Social Neuroscience

ISSN: 1747-0919 (Print) 1747-0927 (Online) Journal homepage: https://www.tandfonline.com/loi/psns20

Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: Spontaneous inferences activate only its core areas

Ning Ma , Marie Vandekerckhove , Frank Van Overwalle , Ruth Seurinck & Wim Fias

To cite this article: Ning Ma , Marie Vandekerckhove , Frank Van Overwalle , Ruth Seurinck & Wim Fias (2011) Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: Spontaneous inferences activate only its core areas, Social Neuroscience, 6:2, 123-138, DOI: 10.1080/17470919.2010.485884

To link to this article: https://doi.org/10.1080/17470919.2010.485884

Published online: 21 Jul 2010.

SOCIAL NEUROSCIENCE, 2011, 6 (2), 123–138

Psychology Press
Taylor & Francis Group

Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: Spontaneous inferences activate only its core areas

Ning Ma, Marie Vandekerckhove, and Frank Van Overwalle
Vrije Universiteit Brussel, Brussels, Belgium

Ruth Seurinck and Wim Fias
Ghent University, Ghent, Belgium

This fMRI study analyzes inferences on other persons' traits, whereby half of the participants were given spontaneous ("read") instructions while the other half were given intentional ("infer the person's trait") instructions. Several sentences described the behavior of a target person from which a strong trait could be inferred (trait diagnostic) or not (trait nondiagnostic). A direct contrast between spontaneous and intentional instructions revealed no significant differences, indicating that the same social mentalizing network was recruited. There was, however, a difference with respect to different brain areas that passed the significance threshold, suggesting that this common network was recruited to a different degree. Specifically, spontaneous inferences significantly recruited only core mentalizing areas, including the temporo-parietal junction and medial prefrontal cortex, whereas intentional inferences additionally recruited other brain areas, including the (pre)cuneus, superior temporal sulcus, temporal poles, and parts of the premotor and parietal cortex. These results suggest that intentional instructions invite observers to think more about the material they read, and consider it in many ways besides its social impact. Future research on the neurological underpinnings of trait inference might profit from the use of spontaneous instructions to get purer results that involve only the core brain areas in social judgment.

Keywords: fMRI; Trait inferences; Spontaneous; Intentional.

INTRODUCTION

As social beings, humans have to understand the behaviors and thoughts of other people around them. When we meet novel persons, we make quick impressions as a first step to knowing them. Sometimes we form these impressions unintentionally, for instance, about the person who just sits near you on the train. At other times, we form impressions while we try to identify traits explicitly; for example, a manager trying to know the personality of a job candidate. What are the neurological processes underlying these two kinds of impression formation? Past neuroscientific research has explored in some detail the brain activity involved in forming trait impressions about another person. The present study investigates whether or not this activity is different when people form an impression spontaneously vs. intentionally.

Intentional trait inference (ITI) refers to impressions made with the explicit goal of identifying the traits of a target person, while spontaneous trait inferences (STIs) are formed without intention or awareness (Uleman, 1999; Uleman, Blader, & Todorov, 2005). According to dual-process models in the social cognition literature, social information processing is based on two general-purpose systems: either an automatic

Correspondence should be addressed to: Frank Van Overwalle, Department of Psychology, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. E-mail: Frank.VanOverwalle@vub.ac.be

This research was supported by an OZR Grant of the Vrije Universiteit Brussel to FVO.

www.psypress.com/socialneuroscience DOI: 10.1080/17470919.2010.485884

associative process or controlled symbolic reasoning (Keysers & Gazzola, 2007; Satpute & Lieberman, 2006; Smith & DeCoster, 2000; Uleman, 1999). The associative system relies on prior knowledge and beliefs, and uses basic cognitive operations such as association, similarity, and memory retrieval to produce primitive judgments quickly and spontaneously. The symbolic system reflects more evolutionarily advanced mechanisms that implement reasoning procedures deliberately and consciously, and largely depends on the application of rule-based procedures and logical standards (De Neys, 2006; De Neys, Vartanian & Goel, 2008). Applied to social inferences, STIs are relatively automatic in the sense that they require little mental effort, and are difficult to suppress or modify, such as when we form a quick negative impression about a person that pushed a homeless child (e.g., aggressive). In contrast, ITIs involve a deliberate attempt to make a relevant social judgment, which requires more mental effort (Uleman et al., 2005), such as when after some more observation and thought it becomes evident that this person pushed to defend himself against stealing (e.g., fair). Behavioral research during the past 30 years has convincingly demonstrated that trait inferences can be activated spontaneously from behavioral descriptions that imply a trait (Winter & Uleman, 1984).

There is now also neurological evidence on the existence of STIs (Mitchell, Cloutier, Banaji & Macrae, 2006; Moran, Heatherton & Kelley, 2009; Van Duynslaeger, Van Overwalle, & Verstraeten, 2007). However, in recent neuroscientific theory, the distinction between spontaneous and intentional modes of thinking has taken a somewhat different perspective in terms of differences between simple cortical networks that are closely related to the senses and actions vs. more complex cortical networks that represent more abstract information of perceptual or executive character (Fuster, 2006). In particular, recent developments suggest a distinction between a cognitively efficient but inflexible capacity for tracking the beliefs and states of others shared by infants and nonhuman chimpanzees vs. a later developing, more flexible but more cognitively demanding ability to understand the mind and traits of others (Apperly & Butterfill, 2009). To illustrate, chimpanzees and young infants can follow eye direction or pointing by someone else in nonmentalistic terms of the importance of a situation or request for action (Doherty, 2006; Gomez, 2007) and they can understand the goals and knowledge of others (Call & Tomasello, 2008). However, they do not understand eye direction or pointing in mentalistic terms (e.g., what the other "wants"), nor do they understand that others may have "false" beliefs that contradict their own beliefs or reality. Van Overwalle and Baetens (2009) pointed to a similar distinction between lower-level perceptual processing subserved by a mirror network vs. higher-level symbolic or abstract processing subserved by the mentalizing network.

In sum, the dual-process approach poses a "puzzling" question: "that human adults seems so quick and efficient in their social interaction . . . yet . . . adults' belief reasoning is [also] relatively effortful and dependent upon limited cognitive resources" (Apperly & Butterfill, 2009, p. 966). The crucial question is thus whether spontaneous (efficient) and intentional (effortful) social reasoning by human adults tap into the same neural circuitry—perhaps to a different degree—or whether they involve different underlying processes and related brain areas. This question is not only relevant for trait inferences, but also extends to inferences and theories on spontaneous social thought in general (Keren & Schul, 2009). The question is particularly timely because most functional magnetic resonance imaging (fMRI) research on trait inferences has been conducted with explicit instructions to infer a trait (see Van Overwalle, 2009) while spontaneous inferences have been relatively neglected.

NEUROSCIENTIFIC STUDIES ON TRAIT INFERENCES

Recent neuroimaging studies using fMRI or event-related potentials (ERPs) suggest that two brain areas are predominantly involved in social inferences of others: the medial prefrontal cortex (mPFC) and temporo-parietal junction (TPJ; see also meta-analysis by Van Overwalle, 2009). These two regions form the core areas of the social mentalizing network. Several studies indicate that the mPFC seems to be essential in the attribution of enduring traits (Harris, Todorov, & Fiske, 2005; Mitchell, Banaji, & Macrae, 2005; Todorov, Gobbini, Evans & Haxby, 2007; Van Duynslaeger et al., 2007). The TPJ is mainly related to temporary goals, intentions, and beliefs of others (Saxe & Powell, 2006; Saxe & Wexler, 2005; for overviews, see Frith & Frith, 2001; Saxe, 2006; Van Overwalle, 2009), although this area is implicated not only in mentalizing, but also in lower-level processes including attention reorientation and self–other distinction (see Decety & Lamm, 2007; Mitchell, 2008).

As noted above, almost all fMRI studies to date that explored trait inferences were conducted using intentional instructions. Nevertheless, a few studies addressed spontaneous inferences. In 2006, a first

fMRI study by Mitchell et al. (2006) addressed the issue of potential differences between spontaneous and intentional trait inferences. 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 revealed only a marginally stronger activation of the mPFC during ITI as opposed to STI for trait-diagnostic descriptions (while this difference was significant for the less interesting, nondiagnostic descriptions). However, the weakness of this result might be due to the fact that the spontaneous ("encode the order") inferences were made with some awareness of trait-relevance, because spontaneous and intentional task instructions were randomly alternated between trials within the same participants, so that the intentional "trait" instructions may have contaminated the nonintentional condition. Another fMRI study (Moran et al., 2009) investigated self-ratings. ITI instructions requested participants to indicate "how much does this trait describe me?" (p. 199) while STI instructions requested to make social desirability judgments on these trait adjectives. Importantly, the STIs were made before the ITIs were requested, so that trait instructions could not contaminate spontaneous processing. The results revealed significantly increased activity under ITI in the mPFC, as well as in the posterior cingulate cortex (PCC) and the right superior premotor cortex (PMC). There was a positive relationship with explicit self-relevance ratings and the mPFC, while implicit self-relevance was not related. Although relevant for trait inferences about the self, this study has little to tell about inferences on others, which is the focal question of the present research.

An ERP study exploring trait inferences of others by Van Duynslaeger et al. (2007) avoided possible contamination between instructions by using two groups of participants. After reading behavioral descriptions that implied a trait, the participants were requested either "to read . . . attentively" (STI) or "to form a trait impression about the agent" (ITI, p. 179). Applying a LORETA source analysis on their ERP data, these researchers found important differences in a time window from 400 ms up to 700 ms that matched the time at which participants inferred the trait (as detected by greater ERP positivity following trait inconsistencies). There was more involvement of the TPJ especially during the early stages of spontaneous inferences (< ~600 ms), while there was more involvement of the mPFC given intentional inferences starting at a later stage (~600 ms). Although the correspondence between LORETA and fMRI localization is quite high for these cortical regions (see Mulert et al., 2004; Vitacco, Brandeis, Pascual-Marqui, & Martin, 2002), given the rougher spatial resolution of LORETA than fMRI, these results need replication with fMRI.

Taken together, although spontaneous and intentional trait instructions induce different processing modes, the results so far suggest that they nevertheless rest on the same mentalizing network including the mPFC and TPJ as proposed by Van Overwalle (2009), and possibly also the PCC. To improve on some methodological issues in prior research, in this experiment we investigate spontaneous vs. intentional trait inferences by comparing between groups of participants rather than within participants. This should avoid any carry-over effects of intentional inferences in contrast to intentional and spontaneous instructions that are manipulated between trials within the same participants. To avoid the limitations of LORETA, we used the fMRI methodology, which provides much higher spatial resolution.

PRESENT RESEARCH AND HYPOTHESES

We used a modified version of Van Duynslaeger et al.'s (2007) ERP paradigm. Participants were given behavioral descriptions which were diagnostic for a trait and they were requested either to make trait inferences about each target agent (ITI) or to read the stimulus material carefully while mentioning nothing about impression formation or traits (STI).

Our prediction is that both spontaneous and intentional instructions lead to trait inferences causing activity in the same core social brain areas comprising the TPJ and mPFC, as documented in the ERP study by Van Duynslaeger et al. (2007) and previous fMRI studies (Van Overwalle, 2009). We also expect some differences: Given that the mPFC is a part of the frontal cortex involved in more controlled executive processing and symbolic reasoning (Keysers & Gazzola, 2007; Miller and Cohen, 2001; Satpute and Lieberman, 2006) and because this brain region has been reported in fMRI studies using mainly explicit instructions (Mitchell et al., 2005; Todorov et al., 2007; Saxe and Powell, 2006; Van Duynslaeger et al., 2007), we expect it to be particularly active during ITI. In contrast, to the extent that the TPJ subserves a truly automatic inference process, it is expected to be equally active under ITI and STI. However, since Van Duynslaeger et al. (2007) found that the TPJ is the most active area under spontaneous instructions—at least during initial

processing stages—it is also possible that the TPJ will be more recruited during STI.

We also used two memory measures taken immediately after the presentation of all stimulus material, as behavioral validation of the fMRI data. The first, a trait-cued recall task developed by Winter and Uleman (1984), is an effective method to validate the occurrence of trait inferences. Enhanced recall cued by the implied trait suggests that trait interpretations were made following the diagnostic behaviors and encoded in memory together with sentences. The second memory measure is a sentence completion task, borrowed from Bartholow, Fabiani, Gratton, and Bettencourt (2001) and Bartholow, Pearson, Gratton, and Fabiani (2003). The participants are given the sentences during scanning but lacking the last, trait-implying word. Again, enhanced recall of these critical words indicates that trait inferences were made and integrated with the diagnostic behaviors. For both these memory measures, we predict better recall for diagnostic trait-implying information as opposed to nondiagnostic trait-irrelevant information, and we also expect that at least one of these measures will show positive correlations with brain activity in the TPJ or mPFC.

METHOD

Participants

Participants were all right-handed, 15 women and 15 men, with ages varying between 18 and 50. 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 (5 women and 10 men) received an STI, while the other half (10 women and 5 men) received an ITI. Informed consent was obtained in a manner approved by the Medical Ethics Committee at the Hospital of University of Ghent.

Procedure and stimulus material

The design and stimulus material were trait-consistent and trait-irrelevant sentences borrowed from ERP studies on trait inferences by Van Duynslaeger et al. (2007) and Bartholow et al. (2001, 2003). Participants read a large number of events that described the behavior of a fictitious target agent and from which a strong trait can be inferred. The events involve positive and negative moral traits. To avoid associations with a familiar and/or existing name, fictitious 'Trek'-like names were used (Bartholow et al., 2001, 2003). For each agent, a series of four behavioral sentences was presented. Each sentence consisted of six words and was presented once in the middle of the screen for 5.5 s. To optimize estimation of the event-related fMRI response, each sentence for an agent was separated by a variable interstimulus interval of 2.5 to 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. The last word of each sentence was the critical one, because it determined the degree of consistency with the previously inferred trait: trait-diagnostic and irrelevant. Trait-diagnostic sentences described positive or negative moral behaviors that were consistent with the inferred trait (for example "Tolvan gave her sister a hug" is consistent with the trait "friendly"). The irrelevant sentences described neutral behaviors (for example "Tolvan gave her mother a bottle").

To make sure that the participants were attending to the task and instructions, in the spontaneous trait instruction (STI) they had to respond for about one-third of the agents on a control question asking whether the agent was a female or not by pressing a response key. This question was asked after all four behavioral sentences on an agent were given. In the ITI, the participants additionally had to respond for all agents that ended with trait-implying sentences whether a given trait was correct or not, and on all irrelevant sentences whether the agent was female or not. Given that gender is automatically induced during sentence reading, this control question interferes minimally with the social inference processes under study.

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 for the incomplete trait 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 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.

Statistical analyses were performed using the general linear model of SPM5 of which the event-related design is modeled using a canonical hemodynamic response function and its temporal derivative. Comparisons of interest were implemented as linear contrasts using a random-effects model. A voxel-based statistical threshold of $p \leq 0.005$ (uncorrected) is used for all comparisons. 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 performed with the small volume correction in SPM5. For all ROI analyses, small volume correction was required to exceed 10 contiguous voxels in extent and based on a sphere of 15 mm radius around the centers (in MRI coordinates) of areas that were identified in the meta-analysis by Van Overwalle and Baetens (2009) as involved in mentalizing: 0 –60 40 (precuneus, PC), ±50 –55 25 (TPJ), 0 50 20 (mPFC); action understanding via mirror neurons: ±50 –55 10 (posterior superior temporal sulcus, pSTS), ±40 –40 45 (anterior intraparietal sulcus, aIPS), ±40 5 40 (PMC); and by Sugiura, Shah, Zilles, and Fink (2006) in person identity, ±45 5 –30 (temporal pole, TP). Analyses of the ROI were conducted using *t*-tests with a threshold of $p < 0.05$, FDR corrected. In addition, the mean % 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

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. Our prediction was that if traits were inferred during reading of the sentences, then these traits would be stored in memory together with the sentences and so facilitate (a) recall of the sentences by the aid of a trait cue and (b) 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 Diagnosticity (diagnostic vs. irrelevant) as within-participants factor. For trait-cued recall, the analysis revealed better recall for diagnostic as opposed to irrelevant sentences, $F(1, 28) = 33.12$, $p < .001$ (Table 1). There were no other effects. For sentence completion, the analysis revealed the same pattern of results indicating better memory for diagnostic as opposed to irrelevant sentences, $F(1, 28) = 102.04$, $p < .001$, and no other effects. This indicates that, as predicted, trait inferences were made to the same extent irrespective of instructions.

TABLE 1
Proportion (%) of correct memory at various measures as a function of instruction and trait diagnosticity

	Spontaneous		*Intentional*	
	Diagnostic	*Irrelevant*	*Diagnostic*	*Irrelevant*
Cued recall	8[a]	2[b]	8[a]	1[b]
Sentence completion	24[a]	7[b]	25[a]	7[b]

Notes: Means in a row sharing the same superscript do not differ significantly from each other according to a Fisher LSD test, $p < .01$.

Whole-brain analysis

Our analytic strategy for the imaging data was as follows. First, to verify whether intentional and spontaneous inferences recruit different or similar brain areas, we conducted a whole-brain, random-effects analysis contrasting the diagnostic > irrelevant conditions. Next, to verify our hypotheses, we conducted an ROI analysis using a small volume correction. In both cases, we calculated FDR corrected significance levels (on the whole brain or the small volumes respectively). For brevity and consistency, we report here only on those areas where the FDR corrected *p*-value surpassed the .10 level in any of these cases.

For the whole-brain analysis, the threshold was set at $p < 0.005$ (to detect all relevant areas also under the shallower, spontaneous processing) with a cluster extent of 10 voxels (Figure 1 and Table 2). Under intentional instructions, the diagnostic > irrelevant contrast revealed significant activation ($p < .05$, FDR corrected) in a large number of brain regions, including the TPJ, mPFC, and PC related to social mentalizing (Van Overwalle, 2009), the STS, aIPS, and PMC related to the mirror system (Van Overwalle & Baetens, 2009), the TP involved in person identity (Sugiura et al., 2006) and social processing (including face recognition and mentalizing; Olson, Plotzker & Ezzyat, 2007) as well as the parahippocampal region and a part of the primary visual cortex. Under spontaneous instructions, this contrast revealed activation in a much more limited number of brain areas, including the TPJ and mPFC, which are core regions of social mentalizing, as well as the TP. These areas fell short of significance after FDR correction ($p < .21$). Note that raising the threshold to a more typical .001 level did not alter the general pattern of results (see regions marked with superscript "a" in Table 2).

Next, we analyzed the critical difference between intentional and spontaneous instructions for the diagnostic > irrelevant contrast. Although more brain regions were involved in intentional than spontaneous inferences as shown above, an ANOVA revealed that the interaction involving Instruction and Diagnosticity (i.e., testing for areas where the diagnostic > irrelevant contrast is larger or smaller for the intentional than for the spontaneous instructions) yielded no significant differences after FDR correction ($p > .70$). This suggests that intentional and spontaneous trait inferences involve the same neural network, although the network is more broadly recruited under intentional than under spontaneous inferences.

A conjunction analysis of the diagnostic > irrelevant contrast on the whole brain showed that there was some overlap in activation under intentional and spontaneous instructions, but surprisingly, no contrast surpassed the FDR correction ($p > .25$). This can also be observed in Figure 1, which shows the unique activation under intentional (red) and spontaneous (green) instruction, with only small and scattered spots of overlap (yellow). When the whole-brain threshold is lowered to $p < .01$ (not shown), the overlap grows larger, especially at the mPFC where the overlap covers almost the entire area activated under STI. This suggests that although the same brain areas are active under the two instructions, their strongest and unique trait-relevant activation (i.e., surpassing the whole-brain significance threshold of $p < .005$) is not always located at exactly the same voxels.

ROI analysis

To explore our specific hypotheses on the regions known to be involved in or related to the mentalizing system (PC, TPJ, mPFC), the mirror system (pSTS, aIPS, PMC), or person identity (TP), we conducted an ROI analysis using small volume correction by drawing spheres of 15 mm radius around the a priori defined centers of these ROIs (see "Method"). We contrasted the diagnostic > irrelevant conditions for each instruction, using *t*-tests with an FDR corrected threshold of $p < 0.05$ and 10 contiguous voxels extent. The results confirm the whole-brain analysis but, as might be expected, show additional ROIs that are significant (see also Table 2 for significance levels). Under intentional instructions, significant diagnostic > irrelevant contrasts ($p < .05$) were found at the medial side and at both lateral sides of the following ROIs: the mPFC, TPJ and PC involved in mentalizing, the STS and aIPS involved in mirroring, and the TP (left TP at $p < .01$) involved in identity processing. Under spontaneous instructions, only the mentalizing ROIs including the mPFC and right TP showed a significant contrast ($p < .05$), while the left TPJ and left TP were marginally significant ($p < .09$). The ANOVA revealed no interaction for any of the ROIs, that is, no significant differences between intentional and spontaneous instructions for the diagnostic > irrelevant contrast. Again, a conjunction analysis showed little overlap in the activation under the two instructions, except for two ROIs: the left pSTS and left TPJ (Table 2). These are indeed the largest spots in Figure 1, marked in yellow to denote overlap.

Apart from the small volume correction analyses, we also extracted percentage change estimations from each ROI, and *t*-tests on these estimates basically replicated the diagnostic > irrelevant contrasts. As can be seen in Figure 1 (bottom), under intentional

TABLE 2

Peak voxel, number of voxels, and *t*-value of the Diagnostic > Irrelevant contrasts from the ROI analysis (*Regions of interest*) and additional regions of the whole brain analysis (*Other regions*), all at threshold $p < .005$ (uncorrected; number of voxels ≥ 10)

	MNI coordinates				
Anatomical label	*x*	*y*	*z*	*Voxels*	*Max t*
Intentional: Diagnostic > Irrelevant					
Regions of interest					
Precuneus (& posterior cingulate)	–2	–58	26	326	5.57**[a]
R posterior STS	42	–42	8	563	6.65**[a]
L posterior STS	–42	–42	14	248	3.78**
R Temporo-parietal junction	44	–62	30	628	6.79**[a]
L Temporo-parietal junction	–58	–60	32	680	5.51**[a]
R Temporal pole	36	0	–20	174	5.93**[a]
L Temporal pole	–50	–6	–24	47	4.65*[a]
R Anterior intraparietal sulcus	44	–38	32	133	5.77**[a]
L Anterior intraparietal sulcus	–36	–30	56	273	4.29**[a]
Medial prefrontal cortex	14	48	16	141	5.81**[a]
Other regions					
R Calcarinus (primary visual cortex)	12	–88	6	286	4.98**[a]
L Calcarinus (primary visual cortex)	–12	–96	2	540	4.42**[a]
L Precentral (PMC)	–38	–26	68	26296	7.56**[a]
Parahippocampal	–4	–18	–16	16	4.44**[a]
R Inferior temporal (STS)	52	–18	–32	78	4.93**
L Superior temporal (STS)	–48	16	–18	58	4.62**
R Inferior frontal (PMC)	54	34	0	135	5.57**[a]
L Superior frontal (PMC)	–20	26	62	108	4.66**[a]
Spontaneous: Diagnostic > Irrelevant					
Regions of interest					
L Temporo-parietal junction	–62	–54	24	24	3.78*
R Temporal pole	44	8	–24	113	4.43**[a]
L Temporal pole	–50	10	–26	34	4.88*
Medial prefrontal cortex	4	54	28	310	4.46**[a]
Conjunction of Intentional + Spontaneous: Diagnostic > Irrelevant					
Regions of interest					
L posterior STS	–42	–58	4	54	2.98**
L Temporo-parietal junction	–58	–58	32	48	3.06**

Notes: Coordinates refer to the MNI (Montreal Neurological Institute) stereotactic space. ROIs 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) and 0 50 20 (mPFC). R = right; L = left; STS = superior temporal sulcus; PMC = premotor cortex. All regions thresholded at $p < .005$, number of voxels ≥ 10; Regions denoted by [a] also significant at whole brain threshold $p < .001$. *$p < .10$, ** $p < .05$, FDR corrected (for *Other regions* corrected after whole brain analysis; for *ROIs* corrected after small volume analysis). This table lists only areas with $p < .10$ after this FDR correction.

instructions, the same ROIs were now also significant ($p < .05$; for the right aIPS, $p = .10$). Under spontaneous instructions, the same ROIs that were significant before were now also significant ($p < .05$; for the right TP, $p < .06$); and the left pSTS was additionally significant ($p < .05$).

An ANOVA on the percentage change estimations with Instruction, Diagnosticity, and Region as factors revealed a main effect of Diagnosticity, $F(1, 28) = 18.35$, $p < .001$, which was confirmed for all ROIs separately ($p < .05$; except right PMC $p < .10$). Crucially, the ANOVA again failed to reveal a significant interaction between Instruction and Diagnosticity, as well as a three-way interaction with Region, F values < 1. Separate ANOVAs confirmed that the interaction between Instruction and Diagnosticity was absent in all ROIs ($p > .30$), replicating the whole-brain analysis reported above. That the main effect of Diagnosticity is not modulated by an interaction with Instruction in none of the ROIs is additional evidence that all the areas show stronger inference activity given diagnostic sentences, independent of instruction. This strengthens the conclusion that a common network is involved.

To summarize the results up to now, the whole-brain and ROI analyses indicate that intentional and spontaneous trait inferences recruit the same neural network involved in social mentalizing. However, the

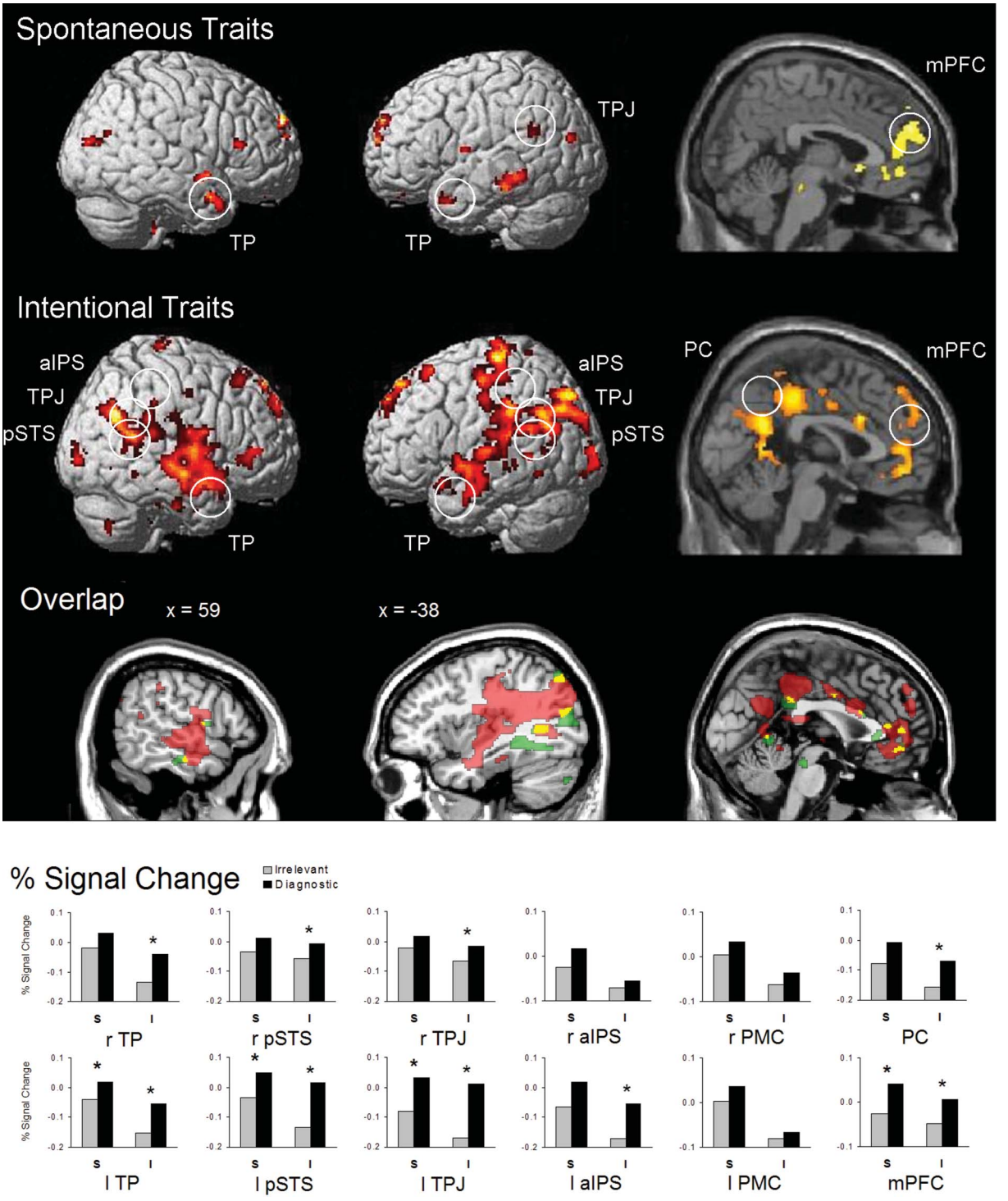

Figure 1. The Diagnostic > Irrelevant contrast under spontaneous and intentional instructions. Whole-brain activation thresholded at $p < .005$ (uncorrected) with at least 10 voxels. Circles indicate ROIs with significant activation after FDR correction. The overlap was created using MRIcro, showing selected areas under intentional (red) or spontaneous (green) instructions at the same whole-brain threshold, and their overlap (yellow). % Signal change was based on ROIs with 15 mm sphere created by MarsBaR. Significant changes ($p < .05$) are indicated by an asterisk above the relevant conditions; gray bar = Irrelevant, black bar = Diagnostic, S = Spontaneous, I = Intentional.

specific voxels that are activated are not identical under the two processing modes. More importantly, the network is activated to a different degree, as it is more strongly recruited under intentional processing than under spontaneous processing. Given a specific threshold, the intentional instructions show more active and extensive clusters of brain activation than spontaneous instructions for trait diagnostic sentences in contrast to irrelevant sentences.

Why do intentional instructions lead to higher activation? Since we presented four behavioral sentences for each agent (as in ERP research on spontaneous vs. intentional trait instructions; e.g., Van Duynslaeger et al., 2007), we can explore to what extent this stronger activation is due to higher elaboration of trait implications in the last sentence compared to the first. An ANOVA with a threshold at $p < 0.05$ and a cluster extent of 10 voxels revealed no significant differences in the diagnostic > irrelevant contrast between the first and last sentences. On the contrary, the TPJ and mPFC are already significantly activated when reading the first sentence (Table 3), while this activation does not show up in the last sentence; and nor do other regions show much activation. This suggests that a large amount of deliberation under intentional instructions occurs already during the first sentence.

Correlations with memory measures

To examine to what extent fMRI activation in some brain regions might be indicative of trait inferences as measured by the memory tasks, we computed Pearson correlations with the peak activation of each ROI in each of the instruction conditions. There were no significant correlations under spontaneous instructions. In contrast, the activation in the PC and right TPJ were significantly correlated with sentence completion of diagnostic sentences (Figure 2). Given that the TPJ is involved in mentalizing, while one of the PC's functions is the retrieval of episodic information, these correlations suggest that making trait inferences and remembering doing so under intentional instructions (but not under spontaneous instruction) lead to enhanced elaboration of the information and, as a consequence, better memory of these traits. Presumably there were no reliable correlations for cued recall because of a floor effect due to the low cued recall performance overall.

TABLE 3
Peak voxel and number of voxels in the regions of interest from the Diagnostic > Irrelevant contrast for the first and last sentence under Intentional instructions ($p < .01$, uncorrected)

	MNI coordinates				
Anatomical label	*x*	*y*	*z*	*Voxels*	*Max t*
First sentence: Regions of interest					
R posterior STS	56	–48	14	27	3.44
R Temporo-parietal junction	56	–48	14	17	3.44
L Temporal pole	–40	2	–32	34	3.69
Medial prefrontal cortex	12	50	20	22	3.20
Fourth sentence: Regions of interest					
L posterior STS	–58	–58	20	15	3.36
L Temporo-parietal junction	–58	–58	20	18	3.36
L Temporal pole	–34	–4	–34	15	3.58

Notes: Coordinates refer to the MNI (Montreal Neurological Institute) stereotactic space. ROIs 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) and 0 50 20 (mPFC). R = right; L = left; STS = superior temporal sulcus.

DISCUSSION

This study explored whether brain activity differs when observers form an impression about another person's traits spontaneously or intentionally. In line with our predictions, spontaneous and intentional trait judgments activated a common mentalizing network. However, spontaneous trait instructions reliably activated only the core mentalizing areas (TPJ and mPFC), whereas the intentional instructions activated many more brain areas that presumably have a more supportive role.

Brain localization of trait inferences

That the TPJ and mPFC were involved under both STI and ITI is in line with previous fMRI research documenting that these areas are implicated in mentalizing about others (Amodio & Frith, 2006; Frith & Frith, 2001; Harris et al., 2005; Saxe, 2006; Saxe & Kanwisher, 2003; Van Overwalle, 2009). Moreover, the lack of significant differences between STI and ITI replicates the results of a similar ERP study on trait inferences by Van Duynslaeger et al. (2007). This study showed that STI and ITI run largely in synchrony and reveal similar ERP amplitudes and location (using LORETA). Nevertheless, the TPJ was more strongly activated under spontaneous instruction during early processing stages (400–600 ms), while, starting at around 600 ms, the mPFC was more strongly activated during intentional trait processing. Because fMRI does not allow for such a millisecond precision, we are unable to replicate these timing differences at early processing stages.

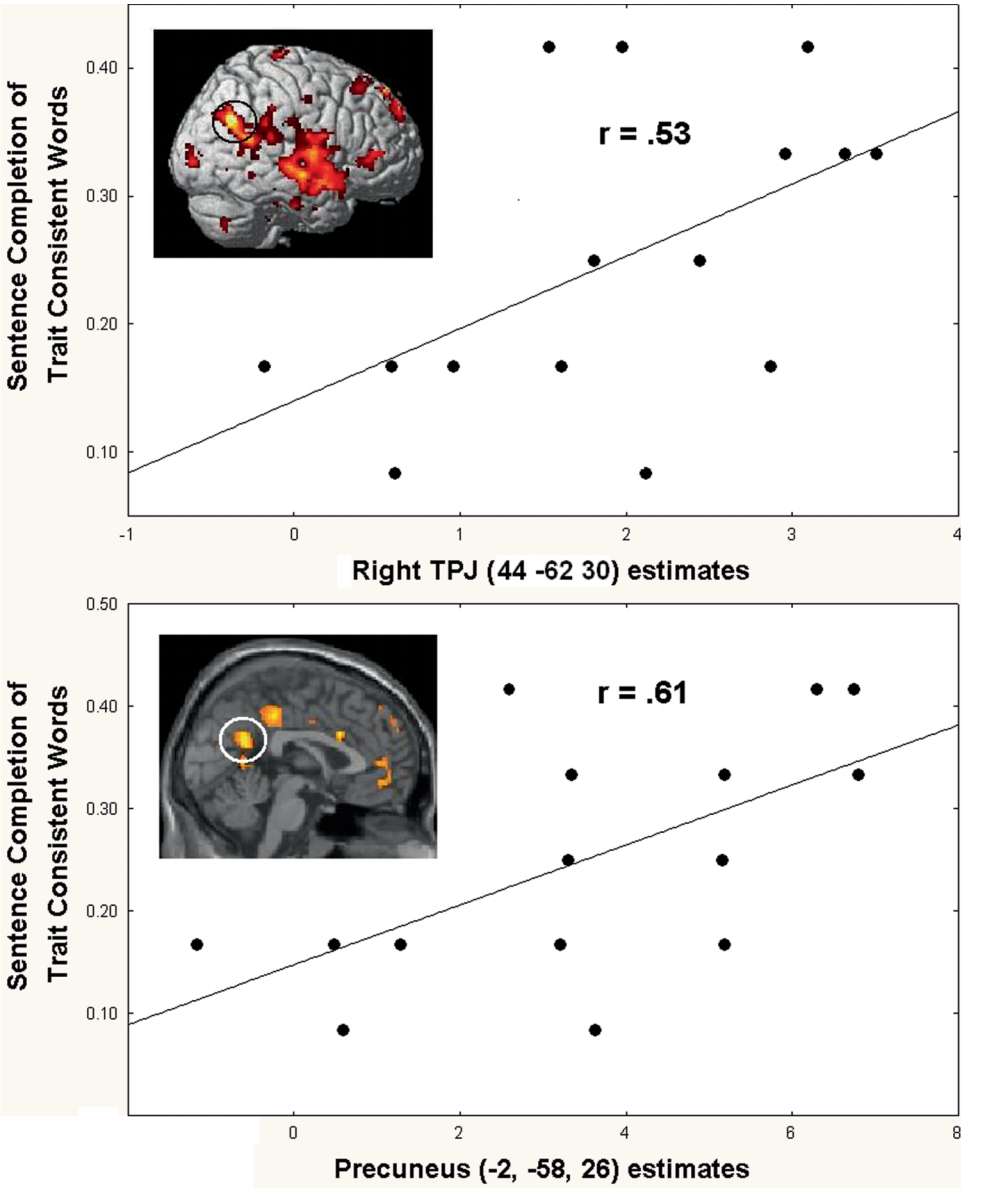

Figure 2. Pearson correlations between memory (proportion correct) for trait-diagnostic words in the sentence completion task and activation (parameter estimates) in the right TPJ and PC.

However, the present study and that of Van Duynslaeger et al. (2007) use very similar materials. Hence, irrespective of methodology, they converge on the conclusion that the TPJ and mPFC are involved in both spontaneous and intentional trait inferences. Given the paucity of ERP research on trait judgments, this overlap in itself is an interesting result.

The present results also extend this earlier ERP research (Van Duynslaeger et al., 2007) as well as previous fMRI research (Mitchell et al., 2006), because it revealed other interesting differences between spontaneous and intentional trait inferences. Although direct differences fell short of significance, after taking a conventional whole-brain threshold ($p < .005$ or $< .001$) we found that additional brain areas were recruited during explicit trait processing. Some of these areas such as the PC (and the nearby PCC) and sometimes the pSTS have been identified as part of the mentalizing network in fMRI research where participants were given explicit instructions to infer characteristics of another person (Amodio et al., 2006; Spiers & Maguire, 2006; Lindner, Hundhammer, Ciaramidaro, Linden & Mussweiler, 2008; Lombardo et al., 2009; Olson et al., 2007; Van Overwalle & Baetens, 2009). The present findings suggest that the explicit instruction to make trait identifications is

responsible for the activation of these areas. In other words, the explicit instructions induce the participants to think more about the material they read during scanning and consider it with respect to additional task-related information in different ways, depending on the demands of the current goal or task. As Uleman (1999, p. 155) stated metaphorically, "spontaneous impression formation processes are part of an underground stream of unconscious thought, while intentional impression formation processes are an aboveground aqueduct system, which is flexibly channeled to destinations consistent with current needs and integrated with other complex processing systems." In line with this perspective, the present study identified both the TPJ and mPFC as the minimally required "underground" network of social thought. The additional brain areas that were activated following intentional instructions provide some clues about the thoughts and deliberations that our participants entertained. One way to view these activations is that these thoughts were to confirm or validate the trait hunches made initially.

Precuneus and posterior cingulate

Validating trait inferences might have occurred by remembering similar events in the past that are highly imaginable or have salient emotional valence. Functional brain imaging studies have revealed that past experiences activate regions of the posterior midline cortex extending from the cuneus/precuneus to bilateral parahippocampal cortices via the PCC (Botzung, Denkova & Manning, 2008; Cabeza & Nyberg, 2000; Krueger, Moll, Zahn, Heinecke & Grafman, 2007; Maddock, 1999; Szpunar, Chan, & McDermott, 2009; Vandekerckhove, Markowitsch, Mertens, & Woermann, 2005). This is in line with the involvement of the dorsal part of the PCC in the processing of familiar aspects of faces and objects (Sugiura et al., 2005). Hence, these posterior midline structures seem to be involved in executive processes such as response selection based on memory during the process of inferring the right trait. Nevertheless, fMRI studies also found that retrieval of emotional compared to neutral stimuli consistently activates the ventral PCC (Maddock, 1999). As such, the activation of this area in the present study may suggest an important role of the emotional content of trait implications, more than had been assumed previously.

Other studies have shown that self-relevant thoughts or self-referential information modulates activation in posterior and anterior cortical midline structures, including the PCC and mPFC (Craik et al., 1999; D'Argembeau, Xue, Lu, Van der Linden, & Bechara, 2008; Johnson et al., 2002; Kelley et al., 2002; Lieberman, Jarcho, & Satpute, 2004; Lou et al., 2004; Mitchell et al., 2005; Moran et al., 2009; Saxe, Moran, Scholz, & Gabrieli, 2006). Saxe et al. (2006) confirmed a substantial overlap in these areas between self-related and mentalizing tasks within participants. Importantly, Moran et al. (2009) suggested that these midline structures are engaged during explicit self-reflection but not during implicit processes, such as when participants are requested to make social desirability judgments on trait adjectives rather than making explicit ratings of self-relevance. These findings reasonably explain that the activations in the PC and PCC under intentional instructions encourage online self-reflection processing.

Mirror system

Validating trait inferences might take place also by attending to specific behavioral details of the information. Rizzolatti and Craighero (2004) suggested that the PMC and aIPS belong to the mirror neuron system in humans which becomes activated on observing another person carrying out actions, through a noninferential, spontaneous simulation process that leads to an understanding of the action and its goal. Although there is some controversy on the exact role of the mirror system (Hickok, 2008; Jacob, 2008), a large meta-analysis on mirror vs. mentalizing systems for understanding others' actions by Van Overwalle and Baetens (2009) documented that the mirror system does not provide support for mentalizing, but rather that the two systems are independent (see also Spunt, Satpute, & Lieberman, in press). The mirror system is responsible for lower-level action identification (e.g., "gave his wife a slap" [as in one of our sentences] to understand "how" someone acts aggressively) and is preferentially engaged when articulated motions of body parts are perceived via visual input, but also through verbal stimuli when perceptual details are made salient. In contrast, the mentalizing system is most active during higher-level understanding of brief verbal or pictorial descriptions of social behaviors (e.g., "having a fight" to understand "what" someone is doing). Consistent with this, Speer, Reynolds, Swallow, and Zacks (2009) found that when readers are in the process of comprehending a narrated activity, neural indices of perceptual and motor representations are activated in the IPS and PMC mirror areas, together with neural indices of mentalizing (TPJ) and memory (PC and PCC). It is very likely that our elaborated material (a series of four sentences for each agent) induced our participants to focus on the whole narrative, especially when given an explicit

instruction. This may have resulted in deeper processing and activation of a broad range of brain areas, involving not only mentalizing brain areas but also mirror areas as in a true story (Speer et al., 2009).

Superior temporal sulcus

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 overall. Allison, Puce, and McCarthy (2000) found that the STS is involved in the analysis of biological motion from visual cues, and more generally from stimuli that signal actions by others and their implied motions. In a meta-analysis by Hein and Knight (2008), it was found that the STS region is part of an integrated neural network that supports not only motion detection, but also speech and mentalizing (theory of mind). Thus, the contribution of STS may extend to perception of meaningful social stimuli verbally within a more general network of mentalizing and trait inference (Mitchell et al., 2006). Although the STS is not involved in the core process of trait inference, it responds to perception of social behaviors and collaborates with the mentalizing system.

Temporal poles

The TP is the anterior portion of the temporal lobe. An interesting finding is that the TP was activated in both STI and ITI. One possible explanation is that the TP is involved while memorizing people's name or identity, since this identity had to be kept in memory during the whole narrative describing an agent's actions. FMRI experiments revealed greater activation in the temporal pole bilaterally during the retrieval of newly learned names especially when they were related to novel semantic material, such as the person's profession (Grabowski et al., 2001; Tsukiura, Mochizuki-Kawai, & Fujii, 2006, 2008).

Alternatively, because temporal activation is related to successful retrieval of episodic memory and life-like events (Cabeza & Nyberg, 2000; Markowitsch, Vandekerckhove, Lanfermann, & Russ, 2003; Vandekerckhove et al., 2005), another interpretation is that the TP utilizes personal memories to comprehend the state of mind of others (Olson et al., 2007). In reviewing neuroimaging research, these authors found systematic activation of the TP in mentalizing tasks. They suggested that the "TP is sensitive to stimuli with socially important narratives, either in the form of a film strip, a comic strip or a story, and to tasks that require one to analyse other agent's emotions, intentions or beliefs" (Olson et al., 2007, p. 1726). The specific role of the TP in the mentalizing processes is to link the recognition of social cues to emotional interpretations and reactions, and this occurs especially with complex social stimuli within a narrative or script.

Memory measures and correlations

To validate that our functional imaging measures reflect the psychological processes under study, namely trait inferences, we correlated our fMRI results with several memory measures from behavioral social cognition research, namely trait-cued recall (Winter & Uleman, 1984) and sentence completion (Bartholow et al., 2001, 2003). The idea behind these measures is that when trait inferences are made (and relevant brain areas are activated), the trait is integrated and memorized together with the behavioral descriptions leading to increasing memory associations, so that the sentences (or sentence words) are remembered better than if no trait is implied.

Consistent with our predictions, memory was better for diagnostic trait-implying information than for trait-irrelevant information irrespective of instructions. The lack of significant differences between STI and ITI in overall memory performance is taken as evidence that trait inferences were made under both instructions and involved similar memory encoding and storage processes. Because trait inferences were requested under ITI instructions, the same level of performance on these two memory measures suggests that traits were attributed to the agents also under STI, and thus reflects more than action identification (Carlston & Skowronski, 1994; Uleman et al., 1999, 2005; Van Overwalle, Drenth, & Marsman, 1999; Todorov & Uleman, 2002). These results replicate recent ERP research involving trait inferences by Van Duynslaeger et al. (2007) and Van Duynslaeger, Sterken, Van Overwalle, & Verstraeten (2008).

Under intentional instructions, the activation in the PC and TPJ was reliably correlated with memory of the diagnostic sentences in a sentence completion task, indicating that better memory for the traits implied by behavioral information is associated with increased brain activation in these two areas. Importantly, the significant correlation with the TPJ confirms the role of this core area in mentalizing. Although a similar correlation would be expected for the mPFC, this area may have been less activated than in earlier research. Indeed, we found that the mPFC was less active during the last sentence (of four), indicating that there was less explicit reasoning about the

trait when the last sentence confirmed the trait inferred in the previous sentences. The stronger correlation with the PC is understandable given its role in episodic memory (Botzung et al., 2008; Krueger et al., 2007; Szpunar et al., 2009). In contrast, no significant correlations with the memory measures appeared under spontaneous trait inference. Given the coarse fMRI timing, spontaneous processes may be too short-lived to be reliably picked up and correlate with memory, while intentional processes require more time to develop and hence may reveal correlations more easily. This contrasts with Van Duynslaeger et al. (2007), who found reliable correlations with the TPJ and mPFC under spontaneous trait inferences at a relatively early time of 600 ms poststimulus.

Some limitations

An important concern about the present results is that our major conclusion—no essential difference between spontaneous and intentional trait processing—is based on a failure to find significant differences. Affirming the null hypothesis is always a cause for prudence, because the lack of significance may be due to lack of power, substantial noise in the material, and so on. Nevertheless, these problematic explanations are less of a concern here. First, the predicted brain areas of the mentalizing system were revealed under a conventional whole-brain threshold level ($p < .005$), suggesting that reduced signal can be ruled out as explanation. Moreover, both a whole-brain analysis and an analysis of percentage signal change failed to reveal a significant difference between instructions.

In addition, some particularities in the design of this study can also be ruled out. Indeed, earlier studies on trait inferences using the same fMRI methodology on a within-participants design (Mitchell et al., 2006) or using the same between-participants design but a weaker localization methodology (based on ERPs; Van Duynslaeger et al., 2007) also failed to uncover significant differences between instructions. Also the behavioral results in terms of trait-relevant memory failed to reveal significant differences. Given all these findings, it is difficult to escape the conclusion that both spontaneous and intentional trait inferences rely greatly on essentially similar neural processes. Perhaps future research might find a way to confirm this overlap based on significant (rather than nonsignificant) differences. For instance, the phenomenon of fMRI adaptation predicts that repeating identical behaviors should lead to weaker trait inferences and brain activation than if such information is not repeated. If both instructions recruit the same network, then this effect should occur even when this information is repeated using different instructions in subsequent trials (e.g., "infer a trait," ITI, vs. "memorize the sentence," STI).

Contrary to much of earlier fMRI research where, typically, one diagnostic sentence about one agent is presented, this research presented a larger number behavioral sentences (four) to describe a single agent. Of course, this provides a greater amount of information from which an agent's traits could be inferred. Perhaps, given that a relevant trait may have been inferred from the first sentence onwards, this may have given the participants ample time to think about the material they were reading. One may argue that this may have accentuated some (nonsignificant) differences between spontaneous and intentional processing. On the other hand, it is also conceivable that once a trait was inferred (under ITI) or the sentence was understood (under STI), the participants were effectively left alone to think about anything they wanted, so that one would expect more irrelevant noise rather than reliable divergence in brain activation. The present data lends credit to the latter explanation because there were no major differences between the first and last sentences.

CONCLUSION

One conclusion stands out: Spontaneous and intentional trait inferences recruit the same mentalizing network, but to different degrees. Under spontaneous processing, only the core areas of social mentalizing, the TPJ and mPFC, are recruited, while under intentional processing additional areas are activated. Although the present results support the view that different brain areas are responsible for different processes during impression formation, the brain is a highly dynamic system. Earlier ERP studies have documented this dynamic aspect, by exploring the brain areas involved in various information processes at different time windows (Van Duynslaeger et al., 2007, 2008; Van der Cruyssen, Van Duynslaeger, Cortoos, & Van Overwalle, 2009). By coupling these previous ERP results for trait inferences (Van Duynslaeger et al., 2007, 2008) with the present study, we can reconstruct the following time line. During spontaneous processing, initial trait detection begins at about 500 ms while the TPJ is more strongly activated; thereafter both the TPJ and mPFC are recruited. Under intentional processing, initial trait detection begins at about the same time while the mPFC is more strongly activated than the TPJ; followed by the recruitment of many more brain

areas of the mentalizing network. We suggest that processing in these additional areas serves to further support and validate the initial trait identification, pointing again to a dynamic interplay between different brain areas during impression formation. This interpretation is in line with a recent critique by Keren and Schul (2009) that many two-system theories in social cognition that imply the involvement of distinct neural networks (see "Introduction") lack evidence. These authors argued that "the apparent differences between the two modes of processing can be explained by assuming that the nature of processing . . . has to do with the goals of the individual as construed at the time he or she uses the information" (p. 547). Thus, differences in activation of brain areas may reveal a divergence in attention to various pieces of information involved in a common mentalizing network. We would stress, however, that the TPJ and mPFC are core areas that are always necessarily involved in trait inferences extracted from behavioral information.

This study presents a novel strategy to reveal the core areas involved in trait inferences. While earlier studies typically used explicit instructions to infer the traits of another person, we provided action descriptions that spontaneously induced trait inferences during mere reading. Thus, the present approach of asking less resulted in activating less, and this allowed us to identify the core areas of trait inference. We suggest that this approach might be profitably used in other areas of social reasoning, to reveal brain areas that really matter, as opposed to other areas that have a supportive, but not crucial, role. Perhaps this spontaneous approach will allow us to better understand what the specialized functionality of each of these brain areas is. Given that trait-implying stories involve both action and traits, future research is needed to determine whether the TPJ and mPFC have different contributions to each.

Manuscript received 14 December 2009
Manuscript accepted 1 April 2010
First published online 21 July 2010

REFERENCES

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 4*, 267–278.

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews, 7*, 268–277.

Apperly, I., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116*, 953–970.

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*, 627–638.

Botzung, A., Denkova, E., & Manning, L. (2008). Experiencing past and future personal events: Functional neuroimaging evidence on the neural bases of mental time travel. *Brain Cognition, 66*, 202–212.

Cabeza, R. & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience, 2000*, 1–47.

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*, 187–192.

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*, 840–880.

Craik, F. I. M., Moroz, T.-M., Moscovitch, M., Stuss, D.-T., Winocur, G., Tulving, E., et al. (1999). In search of the self: A positron emission tomography study. *Psychological Science, 10*, 26–34.

D'Argembeau, A., Xue, G., Lu, Z. L., Van der Linden, M., & Bechara, A. (2008). Neural correlates of envisioning emotional events in the near and far future. *NeuroImage, 40*, 398–407.

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *The Neuroscientist, 13*, 580–593.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological Science, 17*, 428–433.

De Neys, W., Vartanian, O. & Goel, V. (2008). Smarter than we think: When our brains detect that we are biased. *Psychological Science, 19*, 483–489.

Doherty, M. J. (2006). The development of mentalistic gaze understanding. *Infant and Child Development, 15*, 179–186.

Frith, U., & Frith, C. (2001). The biological basis of social interaction. *Current Directions in Psychological Science, 10*, 151–155.

Fuster, J. M. (2006). The cognit: A network model of cortical representation. *International Journal of Psychophysiology, 60*, 125–132.

Gallese V. (2007). Before and below 'theory of mind': Embodied simulation and the neural correlates of social cognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 362*, 659–669.

Gomez, J.-C. (2007). Pointing behaviors in apes and human infants: A balanced interpretation. *Child Development, 78*, 729–734.

Grabowski, T. J., Damasio, H., Tranel, D., Boles Ponto, L. L., Hichwa, R. D. & Damasio, A.R. (2001). A role for left temporal pole in the retrieval of words for unique entities. *Human Brain Mapping, 13*, 199–212.

Harris, L. T., Todorov, A., & Fiske, S. T. (2005). Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind. *NeuroImage, 28*, 763–769.

Hein, G., & Knight, R. T. (2008). Superior temporal sulcus: It's my area: or is it? *Journal of Cognitive Neuroscience, 20*, 2125–2136.

Hickok, G. (2009). Eight problems for the mirror neuron theory of action understanding in monkeys and human. *Journal of Cognitive Neuroscience, 21*, 1229–1243.

Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind and Language, 23*, 190–223.

Johnson, S. C., Baxter, L. C., Wilder, L. S., Pipe, J. G., Heiserman, J. E., & Prigatano, G. P. (2002). Neural correlates of self-reflection, *Brain, 125*, 1808–1814.

Kelley, W. M, Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event-related fMRI study, *Journal of Cognitive Neuroscience, 14*, 785–794.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: From self to social cognition. *Trends in Cognitive Sciences, 11*, 194–196.

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, 4, 533–550.

Krueger, F., Moll, J., Zahn, J. R., Heinecke, A., & Grafman, J. (2007). Event frequency modulates the processing of daily life activities in human medial prefrontal cortex. *Cerebral Cortex, 17*, 2346–2353.

Lieberman, M. D., Jarcho, J. M., & Satpute, A. B. (2004). Evidence-based and intuition-based self-knowledge: An fMRI study. *Journal of Personality and Social Psychology, 87*, 421–435.

Lindner, M., Hundhammer, T., Ciaramidaro, A., Linden, D. E., & Mussweiler, T. (2008). The neural substrates of person comparison: An fMRI study. *NeuroImage, 40*, 963–971.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., Baron-Cohen, S., & MRC AIMS Consortium (2009). Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience, 22*, 1623–1636.

Lou, H. C., Luber, B., Crupain, M., Keenan, J. P., Nowak, M., Kjaer, T. W., et al. (2004). Parietal cortex and representation of the mental self. *Proceedings of the National Academy of Sciences of the United States of America, 101* (17), 6827–6832.

Maddock, R. J. (1999). The retrosplenial cortex and emotion: New insights from functional neuroimaging of the human brain. *Trends in Neurosciences, 23*, 195–197.

Markowitsch, H. J., Vandekerckhove, M. M. P., Lanfermann, H., & Russ, M. O. (2003). Engagement of lateral and medial prefrontal areas in the ecphory of sad and happy autobiographical memories. *Cortex, 39*, 643–665.

Miller, E. K. & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202.

Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex, 18*, 262–271.

Mitchell, J. P., Banaji, R. B., & Macrae, C. N. (2005). The link between social cognition and self-referential thought in the medial prefrontal cortex. *Journal of Cognitive Neuroscience, 17*, 1306–1315.

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.

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., Jäger, L., Schmitt, R., Bussfeld, P., Pogarell, O., Möller, H.-J., et al. (2004). Integration of fMRI and simultaneous EEG: Towards a comprehensive understanding of localization and time-course of brain activity in target detection. *NeuroImage, 22*, 83–94.

Olson, I. R., Plotzker, A., & Ezzyat, Y. (2007). The enigmatic temporal pole: A review of findings on social and emotional processing. *Brain, 130*, 1718–1731.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192.

Rizzolatti, G., & Luppino, G. (2001). The cortical motor system. *Neuron, 31*, 889–901.

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*, 235–239.

Saxe, R., & Kanwisher, N. (2003). People thinking about people: The role of the temporo-parietal junction in theory of mind. *NeuroImage, 19*, 1835–1842.

Saxe, R., Moran, J., Scholz, J., & Gabrieli, J. (2006). Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Social Cognitive and Affective Neuroscience, 1*, 229–234.

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*, 692–699.

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia, 43*, 1391–1399.

Smith, E. R., & DeCoster, J. (2000). Associative and rule-based processing: A connectionist interpretation of dual-process models. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 323–338). London: Guilford.

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*, 989–999.

Spiers, H. J., & Maguire, E. A. (2006). Spontaneous mentalizing during an interactive real world task: An fMRI study, *Neuropsychologia, 44*, 1674–1682.

Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (in press) Identifying the what, why and how of an observed action: An fMRI study of mentalizing and mechanizing during action observation. *Journal of Cognitive Neuroscience*. Advance online publication, doi:10.1162/jocn.2010.21446

Stone, V. E., & Gerrans, P. (2006). What's domain-specific about theory of mind? *Social Neuroscience, 1*(3–4), 309–319.

Sugiura, M., Sassa, Y., Watanabe, J., Akitsuki, Y, Maeda, Y., Matsue, Y., et al. (2006). Cortical mechanisms of

person representation: Recognition of famous and personally familiar names. *NeuroImage*, *31*, 853–860.

Sugiura, M., Shah, N. J., Zilles, K., & Fink, G. R. (2005) Cortical representation of personally familiar objects and places: Functional organization of the human posterior cingulate cortex. *Journal of Cognitive Neuroscience*, *17*, 183–198.

Szpunar, K. K., Chan, J. C., & McDermott, K. B. (2009). Contextual processing in episodic future thought. *Cerebral Cortex*, *19*, 1539–1548.

Todorov, A., Gobbini, M. I., Evans, K. K., & Haxby, J. V. (2007). Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia*, *45*, 163–173.

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology*, *83*, 1051–1064.

Tsukiura, T., Mochizuki-Kawai, H., & Fujii, T. (2006). Dissociable roles of the bilateral anterior temporal lobe in face–name associations: An event-related fMRI study. *NeuroImage*, *30*, 192–199.

Tsukiura, T., Suzuki, C., Shigemune, Y., & Mochizuki-Kawai, H. (2008). Differential contributions of the anterior temporal and medial temporal lobe to the retrieval of memory for person identity information. *Human Brain Mapping*, *29*, 1343–1354.

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141–160). New York: Guilford Press.

Uleman, J. S., Blader, S. L., & Todorov, A. (2005). Implicit impressions. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 362–392). New York: Oxford University Press.

Vandekerckhove, M. M. P., Markowitsch, H. J., Mertens, M., & Woermann, F. (2005). Bi-hemispheric engagement in the retrieval of autobiographical episodes. *Behavioral Neurology*, *16*, 203–210.

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*, 165–184.

Van Duynslaeger, M., Sterken, C., Van Overwalle, F., & Verstraeten, E. (2008). EEG components of spontaneous trait inferences. *Social Neuroscience*, *3*, 164–177.

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*, 174–188.

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*, 829–858.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*, 564–584.

Van Overwalle, F., Drenth, T., & Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or to the action ? *Personality and Social Psychology Bulletin*, *25*, 450–462.

Vitacco, D., Brandeis, D., Pascual-Marqui, R., & Martin, E. (2002). Correspondence of event-related potential tomography and functional magnetic resonance imaging during language processing. *Human Brain Mapping*, *17*, 4–12.

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*, 237–252.