific research and theorizing on this fundamental distinction is surprising, because when observers believe that a behavior is caused mainly by the situation rather than the actor, mentalizing about the actor's internal goals, beliefs and traits seems less necessary. More generally, quickly learning that a behavior is caused by human or physical causes may hold strong evolutionary advantages, because each cause may require distinct responses to reap benefits and avoid harm (e.g. shielding from an aggressive person or bad weather).

The goal of the present research is to begin exploring this neglected area of neuroscientific research, by investigating how observers make a distinction between the person and situation as causes of behaviors, and which brain areas are subserving this process. After briefly reviewing earlier behavioral research, we turn to neuroscientific evidence on mentalizing, and end with a description of the present research.

Advance Access publication 15 February 2012

Person and situation attributions

In his seminal book, Heider (1958) argued that in order to understand the social environment, observers foremost distinguish between 'person' and 'impersonal' causality. If the person is a necessary and sufficient condition to cause the effect, inferences are made to the person. In contrast, if the environment or 'things' influence the person's actions, inferences are made to impersonal causality, more commonly termed situation causality. Person causes may refer to temporary states (e.g. emotions) or enduring dispositions (e.g. traits) and likewise, situational causes may refer to occasional circumstances (e.g. luck) or stable features of the context (e.g. water is wet). Nevertheless, there is a pervasive tendency to overestimate person causes and underestimate situation causes, termed the fundamental attribution bias (Ross, 1977; Gilbert and Malone, 1995). When we observe an agent performing a behavior in a particular situation, it is especially the agent who is salient to us, while the situation remains relatively stable and unnoticed (Jones and Nisbett, 1972; Taylor and Fiske, 1975). Consequently, observers automatically tend to make initial inferences to the agent while ignoring the situation, often in terms of personal dispositions. However, correction of this bias is possible. If sufficient cognitive resources are available, the observer can appraise the situation in more respects and realize that circumstantial constraints may have interfered with or may have determined the agent's behavior, leaving more room for causal inferences to the situation (Gilbert et al., 1988).

Recent models of causal inference, however, argue that when perceiving human behaviors, observers often spontaneously activate multiple co-occurring social inferences on traits, goals and circumstances (Read *et al.*, 1990; Reeder *et al.*, 2002, 2004; Reeder, 2009). There is now a plethora of behavioral research indicating that observers make various social inferences when observing others' actions spontaneously, that is, without explicit intention to do so and unaware of making the inference itself (Uleman, 1999). Such spontaneous inferences stand in contrast to intentional inferences, which are deliberately and consciously made to produce a social judgment. Spontaneous inferences are made about an actor's traits or dispositions (for an overview, see Uleman *et al.*, 1996), an actor's goals (Hassin *et al.*, 2005) or to situational circumstances and causes (e.g. Lupfer *et al.*, 1990; Duff and Newman, 1997; Hassin *et al.*, 2002). More importantly, it has been documented that inferences about traits and situations can be activated

Received 1 September 2011; Accepted 8 February 2012

This research was supported by an OZR Grant GOA68 of the Vrije Universiteit Brussel to Frank Van Overwalle, and performed at GlfMI (Ghent Institute for Functional and Metabolic Imaging).

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

spontaneously at the same time, and co-occur together (Ham and Vonk, 2003; Todd *et al.*, 2011). Even when observers are explicitly requested to intentionally make trait or situation inferences, then both inferences become automatically activated to the same degree. However, this automaticity for making a large range of social inferences is limited. When leading questions are asked in the direction of one of possible causes and when cognitive resources are limited, then inferences tend to be made only about the trait or situation causes alluded to in the question (Todd *et al.*, 2011).

Although it is generally acknowledged that social inferences are often made spontaneously, there is less evidence on the underlying mechanisms that produce these spontaneous inferences. According to dual-process models, there is a large difference between spontaneous and intentional inferences. Spontaneous social inferences are driven by automatic associative processes, whereas intentional social inferences are guided by controlled symbolic reasoning (Uleman, 1999; Smith and DeCoster, 2000). The associative processing system relies on prior knowledge and beliefs, and uses association, similarity and memory retrieval to produce primitive judgments quickly and spontaneously. In contrast, the symbolic system largely depends on the application of rule-based and logical reasoning procedures to produce deliberate and intentional judgments (De Neys, 2006; De Neys *et al.*, 2008).

Neuroscientific evidence

The majority of neurological research on social inferences using fMRI has been conducted under intentional instructions to infer other persons' goals and traits (see meta-analysis by Van Overwalle, 2009). This meta-analytic review documented that making inferences about other persons' mental states strongly recruits the temporo-parietal junction (TPJ) bilaterally. It confirmed that the TPJ is important in detecting goals from observed behavior (Frith and Frith, 2001), in distinguishing between self and others (Frith and Frith, 2001), and in representing beliefs of other persons (Apperly et al., 2004; Saxe and Wexler, 2005). The medial prefrontal cortex (mPFC, including the anterior part of the anterior cingulate; Amodio and Frith, 2006) is another region that often contributes to mentalizing (Lieberman, 2007; Carrington and Bailey, 2009; Mitchell, 2009; Van Overwalle, 2009). This area is most strongly recruited when making trait inferences about the self or others (Amodio and Frith, 2006; Lieberman, 2007; Mitchell, 2009; Van Overwalle, 2009).

Apart from these two 'core' mentalizing areas, several studies also reported activation in the posterior superior temporal sulcus (pSTS), which is a multisensorial area sensitive to biological movement (Saxe *et al.*, 2004b; Lieberman, 2007), the precuneus (PC) that processes episodic memory and imagery about the self and others (Van Overwalle and Baetens, 2009) and the temporal pole (TP) that might be involved in memory-related processing of social information (Olson *et al.*, 2007; Ross and Olson, 2010). In addition, Van Overwalle and Baetens (2009) suggested that the mirror brain network [i.e. anterior intra-parietal sulcus (aIPS) and premotor cortex (PMC)] might also contribute to understanding human behavior, because it is responsible for understanding the goals and means of human movement, that is observed visually or in another perceptually salient manner.

In one of the few fMRI studies that explored competing inferences to person or situation causes, Harris *et al.* (2005) reported activation of pSTS/TPJ, mPFC and PC for person inferences, although they did not report activation for situation inferences. Mitchell *et al.* (2002) compared trait inferences about persons *vs* characteristics of inanimate objects (e.g. glove, mango). They found activation of the pSTS, mPFC and IPS for trait inferences, whereas object inferences (which might be relevant for situation inferences) did not activate any of the mentalizing areas. Researchers also compared causal attributions to the self as opposed to external attributions that combine other persons and the situation. They reported that attributions to external others or the situation activated the pSTS (Blackwood *et al.*, 2003) and the TPJ and PC (Seidel *et al.*, 2010). Taken together, person inferences consistently activate mentalizing brain areas. In contrast, there is little converging evidence on the mentalizing or mirror areas that are involved during inferences on situational pressures or constraints. Moreover, in all these studies, participants were explicitly requested to make causal judgments. As such, all processes were intentional, so that it remains unclear to what extent the same processes and brain areas are involved during spontaneous person and situation inferences.

Recently, there is growing fMRI evidence on the existence of spontaneous social inferences (Mitchell et al., 2006; Van Duynslaeger et al., 2007; Moran et al., 2009; Van der Cruyssen et al., 2009; Van Overwalle et al., 2009; Ma et al., 2011). One leading question was whether spontaneous and intentional inferences tap into the same neural circuitry-perhaps to a different degree-or whether they involve entirely different underlying processes and related brain areas as suggested by the dual-process models (Uleman, 1999; Smith and DeCoster, 2000). In recent neuroscientific theory, the distinction between spontaneous and intentional modes of thinking focuses, respectively, on lower-level fast perceptual processing that is subserved by the mirror network while higher level symbolic or abstract processing is subserved by the mentalizing network (Keysers and Gazzola, 2007; Van Overwalle and Baetens, 2009). Within mentalizing, a distinction had been proposed between an early developing, cognitively fast and efficient, but inflexible capacity for tracking the beliefs and states of others vs a later developing and more flexible, but cognitively demanding ability to understand the beliefs and traits of others (Apperly and Butterfill, 2009).

To date, there is only one fMRI study by Ma *et al.* (2011), which compared spontaneous and intentional inferences on traits between participants, so that leakage of the intentional instructions into the spontaneous condition was avoided. The results revealed a significant overlap in activated brain areas during both types of processing, which involved the TPJ and mPFC as 'core' areas of mentalizing. Event-related potential (ERP) studies using a similar between-participants design corroborate that electrophysiological activity is quite often highest in at least one of these two areas during both spontaneous and intentional inferences of traits (Van Duynslaeger *et al.*, 2007; Van Overwalle *et al.*, 2009) and goals (Van der Cruyssen *et al.*, 2009). Moreover, these ERP studies revealed that the onset of social inferences occurs at about the same time irrespective of their spontaneous or intentional source (200 ms for goals and 600 ms for traits).

Importantly, Ma et al. (2011) further noted that spontaneous trait inferences significantly recruit only the core mentalizing areas, whereas intentional trait inferences additionally recruit other brain areas involved in mentalizing. What critical functions might these additional activations serve in intentional inferences? The authors suggested that these activations may reflect thoughts to confirm or validate the spontaneous trait hunches made initially, in a variety of ways. First, the activation of the PC and posterior cingulate might reflect explicit memories of similar self-referential events (Moran et al., 2009) which aid to understand and validate social judgments based on similar events. Second, verifying social inferences might be facilitated by attending to specific behavioral details subserved by the mirror network in humans. When readers are in the process of comprehending a narrated behavior, it has been found that neural indices of perceptual and motor representations are activated in the mirror areas together with neural indices of mentalizing (Speer et al., 2009; Spunt et al., 2010), although mirror activation does not seem to provide direct input to mentalizing (Van Overwalle and Baetens, 2009).

Situation and person attributions

Third, along the same line of reasoning, intentional instructions may also render biological motion and body parts more salient (e.g. gave 'a kiss', 'a hug', etc.), resulting in more activation of the STS. A fourth possibility, not entertained by Ma et al. (2011), is that activity in the dorsolateral PFC may not necessarily reflect mirror activity (by the PMC), but rather top-down control on mentalizing. Taken together with the ERP data (Van Duynslaeger et al., 2007; Van der Cruyssen et al., 2009; Van Overwalle et al., 2009), the results of Ma et al. (2011) seem to suggest that intentional instructions exert their influence only after a common (spontaneous) process produced an initial inference, unlike dual-process models, which predict a sharp distinction between them. Specifically, intentional instructions seem to invite observers to think more deeply about the material they read, and consider it in more different ways to verify the inference made (e.g. to verify an initial spontaneous inference with more details as laid down in memories). This is in line with the general view that higher-level mentalizing may involve cognitively demanding thinking processes (cf Apperly and Butterfill, 2009). However, there is as yet no fMRI evidence indicating that these results can be extended to other types of causal inferences, such as situational pressures or a person in general (and not only traits).

Present research and hypotheses

As noted earlier, this study investigates how observers make a distinction between person and situation causes of events. Moreover, we explore to what extent these inferences are subserved by similar brain processes under spontaneous or intentional circumstances. To do so, we present short behavioral descriptions that imply either the actor or the situation as the main cause of the behavior, and request the participants to familiarize themselves with the material (spontaneous inferences) or to decide which cause is due to the person or the situation (intentional inferences).

Given earlier research on trait inferences by Ma et al. (2011), showing an overlap in the core areas of mentalizing under both instructions, we predict here also that attributions made under spontaneous and intentional instructions will show a significant overlap in core mentalizing areas such as the TPJ and mPFC, and perhaps also in additional mentalizing areas such as the pSTS, TP and PC. This stands in contrast to predictions from classic dual-process theories, which assume totally different processes as discussed earlier (De Neys, 2006; De Neys et al., 2008). In addition, we predict stronger mPFC activity following sentences that imply the person as the cause, because many person causes may involve trait inferences which recruit this area (see also Van Overwalle, 2009). We also predict that the fundamental attribution bias, or the tendency to focus attributions automatically to the actor, will show up mainly in the spontaneous condition. This prediction is based on the robust finding from behavioral research that when there is limited mental capacity or motivation, observers often fall trap to this bias (e.g. Gilbert et al. 1988; Todd et al., 2011). Since many person causes may involve a trait, this will result in more mPFC activation in this condition. In contrast, this bias and related mPFC activity will be weaker under intentional instructions because these instructions (i) focus explicitly on situational circumstances besides the person as possible alternative cause (see also Todd et al., 2011) and (ii) therefore motivate the participants to exert sufficient cognitive attention and effort to correct the fundamental bias (Gilbert et al., 1988). With respect to mentalizing or mirror areas besides the mPFC, we simply explore their activation under person and situation causes both under spontaneous and intentional instructions given the lack of firm evidence on these types of causes.

METHOD Particinants

Participants

Thirty-four right-handed Dutch-speaking participants were recruited for this study. Half of them (8 men, 9 women) participated in the intentional condition, the other half (3 men and 14 women) participated in the spontaneous condition. Their age ranged from 18 to 27 years, with a mean age of 21.76 years in the intentional condition and 20.76 years in the spontaneous condition. In exchange for their participation, participants were paid €10 and received a CD with their structural scanning images.

The participants were recruited via mailing lists of the health department of the University Hospital Ghent, and complied with the following selection criteria: no internal metal objects or artificial implants, no dental brackets or other important dentures, no increased risk for epileptic attacks, nor psychiatric diagnose and no pregnant women or women giving breastfeeding. None of the participants reported an abnormal neurological history and all had normal, or corrected-to-normal, vision. The informed consent was obtained in a manner approved by the Medical Ethics Committee of the University Hospital Ghent (where the study was conducted) and the Free University of Brussels (of the primary investigator FVO).

Note that although assignment between instruction conditions was randomly determined, gender of the participants was not equally distributed across the conditions. However, gender was of little relevance in our stimulus material (in <10% of the sentences gender could be inferred), and this imbalance had minimal effects as an analysis with gender as covariate did not alter our results meaningfully [all the regions of interest (ROI) in Tables 2 and 3 remained significant, except one PMC cluster with nine voxels and its conjunction].

Stimulus material

The stimulus material consisted of 80 experimental sentences describing several mundane events, divided in two conditions. Half of the sentences implied the situation as the cause of the event (e.g. Gabril changes the ink—implying that the ink holder was empty) and were presented in the situation condition. The other half implied the person as the cause (e.g. Jun gives a bouquet at arrival—implying that he or she was romantic) and were provided in the person condition (see Appendix A1 for all sentences and implied causes). The sentences in the two conditions (situation and person) were balanced on two relevant characteristics: each condition consisted of 27 sentences with a positive valence and 13 with a negative valence, with a mean valence of 6.0 (on a 0–10 rating scale), and each condition had the same mean number of words per sentence (= 5.1). Moreover, as can be seen in the examples, 'star trek-like' names were used, to make sure that there were no similarities with familiar others of the participants.

The experimental sentences were selected out of an initial pool of 504 sentences. In a first pilot study, participants (N=143) from the same population as the main study were asked to write down for each sentence the most likely cause of the event. This cause could be a characteristic of the situation, of the actor or both: 'Read each sentence very carefully, try to imagine the event and try to think about a cause for it. Do you think this event tells something about a characteristic of the situation or about a characteristic of the person? If you can find a cause in both the situation and the person you can fill in both. Try to do this as spontaneous as you can'. When $\geq 69\%$ of the students agreed about their attribution in one of the two categories (situation or person) and agreed for <35% in the other category, the sentence was selected for a second pilot study. In this study, participants (N=120)had to read the event and the cause obtained from the first pilot study. Participants answered to what degree the cause was (i) a good explanation of the event on a rating scale going from 0 = totally wrong to

10 = very good, (ii) due to/involved the person or the situation on a rating scale going from 0 = person to 10 = situation and (iii) positive or negative on a rating scale ranging from 0 = negative to 10 = positive. If the cause was seen as a good explanation (i.e. a score of ≥ 6.5) and if the cause was ranged in the same (situation or person) category as in the first pilot, then the sentence was selected for the main experiment. Finally, 40 sentences in each category were chosen, in such a way that the mean valence score (from the third question) and mean number of words were equal in both conditions. Most of the implied causes were not close semantic associates of the words in the sentences (see Appendix A1).

Apart from the experimental sentences, we also created 40 sentences for the baseline condition, which consisted of a semantic truth judgment in which participants had to judge whether the event description was true or false or unknown by the participant (e.g. 'Brussels is the main city of Belgium'). In this condition, the mean number of words was 6.3; all were neutral sentences. Note that although a minority (8) of the truth sentences involved persons (e.g. 'Salvador Dali was born in Spain'), all of these sentences involved nonmental physical facts, which are typically taken as baseline in mentalizing studies (Van Overwalle, 2009).

Procedure

During the structural scanning, the participants received written instructions about the experiment. In the intentional condition, they were asked to make attributions to either situation or person causes after reading each sentence. Several descriptions and examples (e.g. 'Being a hard worker = cause in person', 'danger = cause in situation', etc.) were given to familiarize the participants with this distinction. In contrast, in the spontaneous condition, they were only asked to read the sentences very carefully because questions about them would be asked after the scanning. Nothing was mentioned about causes or inferences.

During the functional scanning, in the intentional condition, each experimental trial started with a different task instruction per condition (2s): 'search cause in person' (person condition) or 'search cause in situation' (situation condition), followed by a fixation cross (jittered between 3 and 5 s). The sentence describing the event was presented for 5.5 s, giving the participants ample time to make a specific causal attribution to a characteristic of the person or the situation. After a fixation cross (0.5 s), the question appeared: 'Is the cause X?', with X being the cause selected from the pilot studies. Participants had to press the appropriate response button on a scale ranging from 1 = no, 2 = a little, 3 = a lot and 4 = totally. For instance, in the situation condition, after presenting the sentence 'Maldron earns a salary', the question was 'Is the cause: work?' (Figure 1). The responses during scanning show a mean of 3.32, confirming that most participants thought about the same attribution as the one selected during the pilot study.

In the spontaneous condition, each trial started with a fixation cross but no task instructions (jittered between 5 and 7 s). The remainder of the procedure was the same, except that no question was asked after the experimental (situation and person) sentences (Figure 1). To make sure that participants would remain attentive throughout the whole experiment, 20 filler sentences and questions were added. These sentences were similar to the sentences describing the events (e.g. 'Rekon dreams while he sleeps'), except that the actor's gender was made evident through the use of he, her, his or hers pronouns. The sentence was

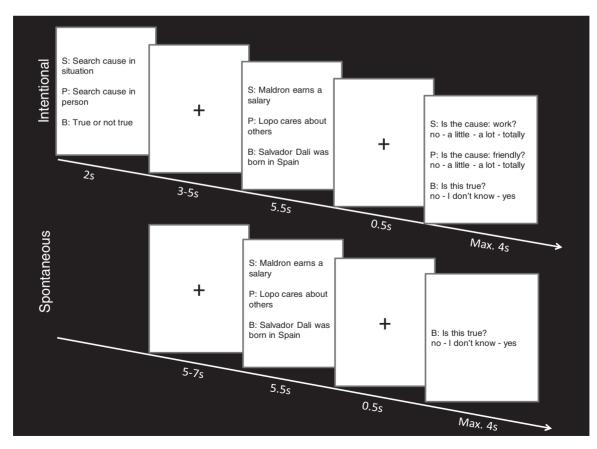


Fig. 1 Stimulus material presented in the intentional and spontaneous condition. S, P and B denote the instructions and questions given for situation, person and baseline conditions. Per trial, only one instruction, sentence and question were presented.

followed by a question about the gender of the actor (i.e. 'is [actor] a male?'). Participants had to respond by pressing the appropriate button ranging from 1 = no, 2 = I don't know to 3 = yes. The mean number of words of these sentences was 5.4. The sentences consist of seven negative, seven neutral and six positive sentences.

In the baseline condition, each trial started with a 'true or not true' task instruction in the intentional condition. In the spontaneous condition, instead, these trails started with a fixation cross. After presentation of the sentence, the following question appeared in both the intentional and the spontaneous condition: 'Is this true?'. Participants responded by pressing the appropriate button ranging from 1 = no, 2 = I don't know to 3 = yes (Figure 1).

Memory measures

After the participants had left the scanner, they all received two memory tasks, typically used to measure spontaneous inferences. This was to verify that they had made situation and person attributions under both the intentional and spontaneous instructions. All participants got, in the same order, a relearning task, a sentence completion task and the final recall for the relearning task.

First, participants had to learn 40 pairs of names plus words. All those names were names from the experimental sentences, but only half of the words were person attributions (10) and situation attributions (10) implied during the experiment together with the same actor name. The other 20 words were person (10) and situation (10) characteristics that were never implied. This means that the previously implied actor-attribution pairs were seen for the second time (= relearning items), while the other actor-attribution pairs were new (= new items). After a short interference task to empty short-time memory (about 5 min), participants completed the relearning task. The 40 names were presented again and the participants had to complete the pairs by adding the person or situation attribution word previously presented during the relearning task. The basic idea is that if participants made attributions associated with the actor, relearning these actor-attribution pairs will reinforce existing memory traces and hence relearning pairs will be remembered better than new pairs (Nelson, 1985; Carlston and Skowronski, 1994; Carlston et al., 1995).

The interference task actually consisted of a sentence completion memory task. In this task, participants had to fill in the last word(s) of incomplete experimental sentences that were critical in implying a person or situation attribution. The basic idea is that if participants make situation or person attributions about the described behavior or event, the words that strongly implied these attributions will be better remembered.

Several random versions of these two memory measures were created, so that the sentences in the sentence completion task were never the ones used in the relearning task. No significant differences emerged between versions (all Fs < 1.76) so that these were ignored in the analyses.

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 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. Foam cushions were placed within the head coil to minimize head movements. We first collected 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 echo planer pulse sequence (Repetition Time (TR) = 2 s; Echo-Time (TE) = 33 mm; $3.5 \text{ mm} \times 3.5 \text{ mm} \times 4.0 \text{ mm}$ resolution).

Image processing and statistical analysis

The fMRI data were preprocessed and analyzed using SPM5 (Wellcome Department of Cognitive Neurology, London). 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), which approximates Talairach and Tournoux atlas space. Normalized data were then spatially smoothed (6-mm full-width at half-maximum, FWHM) using a Gaussian kernel. Finally, realigned data were examined, using the Artifact Detection Tool software package (ART; http://web.mit.edu/ swg/art/art.pdf; http://www.nitrc.org/projects/artifact_detect/), for excessive motion artifacts and for correlations between motion and experimental design, and between global mean signal and experimental design. Outliers were identified in the temporal difference series by assessing between-scan differences (Z-threshold: 3.0, scan to scan movement threshold 0.45 mm; rotation threshold: 0.02 radians). These outliers were omitted in the analysis by including a single regressor for each outlier (bad scan). No correlations between motion and experimental design or global signal and experimental design were identified.

Statistical analyses were performed using the general linear model of SPM8 (Wellcome Department of Cognitive Neurology, London, UK) of which the event-related design was modeled with two regressors for each condition (stimuli and response), time-locked at the presentation of the sentence and response scale, using a canonical hemodynamic response function and its temporal derivative (with event duration set to 0, which results in an estimation of the default hemodynamic response function). Six directions of motion parameters from the realignment step as well as outlier time points (defined by ART) were included as nuisance regressors. We used a default high-pass filter of 128 s and serial correlations were accounted for by the default autoregressive AR(1) model. The statistical analyses involved first-level single participant analyses followed by second-level group analyses, all time-locked at the presentation of the sentence. Comparisons of interest were implemented at the group level as linear contrasts using a random-effects model. A voxel-based statistical threshold of $P \le 0.005$ (uncorrected) was used for all comparisons. Statistical comparisons between conditions were conducted using a full factorial analysis of variance (ANOVA) procedure on the parameter estimates associated with each trial type. Conjunction analyses were performed combining the situation > baseline and person > baseline contrasts for each of the intentional and spontaneous instruction conditions separately. Also, two conjunction analyses were performed combining the intentional and spontaneous analyses in each of the situation and person conditions separately. A final conjunction analysis combined all four contrasts (intentional situation > baseline, intentional person > baseline, spontaneous situation > baseline and spontaneous person > baseline).

ROI analyses were performed with the small volume correction in SPM8. For all ROI analyses, small volume correction was required to exceed 10 contiguous voxels in extent and based on a sphere of 8 mm radius around the centers (in MRI coordinates) of areas that were identified in the meta-analysis by Van Overwalle (2009) and Van Overwalle and Baetens (2009) as involved in mentalizing: 0, -60, 40 (PC), ± 50 , -55, 25 (TPJ), 0, 50, 20 (mPFC), 0, 50, 35 (dmPFC), 0, 50, 5 (vmPFC); action understanding via mirror neurons: ± 50 , -55, 10 (pSTS), ± 40 , -40, 45 (anterior intraparietal sulcus, aIPS), ± 40 ,

5, 40 (PMC); and by Sugiura *et al.* (2006) as involved in person identity, ± 45 , 5, -30 (TP). Analyses of the ROI were conducted using *t*tests with a threshold of *P* < 0.05, Family-wise error (FWE) corrected. In addition, the mean percent signal change in each ROI was extracted using the MarsBar toolbox (http://marsbar.Sourceforge.net), and analyzed using *t*-tests with a threshold of *P* < 0.05.

RESULTS

Memory measures

To investigate whether situation and person attributions were made under both intentional and spontaneous instructions, participants completed two memory tasks, relearning and sentence completion. The logic behind relearning is that participants who made an attribution spontaneously while reading the sentences will remember that attribution better after relearning the same actor-attribution couple in comparison with learning a new actor-attribution couple. The logic behind sentence completion is that the critical words that implied the attribution should be remembered also, although this test is evidently more indirect than (re)learning the attribution itself.

For relearning, a repeated measures analysis of variance (ANOVA) was conducted, with Instruction (intentional *vs* spontaneous) as between-participants factor and relearning (relearning *vs* new items) and Sentence Type (person *vs* situation) as within-participants factors. Only a significant main effect of Relearning was found, F(1, 32) = 7.51, P = 0.01, $\eta_p^2 = 0.19$, with better recall for the relearned items than for the new items (Table 1). No other main or interaction effects were found. This indicates that irrespective of instruction conditions, participants remembered both types of attributions much better than new words, clearly suggesting that situation and person attributions were made to the same extent (see also Carlston and Skowronski, 1994).

Sentence completion was coded as accurate if the last critical word was completed with the originally presented word(s) or a synonym (Bartholow et al., 2001, 2003). The completion scores were analyzed with a similar ANOVA, with Instruction (intentional vs spontaneous) as between-participants factor and Sentence Type (person vs situation) as within-participants factor. The results showed only a main effect of instruction, F(1, 32) = 18.07, P < 0.001, $\eta_p^2 = 0.36$, indicating that participants in the intentional condition could remember significantly more critical words than participants in the spontaneous condition (Table 1). This result contradicts the analysis of the relearning task where no effect of instruction was found. However, since sentence completion is only an indirect test of making attribution judgments, we can safely assume that participants made person and situation attributions to the same extent under both instructions, but that under spontaneous instructions, they remembered less of the critical words implying these attributions.

fMRI analysis

For the whole-brain analyses, the threshold was set at P < 0.005 with a cluster extent of 10 voxels. This threshold is somewhat more lenient than usual because spontaneous instructions tend to activate ROI less,

as participants' attention is more diffuse and not necessarily focused on mentalizing (see also Ma *et al.*, 2011). Additionally, ROI analyses (P < 0.05, FWE corrected) were conducted with spheres of 8-mm radius around a priori defined centers known to be related to the mentalizing network (PC, TPJ, TP, dmPFC, mPFC and vmPFC) and the mirror network (pSTS, aIPS and PMC) based on the mean coordinates from the meta-analysis by Van Overwalle (2009) and Van Overwalle and Baetens (2009) or other relevant studies for areas not included in this meta-analysis (Sugiura *et al.*, 2006). Given that the whole-brain analyses merely confirmed the results of these ROI analyses, we only report the ROI analyses (P < 0.05 corrected) and additional regions that survived the same P < 0.05 corrected threshold after the whole-brain analyses.

After comparing each separate condition against the truth baseline, it turned out that the activation pattern of situation and person attributions was quite similar, but that there were larger differences between intentional and spontaneous instructions. This similarity was confirmed by a conjunction analysis, which showed a great overlap across situational and person attributions (Table 2). In the interest of brevity, we therefore discuss first the similarities between situation and person inferences, and subsequently report deviations from this pattern for single situation or person comparisons.

Similarities between situation and person attributions.

Under intentional instructions, the Situation + Person conjunction revealed activation in the following ROIs: PC, pSTS bilaterally and TPJ bilaterally. These ROIs were also recruited under each individual Situation > Truth and Person > Truth contrast (Table 2 and Figure 2). Under spontaneous instructions, the Situation + Person conjunction revealed activation in the same ROIs: PC, pSTS bilaterally and TPJ bilaterally; and additionally in the (ventro)mPFC. These ROIs were also recruited under each individual Situation > Truth and Person > Truth contrast (Table 2 and Figure 2). In addition, the conjunction revealed for intentional as well as spontaneous instructions additional activation in identical areas: the mid-temporal gyrus, precentral gyrus and insula, all at the left hemisphere.

Direct comparisons between intentional and spontaneous instructions (bottom part of Table 2) revealed some differences. The most relevant differences in the ROIs are that intentional instructions invite more activation of the mirror areas and that spontaneous instructions recruit more strongly the right TPJ and mPFC. Most of these differences are shared by both Situation and Person causes, which is confirmed by a conjunction analysis of the Person and Situation conditions.

To summarize, the PC, pSTS and TPJ appear to be common 'core' mentalizing areas that are activated under all conditions. In addition, the mid-temporal gyrus, precentral gyrus and insula were also frequently recruited. A conjunction analysis taking together the intentional and spontaneous instructions mainly confirms this analysis, with the exception of the PC and mid-temporal areas which do not overlap. Furthermore, the mPFC is activated only under spontaneous instructions.

Table 1 Correct memory (in percent) in relearning and sentence completion as a function of instruction and judgment

	Intentional			Spontaneo	Spontaneous			Situation			Person		
	Old	New	Total	Old	New	Total	Old	New	Total	Old	New	Total	
Relearning Completion	12.4ª	8.8 ^b	10.6 ^{a,b} 25.9 ^c	12.1ª —	7.7 ^b	9.9 ^{a,b} 12.4 ^d	12.4 ^e	7.9 ^f	10.2 ^{e,f} 19.9 ^g	12.1 ^{e,f}	8.5 ^{e,f}	10.3 ^{e,f} 18.4 ^g	

Note: Rows with different superscripts differ significantly from each other, $P \leq .05$.

Table 2 Contrasts of situation and person (> truth baseline) and conjunctions under intentional and spontaneous instructions separately as well as under direct comparison between intentional and spontaneous instructions

	Situation > Truth					Person > Truth				Situation >Truth and Person >Truth					
Anatomical label	x	у	Ζ	Voxels	Max t	X	у	Ζ	Voxels	Max t	x	у	Ζ	Voxels	Max t
Intentional															
Regions of interest															
РС	-6	-64	42	32	6.05***	-4	-66	42	33	6.25***	-6	-64	42	31	6.05***
R pSTS	52	-56	16	237	5.15***	52	-56	14	34	3.14**	52	-56	14	34	3.14**
L pSTS	-56	-50	8	252	6.54***	-56	-50	8	182	5.27***	-56	-50	8	182	5.27***
R TPJ	46	-56	20	95	4.68***	46	-56	20	5	2.84*	46	—56	20	5	2.84*
l TPJ	-52	-58	20	156	5.29***	-52	-50	20	79	4.13***	-52	-50	20	79	4.13***
Other Regions															
L IFG	-50	32	-4	636	5.41***	-48	28	4	406	4.97**					
L Mid-Temporal						-50	-42	0	1715	5.43***	-50	-42	0	1696	5.43***
L Precentral	-28	-26	58	21483	8.12***	-30	-24	56	2764	5.63***	-30	-24	56	2751	5.63***
L Insula						-42	-22	18	406	5.16**	-42	-22	18	403	5.16**
L Cingulate						-4	-16	30	467	5.11**					
R Cerebellum	12	-50	-18	626	5.31***										
Spontaneous		50		020	5151										
Regions of interest															
PC	8	-60	40	26	3.68**	6	—56	40	113	4.11***	8	-60	40	26	3.68**
R pSTS	50	—54	18	156	5.77***	50	—56	16	139	5.52***	50	—56	16	137	5.52***
L pSTS	-48	—54 —60	14	150	5.25***	-48	—50 —60	14	145	5.38***	-48	—50 —60	10	137	5.25***
R TPJ	-48 48	—00 —56	14	247	5.97***	-48 -48	—00 —50	32	241	5.83***	48 48	—00 —56	20	236	5.78***
L TPJ					4.66***					4.96***			20		4.40***
	-48	-58	18	70		-48	-58	20	88	4.96**** 4.41***	-46	-56		43	
mPFC	6	52	16	37	3.97***	6	54	18	57		6	52	16	37	3.97***
vmPFC	4	54	10	14	3.60**	6	52	10	20	3.87***	6	54	8	14	3.55**
Other regions															
L Mid-Temporal											-48	-60	14	555	5.25**
R Supra-Marginal	56	-28	26	4948	6.33***										
L Precentral	-28	-24	54	4098	5.18**	-28	-26	58	1830	4.87**	-28	-26	58	1674	4.87**
L Insula	-36	-18	18	4618	6.84***	-38	-18	16	3029	5.70***	-38	-18	16	2847	5.70***
Conjunction: intentional and	spontaneo	us													
Regions of Interest															
PC						-6	-64	42	26	3.59**					
R pSTS	52	-56	16	156	5.15***	52	-56	14	34	3.14**	52	-56	14	34	3.14**
L pSTS	-48	-58	14	156	5.00***	-50	-56	16	95	3.81***	-50	-56	16	94	3.81***
R TPJ	46	-56	20	95	4.68***	46	-56	20	5	2.84*	46	-56	20	5	2.84*
l TPJ	-48	-58	20	33	4.32***	-50	-54	18	27	3.57**	-50	—54	18	11	3.47**
Other regions															
R Supra-Marginal	66	-24	28	1562	5.23**										
L Precentral	-28	-24	54	3492	5.18**	-28	-26	58	877	4.87**	-28	-26	58	844	4.87**
L Insula	-38	-18	20	2356	6.01***	-42	-22	18	342	4.90**	-42	-22	18	339	4.90**
Intentional > spontaneous	50	10	20	2550	0.01	72	22	10	542	1.70	12	22	10	557	1.70
Regions of interest															
L pSTS	—56	—50	8	18	3.15**										
					3.15***										
L aIPS	-42	-44	42	13	3.03** 4.15***	24	A	24	71	∂ /1¥¥	37	A	24	71) /1 ₩¥
R PMC	36	4	36	66 176		36	4	36	21	3.41**	36	4	36	21	3.41**
L PMC	-44	4	46	1/6	4.45***	-44	4	46	9	2.92*	-44	4	46	9	2.92*
Other regions					1.02.99										
L Inferior Parietal	-34	-58	42	643	4.93**										
R Brainstem	10	-22	-16	140	5.03**										
L Inferior Frontal Tri	—50	22	22	3148	6.20***	—50	22	22	427	4.82**	—50	22	22	427	4.82**
L Suppl. Motor Area	-8	10	62	4463	7.64***	-6	10	62	959	5.82***	-6	10	62	956	5.82***
Spontaneous > intentional															
Regions of interest															
R TPJ	50	-54	34	28	3.37**	50	-52	32	107	4.48***	50	-54	34	27	3.37**
vmPFC	6	54	8	88	4.70***	6	52	10	174	4.82***	6	54	8	82	4.70***
mPFC	6	50	16	205	4.90***	6	50	16	251	5.11***	6	50	16	202	4.90***
dmPFC	-					Ő	50	28	15	3.00**	-				
Other regions						č	50	20							
R Angular						50	-66	42	977	5.19**					
R Superior Frontal	18	30	52	963	5.51***	20	30	52	1291	5.89***	18	30	52	900	5.44***
n superior frontal	10	50	52	202	5.51	20	50	52	1271	5.07	10	50	52	200	5.77

Notes: Coordinates refer to the MNI (Montreal Neurological Institute) stereotaxic space. ROIs are spheres with 8 mm radius around 0, -60, 40 (PC), ± 50 , -55, 10 (pSTS), ± 50 , -55, 25 (TPJ), ± 40 , 5, 40 (PMC), 0, 50, 5 (vmPFC), 0, 50, 20 (mPFC) and 0, 50, 35 (dmPFC). R = right; L = left; PC = precuneus; STS = superior temporal sulcus; TPJ = temporo-parietal junction; mPFC = medial prefrontal cortex; vmPFC = ventromedial prefrontal cortex; dmPFC = dorsomedial prefrontal cortex. All regions thresholded at *P* < 0.005, with a minimum cluster threshold of 10 (or 5 if > 10 elsewhere on the same row). Only significant regions after FWE correction are listed. **P* < 0.10, ***P* < 0.05, ****P* < 0.05, FWE corrected (for other regions corrected after whole-brain analysis; for ROIs corrected after small volume analysis).

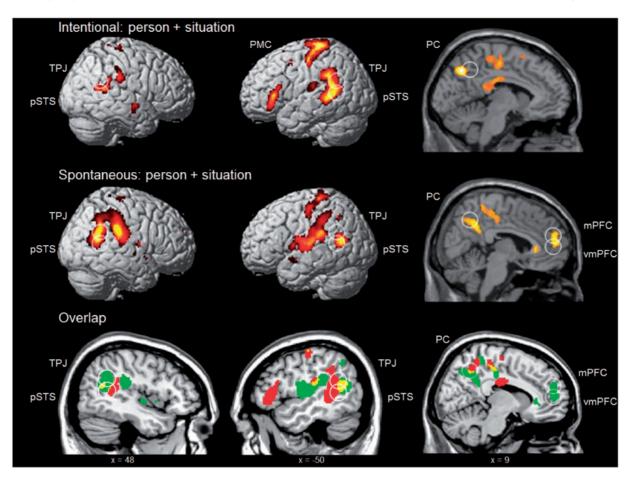


Fig. 2 Conjunction of person > truth and situation > truth under intentional and spontaneous instructions. Whole-brain activation thresholded at P < 0.005 (uncorrected). Circles indicate ROIs with significant activation, P < 0.05, FWE corrected. The overlap was created using MRIcroN, showing the significantly activated areas under intentional (red) and spontaneous (green) instructions at the same whole-brain threshold, and their overlap (yellow).

Differences between situation and person attributions

Situation attributions made under intentional instructions generated stronger activation in the pSTS bilaterally, TPJ bilaterally and the TP bilaterally, together with other ROIs related to the mirror network such as the left aIPS and right PMC (Table 3 and Figure 3). There is also higher activation in a large area of the left mid-frontal cortex extending to the posterior medial frontal cortex (pmFC, related to conflict monitoring and cognitive control; Botvinick *et al.*, 2004) and the fusiform area (related to face and word processing; FitzGerald and Folan-Curran, 2002), which may reflect some difficulty in interpreting the situation information or making a situation attribution. In contrast, person attributions generated under spontaneous instruction elicited more activation in the PC and dmPFC, two areas related to mentalizing about other persons (Van Overwalle, 2009).

Most of these differences between situation and person attributions were confirmed by an additional analysis of signal changes of each ROI. We conducted an ANOVA with Instruction (intentional *vs* spontaneous) as between-participants factor, and Sentence Type (situation *vs* person) as within-participants factor. Under intentional instructions, situation attributions activated more strongly the pSTS and TPJ bilaterally than person attributions (Figure 3). In addition, there was more activation in the left aIPS [$F(1, 16) = 13.00, P = 0.002, \eta_p^2 = 0.45$] and right PMC [$F(1, 16) = 4.61, P = 0.048, \eta_p^2 = 0.22$]. Under spontaneous instructions, there was more activation for person than situation attributions in the dmPFC and nonsignificantly so in the PC. In addition, there was more activation in the left PMC $[F(1, 16) = 6.81, P = 0.019, \eta_p^2 = 0.30].$

Correlations with memory measures

To examine which brain activation could be indicative of situation or person attributions as measured by the memory tasks, we computed Pearson correlations between the percentage signal change of our ROIs (that were significant in the earlier ROI analysis) and the memory scores. Under intentional instructions, a positive correlation was found between activation in right TPJ for situation attributions (r=0.54, P=0.025) and the relearning memory score (Figure 4).

DISCUSSION

This research was motivated by the question which brain processes underlie observers' perception and interpretation of another person's behavior. Prior research on mentalizing (or theory of mind) emphasized how observers keep track of other persons' mind and beliefs, without investigating the beliefs of the observers themselves. Especially important in this question is the distinction between person and situation, because if a person is causally involved, perceivers might become more interested in his or her mind and beliefs. We presented short behavioral descriptions that implied either the actor or the situation as the cause, and asked participants explicitly to provide a person or situation cause, or allowed them to make these inferences spontaneously.

 Table 3
 Contrasts between situation and person under intentional and spontaneous instructions separately as well as under direct comparison between intentional and spontaneous instructions (interaction) and conjunction

	Situation > Person							
Anatomical label	x	у	Ζ	Voxels	Max t			
Intentional								
Regions of interest								
R pSTS	54	-52	16	32	2.88*			
L pSTS	-52	-60	6	49	3.78***			
R TPJ	50	-62	26	26	3.52**			
l TPJ	-54	-58	22	27	3.27**			
L aIPS	-46	-42	46	61	2.89*			
R PMC	40	10	44	34	3.08**			
R TP	46	8	-24	79	3.96***			
L TP	-44	2	-24	11	3.97***			
Other regions								
L Fusiform	-26	-34	-20	30562	9.70***			
L Mid Frontal	-22	12	60	1036	4.97**			
Intentional > spontaneous (Interaction)								
Regions of interest								
R TPJ	48	-60	22	3	2.73*			
l TPJ	-52	-60	22	18	3.15**			
L aIPS	-40	-42	44	63	3.17**			
R TP	48	10	-24	54	3.87***			
R PMC	40	10	44	37	3.24**			
L PMC	-46	8	42	39	3.20**			
dmPFC	-4	44	34	3	2.81*			
Other regions								
Vermis	4	-56	-6	1300	5.10**			
L Thalamus	-4	-20	10	4117	5.57***			
L Mid-Temporal	-46	_4	-22	93	4.64*			
Spontaneous			(none)					
Intentional and spontaneous (Conjunction)			(none)					
Spontaneous > intentional (Interaction)			(none)					
spontaneous > intentional (interaction)	Person	> Situatior						
Spontaneous	1 CI JOIL		1					
Regions of interest								
PC	-2	-64	42	12	2.99**			
dmPFC	-2 -2	-04 54	32	76	3.27**			
Intentional	-2	JH	(none)	70	5.21			
Intentional and spontaneous (Conjunction)			(none)					
intentional and spontaneous (conjunction)			(none)					

Note: Coordinates refer to the MNI (Montreal Neurological Institute) stereotaxic space. RIOs are spheres with 8 mm radius around 0, -60, 40 (PC), \pm 50, -55, 10 (pSTS), \pm 50, -55, 25 (TP)), \pm 45, 5, -30 (TP), \pm 40, -40, 45 (aIPS), \pm 40, 5, 40 (PMC), 0, 50, 35 (dmPFC). Re right; L = left; PC = precuneus; STS = superior temporal sulcus; TPJ = temporal pole; aIPS = anterior intraparietal sulcus; PMC = premotor cortex; dmPFC = dorsomedial prefrontal cortex. All regions thresholded at *P* < 0.005. Only significant regions after FWE correction are listed. ^aStatistically, the Intentional > Spontaneous and Spontaneous > Intentional contrasts for the Person > Situation comparison are identical to the reverse contrasts of the Situation > Person

*P < 0.10, **P < 0.05, ***P < 0.005, FWE corrected (for other regions corrected after whole-brain analysis; for ROIs corrected after small volume analysis).

As predicted, most of the mentalizing brain areas (i.e. pSTS, TPJ and PC) were recruited under both spontaneous and intentional instructions, irrespective of whether the person or the situation was implied as the cause. That there were no essential differences between inferences to the person or the situation was unexpected, and is even more remarkable given the different content of the person and situation sentences. However, we found some differences between intentional and spontaneous instructions. Most interesting was the finding that the mPFC was activated under spontaneous inferences, but not under intentional inferences, unlike earlier fMRI research on trait inferences which found mPFC activity under both instructions (Ma *et al.*, 2011). Moreover, under intentional instructions, situation attributions activated the common mentalizing areas (i.e. without mPFC) more than person attributions, and recruited also some additional areas (e.g. right PMC, left aIPS, left fusiform gyrus and left mid frontal gyrus extending to the pmFC). Under spontaneous instructions, person attributions activated the dmPFC and PC more than situation attributions. In fact, the signal change analysis indicated that the mPFC was most robustly activated when participants spontaneously read about events that implied the person as the cause.

How can we explain these results and how do they accord with previous fMRI research? In general, our results which revealed activation in several areas of the mentalizing network are largely consistent with previous research on social inferences and theory of mind. The robust activation of the TPJ in the present study is in general agreement with the key role of this area in theory of mind tasks that investigate how observers represent the transitory beliefs of other persons (Saxe and Powell, 2006; Lieberman, 2007; Carrington and Bailey, 2009; Van Overwalle, 2009), others' goals (e.g. den Ouden et al., 2005) and others' perspective (e.g. Aichhorn et al., 2006). It is also consistent with earlier studies that explored causal interpretations of behavior by contrasting attributions to the person against all other possible causes (Harris et al, 2005) or against the self (Blackwood et al., 2003; Seidel et al., 2010). Our results extend these earlier findings by revealing that the TPJ and pSTS were also recruited while encoding descriptions that implied the situation. The present study is the first to analyze situation attributions, because previous research either focused on objects in the situation (Mitchell et al., 2002), conflated situation and person into a single category of external attributions (Blackwood et al., 2003; Seidel et al., 2010), or did not report coordinates of situational attributions (Harris et al., 2005).

Another contribution of the present study is that we investigated not only inference processes while observers intentionally made person and situation inferences, but also when they did so spontaneously. Our memory results confirmed that attributions were made to the person and situation under both instructions. More importantly, as noted earlier, most active areas under intentional instructions were also recruited under spontaneous instructions. These findings suggest that there is a common process mechanism underlying both spontaneous and intentional person and situation inferences, in line with earlier research on trait inferences (Ma et al., 2011). Given that most ERP research up till now revealed that the first neural signature of social inferences are elicited at about the same time irrespective of their spontaneous or intentional source (Van Duynslaeger et al., 2007; Van der Cruyssen et al., 2009; Van Overwalle et al., 2009), it is very likely that spontaneous inferences support and inform intentional inferences. Stated differently, both processes share a common spontaneous onset, which can evolve into an intentional inference. Taken together with the present study, the current evidence seems to contradict many dual-process models which assume that spontaneous and intentional inferences are guided by different processes (Uleman, 1999; Smith and DeCoster, 2000; De Neys, 2006; De Neys et al., 2008).

However, our results also revealed some differences. They confirmed our expectation that substantial mPFC activation would be revealed under spontaneous inferences, because this condition increases the likelihood of a fundamental attribution bias to the person (and his or her traits) as a result of the lower or less focused mental effort exerted during spontaneous processing. Surprisingly, as noted earlier, there was no mPFC activation during intentional reasoning, in contrast to prior research. A possible reason is that prior (intentional) studies documenting a reliable amount of mPFC activation, focused on trait inferences (Mitchell *et al.*, 2002; Harris *et al.*, 2005; Ma *et al.*, 2011). In contrast, our person causes also included temporary features such as goals, wishes and states, which may have recruited more activation in the TPJ than the mPFC (Van Overwalle, 2009). Moreover, studies that did not report mPFC activation conflated person and situation inferences, which makes a search for other potential differences and

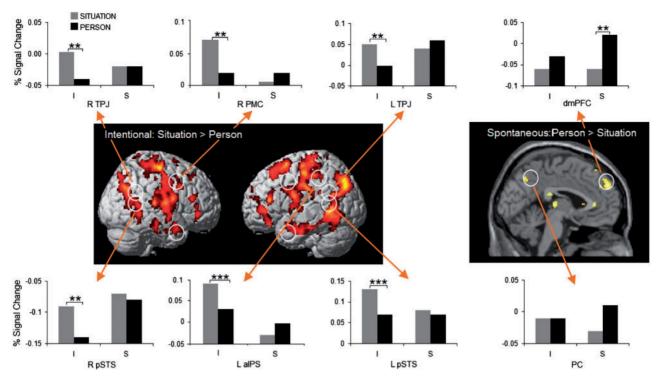


Fig. 3 Situation > person contrast under intentional instructions and person > situation contrast under spontaneous instructions. Whole-brain activation threshold at P < 0.005 (uncorrected). Circles indicate ROIs with significant activation, P < 0.05, FWE corrected. The bars show the percentage signal change (PSC), based on ROIs with 8-mm sphere created by MarsBaR. Gray bar = situation, black bar = person, I = intentional, S = spontaneous. Significant changes between situation and person sentences are indicated by **P < 0.05, ***P < 0.05.

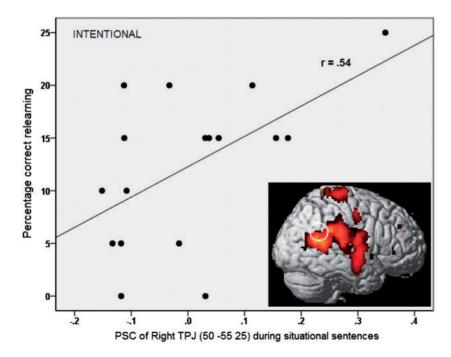


Fig. 4 Pearson correlation between PSC in right TPJ during situation-implying sentences and correctly remembered items under intentional instructions.

explanations difficult (Blackwood *et al.*, 2003; Seidel *et al.*, 2010). However, the present lack of mPFC activation during intentional reasoning accords well with the general tenet of behavioral research indicating that mental effort to think more deeply about the material and to search more extensively for a cause (including a situation cause) invites observers to avoid the fundamental attribution bias (Gilbert *et al*, 1988; Gilbert and Malone, 1995). Indeed, in our intentional instructions, participants were invited at each trial to search either for a person or a situation cause. Throughout the experiment, they were thus reminded of the possibility of causes other than the person, and

Situation and person attributions

this may have reduced the fundamental attribution bias and the activation of the mPFC. This reluctance did not occur during the spontaneous instructions, leading to a greater bias and more mPFC activation. Moreover, note that according to Van Overwalle (2009), the mere contemplation of a potential person cause under intentional instructions should not activate the mPFC right away (but only the TPJ), because mPFC activation is typical for inferences to a person's traits.

We also found that under intentional instructions, situation attributions activated more areas and to a greater extent than person attributions. In contrast, under spontaneous instructions, no more areas were activated given situation attributions, but the dmPFC and PC were more strongly activated given person attributions compared to situation attributions. This might indicate that situation attributions are facilitated under more extended and controlled processing, while person attributions are already extensive under automatic and spontaneous processing. Intentional instructions on situation causes may have motivated the participants to think longer and deeper about the event and possible causes, and to focus more on behavior to verify and validate the initial situation attribution made spontaneously. This again speaks against extant dual-process models (Uleman, 1999; Smith and DeCoster, 2000; De Neys, 2006; De Neys et al., 2008), and seems to point to a situation attribution process in which spontaneous and intentional instructions initiate a shared inference process in an initial phase, which is then prolonged under intentional instructions allowing observers to verify and validate their initial hunch (see also Ma et al., 2011). Unlike situation attributions, person causes under intentional instructions did not invite observers to such additional thinking, but rather prevented them to fall trap to the attribution bias.

Furthermore, the increased activation of the pmFC—responsible for conflict detection and cognitive control (Botvinick *et al.*, 2004)—during intentional situation attributions may underscore the notion that participants were acutely aware of the conflictive implications of person or situation causes. An additional point of support for this reasoning is that the left frontal inferior gyrus, known to be involved in selecting among competing alternatives from semantic memory and response inhibition (Thompson-Schill *et al.*, 1997; Aron *et al.*, 2004), was activated in the intentional but not spontaneous condition. Under spontaneous instructions, this mixed information might have been equally apparent (cf. similar relearning memory effects after scanning), although as noted earlier, situation causes may have been 'ignored' as relevant, leading to more person attributions (activating the mPFC) in line with the fundamental attribution bias.

Limitations and future challenges

This study has some potential limitations. One limitation is that the implied causes were sometimes semantically associated with the sentences presented. Although close semantic associates were rare (<7%), the fundamental question remains whether the causal judgments in these and other sentences were resolved by semantic reasoning rather than causal inference. We believe not. Take the example of a person 'going to the hospital'. Although the implied cause 'ill' is semantically related to 'hospital', it is still necessary to appreciate the unfolding of the whole event to make an appropriate causal inference, because other events with the same semantic associate could lead to different attributions. To illustrate, leaving the hospital might imply 'health' or 'cured', attacking the hospital might imply 'pro-live protests' and so on. Consequently, in both spontaneous and intentional conditions, understanding of the unfolding of the events was necessary to make causal inferences.

Another limitation is that sentences were only matched on valence and word count. Hence, it is possible that additional features in the material may explain our results. For instance, because we did not control for temporal stability, it is possible that the greater number of stable person dispositions *vs* unstable situation characteristics in our sentences might be responsible for the increase in mPFC activation for person relative to situation causes, rather than anything inherent to person causes themselves. Nevertheless, the fact that this effect was observed only under spontaneous instructions but not under intentional instructions speaks against an explanation in terms of features of the material.

Our results leave a number of questions unanswered. One of the issues is that spontaneous inferences may occur quickly and further inform more deliberate decisions taking into account more information (Gilbert et al., 1988; Reeder et al., 2002, 2004; Todd et al., 2011). To check this assumption, ERP measures could give some precise indication of the timing of the activations under spontaneous vs intentional instructions. Moreover, ERP measures could also help to disambiguate co-occurrence (Reeder et al., 2002, 2004) and correction theories (Gilbert et al., 1988) of spontaneous inferences. These theories predict that observers activate spontaneously either many relevant causes (Reeder et al., 2002, 2004) or only person causes (Gilbert et al., 1988) before deliberative processing. Although the large overlap between situation and person inferences in the present study seems to favor a co-occurrence account (Reeder et al., 2002, 2004), it is still unclear whether observers spontaneously encoded both possible inferences at about the same time as the co-occurrence account suggests. This question contains a timing component, measurable with ERP.

Recent research suggests that the mPFC responds more to ambiguous compared to unambiguous material (Jenkins and Mitchell, 2009). Thus, in this study, the mPFC might have been activated more because of the inherent ambiguity of the spontaneous instruction rather than the fundamental attribution bias as we hypothesized. Although our pilot study was conducted to make sure that our sentences were unambiguously implying either a person or situation inference, making spontaneous inferences relatively straightforward, we may have been less successful on this point. Hence, we cannot completely rule out this alternative ambiguity explanation. Follow-up research could try to implicate situations or persons more strongly as an unambiguous causes, for instance by implicit primes.

It is also interesting to note that across attribution types and instructions, there was robust activation in a number of areas not well-documented in social cognitive inference: the mid-temporal gyrus, precentral gyrus and insula, all at the left hemisphere (Figure 2). Since these activations may be due to some particularities of the present material (as it was prepared to include person or situation attributions, but not both), any interpretation is premature. We simply note that the mid-temporal gyrus is involved in accessing word meaning while reading and memory, that the precentral gyrus is the location where motor nerves depart (FitzGerald and Folan-Curran, 2002) and hence may be involved in imagining of motor actions, and that the insula is not only involved in somatosensorial information processing relevant for emotion responding, but also involved in memory processing (Ross and Olson, 2010) and language (FitzGerald and Folan-Curran, 2002). Future research should establish whether the activation of these areas is in any way replicated.

CONCLUSION

Our main finding was that observers' own beliefs about the causality of an actor's behavior in terms of this actor or the situation rely on similar mentalizing brain areas as inferring another person's beliefs. This was the case when these inferences were made intentionally or spontaneously, with the exception that the mPFC was not involved during the intentional instructions. Moreover, intentional inferences invited more elaborate and extensive processing of situational information. In contrast, spontaneous inferences increased activation of the mPFC, presumably reflecting the tendency to overestimate personal causality (i.e. the fundamental attribution bias).

Conflict of Interest

None declared.

REFERENCES

- Aichhorn, M., Perner, J., Kronbichler, M., Staffen, W., Ladurner, G. (2006). Do visual perspective tasks need theory of mind? *NeuroImage*, 30, 1059–68.
- Amodio, D.M., Frith, C.D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, 7, 268–77.
- Apperly, I.A., Butterfill, S.A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953–70.
- Apperly, I.A., Samson, D., Chiavarino, C., Humpreys, G.W. (2004). Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false belief task with reduced language and executive demands. *Journal of Cognitive Neuroscience*, 16, 1773–84.
- Aron, A.R., Robbins, T.W., Poldrack, R.A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Science*, 8, 170–7.
- 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.
- Blackwood, N.J., Bentall, R.P., Ffytche, D.H., Simmons, A., Murray, R.M., Howard, R.J. (2003). Self-responsibility and the self-serving bias: an fMRI investigation of causal attributions. *NeuroImage*, 20, 1076–85.
- Botvinick, M.M., Cohen, J.D., Carter, C.S. (2004). Conflict monitoring and anterior congulate cortex: an update. *Trends in Cognitive Sciences*, 8, 539–546.
- Carlston, D.E., Skowronski, J.J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, 66, 840–56.
- Carlston, D.E., Skowronski, J.J., Sparks, C. (1995). Savings in relearning. II. On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, 69, 420–36.
- Carrington, S.J., Bailey, A.J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, 30, 2313–35.
- den Ouden, H.E.M., Frith, U., Frith, C., Blakemore, S.J. (2005). Thinking about intentions. *NeuroImage*, 28, 787–96.
- De Neys, W. (2006). Dual processing in reasoning: two systems but one reasoner. Psychological Science, 17, 428–33.
- 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.
- Duff, K.J., Newman, L.S. (1997). Individual differences in the spontaneous construal of behavior: idiocentrism and the automatization of the trait inference process. *Social Cognition*, 15, 217–41.
- FitzGerald, M.J.T., Folan-Curran, J. (2002). Clinical Neuroanatomy and Related Neuroscience. London, UK: Elsevier.
- Frith, U., Frith, C. (2001). The biological basis of social interaction. Current Directions in Psychological Science, 10, 151–5.
- 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 busyness: when person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, 54, 733–40.
- Ham, J., Vonk, R. (2003). Smart and easy: co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*, 434–47.
- Harris, L.T., Todorov, A., Fiske, S.T. (2005). Attributions on the brain: neuro-imaging dispositional inferences, beyond theory of min. *NeuroImage*, 28, 763–9.
- Hassin, R.R., Aarts, H., Ferguson, M.J. (2005). Automatic goal inferences. Journal of Experimental Social Psychology, 41, 129–40.
- Hassin, R.R., Bargh, J.A., Uleman, J.S. (2002). Spontaneous causal inferences. Journal of Experimental Social Psychology, 38, 512–22.
- Heider, F. (1958). The Psychology of Interpersonal Relations. New York, NY: Wiley.
- Jenkins, A.C., Mitchell, J.P. (2009). Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. *Cerebral Cortex*, 20, 404–10.
- Jones, E.E., Nisbett, R.E. (1972). The actor and the observer: divergent perceptions of the causes of behavior. In: Jones, E.E., Kanouse, D.E., Kelley, H.H., Nisbett, R.E., Valins, S., Weiner, B., editors. Attribution: Perceiving the Causes of Behavior. Morristown, NJ: General Learning Press, pp. 79–94.

- Keysers, C., Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in Cognitive Science*, 11, 194–6.
- Lieberman, M.D. (2007). Social cognitive neuroscience: a review of core processes. Annual Review of Psychology, 58, 259–89.
- Lupfer, M.B., Clark, L.F., Hutcherson, H.W. (1990). Impact of context on spontaneous trait and situational attributions. Journal of Personality and Social Psychology, 58, 239–49.
- 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*, 1, 1–16.
- Mitchell, J.P. (2009). Social psychology as a natural kind. Trends in Cognitive Sciences, 13, 246-51.
- Mitchell, J.P., Cloutier, J., Banaij, 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., Heatherton, T.F., Macrea, C.N. (2002). Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences USA*, 99, 15238–43.
- 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.
- Nelson, O.T. (1985). Ebbinghaus's contribution to the measurement of retention: savings during relearning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 472–9.
- Olson, I.R., Plotzker, A., Ezzyat, Y. (2007). The enigmatic temporal pole: a review of findings on social and emotional processing. *Brain*, 130, 1718–31.
- Read, S.J., Jones, D.K., Miller, L.C. (1990). Traits as goal-based categories: the importance of goals in the coherence of dispositional categories. *Journal of Personality and Social Psychology*, 58, 1048–61.
- Reeder, G.D. (2009). Mindreading: judgments about intentionality and motives in dispositional inference. *Psychological Inquiry*, 20, 1–18.
- Reeder, G.D., Kumar, S., Hesson-McInnis, M.S., Trafimow, D. (2002). Inferences about the morality of an aggressor: the role of perceived motive. *Journal of Personality and Social Psychology*, 83, 789–803.
- Reeder, G.D., Vonk, R., Ronk, M.J., Ham, J., Lawrence, M. (2004). Dispositional attribution: multiple inferences about motive-related traits. *Journal of Personality and Social Psychology*, 86, 530–44.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: distortions in the attribution process. In: Berkowtiz, L., editor. Advances in Experimental Social Psychology. New York, NY: Academic, pp. 174–221.
- Ross, L.A., Olson, I.R. (2010). Social cognition and the anterior temporal lobes. *NeuroImage*, 49, 3452–462.
- Saxe, R., Powell, L.J. (2006). It's the thought that counts. Psychological Science, 17, 692-9.
- Saxe, R., Carey, S., Kanwisher, N. (2004a). Understanding other minds: linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, 55, 87–124.
- Saxe, R., Wexler, A. (2005). Making sense of another mind: the role of the right temporo-parietal junction. *Neuropsychologia*, 43, 1391–9.
- Saxe, R., Xiao, D.-K., Kovacsc, G., Perrett, D.I., Kanwisher, N. (2004b). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, 42, 1435–46.
- Seidel, E., Eickhoff, S.B., Kellermann, T., et al. (2010). Who is to blame? Neural correlates of causal attribution in social situations. *Social Neuroscience*, 5, 335–50.
- Smith, E.R., DeCoster, J. (2000). Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4, 108–31.
- Speer, N.K., Reynolds, J.R., Swallow, K.M., Zacks, J.M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science*, 20, 989–99.
- Spunt, R.P., Falk, E.B., Lieberman, M.D. (2010). Dissociable neural systems support retrieval of how and why action knowledge. *Psychological Science*, 21, 1593–6.
- Sugiura, M., Shah, N.J., Zilles, K., Fink, G.R. (2006). Cortical representation of personally familiar objects and places: functional organization of the human posterior cingulated cortex. *Journal of Cognitive Neuroscience*, 17, 183–98.
- Taylor, S., Fiske, S. (1975). Point of view and perceptions of causality. *Journal of Personality* and Social Psychology, 32, 439–445.
- Thomson-Schill, S.L., D'Esposito, M., Aguirre, G.K., Farah, M.J. (1997). Role of the left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proceedings* of the National Academy of Sciences USA, 94, 14792–7.
- Todd, A.R., Molden, D.C., Ham, J., Vonk, R. (2011). The automatic and co-occurring activation of multiple social inferences. *Journal of Experimental Social Psychology*, 47, 37–49.
- 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, NY: Guilford Press, pp. 141–60.
- 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. San Diego, CA: Academic Press, pp. 211–79.

Situation and person attributions

Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A., Van Overwalle, F. (2009). ERP time course and brain areas of spontaneous and intentional goal inferences. *Social Neuroscience*, 4, 165–84.

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.

APPENDIX A

Table A1 List of experimental sentences that implied either the person or the situation as the cause.

Person-implied sentences	Person cause			
can work well together	social			
gives a bouquet at arrival	romantic			
joins the conversation	social			
meets with open arms	warm			
puts away the mess	orderly			
always remains quiet	calm			
listens to their problems	compassionate			
takes care of the homeless	helpful			
smiles with pleasure	happy			
works hard to assist others	helpful*			
doesn't lose his courage	persistent			
thinks that the future is beautiful	optimistic			
looks interested to the cultural info	cultural*			
gives about others	kind			
buys his girlfriend flowers	romantic			
likes looking inside	curious			
reaches for a can of soda	thirsty			
listens to the reporter	interested			
bites in the apple	hungry			
tells the truth	honest			
carries the luggage of the children	helpful			
talks to his colleagues	social			
loves children	motherly			
enjoys going out	uncontrolled			
avoids accidents	careful			
makes 100 push-ups	sportive			
thinks about his girlfriend	in love			
pushes the invalid	brutal			
takes advantage of others	stingy			
talks all the time	talkative**			
enjoys the bloodbath	psychopath			
plays with her feelings	playboy			
never talks to someone	asocial			
hits a young girl	aggressive			
stares at showering children	pedophile			
kicks the corpse	rage			
leans against the handrail	tired			
goes to the hospital	ill			
sleeps on the sofa	tired			
looks at child pornography	pedophile			

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

Van Overwalle, F., Van den Eede, S., Baetens, K., Vandekerckhove, M. (2009). Trait inferences in goal-directed behavior: ERP timing and localization under spontaneous and intentional processing. *Social Cognitive and Affective NeuroScience*, 4, 177–90.

Table A1 Continued

Situation-implied sentences	Situation cause				
earns a salary	work				
gets a present	birthday				
puts the plates on the table	mealtime				
writes a report	school assignment				
searches for eggs with the kids	Easter				
avoids her work on public holidays	home with family				
earns money every month	work				
swims in the Mediterranean	holiday				
drinks at the wedding	party				
listens to the singing	nice singing*				
replaces the ink	empty holder				
enjoys the hot drink	it's cold				
can go on a holiday	has leave**				
cuts into the onion	cooking				
knocks on the nail	construction works				
pushes against the door	opening the door				
enters the waiting room	appointment				
jumps over the puddle	wet				
searches a winter jacket	it's cold				
kicks the ball	soccer				
pushes the swing	playing				
gets off the train	reaching destination				
shows his ticket	getting access				
runs over the pedestrian crossing	traffic rules				
throws the dice on a white-black board	game				
undresses in the locker room	sports				
makes a special dinner	visit				
replaces the tire	flat tire				
pulls the brake	danger				
avoids the dunghill	stinking				
shives from the wind	it's cold				
pushes the car	breakdown				
buy a bandage	accident				
trembles in the sea	water is cold				
pushes the motorbike	breakdown				
hides for the storm	dangerous				
takes the super glue	something broken				
runs away from the rain	wet				
talks loudly during the move	much noise				
is uninterested in the documentary	dull				

Best possible translation from Dutch. Asterisks denote close semantic associations between sentence and implied cause due to *same words **synonyms in Dutch.

(continued)