Distinct neural correlates of social categories and personality traits

Laurens Van der Cruyssen *, Elien Heleven, Ning Ma, Marie Vandekerckhove, Frank Van Overwalle *

Vrije Universiteit Brussel, Brussel, Belgium

A B S T R A C T

Does the processing of social category-related versus trait-related information generate a different pattern of brain activation? In this fMRI study, we compared the processing of behaviors performed by a member of a social category versus an individual, both possessing similar personality traits. Based on previous behavioral studies we predicted that the processing of social category-related information would recruit more activation in brain areas related to mentalizing than individual trait-related information. Participants read sentences describing behaviors performed by a member of a social category (of which the stereotype involves a given trait) or by an individual possessing the same trait. These behavioral sentences varied on both valence (positive versus negative) and consistency (consistent versus inconsistent) with regard to the social category or trait. The results revealed that social category-related behavioral information showed more activation in mentalizing areas (medial prefrontal cortex, anterior temporal lobe, bilateral temporo-parietal junction, posterior cingulate cortex) than trait-related information. This increased activation is interpreted in terms of the impact of autobiographical memories, greater variance among members of social categories than individual traits, a higher construal level (i.e., abstractness), and larger perceived group size. Additionally, inconsistent as opposed to consistent information showed more activation in the right temporo-parietal junction and left lingual gyrus.

© 2014 Elsevier Inc. All rights reserved.

Distinct neural correlates of social categories and personality traits

In a world of continuously increasing complexity, assigning individuals to different categories has become vital to allow flexible maneuvering through our social environment. In the present study we focus on two fundamental types of social constructs – social categories and personality traits – which have generated ample research over the past decades (Andersen et al., 1990; Brewer, 1996; Contreras et al., 2012; Fiske and Neuberg, 1990; Hamilton and Sherman, 1996; Macrae et al., 1994; Mitchell et al., 2009; Winter and Uleman, 1984). Social categories are groups of people that carry a socially shared label, and tend to be characterized by specific features like occupations and belief systems, while personality traits are concepts used to describe the essential qualities of individuals, often inferred from specific behavior (Andersen and Klatzky, 1987).

Theories like the Continuum Model (Fiske et al., 1999), the Dual-Process model of Impression formation (Brewer et al., 1999), the Parallel Distribution Satisfaction model of person perception (Kunda and Thagard, 1996), and even Social Identity Theory (Tajfel, 1981), pay special attention to the way social categories influence impression formation and updating on other persons. Several studies have repeatedly demonstrated that knowledge of social categories and their stereotypical characteristics often have more impact on impressions about other persons than individual trait knowledge (Bodenhausen et al., 1999). In contrast to the wealth of theories and research in the behavioral literature on the role of social categories and personality traits during impression formation (Andersen et al., 1990; Hamilton and Sherman, 1996; Kunda and Thagard, 1996; Fiske and Neuberg, 1990; Brewer, 1988), direct comparisons between these two types of constructs are, to the best of our knowledge, much less investigated in current social neuroscientific research. Some studies have touched upon related topics, such as the difference between categorization and individuating processes in the brain (Mason & Macrae, 2004), stereotypical judgments (Quadflieg et al., 2009), trait and evaluative representations during stereotype judgments (Gilbert et al., 2012), and inferring mental states from groups (Contreras et al., 2013).

In line with this earlier work, the aim of this study is to investigate directly the differences and similarities in brain activations during the processing of behavioral descriptions of individuals in the context of social categories and personality traits. We use behavioral theories about these issues to make predictions about activation patterns in the human brain.

Differences between social categories and traits

Social categories and traits appear to share a similar function, as both social constructs provide information which allows us to predict future
behavior of a target person, as well as to coordinate our own behavior, cognition and affect towards targets of specific categories (e.g. thieves) or personality types (e.g. dishonest people) (Fiske et al., 1999; Bodenhausen et al., 1999). As the example demonstrates, social categories and individuals may even share common traits, such as dishonesty. Nevertheless, social categories and traits also differ with regard to some specific characteristics.

It is evident that social categories refer to group-membership while traits suggest individual attributes. Andersen et al. (1990) further pointed out that social categories tend to be more idiosyncratic and specific, and hence more informative and imaginable than personality traits. Social categories and traits also differ with regard to abstraction, and here Andersen et al. (1990) offered two possible options. On the one hand, social categories can be viewed as concrete and specific categories (e.g., nurses) that are subordinate in a hierarchical relationship to traits, which are more general and abstract (e.g., caring). On the other hand, social categories are often considered as an organizing entity of group members (e.g., nurses), combining different traits (e.g., caring) as well as other types of individual information like physical attributes (e.g., dressed in white) or demographic characteristics (e.g., women). In this sense, traits are basic and concrete, and subordinate to more general social categories (Andersen et al., 1990; Macrae et al., 1994).

The latter view corresponds to a narrow definition of personality traits, in which traits only convey the behavior they summarize (Andersen et al., 1990) and a broader definition of social categories as a more complex concept that involves a higher level of abstractness, also termed level of construal (Trope and Liberman, 2010).

A crucial question is, do these differences play a role in how social categories and traits are used in impression formation? As noted earlier, behavioral research has repeatedly demonstrated that knowledge of social categories and their stereotypical trait characteristics have more impact on impression formation than individual trait knowledge (Bodenhausen et al., 1999; Fiske et al., 1999). In a study directly comparing social categories and traits, Andersen et al. (1990) documented that social categories are more influential than traits. These authors presented behavioral descriptions of a person who was depicted by either a social category (e.g., politician) or a trait (e.g., daring) and performed a mundane act (e.g. The politician/The daring type opened the drawer) or experienced a common state (e.g. The politician/The daring type felt a draft). Participants had to judge whether the target person was likely to exhibit such behavior or experience such a state. The results revealed that when judging the likelihood of the behavior or state, the response was faster following social categories compared to traits, indicating that social categories function as a more efficient prime than traits. Social categories also appeared to provide a memory advantage over traits in a cued recall task, where the presented behaviors or states served as cues. These results were interpreted as indicating that social categories are more accessible and semantically richer than traits, encompassing traits as one of the many features that make up a single social category (Macrae et al., 1994).

Neurophysiological findings on social categories and traits

Recent event-related-potential (ERP) studies support the greater influence of social categories during impression formation, in particular with respect to processing speed. The results indicate a very rapid detection of race and gender from visual information (Ofan et al., 2011; White et al., 2009). In contrast, personality traits which are usually inferred from verbal behavioral descriptions tend to require more time (Vanduyneelaeger et al., 2007; Baetens et al., 2011).

Neuroimaging studies focused on the brain areas involved in impression formation, using verbal descriptions of trait-implicating behaviors (Ma et al., 2012; Mende-Siedlecki et al., 2012) or a stereotype (political affiliation of the protagonist, see Cloutier et al., 2011), revealed substantial overlap for trait and stereotypical processing in a number of brain areas belonging to the mentalizing network. This network is involved in the understanding of other person’s behavior as driven by internal mental states like thoughts, emotions and beliefs (Van Overwalle, 2009; Spreng et al., 2008). It encompasses various medial areas including the medial prefrontal cortex (mPFC), posterior cingulate cortex, and precuneus, as well as areas in the lateral temporal lobe, including the temporo-parietal junction (TPJ), posterior superior temporal sulcus (pSTS), and anterior temporal lobe (aTL). This social mentalizing network shows considerable overlap with other networks that share similar functionalities, such as the default-network (Raichle et al., 2001; Spreng et al., 2008), the autobiographical memory network (Svoboda et al., 2006), the evaluative network (Legrand and Ruby, 2009) and the task-negative network (Fox et al., 2005; see Spreng, 2012 for a discussion on this classification).

Beyond the processing of individual mental states, activation in the mentalizing network has also been observed when making judgments about social groups, such as about their typical characteristics and features (Contreras et al., 2012; Mitchell et al., 2009; Quadflieg et al., 2009), social identity (Volz et al., 2009), ingroup loyalty (Baumgartner et al., 2012), social interaction (Lahnakoski et al., 2012), group-based emotional responses (Harris et al., 2007), and when making evaluative and trait judgments related to group stereotypes (Gilbert et al., 2012).

However, given the lack of direct comparisons in previous neuroimaging research, it is still unclear whether the mentalizing system is more strongly involved in the processing of behavioral information with respect to social categories or individual traits. In line with behavioral research reviewed earlier, there are several arguments supporting the prediction that processing information about groups as opposed to individuals recruits more activity in the mentalizing network.

First, social categories appear to be more influential in the process of impression formation. Upon activation they provide a wealthier source of information than personality traits, which aids in interpreting behaviors. The processing of social categories may facilitate the retrieval of multiple autobiographical memories of individual contacts with category members, leading to increased accessibility of typical group exemplars as opposed to a single individual. Such elaborate processing of autobiographical memories may lead to increased activation in areas associated with the mentalizing system (Holland et al., 2011). Second, social categories have more variability and complexity than individuals. They have higher idiosyncratic, specific and complex representations, due to the higher number of individuals involved and the sometimes broad social categories they represent (e.g., woman). This higher complexity and variability may demand a higher level of information processing, leading to increased activation of the mentalizing network (Meyer et al., 2012; Mitchell, 2010). Lastly, social categories reflect a higher level of abstractness or construal (Trope and Liberman, 2010). Recent imaging research demonstrated that increasing the level of construal (e.g., by focusing on trait versus visual characteristics of a person) increases the activation of the mentalizing network, and the mPFC in particular (Baetens et al., 2014). Based on a similar logic, we might expect an increase of activation when moving the level of abstractness from traits to social categories.

The present research

In the present study, we explore how social categories and traits have a differential impact on the processing of behavioral descriptions. We briefly presented a personality trait (e.g. friendly) or social category (e.g. nurse) that implied the same trait, and then provided behavioral sentences that were either consistent or inconsistent with the social category or trait. Participants were requested to judge to what extent the behavior was applicable to the social category or individual trait. Pilot testing ensured that the behavioral descriptions were on average equally applicable in relation to the social category or the individual trait, and the analyses further controlled for remaining differences in applicability.
Based on previous findings, we expect that the processing of behavioral descriptions, primed by either a social category or a trait, will activate brain areas associated with the mentalizing network. Given that social categories seem to reflect a richer and more complex high-level construct (Andersen et al., 1990; Macrae et al., 1994), we also expect that social categories recruit the mentalizing network more strongly than traits. In addition, when behavioral descriptions are inconsistent as opposed to consistent with the group or trait, we expect stronger attempts to resolve the inconsistency, resulting in increased activation in the mentalizing network and/or the conflict monitoring network (Ma et al., 2012; Cloutier et al., 2011; Mende-Siedlecki et al., 2012). This latter, domain-general network is involved in monitoring and detecting conflicts between multiple inputs (subserved the dorsal anterior cingulate), which are resolved by modifying the attention and weight given to the conflicting inputs (subserved by the lateral PFC; Botvinick, 2007; Shenhav et al., 2013).

Method

Participants

Eighteen right-handed native Dutch-speaking undergraduate and graduate students from the University of Ghent (13 women, ages ranging between 18 and 27, mean age 21.36 years) participated in exchange for 10 euro. Participants had 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 Ghent University (where the study was conducted) and of the Vrije Universiteit Brussel (of the principal investigator Frank Van Overwalle).

Procedure and stimulus material

Before entering the scanning room, the participants received written and oral instructions. During fMRI scanning, a personality trait (e.g. friendly) or a social category (e.g. kindergarten teacher) of positive or negative valence was presented for a duration of 2 s (Fig. 1). The participants were instructed to visualize a prototypical person possessing the trait or belonging to the social category. This was followed by a behavioral description of the person that was either consistent (e.g. smiles to the baby) or inconsistent (e.g. leaves the children unsupervised), for a duration of 6 s (see Appendix A for illustrative sentences). As such, the design of this study involved 3 within-participants factors, involving Type (social categories versus traits), Consistency (consistent versus inconsistent) and Valence (positive versus negative), resulting in eight conditions (see Appendix A for example sentences in each condition).

After each presentation of a behavioral sentence, the participants rated the applicability of the behavioral description to the trait or social category on a 4-point scale (ranging from 1 = very unlikely to 4 = very likely) by pressing a button on a 4-button response box with their left hand. At the beginning of each trial we asked to visualize a typical exemplar in order to avoid a purely semantic answer which could be driven only by shallow semantic associations. Another goal of this manipulation was to constrain the judgment of the behavior to an individual for both types of constructs. There were in total 160 experimental trials segregated into 8 blocks of 40 trials, separated by 30-second pauses. The experimental trials were preceded by 10 practice trials. After scanning, each participant was debriefed and thanked.

During pretesting, the behavioral sentences were rated \( n = 108 \) on two 7-point scales measuring the applicability of the social category or trait (1 = very unlikely, 7 = very likely) and valence (1 = negative, 7 = positive). Consistent behavioral descriptions were selected when they received an average applicability rating of 5 or higher, while inconsistent behavioral descriptions had an average rating of 3 or lower. Positive behavioral descriptions had an average valence rating of 5 or higher, while negative behavioral descriptions had an average rating of 3 or lower. Applicability ratings did not differ significantly between social categories \( (M_T = 3.99) \) and traits \( (M_T = 3.87, \text{ns}) \). This was also the case for valence ratings for social categories \( (M_T = 3.91) \) and traits \( (M_T = 3.89, \text{ns}) \). The length of the sentences did not differ across a positive valence \( (M_T = 24.38) \) or negative valence \( (M_T = 23.30, \text{ns}) \), or across consistent \( (M_T = 24.35) \) or inconsistent \( (M_T = 23.33, \text{ns}) \) behaviors, and ranged between 4 and 7 words. Log10 word-frequency was assessed through use of the SUBTLEX-NL database (Keuleers et al., 2010) and was also equivalent for social categories and traits across valence \( (M_T = 19421.64 \text{ and } M_C = 21897.97, \text{ns}) \) and consistency \( (M_C = 18615.12 \text{ and } M_C = 22704.49, \text{ns}) \) of the behaviors.

During a post-test, consisting of a shortened version of the current experiment (40 trials), eleven participants (9 women, ages ranging

---

**Fig. 1.** Schematic illustration of stimulus presentation.
Statistical analysis correlations were accounted for by the default auto-regressive resulting from our within-participants design involving Type (social participant analyses with a regressor for each of the eight conditions nuisance regressors. We used a default high-pass step as well as outlier time points (de
relations between motion and experimental design, and between global org/projects/artifact_detect/), for excessive motion artifacts and for cor-
experimental design or global signal and experimental design were
single regressor for each outlier. No correlations between motion and
default, these outliers were omitted from the analysis by including a
using the following criteria in ART (Z-threshold: 3.0 mm, scan to scan
temporal differences series by assessing between-scan differences
MNI), which approximates Talairach and Tournoux atlas space. Normal-
with each participant
and across runs to correct for head movement, and co-registered
tion time between slices for each whole-brain volume, realigned within
and artifact. Functional data were corrected for differences in acquisi-
tion time between slices for each whole-brain volume, realigned within
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) running under Windows XP. Foam cushions were placed within the head coil to minimize head movements. We first collected a high-resolution T1-weighted structural scan (MP-RAGE) followed by one functional run (30 axial slices; 4 mm thick; 1 mm skip). Functional scanning used a gradient-echo echo-planar pulse sequence (TR = 2 s; TE = 33 ms; 3.5 × 3.5 × 4.0 mm in-plane resolution).

Imaging procedure

Images were collected using a 3 Tesla Magnetom Trio MRI scanner system (Siemens medical Systems, Erlangen, Germany), with 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
and also differed with regard to familiar persons (4), including family,
friends and famous people. The questions further involved describing specific characteristics of the person, including (5) face and (6) clothing, whether participants (7) were able to create this impression within the timeframe of 2 seconds, and whether this timing differed for social categories and traits (8).

Statistical analysis

The statistical analysis of the fMRI data involved first-level single participant analyses with a regressor for each of the eight conditions resulting from our within-participants design involving Type (social categories versus traits), Consistency (consistent versus inconsistent) and Valence (positive versus negative) (see Appendix A for individual conditions). These regressors were time-locked at the onset of the behavioral sentence presentation. In addition, there were 6 movement artifact regressors, and a variable amount of artifact regressors determined by ART (Gabrieli Lab, 2009; Whitfield-Gabrieli, Mozes, & Castanon, MIT). After defining the conditions in SPM8 (Wellcome Department of Cognitive Neurology, London, UK), a reparameterization procedure (developed by Lindquist and Wager, 2007) was applied to estimate the latency, height and width of the hemodynamic response function for each block. This semi-parametric procedure models the hemodynamic response function for the data at the onset of each trial for each voxel individually, to allow estimation of the amplitude for each condition. The analysis returns Area Under the Curve (AUC) images that take into account the time to reach maximum amplitude post stimulus onset, and the duration of the hemodynamic response. AUC images were created for each condition at subject level and subsequently used at the group-level.

At the group second-level, analyses were performed on the parameter estimates (regressors) associated with each of the eight conditions using a random-effects model. The statistical analysis involved 3 within-participants factors (Type, Consistency and Valence). Statistical comparisons between conditions were tested by conducting an analysis of variance (ANOVA) on the parameter estimates associated with each trial type using the default contrasts provided by SPM, and testing specific comparisons of interest with one sample t-tests. To confirm that both types of constructs activated similar regions, we ran an analysis testing a conjunction null hypothesis for activation shared by social categories (≤ implicit baseline) and traits (≤ implicit baseline). A voxel-based statistical threshold of p ≤ 0.05 (FWE cluster-corrected) was used for all analyses with a minimum cluster extent of 10 voxels.

To verify the time course of the activation in mentalizing and conflict monitoring regions revealed by the whole brain analysis, regions of interest (ROIs) were analyzed. The ROIs involved a sphere of 8 mm radius around the peaks of the contrasts social categories > traits and traits > social categories, and of the contrasts inconsistent > consistent and consistent > inconsistent. We estimated a Finite Impulse Response (FIR) model of the hemodynamic response at the subject-level, and extracted the time course and the mean percentage signal change from each ROI using the MarsBar toolbox (http://marsbar.sourceforge.net).

We defined in this analysis the presentation of the behavioral sentences as occurring at time 0; of the (earlier) presentation of the social construct type at time −2, and of the applicability question at time 6 (see Fig. 1). To capture the hemodynamic response peak which is expected to follow at about 6 seconds after stimulus presentation, we calculated the average percentage signal change for the social construct type during a window of 2 to 4 seconds, of the behavioral sentences during a window of 6 to 8 seconds, and of the applicability judgment during a window of 12 to 14 seconds. This allowed to differentiate between hemodynamic peaks involved in the processing of the social construct type, behavioral sentences and the applicability judgment respectively, and analyzed them using one sample t-tests (Miezin et al., 2000).

Finally, we tested each region’s involvement during the different tasks by testing the average percentage signal change against the session mean for each condition.

Results

Behavioral results

An ANOVA with Type (social category vs trait), Consistency (consistent vs. inconsistent) and Valence (positive vs. negative) as within-participants factors was performed on the mean applicability ratings per condition. The results revealed a significant main effect of Consistency F(1,136) = 60.94, p < 0.001, indicating that applicability was rated higher when consistency was high. No other main or interaction effects were significant.
The results of the post-test revealed that participants (n = 11) imagined the person in a specific situation, and that the impression they had in mind differed for traits and social categories. The person described by a social category was easy to visualize, concrete and well-defined, and tended to wear specific clothing. Conversely, the person described by a trait was difficult to imagine, abstract and fuzzy. Five participants thought of familiar persons while creating an impression of the person, in line with the presented social category or trait. Additionally, all participants reported that when judging the applicability of a behavior by a member of a social category, they also considered atypical members as being part of the social category. Finally, all native Dutch speaking participants (n = 7; excluding participants with bilingual background just like in the FMRI experiment) indicated that they were able to come up with a visual impression of the person within the 2 second timeframe in the majority of the trials, and this did not differ between social categories or traits. The instances in which these participants were unable to form an image within this timeframe were all due to being personally less familiar with the social category or trait, so that only after reading the subsequent behavior they had sufficient information to form an impression.

**Imaging data**

**ANOVA**

We conducted a whole-brain analysis, in which we performed an ANOVA on the same design (i.e., Type x Consistency x Valence). The results revealed significant main effects for Type and Consistency, as well as a significant interaction between Consistency and Valence. No other significant main effects or interactions were found. We next analyzed specific contrasts testing the differences between Construct Type and Consistency, and finally explored the interaction between Consistency and Valence.

**Social construct type T-contrasts**

We tested our first hypotheses by computing contrasts between social categories and traits (Table 1 and Fig. 2). As expected, social categories recruited the mentalizing areas in the midline (dorsal mPFC, orbitofrontal cortex, posterior cingulate cortex) and temporal cortices (TPJ and aTL) more strongly than traits. In addition to this, social categories also recruited the bilateral fusiform gyrus, the cerebellum, and left temporal pole, more than traits. In contrast, traits recruited the anterior part of the middle cingulate cortex, left inferior parietal cortex, and precuneus more strongly than social categories.

**Social construct type ROI analysis**

To verify the time course of the activation revealed by the whole-brain, we performed a Region of Interest (ROI) analysis on an 8 mm sphere surrounding the peak voxel in the activated cluster of the contrasts between social categories and traits (social categories > traits and traits > social categories), and extracted the percentage signal change. As detailed in the method section, we extracted the signal of the social construct type at 2–4 s, the behavioral sentence at 6–8 s, and applicability rating at 12–14 s. Note, however, that because these elements are all nested within one another within a trial, the activity of one element is necessarily tied to the activity of another element. Therefore, these results should be interpreted with caution. In the 2–4 seconds time window corresponding to the presentation of the social construct type, t-tests revealed higher activation of social categories than traits in the posterior cingulate cortex (PCC), left anterior inferior parietal sulcus (aIPS), orbitofrontal cortex, left parahippocampal gyrus and left aTL (Table 1). In the 6–8 seconds time window following the presentation of the behavioral sentence, all core regions of the mentalizing network and the cerebellum were recruited. During the 12–14 second time window corresponding to the expected hemodynamic response peak of the applicability judgment task, a subset of these mentalizing regions still showed a significant difference in activation, including the dorsal mPFC, bilateral aTL and bilateral TPJ. The comparison of traits versus social categories revealed no significant differences in the 2–4 time window, increased activations during the 6–8 time window for the left aIPS, anterior part of the middle cingulate cortex and precuneus, and only the precuneus showing increased activation in the 12–14 seconds time window.

**Social construct type ROI analysis against implicit baseline**

To test whether the regions revealed by each contrast were significantly activated or deactivated compared to baseline for each condition (social categories > implicit baseline and traits > implicit baseline), we computed the difference between the average of each time window and the session mean. The results (Table 2) revealed that social categories showed significant activation above the session mean and the precuneus showing increased deactivation below the session mean in the posterior cingulate cortex in all time windows, and in the left inferior parietal sulcus, middle cingulate cortex, precuneus and left cerebellum lobule VIIa Crus II during limited time windows. Traits revealed significant activation above the session mean in the anterior part of the middle cingulate cortex in all time windows, and in the left temporal pole during some time windows. Traits also showed significant deactivation below the session mean in the right inferior parietal sulcus in all time windows, and the posterior cingulate cortex, precuneus and right parahippocampal gyrus during limited time windows. The dmPFC and left parahippocampal gyrus showed variable increasing and decreasing patterns.

**Consistency T-contrasts**

We tested our second hypothesis by computing contrasts between consistent and inconsistent conditions (Table 1). The results showed increased activation for inconsistent compared to consistent sentences in the right TPJ and left lingual gyrus, and increased activation in the right lingual gyrus, bilateral postcentral gyrus and right cuneus for consistent sentences (Fig. 3).

**Consistency ROI analysis**

These results were confirmed by a ROI analysis analogous to the one performed on the social construct type (Table 2). Inconsistent versus consistent sentences revealed increased activation in the right TPJ during the 6–8 seconds time window (following the presentation of the behavior), and increased activation in the left lingual gyrus during the 12–14 seconds time window (following the judgment task). Consistent versus inconsistent sentences showed increased activation in the right lingual gyrus and bilateral postcentral gyrus during the 6–8 seconds time window, and only the right lingual gyrus showed increased activation for consistent sentences during the 12–14 seconds time window. There were no significant differences during the 2–4 time window (corresponding to the presentation of the social construct type: social category or trait).

**Consistency ROI analysis against implicit baseline**

To test whether the regions revealed by each contrast were activated or deactivated significantly compared to baseline for each condition, we computed the difference between the average of each time window and the session mean (inconsistent > implicit baseline and consistent > implicit baseline). The results (Table 2) revealed that inconsistent sentences showed significant activation above the session mean in the right precentral gyrus and left postcentral gyrus during limited time windows, while showing significant deactivation below the session mean in the right angular gyrus, right superior occipital gyrus and left postcentral gyrus during limited time windows,
and variable increasing and decreasing patterns in the left postcentral gyrus and bilateral lingual gyrus. Consistent sentences showed significant activation above the session mean in the right precentral gyrus and left postcentral gyrus during limited time windows, while showing deactivation in the right angular gyrus and left lingual gyrus during limited time windows. The right lingual gyrus and right superior occipital gyrus showed variable increasing or decreasing patterns.
Interaction consistency and valence T-contrasts

Third, we computed contrasts testing the interaction between Consistency and Valence (Table 1). The results revealed increased activation for negative consistent sentences, compared to positive consistent sentences, in the left inferior frontal gyrus, and increased activation for negative inconsistent sentences, compared to positive

**Fig. 2.** Brain activation related to the processing of social construct type, increased activations for social categories compared to traits are shown in red, and increased activations for traits compared to social categories are shown in blue. A) Whole-brain contrasts thresholded at \( p < .05 \) (FWE-corrected) with a minimum cluster extent of 10 voxels. B) Finite impulse response time course in percentage signal change per TR of 2 seconds from \(-2\) seconds to 30 seconds following the presentation of the behavioral sentences for the regions of interest (ROIs) indicated by the arrows in A.
inconsistent sentences, in the left middle temporal gyrus and bilateral lingual gyrus. The other contrasts did not reveal increased activations.

**Conjunction of social construct types**

The previous analyses suggest that although differences appear between social categories and traits, both social constructs seem to recruit
similar networks. To explore shared activation for traits and social categories, we ran a conjunction analysis across traits and social categories, each contrasted against an implicit baseline. Results revealed that both traits and social categories share activation in the bilateral occipital lobe, left inferior frontal gyrus and left precentral gyrus (Table 3). The conjunction analysis failed to support our expectations that identical mentalizing regions are activated during both social categories and traits. However, these results should be treated with caution, because this conjunction is compared against an implicit baseline (reflecting the average activation of categories and traits) rather than a non-mentalingizing control condition, so that many more areas may actually be shared.

### Discussion

The present study explored whether the underlying neural process linking behaviors to social categories or traits is similar. We reasoned that processing behavioral information associated with social categories in comparison with traits, would require increased involvement of the mentalizing network. Our results partially confirmed this prediction. The analysis revealed increased activation for social categories in areas of the mentalizing network, including midline (dmPFC, orbitofrontal cortex, PCC) and temporal structures (TPJ and ATL) during behavioral processing. Furthermore, we predicted that behaviors which are inconsistent with these social categories or traits would reveal increased activation in regions associated with the mentalizing or conflict monitoring networks. Our analysis of the time course of the activations suggests that of aforementioned clusters, only the right TPJ showed increased activation during the processing of behavior inconsistent with the presented social category or trait. Although this time analysis should be interpreted with caution because all elements in a trials are inherently nested, it seems sensible to assume increased TPJ activation after the presentation of an inconsistent behavior. As social categories and traits have similar functions, such as making causal attributions and predicting future behavior (Jackson et al., 1993), we expected both social constructs to activate the mentalizing system. Although the

### Table 2

<table>
<thead>
<tr>
<th>ROI x</th>
<th>y</th>
<th>z</th>
<th>Percentage signal change 2–4 s</th>
<th>Percentage signal change 6–8 s</th>
<th>Percentage signal change 12–14 s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trait Category</td>
<td>Trait</td>
<td>Trait</td>
<td>Trait</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>Trait</td>
<td>Trait</td>
<td>Trait</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emotional</td>
<td>Trait</td>
<td>Trait</td>
<td>Trait</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dmPFC</td>
<td>10</td>
<td>56</td>
<td>24</td>
<td>-0.015</td>
<td>-0.026</td>
</tr>
<tr>
<td>dFPFC</td>
<td>12</td>
<td>46</td>
<td>-5</td>
<td>0.105</td>
<td>0.111</td>
</tr>
<tr>
<td>Orbitofrontal cortex</td>
<td>0</td>
<td>30</td>
<td>-28</td>
<td>0.110</td>
<td>0.007</td>
</tr>
<tr>
<td>Right anterior temporal pole</td>
<td>-44</td>
<td>10</td>
<td>-42</td>
<td>0.117</td>
<td>0.080</td>
</tr>
<tr>
<td>Right anterior temporal lobe</td>
<td>-58</td>
<td>-8</td>
<td>-26</td>
<td>0.053</td>
<td>0.003</td>
</tr>
<tr>
<td>Right parahippocampal gyrus</td>
<td>26</td>
<td>30</td>
<td>-20</td>
<td>0.018</td>
<td>-0.023</td>
</tr>
<tr>
<td>Left parahippocampal gyrus</td>
<td>-24</td>
<td>-32</td>
<td>-22</td>
<td>0.041</td>
<td>-0.040</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>6</td>
<td>-52</td>
<td>-48</td>
<td>0.059</td>
<td>0.021</td>
</tr>
<tr>
<td>Right inferior parietal sulcus</td>
<td>-50</td>
<td>-30</td>
<td>62</td>
<td>-0.092</td>
<td>-0.130</td>
</tr>
<tr>
<td>Right inferior parietal lobe</td>
<td>-46</td>
<td>-68</td>
<td>34</td>
<td>0.048</td>
<td>*</td>
</tr>
<tr>
<td>Posterior cingulate cortex</td>
<td>14</td>
<td>-86</td>
<td>-40</td>
<td>0.004</td>
<td>-0.029</td>
</tr>
<tr>
<td>Right superior frontal gyrus</td>
<td>12</td>
<td>-88</td>
<td>-38</td>
<td>-0.055</td>
<td>*</td>
</tr>
</tbody>
</table>

### Table 3

<table>
<thead>
<tr>
<th>Conjunction of inference type</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>k</th>
<th>max</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trait &gt; social category</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inconsistent &gt; consistent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right angular gyrus</td>
<td>44</td>
<td>-54</td>
<td>36</td>
<td>-0.055</td>
<td>*</td>
<td>-0.074</td>
</tr>
<tr>
<td>Left lingual gyrus</td>
<td>-10</td>
<td>-78</td>
<td>12</td>
<td>-0.060</td>
<td>*</td>
<td>-0.036</td>
</tr>
<tr>
<td>Consistent &gt; inconsistent</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right precentral gyrus</td>
<td>58</td>
<td>-8</td>
<td>46</td>
<td>0.047</td>
<td>*</td>
<td>0.074</td>
</tr>
<tr>
<td>Left postcentral gyrus</td>
<td>-50</td>
<td>-12</td>
<td>54</td>
<td>0.042</td>
<td>*</td>
<td>0.087</td>
</tr>
<tr>
<td>Right lingual gyrus</td>
<td>14</td>
<td>-72</td>
<td>0</td>
<td>-0.092</td>
<td>*</td>
<td>-0.047</td>
</tr>
<tr>
<td>Right superior occipital gyrus</td>
<td>20</td>
<td>-80</td>
<td>30</td>
<td>-0.069</td>
<td>*</td>
<td>-0.040</td>
</tr>
</tbody>
</table>

### Note

- x, y, and z = Montreal Neurological Institute coordinates of the peak values in the left-right, posterior-anterior, and inferior-superior axes respectively; P-values corresponding to the difference in average percentage signal change for time window 2–4 seconds, 6–8 seconds and 12–14 seconds respectively. Positive values denote activations above the session mean baseline, while negative values denote deactivations below the session mean baseline.
- p < .05; **p < .01; ***p < .001.
mentalizing system was activated for both constructs, our conjunction analysis did not provide support for our hypothesis that the activated clusters are identical. However, given that these conjunction results were contrasted again an implicit baseline (implicating the average activation of categories and traits), this analysis may underestimate the true overlap in mentalizing activity across categories and traits.

To our knowledge, this is the first fMRI-study that directly compared the distinctive impact of social categories and traits on behavioral information processing. The current results point towards an increased involvement of the mentalizing network during the processing of behaviors related to social categories compared to traits. This differential activation of the mentalizing network may be explained by several important features of social categories and traits.

Previous behavioral studies have shown that during the process of impression formation, social categories seem to facilitate this process more than individual traits. Compared to traits, social categories provide a complex cluster of information which leads to increased mentalizing (Andersen et al., 1990; Macrae et al., 1994). Impressions of social categories are also based on a greater number of an individual’s autobiographical memories of contact with its members (Andersen et al., 1990). This is one of the possible reasons why social categories are also higher in variability, because members may share group characteristics to a lesser or larger degree (Andersen et al., 1990). Lastly, social categories tend to refer to a group of people rather than a single individual, and are thus higher in number. All these factors may lead to increased mentalizing.

Most of our arguments for increased activation during social category processing in comparison with trait processing are supported by the neuroimaging literature. For instance, increased activation of the mPFC has been found when a higher number of social targets have to be remembered (Meyer et al., 2012), when there are more targets in one’s social network (Lewis et al., 2011), and when a higher level of abstractness is required in the representation of social targets (Baetens et al., 2014). Heightened activation for social categories compared to traits in the anterior temporal lobe also corresponds to findings that this region is involved in the retrieval and representation of abstract social semantic knowledge (Olson et al., 2013; Zahn et al., 2007). Increased activation of the precuneus and adjacent posterior cingulate cortex has also been linked to increases in retrieval of autobiographic memories (Cavanna and Trimble, 2006; Krienen et al., 2010). This region is strongly interconnected with the prefrontal region (Cavanna and Trimble, 2006).

The results from our post-test indicate that even though groups as a concept can be considered as more abstract than personality traits, the participants found it easier to generate a visual impression of a representative of a social category, than they did for a trait. The impression generated of a social category was also described as more concrete and specific, than that of a trait. This suggests that the increased activation for social categories is perhaps not due to their higher level of abstraction than traits, but due to their association with a rich knowledge base of specific information, which therefore triggers more elaborate processing.

Interestingly, in contrast to the general pattern of increased mentalizing activity for social categories, the left inferior parietal lobule (IPL; adjacent to the TPJ) and the left inferior parietal sulcus (IPS), showed increased activity during trait processing. The IPL and IPS have been linked to the mirror network related to action understanding, while the TPJ is generally considered as part of the mentalizing network (Spunt and Lieberman, 2014; Van Overwalle and Baetens, 2009). We have currently no definite explanation for the increased IPL and IPS activation during trait inferences, except that perhaps traits more than social categories may elicit simulations of concrete behaviors, and so activate the mirror network. We also found increased activation in the fusiform gyrus for social categories. Given the previously reported association of this region to face processing (Kanwisher, 2010), this activation tentatively suggests increased ease at imagining typical faces fitting social categories compared to traits (as no pictures were presented in this study).

A limitation of this study is that social categories may induce more reasoning in terms of differences between group members, while traits may focus more on a single individual. Although we instructed participants to visualize a typical member/individual of a group, this representation may have induced dissimilarities that we have interpreted as caused only by social categories versus traits but may have been due to other factors like increased ambiguity, or increased freedom and variation in retrieving information. By asking the participants to visualize a typical exemplar we were able to avoid them generating a purely semantic answer which would be driven only by shallow semantic associations. Additionally, by allowing the participants to visualize a typical exemplar before presenting them the behavioral sentences we tried to make this second task easier for the participants and somewhat more similar among them, by guiding them by our instructions. Thus, our intention was to constrain the participants in some way, rather than inducing more freedom and variation. We would also like to argue that when visualizing a person having a certain trait or belonging to a certain group, there is no easy way to further constrain this because retrieving exemplars of a social category or a trait is a part of the perception process and differs for each individual. A different approach, that would’ve given us more experimental control over the retrieval processes, like asking participants to only think of people they know personally, would have created an unnatural experimental context, as participants would have to inhibit information they’ve been retrieving because it did not fit the requirements of the task.

Other differences may have impacted our results as well, such as the syntactical differences between trait adjectives (e.g. athletic) versus stereotype nouns (e.g. athlete). Evidently, more research is needed to single out the processes underlying the increased activation of the mentalizing network during the processing of social categories versus traits.

A methodological limitation of the current study that places certain limits on the interpretation of the results concerns the differential impact of social categories and trait on the different tasks presented. By using different time windows in the ROI analysis, we were able to detect the differential impact of social categories and trait on each task to a large extent. Mainly the difference between the visualization task and the behavioral processing appears difficult as the participants only had two seconds to perform the visualization task and the time windows of 2–4 seconds and 6–8 seconds might not fully allow us to distinguish between these two tasks. At this time the activation in the 2–4 seconds time window might be related to the differential impact of social categories and traits on the visualization task, or to the magnitude of the effect of the manipulation. Future research will have to look at the specific impact of social categories and trait on semantic retrieval processes in the brain.

Conclusion

In this study we found that both social categories and traits activate the mentalizing network, yet they seem to do so to a different extent. It seems acceptable to assume that the amount of mentalizing is proportional to the type of information that is provided, being more active for group social categories than for individual traits.

Appendix A. Examples of stimulus material

<table>
<thead>
<tr>
<th>Condition</th>
<th>Trait/category</th>
<th>Behavioral description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trait</td>
<td>Consistent</td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td>Inconsistent</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>Consistent</td>
<td>Positive</td>
</tr>
<tr>
<td>category</td>
<td>Inconsistent</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Positive</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Negative</td>
</tr>
</tbody>
</table>

doubt bonus | insults his guests |