doi:10.1093/scan/nst098 SCAN (2014) 9, 1185–1192

Traits are represented in the medial prefrontal cortex: an fMRI adaptation study

Ning Ma,[1] Kris Baetens,[1] Marie Vandekerckhove,[1] Jenny Kestemont[1], Wim Fias,[2] and Frank Van Overwalle[1]
[1]Department of Psychology, Vrije Universiteit Brussel, Brussels, Belgium and [2]Department of Psychology, Ghent University, Ghent, Belgium

Neuroimaging studies on trait inference about the self and others have found a network of brain areas, the critical part of which appears to be medial prefrontal cortex (mPFC). We investigated whether the mPFC plays an essential role in the neural representation of a trait code. To localize the trait code, we used functional magnetic resonance imaging (fMRI) adaptation, which is a rapid suppression of neuronal responses upon repeated presentation of the same underlying stimulus, in this case, the implied trait. Participants had to infer an agent's (social) trait from brief trait-implying behavioral descriptions. In each trial, the critical (target) sentence was preceded by a sentence (prime) that implied the same trait, the opposite trait, or no trait at all. The results revealed robust adaptation from prime to target in the ventral mPFC only during trait conditions, as expected. Adaptation was strongest after being primed with a similar trait, moderately strong after an opposite trait and much weaker after a trait-irrelevant prime. This adaptation pattern was found nowhere else in the brain. In line with previous research on fMRI adaptation, we interpret these findings as indicating that a trait code is represented in the ventral mPFC.

Keywords: trait; mPFC; fMRI adaptation

INTRODUCTION

How we form impressions on trait characteristics of other people is one of the central concerns of social cognition. As a process of interpersonal judgment, it involves different steps, including collecting information, integrating it and forming a trait judgment (Fiske and Taylor, 1991). Traits are enduring personality characteristics that tell us what kind of a person someone is, and involves the capacity to remember the behavior of an agent over a long stretch of time under multiple circumstances, and to recognize the common goal in these behaviors (Van Overwalle, 2009).

Uncovering the neurological underpinnings of the trait inference process became an important topic in the emergent field of social neuroscience. A recent meta-analysis of social neuroscience studies using functional magnetic resonance imaging (fMRI) led to the conclusion that trait inference involves a network of brain areas, termed the mentalizing network (Van Overwalle, 2009). It was suggested that in this mentalizing network, the temporo-parietal junction (TPJ) is involved in the understanding of temporary behaviors and beliefs, while the medial prefrontal cortex (mPFC) integrates this social information at a more abstract level, such as the actor's traits. Several fMRI studies have confirmed that the mPFC is most critical for trait inferences (Harris *et al.*, 2005; Mitchell *et al.*, 2005, 2006a; Todorov *et al.*, 2007; Ma *et al.*, 2011; Moran *et al.*, 2011). In addition, other studies showed a supporting role for the TPJ in identifying and understanding other's behaviors that imply various traits (Ma *et al.*, 2011, 2012a, 2012b).

Current neuroscientific research on traits is focused mainly on the brain areas involved in the process of trait inference (see Van Overwalle, 2009). So far, research neglected the neural basis of traits, that is, which neurons or neuronal ensembles represent a trait code. These codes or representations can be defined as distributed memories in neural networks that encode information and, when activated, enable access to this stored information (Wood and Grafman, 2003). The aim of this paper is to uncover the location of this trait code (Northoff and Bermpohl, 2004). We hypothesize that a neural code of higher level traits is located at the mPFC, and that this area is receptive only to traits and remains relatively unresponsive to lower-level action features such as different behaviors, event scripts and agents that exemplify and possess the trait (Wood and Grafman, 2003; Wood *et al.*, 2005; Krueger *et al.*, 2009). Our hypothesis is in line with the structured event complex framework by Krueger *et al.* (2009) who argued that the mPFC represents abstract dynamic summary representations that give rise to social event knowledge. To date, no single fMRI study explored whether a trait code is located in the mPFC, over and above its role in the process of forming a trait inference.

To localize the representation of a trait code independent from representations related to action components from which a trait is abstracted, we applied an fMRI adaptation paradigm. The fMRI adaptation (or repetition suppression) refers to the observation that repeated presentations of a sensory stimulus or concept consistently reduce the fMRI responses relative to presentations of a novel stimulus (Grill-Spector *et al.*, 2006). fMRI adaptation can potentially arise from neural fatigue, increased selectiveness in responding or decreased prediction error to the same stimulus (Grill-Spector *et al.*, 2006). Irrespective of these explanations, adaptation has generally been taken as evidence for a neural representation that is invariant to the differences between those stimuli, whereas recovery from adaptation implies selectivity of the neural population to a specific stimulus or conceptual attribute. The adaptation effect has been demonstrated in many perceptual domains, including the perception of colors, shapes, and objects, and occurs in both lower and higher level visual areas and conceptual domains (Grill-Spector *et al.*, 1999; Thompson-Schill *et al.*, 1999; Kourtzi and Kanwisher, 2000; Engel and Furmanski, 2001; Grill-Spector and Malach, 2001; Krekelberg *et al.*, 2006; Bedny *et al.*, 2008; Devauchelle *et al.*, 2009; Roggeman *et al.*, 2011; Diana *et al.*, 2012; Josse *et al.*, 2012). Recently, fMRI adaptation has also been found during action observation (Ramsey and Hamilton, 2010a, 2010b), action word reading (Yee *et al.*, 2010) and trait judgments of other persons similar to the self (Jenkins *et al.*, 2008).

If these characteristics of fMRI adaptation also apply to traits, we can isolate the critical brain area that is responsible for the representation of a trait code. Moreover, if these traits are inferred from different behavioral descriptions that have little semantic or conceptual

Received 2 February 2013; Revised 12 June 2013; Accepted 13 June 2013
Advance Access publication 18 June 2013
This research was supported by an OZR Grant (OZR1864BOF) of the Vrije Universiteit Brussel to F.V.O. This research was conducted at GIfMI (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

associations except for the implied trait, this would strengthen the notion that this trait code is involved in abstracting out the shared trait implication from varying lower-level behavioral information, and not due to some lower-level visual or semantic similarity between the descriptions.

This study tested fMRI adaptation of traits by presenting a behavioral trait-implying description (the prime) followed by another behavioral description (the target; see also Jenkins *et al.*, 2008). We created three conditions by preceding the target description (e.g. implying honesty) by a prime description that implied the same trait (e.g. honesty), implied the opposite trait (e.g. dishonesty), or implied no trait at all (i.e. trait-irrelevant). Basically, we predict a stronger adaptation effect when the overlap in trait implication between these two behavioral descriptions is large, and a weaker adaptation effect when the trait overlap is small. Specifically, when the prime and target description are similar in content and valence, this would most strongly reduce the response in the mPFC. Thus, if a behavioral description of a friendly person is followed by a behavioral description of another friendly person, we expect the strongest fMRI adaptation. To the extent that opposite behaviors involve the same trait content but of opposite valence (e.g. when a behavioral description of an unfriendly person is followed by a behavioral description of friendly person), we expect weaker adaptation. Alternatively, it is possible that the brain encodes these opposing traits as belonging to the same trait concept, leading to little adaptation differences. Finally, the least adaptation is expected when a target description is preceded by a prime that does not imply any trait. However, note that because the experimental task requires to infer a trait under all conditions, we expect some minimal amount of adaptation even in the irrelevant condition. Given that traits are assumed to be represented in a distributed fashion by neural ensembles which partly overlap rather than individual neurons, a search for possible traits under irrelevant conditions may spread activation to related trait codes, causing some adaptation. Hence, it is important to recognize that adaptation under trait conditions only reflects a trait code, whereas a generalized adaptation effect across all conditions reflects an influence of a trait (search) process. Moreover, note that to avoid confounding trait adaptation with the presence of an actor, all behavioral descriptions involved a different actor in this study.

METHODS

Participants

Participants were all right-handed, 14 women and 3 men, with ages varying between 18 and 30 years. In exchange for their participation, they were paid €10. Participants reported no abnormal neurological history and had normal or corrected-to-normal vision. Informed consent was obtained in a manner approved by the Medical Ethics Committee at the Hospital of University of Ghent (where the study was conducted) and the Free University Brussels (of the principal investigator F.V.O.).

Procedure and stimulus material

The stimulus sentences were borrowed from earlier studies on trait inference using fMRI (Ma *et al.*, 2011, 2012a) and event-related potential (ERP) (Van Duynslaeger *et al.*, 2007). We created the following four conditions: similar, opposite, irrelevant and singleton. Participants read two sentences concerning different agents who were engaged in behaviors that implied positive or negative moral traits. The positive or negative traits were counterbalanced across conditions. The target sentence (e.g. 'Tolvan gave her brother a compliment' to induce the trait friendly) was preceded by a prime sentence that implied the same trait (Similar condition, e.g. 'Calpo gave her sister a hug'), the opposite trait (Opposite condition, e.g. 'Angis gave her mother a slap'), or no trait at all (Irrelevant condition, e.g. 'Jun felt a quite fresh breeze'). After each trial of two sentences, participants were instructed to infer the agent's trait from the last (target) sentence and indicated by pressing button whether a given trait applied to the target description. The trait displayed was either the implied trait or its opposite, so that half of the correct responses was 'yes', and the other half was 'no'. To avoid that participants would ignore the (first) prime sentence and pay attention only on the (second) target sentence, we added a Singleton condition consisting of a single trait-implying behavioral sentence, immediately followed by a trait question. Hence, during the first sentence of any trial, the participants could not predict whether a question would or would not appear afterwards, so that carefully reading was always necessary. There were 20 trials in each condition.

To avoid associations with a familiar and/or existing name, fictitious 'Star Trek'-like names were used (Ma *et al.*, 2011, 2012a, 2012b). To exclude any possible adaptation from the agent, the agents' names differed in all sentences. All the sentences were in Dutch and consisted of six words (except eight sentences with seven words) that were presented in the middle of the screen for a duration of 5.5 s. To optimize estimation of the event-related fMRI response, each prime and target sentence 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. After each trial, a fixation cross was shown for 500 ms and then the trait question appeared until a response was given. We presented one of four versions of the material, counterbalanced between conditions and participants.

Imaging procedure

Images were collected with a 3 Tesla 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 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 echoplanar pulse sequence (TR = 2 s; TE = 33 ms; $3.5 \times 3.5 \times 4.0$ mm in-plane resolution).

Image processing and statistical analysis

The fMRI data were preprocessed and analyzed using SPM5 (Wellcome Department of Cognitive Neurology, London, UK). For each functional run, data were preprocessed to remove sources of noise and artifacts. 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) using a Gaussian kernel. Afterwards, 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 the experimental design. Outliers where identified in 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.

Next, single participant (1st level) analyses were conducted. Statistical analyses were performed using the general linear model of SPM5 of which the event-related design was modeled with one regressor for each prime and target sentence for each condition, time-locked at the presentation of the prime and target sentences and convolved with a canonical hemodynamic response function (with event duration assumed to be 0 for all conditions). Six motion parameters from the realignment as well as outlier time points (identified by ART) were included as nuisance regressors. The response of the participants was not modeled. We used a default high-pass filter of 128 s and serial correlations were accounted for by the default autoregressive AR(1) model.

For the group (2nd level) analyses, we conducted a whole-brain analysis with a voxel-based statistical threshold of $P \leq 0.001$ (uncorrected) with a minimum cluster extent of 10 voxels. Statistical comparisons between conditions were conducted using t tests on the parameter estimates associated with each trial type for each subject, $P < 0.05$ (cluster-level corrected). We defined adaptation as the contrast (i.e. decrease in activation) between prime and target sentence. This adaptation contrast was further analyzed in a conjunction analysis (combining all trait conditions) to identify the brain areas commonly involved in the trait inference process, and more critically, in an interaction analysis (with a Similar > Irrelevant contrast) to isolate the brain areas involved in a trait code. To further verify that the brain areas identified in the previous analysis showed the hypothesized adaptation pattern, we computed the percentage signal change. This was done in two steps. First, we identified a region of interest (ROI) as a sphere of 8 mm around the peak coordinates from the whole-brain interaction as described earlier. Second, we extracted the percentage signal change in this ROI from each participant using the MarsBar toolbox (http://marsbar.sourceforge.net). We also calculated an adaptation index as the percentage signal change of prime minus target condition. These data were further analyzed using t tests with a threshold of $P < 0.05$.

RESULTS

Behavioral results

A repeated-measure analysis of variance test was conducted on the reaction times (RT) and accuracy rates from the four conditions (Table 1). The RT data revealed a significant effect of trait condition, $F\ (1,\ 16) = 12.89$, $P < 0.001$. Participants responded more quickly in the Similar and Irrelevant conditions as compared with the Opposite and Singleton conditions. The accuracy rate data did not revealed any significant difference among conditions, $F\ (1,\ 16) = 0.074$, $P = 0.47$.

fMRI results

Our analytic strategy for detecting an adaptation effect during trait processing was as follows. First, we conducted a whole-brain, random-effects analysis contrasting prime > target trials in the Similar, Opposite and Irrelevant conditions, followed by a conjunction analysis (to identify a common trait inference process) and a Similar > Irrelevant interaction (to isolate the trait code). Second, to verify that the areas representing the trait code showed the hypothesized adaption pattern, we defined a ROI centered on the peak value and calculated the percentage signal change.

Table 1 RT and accuracy rate from behavioral performance

Condition	Similar	Opposite	Irrelevant	Singleton
RT (ms)	1359_a	1409_b	1327_a	1439_b
Accuracy rate (%)	80.0_a	79.9_a	80.7_a	81.5_a

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

The whole-brain analysis of the prime > target contrast revealed significant adaptation effects ($P < 0.05$, cluster-level corrected) in the mPFC, and most strongly in the ventral part of the mPFC, as well as in the precuneus (Table 2). This adaptation effect was observed in all three experimental (Similar, Opposite and Irrelevant) conditions, and also in a conjunction analysis of the three conditions. The finding that adaptation was even found under the irrelevant trait condition is consistent with the idea that some minimal amount of a trait inference process takes place given the explicit instructions to infer a trait. Other areas also showed adaptation effects in one or more experimental conditions (Table 2). However, these effects failed to survive any conjunction analysis. This suggests that these additional adaptation effects are due to idiosyncratic lower-level features that differ for each trait condition (e.g. the same goal given a similar trait but not an opposite trait, the same episodic memory for similar and opposite traits, but not for trait irrelevant descriptions).

To identify the brain areas involved in the trait code, we conducted a whole-brain interaction analysis of the prime > target contrast with all plausible Similar > Irrelevant contrasts, that is, with or without the Opposite condition (Table 2). In all these interactions, the ventral mPFC was the only brain area implicated. This confirms our hypothesis that this mPFC area represents the trait code.

To verify that this mPFC area reveals the predicted effect of adaptation and, more crucially, that this adaptation effect is largest for trait diagnostic as opposed to irrelevant information, we calculated an adaptation index using a ROI centered at the whole-brain interaction (with MNI coordinates −6, 42, −14), by subtracting the percentage signal change in the target sentence from the prime sentence (Figure 1). The adaptation index in the vmPFC clearly showed the predicted pattern: the strongest adaptation was found in the Similar condition, becoming nonsignificantly weaker in the Opposite condition and almost negligible in the Irrelevant condition. Post hoc one-sided t tests revealed, in comparison with the Irrelevant condition, a stronger adaptation of the Similar condition ($P < 0.001$) and the Opposite condition ($P < 0.05$). There was no difference between the Similar and Opposite conditions ($P > 0.15$).

To ensure that the mPFC was involved only in adaptation (i.e. decrease of activation), we also conducted a whole-brain analysis of the reverse target > prime contrast in the Similar, Opposite and Irrelevant conditions. The results revealed a series of brain areas that were more strongly recruited during the presence of the target sentence among the three conditions, including the precuneus, bilateral insula, anterior cingulate cortex, left inferior frontal gyrus, left superior parietal cortex, left middle temporal gyrus and right lingual gyrus (Table 3). Importantly, there was no significant mPFC activation.

DISCUSSION

Trait inference is an important component of social interactions in our daily life. Neuroimaging studies on this topic have implicated the mPFC as an area in a social mentalizing network that is most essentially involved in trait inference (Ma *et al.*, 2012b; for a review, see Van Overwalle, 2009). Although most studies in this domain provided

Table 2 Adaptation (prime > target contrast) effects from the whole-brain analysis

Anatomical label	Similar					Opposite					Irrelevant				
	x	*y*	*z*	Voxels	Max *t*	*x*	*y*	*z*	Voxels	Max *t*	*x*	*y*	*z*	Voxels	Max *t*
Prime > target contrasts															
Ventral mPFC	4	46	−6	2799	7.17***$_a$	2	48	−4	2169	6.02***$_a$	4	50	−2	1129	5.16***$_a$
R. postcentral											62	−20	30	193	4.71*
L. inferior parietal											−64	−28	34	288	5.71**$_a$
Cingulate											−8	−32	48	217	4.04**
R. parahippocampal	38	−32	−16	179	4.26*										
R. posterior cingulate (Precuneus)	16	−50	20	663	5.39***$_a$	10	−52	22	756	5.15***$_a$	14	−52	22	272	4.35**
R. angular gyrus	44	−76	34	153	5.43*$_a$						44	−74	34	225	4.99**
L. angular gyrus	−44	−78	34	200	5.13*$_a$										
L. mid-occipital											−40	−80	38	348	6.55**$_a$
	Similar and opposite					Similar and opposite and irrelevant									
Conjunction of prime > target contrasts															
Ventral mPFC	2	48	−4	2028	6.02***$_a$	4	50	−2	1010	5.16***$_a$					
Precuneus	12	−50	20	520	5.02***	14	−52	22	222	4.35**					
	With similar > irrelevant					With similar + opposite > irrelevant					With similar > opposite + irrelevant				
Interaction of prime > target contrast															
Ventral mPFC	−6	42	−14	280	4.54**	−6	42	−14	131	4.54**	14	28	−14	299	4.37**

Coordinates refer to the MNI (Montreal Neurological Institute) stereotaxic space. All clusters thresholded at $p < 0.001$ with at least 10 voxels. The Similar + Opposite > irrelevant contrast was implemented as [2, 1, −3] and the Similar > Opposite + Irrelevant contrast as [3, −2, −1]. Only significant clusters are listed.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (cluster-corrected; subscript 'a' denotes $P < 0.05$, FWE-corrected also).

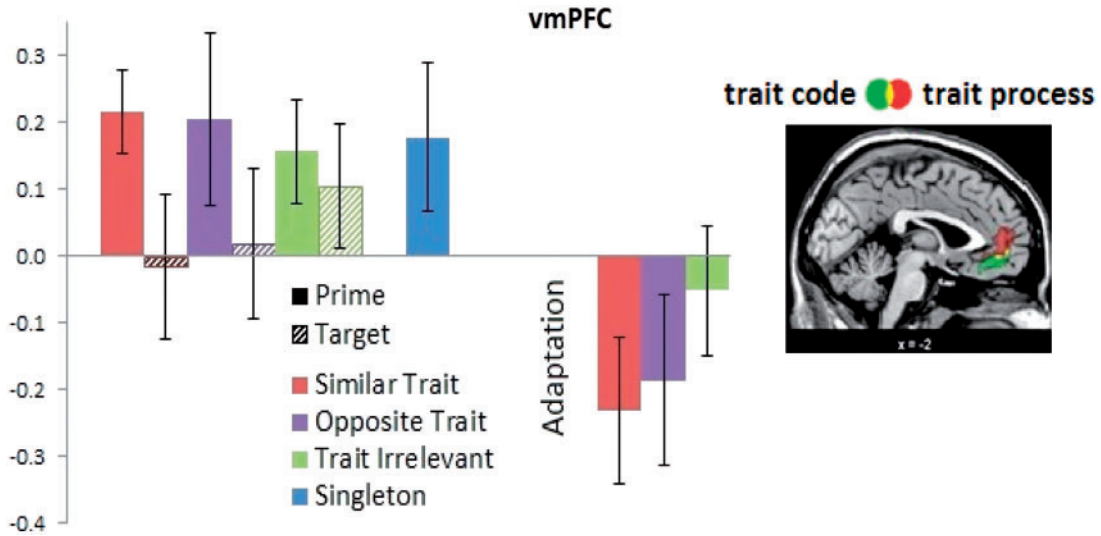

Fig. 1 Percent signal change in the ventral prefrontal cortex for the prime and target sentences in all conditions, and for the adaptation index (target − prime condition) based on the mPFC ROI (with MNI coordinates −6, 42, −14). The inset depicts the whole-brain interaction reflecting the trait code (green), the whole-brain conjunction reflecting a common trait inference process (red) and their overlap (yellow).

evidence that traits are processed in this area, we took a representational approach by exploring to what extent the mPFC represents a trait code for identifying and inferring traits, using an fMRI-adaptation paradigm. fMRI adaptation has not been used previously to study trait representations (except when involving the self, Jenkins *et al.*, 2008), and the interpretation of adaptation differs from the interpretation of traditional fMRI subtraction studies. Adaptation relies on the assumption that neuronal firing tends to be attenuated when a stimulus is presented repeatedly, and so reveals the neuronal population that codes for the invariant features of this stimulus. In contrast, traditional fMRI studies reveal activation in all areas subserving stimulus processing, that is, areas that are involved in essential invariant features of a stimulus as well as in less relevant and variable features.

Adaptation to traits

In this study, participants inferred traits of others while reading behavioral sentences that strongly implied a trait, after they had read sentences that involved the same trait, an opposite trait or trait-irrelevant information. The results revealed evidence for fMRI adaptation in the mPFC, which reached significance in the ventral part as well as the precuneus. However, only the ventral part of mPFC showed adaptation

Table 3 Results of target > prime contrast from the whole-brain analysis

Anatomical label	Similar					Opposite					Irrelevant				
	x	*y*	*z*	Voxels	Max *t*	*x*	*y*	*z*	Voxels	Max *t*	*x*	*y*	*z*	Voxels	Max *t*
Target > prime contrasts															
L. inferior frontal											−44	46	0	690	4.92***
L. insula											−26	26	0	8590	8.61***$_a$
R. insula	34	22	−2	21 433	9.49***$_a$						32	24	0	4279	7.21***$_a$
Posterior mFC						2	16	50	25 376	10.71***$_a$					
Anterior cingulate											−8	−12	6	234	4.90**
L. superior temporal											−48	−26	−10	1435	5.35***$_a$
R. superior temporal	50	−22	−12	342	4.36**	48	−22	−14	1092	6.84***$_a$					
L. superior parietal	−30	−56	46	5597	8.82***$_a$	−30	−56	50	9438	8.84***$_a$	−28	−56	50	2704	7.37***$_a$
R. superior parietal	32	−58	48	1608	7.69***$_a$						28	−56	46	1034	6.26***$_a$
L. fusiform	−32	−58	−32	209	5.15*$_a$										
R. fusiform	36	−62	−30	587	5.63***$_a$	32	−60	−32	3205	6.59***$_a$	38	−62	−30	487	4.82***
L. posterior cingulate						−30	−64	8	233	4.70**					
R. posterior cingulate						14	−66	14	217	4.24**					
R. lingual	10	−78	−38	472	5.10***$_a$						12	−82	−32	261	4.19**
L. lingual	−8	−80	−28	363	5.58***$_a$										
R. cuneus											14	−94	0	332	5.27**$_a$
L. cuneus											−10	−98	2	368	4.64***
	Similar and opposite traits					Similar and opposite and irrelevant									
Conjunction of target > prime contrasts															
L. inferior frontal						−44	46	0	659	4.92***					
L. insula						−26	26	0	8111	8.58***$_a$					
R. insula	34	22	−2	19 957	9.49***$_a$	32	24	0	3949	7.21***$_a$					
Anterior cingulate						−8	−12	6	202	4.90*					
R. superior temporal	50	−22	−12	339	4.36**										
L. middle temporal						-60	-40	0	1179	5.27***$_a$					
L. superior parietal	−30	−56	48	5329	8.76***$_a$	−28	−56	50	2146	7.37***$_a$					
Precuneus						−4	−64	50	287	5.03**					
R. lingual	10	−78	−38	466	5.10***$_a$	12	−82	−32	248	4.19**					
L. lingual	−8	−80	−28	363	5.58***$_a$										
	With opposite > irrelevant														
Interaction of target > prime contrast															
R. mid frontal	44	10	52	359	4.31***										
R. superior parietal	42	−58	50	368	4.09***										

Coordinates refer to the MNI (Montreal Neurological Institute) stereotaxic space. All clusters thresholded at $P < 0.001$ with at least 10 voxels. Only significant clusters are listed.
*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$ (cluster-corrected; subscript 'a' denotes $P < 0.05$, FWE corrected also).

in the trait-diagnostic (Similar and Opposite) conditions while adaptation was negligible in the Irrelevant condition, as revealed by the whole-brain interaction (Figure 1). As predicted, the adaptation effect in the mPFC decreased given less overlap with the initial trait: The largest adaptation was demonstrated when the preceding description implied the same trait, slightly weaker given an opposite trait and almost negligible given trait-irrelevant descriptions. Interestingly, the finding that similar and opposite traits show approximately the same amount of adaptation demonstrates that a trait and its opposite seem to be represented by a highly similar and overlapping neural populations in the mPFC. This is in line with research on the schema-plus-tag model, in which a negated trait is represented as the original (true) trait with a negation tag. For instance, stating that a person is not romantic often makes one think of romantic behaviors and then negates them (Mayo *et al.*, 2004). Moreover, this decrease in the mPFC is similar to gradients that have been observed for letter and word processing (Vinckier *et al.*, 2007), number processing (Roggeman *et al.*, 2011) and to gradients for object processing more generally (Grill-Spector *et al.*, 1999). Crucially, this adaptation effect was not found in other brain areas. These findings confirm that mPFC, and especially its ventral part, is an essential brain area for the representation of a trait code. In sum, the current findings seem to support the representational view that the mPFC not only supports trait processing but also represents the code that identifies traits.

Previous theoretical approaches have suggested a similar representational function of the mPFC. Forbes and Grafman (2010) suggested that the primary role of the PFC is the representation of action and guidance of behavior (Barbey *et al.*, 2009). They argued that series of events form a script that represent a set of goal-oriented events, that is sequentially ordered and guides behavior and perceptions, and refer to this as a structured event complex (Grafman, 2002; Wood and Grafman, 2003; Barbey *et al.*, 2009). There is a similar history in the social psychology literature that conceives traits as abstracted instances of goal-directed behaviors (see also Read, 1987; Read *et al.* 1990, Reeder *et al.* 2004; Reeder, 2009), and recent behavioral and neural evidence has revived the notion that goals are primary, and traits are secondary inferences (Van Duynslaeger *et al.*, 2007; Van der Cruyssen *et al.*, 2009; Ma *et al.*, 2012b; Malle and Holbrook, 2012; Van Overwalle *et al.*, 2012). In a somewhat different view, Mitchell (2009) proposed that individuals may decipher other minds by making use of one's own knowledge of self as the basis for understanding others. He suggested that perceivers can use their own mental traits as proxies for other

minds, and 'simulate' or 'project' their own traits on the other person to make inferences about the other person. Both accounts assume that there exists a repository for a trait code, either in a general format (Forbes and Grafman, 2010) or in reference to the self (Mitchell, 2009). This perspective on the vmPFC is also in line with connectionist approaches to person perception that view processing and representation as integral aspects of brain functioning (Read and Marcus-Newhall, 1993; Read and Montoya, 1999; Van Overwalle and Labiouse, 2004).

Trait code in the ventral mPFC

Our study demonstrates that a trait code is represented in the ventral part of mPFC. The ventral mPFC has been linked to mentalizing about persons perceived to be similar to the self, while the dorsal area has been associated with mentalizing about people that are dissimilar from oneself (Mitchell *et al.*, 2006b; Van Overwalle, 2009). The ventral location of the trait code is consistent with theorizing which posits that this ventral area accounts for the continuous representation of self-referential stimuli which are used as proxy to 'simulate' or 'project' our own traits for judging other individuals (Northoff and Bermpohl, 2004; Mitchell, 2009). Alternatively, given that in this experiment the specific agent was less relevant to infer the trait from the behavioral descriptions, it is possible that participants used self-related representations for judging the traits, thus activating only the ventral part of the mPFC (Van Overwalle, 2009; D'Argembeau and Salmon, 2012).

The present findings leave open a crucial question about the relationship between traits and valences, and the role of the ventral mPFC in this interplay, whereas the dorsal mPFC has been associated with more cognitive controlled operations, the ventral area is connected anatomically to striatal, limbic, and midbrain regions related to emotional processes (Northoff *et al.*, 2006). Several neuroimaging studies revealed that the ventral mPFC is recruited during the regulation of emotional processing, such as regulating emotional responses (Quirk and Beer, 2006; Olsson and Ochsner, 2008; Etkin *et al.* 2011; Roy *et al.* 2011), affective mentalizing (Sebastian *et al.*, 2012) and reward-related processing (Van Den Bos *et al.*, 2007). In fact, human social and emotional behaviors are highly intertwined in many cases and it is difficult to engage in social processing or interaction without emotion. Consequently, social and emotional processing may have shared representations in the brain (Ochsner, 2008; Olsson and Ochsner, 2008). In this study, the stimuli are a set of social behaviors that have positive or negative valence. Recall that the adaptation effect decreased linearly when the trait-implying target sentence was preceded by behavioral information that implied a similar, opposite or no trait. Alternatively, one may view this adaptation pattern as revealing repetition of the same, the opposite or a neutral valence, implicated by the behavior. It is always the case that similar target traits are similar in valence to the prime, and that opposite target traits are opposite in valence. This suggests that the present adaptation effect in the ventral mPFC may be related to evaluative processing when people make social inferences, rather than the content of inferred traits per se. However, because the adaptation effect did not differ significantly between similar and opposite traits, a valence interpretation is not very likely, but cannot excluded entirely. Another possibility is that the ventral mPFC does both, representing a trait code and responding to the magnitude of valence. Nevertheless, future studies are needed to disentangle the contribution of specific traits or their underlying valence on the adaptation effect in the mPFC. Novel research at our lab seems to exclude these alternative valance explanations and confirms that only the trait is coded in the vmPFC.

Having established evidence for the representation of a trait code in the mPFC, we might speculate how this trait code interacts with other brain areas. We suggest that the ventral part of mPFC may act as an amodel hub or convergence area (Patterson *et al.*, 2007; Forbes and Grafman, 2010; Harada *et al.*, 2010; Woollams, 2012), forming ingoing links to connected brain areas such as the TPJ, to receive information on trait attributes such as behavioral goals and exemplary trait-evoking situations or scripts. This hub function may also form outgoing links to adjacent brain areas such as the dorsal mPFC, to transfer the integrated trait information for further evaluation and judgment about unfamiliar persons (Northoff and Bermpohl, 2004; Van Overwalle, 2009; Moran *et al.*, 2011; Frith and Frith, 2012).

Limitations

The strong adaptation effect in all three conditions (including the irrelevant condition) of this study is consistent with the notion that a common trait inference process took place under all conditions, which is not surprising given the explicit instruction to make a trait inference. Assuming trait coding by partially overlapping neural ensembles, an inference process whereby a plausible trait is searched for may have leaked activation to related trait codes, resulted in an adaptation effect also under irrelevant conditions. However, critically, this processing account cannot explain the adaptation effect in the mPFC that was significantly stronger in diagnostic (Similar and Opposite) conditions as opposed to irrelevant conditions.

Another possible criticism may reflect the different processing of prime and target sentences. In the three trait-repetition conditions, participants may ignore the trait information in the prime sentences, even though 25% of the trials (the singleton condition) invited participants make a judgment of agents' traits in prime sentence. Nevertheless, one may expect a more automatic information processing mode for prime sentences and a more controlled mode for target sentences. This may potentially have caused a greater involvement of the ventral part of mPFC during prime sentences and of the dorsal part of mPFC during target sentences (Lieberman, 2007). However, because no dorsal mPFC activation was revealed in the target > prime contrast, this explanation is very unlikely. Another consequence might be that prime sentences were processed in a more internally oriented default mode manner, and target sentences in a more task-oriented manner during the preparation of a response. According to default mode theory (Raichle *et al.*, 2001), such task-oriented preparation may lead to mPFC deactivation during the target sentences. However, a default mode is typically created by putting participants at rest (Spreng *et al.*, 2009; Schilbach *et al.*, 2012), while in our experiment they were continuously reading and responding in all conditions. Moreover, the responses involved social-cognitive processes which typically increase rather than decrease default mode activation.

Although fMRI adaptation is often interpreted as suggestive of an invariant neural code, adaptation may reflect not only bottom–up building of neural fatigue or facilitation but also top–down automatic tuning of neuronal excitation. Our result might be due to attentional or expectation confounds, which may also lead to decreased fMRI signals. However, this is unlikely. The locus of the present adaptation effect is in the mPFC, which does not have a specific role in attention. Furthermore, our experiment used a one-back adaptation design, where some descriptions function as 'prime' and others as 'target.' Although participants were probably aware of this sequence, they could not predict which target description (similar, opposite or irrelevant) would appear after the prime. This rules out an attention or expectation account.

CONCLUSION

Although the neuronal mechanism underlying the fMRI adaptation effect is not entirely clear at this stage in social neuroscience,

the present adaptation paradigm offered for the first time evidence for the representation of a trait code in the ventral mPFC, over and above its role in the processing of trait information. Although it is still unclear whether this adaptation effect is driven by the specific content of the trait or by its valence, this finding opens a novel perspective on the functionality of the mPFC in social mentalizing. Rather than simply processing social information, the mPFC may be an important hub of social cognition, integrating multi-modal low-level behavioral and contextual information with high-level knowledge on individuals and social networks marked by their trait-relevance.

REFERENCES

Barbey, A.K., Krueger, F., Grafman, J. (2009). Structured event complexes in the medial prefrontal cortex support counterfactual representations for future planning. *Philosophical Transactions of the Royal Society Biological Sciences*, *364*, 1291–300.

Bedny, M., McGill, M., Thompson-Schill, S.L. (2008). Semantic adaptation and competition during word comprehension. *Cerebral Cortex*, *18*, 2574–85.

Engel, S.A., Furmanski, C.A. (2001). Selective adaptation to color contrast in human primary visual cortex. *Journal of Neuroscience*, *21*, 3949–54.

Etkin, A., Egner, T., Kalisch, R. (2011). Emotional processing in anterior cingulate and medial prefrontal cortex. *Trends in Neurosciences*, *15*, 85–93.

D'Argembeau, A., Salmon, E. (2012). The neural basis of semantic and episodic forms of self-knowledge: insight from functional neuroimaging. In: López-Larrea, C., editor. *Sensing in Nature*. New York: Springer Science+Bussiness Media, pp. 276–90.

Devauchelle, A., Oppenheim, C., Rizzi, L., Dehaene, S., Pallier, C. (2009). Sentence syntax and content in the human temporal lobe: an fMRI adaptation study in auditory and visual modalities. *Journal of Cognitive Neuroscience*, *21*, 1000–12.

Diana, R.A., Yonelinas, A.P., Ranganath, C. (2012). Adaptation to cognitive context and item information in the medial temporal lobes. *Neuropsychologia*, *50*, 3062–9.

Fiske, S.T., Taylor, S.E. (1991). *Social Cognition*. New York: McgrawHill Higher Education.

Forbes, C.E., Grafman, J. (2010). The role of the human prefrontal cortex in social cognition and moral judgment. *Annual Review Neuroscience*, *33*, 299–324.

Frith, C.D., Frith, U. (2012). Mechanisms of social cognition. *Annual Review Psychology*, *63*, 287–313.

Grafman, J. (2002). The structured event complex and the human prefrontal cortex. In: Stuss, D.T.H., Knight, R.T., editors. *Principles of Frontal Lobe Function*. Oxford/New York: Oxford University Press, pp. 616.

Grill-Spector, K., Henson, R., Martin, A. (2006). Repetition and the brain: neural models of stimulus specific effects. *Trends in Cognitive Science*, *10*, 14–23.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*, 187–203.

Grill-Spector, K., Malach, R. (2001). fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychology*, *107*, 293–321.

Harada, T., Li, Z., Chiao, J.Y. (2010). Differential dorsal and ventral medial prefrontal representations of the implicit self modulated by individualism and collectivism: an fMRI study. *Social Neuroscience*, *5*, 257–71.

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

Jenkins, A.C., Macrae, C.N., Mitchell, J.P. (2008). Repetition suppression of ventromedial prefrontal activity during judgments of self and others. *Proceedings of the National Academy of Sciences U S A*, *105*, 4507–12.

Josse, G., Joseph, S., Bertasi, E., Giraud, A.-L. (2012). The brain's dorsal route for speech represents word meaning: evidence from gesture. *PLoS One*, *7*(9), e46108.

Kourtzi, Z., Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience*, *20*, 3310–8.

Krekelberg, B., Boynton, G.M., Wezel, R.J.A. (2006). Adaptation: from single cells to BOLD signals. *Trends in Neurosciences*, *29*, 250–6.

Krueger, F., Barbey, A.K., Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, *13*, 103–9.

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

Ma, N., Vandekerckhove, M., Baetens, K., Van Overwalle, F., Seurinck, R., Fias, W. (2012a). Inconsistencies in spontaneous and intentional trait inferences. *Social, Cognitive and Affective Neuroscience*, *7*, 937–50.

Ma, N., Vandekerckhove, M., Van Hoeck, N., Van Overwalle, F. (2012b). Distinct recruitment of temporo-parietal junction and medial prefrontal cortex in behavior understanding and trait identification. *Social Neuroscience*, *7*, 591–605.

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

Malle, B.F., Holbrook, J. (2012). Is there a hierarchy of social inferences? The likelihood and speed of inferring intentionality, mind, and personality. *Journal of Personality and Social Psychology*, *102*, 661–84.

Mayo, R., Schul, Y., Burnstein, E. (2004). "I am not guilty" vs "I am innocent": successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, *40*, 433–49.

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

Mitchell, J.P., Cloutier, J., Banaji, M.R., Macrae, C.N. (2006a). Medial prefrontal dissociations during processing of trait diagnostic and nondiagnostic person information. *Social, Cognitive and Affective Neuroscience*149–55.

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

Mitchell, J.P. (2009). Inferences about other minds. *Philosophical Transactions of the Royal Society Biological Sciences*, *364*, 1309–16.

Moran, J.M., Lee, S.M., Cabrieli, J.D. (2011). Dissociable neural systems supporting knowledge about human character and appearance in ourselves and others. *Journal of Cognitive Neuroscience*, *23*, 2222–30.

Northoff, G., Bermpohl, F. (2004). Cortical midline structures and the self. *Trends in Cognitive Sciences*, *8*, 102–7.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., Panksepp, J. (2006). Self-referential processing in our brain—a meta-analysis of imaging studies on the self. *NeuroImage*, *31*, 440–57.

Patterson, K., Nestor, P.J., Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, *8*, 976–87.

Ochsner, K.N. (2008). The social-emotional processing stream: five core constructs and their translational potential for schizophrenia and beyond. *Biological Psychiatry*, *64*, 48–61.

Olsson, A., Ochsner, K.N. (2008). The role of social cognition in emotion. *Trends in Cognitive Sciences*, *12*, 65–71.

Quirk, G.J., Beer, J.S. (2006). Prefrontal involvement in the regulation of emotion: convergence of rat and human studies. *Current Opinion in Neurobiology*, *16*, 723–7.

Ramsey, R., Hamilton, F.C. (2010a). Triangles have goals too: understanding action representation in left aIPS. *Neuropsychologia*, *48*, 2773–6.

Ramsey, R., Hamilton, F.C. (2010b). Understanding actors and object-goals in the human brain. *NeuroImage*, *50*, 1142–7.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., et al. (2001). A default mode of brain function. *Proceedings of National Academy of Science U S A*, *98*, 676–82.

Read, S.J. (1987). Constructing causal scenarios: a knowledge structure approach to causal reasoning. *Journal of Personality and Social Psychology*, *52*, 288–302.

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.

Read, S.J., Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: a parallel distributed processing account. *Journal of Personality and Social Psychology*, *65*, 429–47.

Read, S.J., Montoya, J.A. (1999). An autoassociative model of causal reasoning and causal learning: reply to Van Overwalle's critique of Read and Marcus-Newhall (1993). *Journal of Personality and Social Psychology*, *76*, 728–42.

Reeder, G.D. (2009). Mindreading: judgments about intentionality and motives in dispositional inference. *Psychological Inquiry*, *20*, 1–18.

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.

Roggeman, C., Santens, S., Fias, W., Verguts, T. (2011). Stages of non-symbolic number processing in occipito-parietal cortex disentangled by fMRI-adaptation. *Journal of Neuroscience*, *31*, 7168–73.

Roy, M., Shohamy, D., Wager, T.D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in Cognitive Sciences*, *16*, 147–56.

Schilbach, L., Bzdok, D., Timmermans, B., et al. (2012). Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition. *PLoS One*, *7*, e30920.

Sebastian, C.L., Fontaine, M.G., Bird, G., et al. (2012). Neural processing associated with cognitive and affective theory of mind in adolescents and adults. *Social Cognitive and Affective NeuroScience*, *7*, 53–63.

Spreng, R.N., Mar, R.A., Kim, A.S. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *Journal of Cognitive Neuroscience*, *21*, 489–510.

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

Thompson-Schill, S.L., D'Esposito, M., Kan, I.P. (1999). Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron*, *23*, 513–22.

Van Den Bos, W., McClure, S.M., Harris, L.T., Fiske, S.T., Cohen, J.D. (2007). Dissociating affective evaluation and social cognitive processes in the ventral medial prefrontal cortex. *Cognitive, Affective, & Behavioral Neuroscience*, *7*, 337–46.

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.

Van Overwalle, F., Labiouse, C. (2004). A recurrent connectionist model of person impression formation. *Personality and Social Psychology Review*, *8*, 28–61.

Van Overwalle, F., Van Duynslaeger, M., Coomans, C., Timmermans, B. (2012). Spontaneous goal inferences are often inferred faster than spontaneous trait inferences. *Journal of Experimental Social Psychology*, *48*, 13–8.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J.P., Sigman, M., Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron*, *55*, 143–56.

Wood, J.N., Grafman, J. (2003). Human prefrontal cortex: processing and representational perspectives. *Nature Reviews Neuroscience*, *4*, 139–47.

Wood, J.N., Knutson, K.M., Grafman, J. (2005). Psychological structure and neural correlates of event knowledge. *Cerebral Cortex*, *15*, 1155–61.

Woollams, A.M. (2012). Apples are not the only fruit: the effects of concept typicality on semantic representation in the anterior temporal lobe. *Frontiers in Neuroscience*, *6*, 85.

Yee, E., Drucker, D.M., Thompson-Schill, S.L. (2010). fMRI-adaptation evidence of overlapping neural representations for objects related in function or manipulation. *Neuroimage*, *50*, 753–63.