NeuroImage 156 (2017) 155–165

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Neural correlates of believing

Xiaochun Han, Ting Zhang, Shiyu Wang, Shihui Han*

School of Psychological and Cognitive Sciences, PKU-IDG/McGovern Institute for Brain Research, Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China

ARTICLE INFO

Keywords:
Believe
fMRI
Self
Anterior insula
Inferior frontal cortex

ABSTRACT

Beliefs provide a fundamental cognitive basis for human behavior. But how the brain believes remains a mystery. We investigated the neural underpinnings of believing by scanning healthy adults using functional magnetic resonance imaging when they made yes/no responses to the questions whether they *believe* or *think* that a trait adjective describes themselves or a celebrity. We found that, relative to thinking, believing was characterized with better memory of self-related adjectives. Moreover, believing (vs. thinking) was associated with stronger activations in the left anterior insula/inferior frontal cortex, stronger functional connectivity between the medial prefrontal cortex and left occipital cortex during judgments of one's own personality traits, and stronger intrinsic connectivity between the left occipital cortex and the left anterior insula/inferior frontal cortex. Our findings shed new light on the neurocognitive processes that characterize believing as a mental process in healthy adults.

Introduction

Believing seems the most mental thing we do — Bertrand Russel (1921), p.231).

Beliefs, either religious or secular, are fundamentally important for human societies and influence our behaviors tremendously. The study of beliefs has captivated researchers from philosophy, psychology, and others fields for centuries (Bogdan, 1986). Recent brain imaging studies have shown increasing interests in neural correlates of beliefs (Harris et al., 2008; Kapogiannis, 2009; Seitz and Angel, 2012; Krueger and Grafman, 2013) and how beliefs modulate neural substrates of other cognitive/affective processes (e.g., Han et al., 2008, 2010). The brain imaging approach takes the notion of belief as a mental construct (Kapogiannis et al., 2009; Harris et al., 2009) or experience (Azari et al., 2001; Beauregard and Paquette, 2006), and has revealed the association of multiple brain regions including the medial prefrontal cortex (mPFC) with beliefs. However, beliefs involve both mental representations and assessment of meaningful information (Bogdan, 1986) and *believe* as a process is a fundamental human brain function that happens frequently in everyday lives (Angel and Seitz, 2016). Surprisingly, there has been little empirical research on the neural substrates of believing as a process, and how the brain believes remains a mystery.

It is difficult to unravel the neurocognitive mechanisms of believing due to the lack of a hypothesis of the underlying processes. It is also challenging to isolate the neurocognitive processes of believing from the neural representation of mental contents for believing by designing a control condition that can be compared with believing while perceptual/cognitive/affective processes are well controlled. It has been recently proposed that the believing process is connected with personal relevance, deals with a set of knowledge with a hierarchically organized structure, and has social and personal adaptive functions (Sugiura et al., 2015). Based on this proposition, the current work examined the neurocognitive processes underlying believing by integrating functional magnetic resonance imaging (fMRI) and a well-established self-referential task (Rogers et al., 1977) that has been widely used in brain imaging studies (Kelley et al., 2002; Macrae et al., 2004; Northoff et al., 2006; Hu et al., 2016). The self-referential task requires an individual to make judgments whether a number of trait adjectives can describe oneself and thus engages the process of personally relevant knowledge that is critical for social adaptation.

Two main findings came out of the previous studies using the self-referential task. First, trait adjectives used for Self-judgments were remembered better than those used for trait judgments of a celebrity (Rogers et al., 1977), reflecting more elaborated encoding of self-relevant information. Second, trait judgments of oneself vs. a celebrity activated the mPFC and other brain regions (Kelley et al., 2002; Macrae et al., 2004; Ma and Han, 2011), reflecting unique neural representations of the self. However, the previous studies asked participants to make yes/no responses to the questions whether an adjective describes oneself or a celebrity and such ambiguous task instructions cannot disentangle the believing process from other mental processes during

---

* Correspondence to: School of Psychological and Cognitive Sciences, Peking University, No. 52, Haidian Street, Beijing 100080, China.
E-mail address: shan@pku.edu.cn (S. Han).

http://dx.doi.org/10.1016/j.neuroimage.2017.05.035
Received 11 March 2017; Accepted 16 May 2017

trait judgments. The current work explicitly asked two independent subject groups to respond to the questions whether they *believe* or *think* that an adjective can describe oneself (or a celebrity), respectively. *Believe* and *think* are regarded as the most similar mental processes in lay opinions (Allen et al., 1990) and thus can be mutually compared to disclose the distinct neurocognitive processes. To compare the brain activity from the two subject groups allowed us to discover the neurocognitive processes that distinguish *believe* from *think* and to identify the neural underpinnings of believing while controlling perceptual, cognitive, affective and motor processes. Participants also judged the valence (positive or negative) of trait adjectives as a low-level control condition to exclude influences of semantic processing and motor responses. Both brain activations and functional connectivity between specific brain regions were analyzed to examine the neurocognitive processes that distinguish between *believe* and *think*. According to the lay opinions of taking *believe* and *think* as similar mental processes (Allen et al., 1990), participants may be confused by the task demands of believing and thinking if being asked to alter frequently between *believe* and *think* and this would reduce the effect of task manipulations. Therefore, the present fMRI study employed a between-subject design to enhance the effect of task manipulation (*believe* vs. *think*) on the underlying brain activity.

Methods

Participants

Seventy-two participants were recruited in the present study (Believe group: N=36, mean age=22.19, SD=2.32 yrs, 18 males; Think group: N=36, mean age=22.44, SD=2.52 yrs, 18 males). The sample size was estimated based on the effect size in our previous work (Ma et al., 2014). All participants were undergraduate and graduate students and were paid for their participation. All self-reported to be unaffiliated with any religion. All were right-handed, had normal or corrected-to-normal vision, and reported no abnormal neurological history. Informed consent was obtained from all participants before scanning. This study was approved by the local ethics committee at the School of Psychological and Cognitive Sciences, Peking University.

Stimuli and procedure

We selected 288 trait adjectives from a personality trait adjective pool (Liu, 1990). Half of the trait adjectives were used in the self-referential task (old words) during fMRI scanning and half used as new words in the surprising memory test after scanning. Both old and new words consisted of 72 positive and 72 negative trait adjectives. Each trait adjective consisted of 2 Chinese characters. A block design was used in 2 functional scans. Each scan consisted of 6 blocks of 12 trials (half positive and half negative adjectives) and two successive blocks were separated with 8 s fixation, as illustrated in Fig. 1. On each trial a trait adjective, which subtended a visual angle of 2.72°×1.28°(width×height) at a viewing distance of 80 cm, was presented at the centre of the screen below a cue word (2.0°×1.0°) for 2250 ms followed by a fixation of 750 ms. During each scan participants performed trait judgments on the self, a gender-matched celebrity, or word valence in two blocks of trials. Different judgment tasks were denoted by the cue word (i.e., Self, a celebrity's name, or "Positive Negative") and performed in a random order. During trait judgments participants from the Believe Group were asked to make a yes or no response, by pressing one of two buttons using the right index and middle finger, to the question "Do you *believe* that the trait adjective describe you (or a celebrity)?". All aspects were the same for the Believe Group and the Think Group except that participants from the Think Group were asked to respond to the question "Do you *think* the trait adjective describe you (or a celebrity)?". During valence judgments participants from both subject groups were asked to identify whether a trait adjective is positive or negative.

Participants were asked to complete a 'surprising' memory test after fMRI scanning. The old trait adjectives used during fMRI scanning were intermixed with the new trait adjectives for the memory test that required identification of old vs. new items presented in a random order by pressing one of two buttons. Corrected recognition scores were calculated by subtracting the false alarm rate from the hit rate. All participants completed the Self-Construal Scale (Singelis, 1994) before fMRI scanning to control potential influences of self-esteem and cultural traits on their brain activities as the previous studies have shown cultural influences on brain activities underlying trait judgments (e.g., Ma et al., 2014).

fMRI data acquisition and analysis

Brain images were acquired using a 3.0 T Siemens scanner with a standard head coil. Functional images were acquired by using T2-weighted, gradient-echo, echo-planar imaging (EPI) sequences sensitive to BOLD contrast (64×64×32 matrix with 3.75×3.75×5 mm^3 spatial resolution, repetition time =2000 ms, echo time =30 ms, flip angle =90°, field of view =24×24 cm). A high-resolution T1-weighted structural image (256×256×144 matrix with a spatial resolution of 1×1×1.33 mm, TR=2530 ms, TE=3.37 ms, inversion time (TI) =1100 ms, FA=7°) was subsequently acquired.

Functional images were preprocessed using SPM8 (the Wellcome Trust Centre for Neuroimaging, London, UK). Head movements were corrected within each scan and six movement parameters (translation; x, y, z and rotation; pitch, roll, yaw) were extracted for further analysis in the statistical model. The functional images were resampled to 3×3×3 mm^3 voxels, normalized to the MNI space and then spatially smoothed using an isotropic of 8 mm full-width half-maximum (FWHM) Gaussian kernel. Fixed effect analyses were first conducted by applying a general linear model (GLM) to fMRI data. All four conditions (i.e., Self, Celebrity, Valence, and rest (i.e., the 8-s interval between two blocks of trials)) were included in the model. The design matrix also included the realignment parameters to account for any residual movement-related effect. A box-car function were used to convolve with the canonical hemodynamic response in each condition. The whole-brain random effect analyses were conducted to reveal brain regions that were involved in Self-judgments vs. Valence-judgments and Celebrity-judgments vs. Valence-judgments in the Believe and Think groups, respectively. Brain activations were defined using a threshold of cluster-level $p < 0.05$, FWE corrected.

We conducted the psychophysiological interaction analysis (PPI) (Friston et al., 1997) to identify brain regions that showed significantly increased covariation (i.e. increased functional connectivity) with the seed brain regions observed in the contrast of Self- vs. Valence-judgments (e.g., left AI/IFG and mPFC) and with the seed brain regions observed in the contrast of Celebrity- vs. Valence-judgments (e.g., left temporal pole/middle temporal cortex (TP/MTC) and mPFC). The coordinates of the peak voxel from the contrast images of Self- vs. Valence-judgments or Celebrity- vs. Valence-judgments of the Believe and Think groups were used to define the seed region for the PPI analyses. The region of interest (ROI) was defined in each participant as a sphere with 5-mm-radius centered at the peak voxel of the seed regions. The time series of each ROI were then extracted and the psychophysiological interaction regressor was calculated as the element-by-element product of the mean-corrected activity in the ROI and the vector coding for differential task effects of Self- vs. Valence-judgments or Celebrity- vs. Valence-judgments. The psychophysiological interaction regressors reflected the interaction between psychological variable (Self- vs. Valence-judgments or Celebrity- vs. Valence-judgments) and the activation time course of the seed regions. The individual contrast images reflecting the effects of the psychophysiological interaction between the seed brain regions and other brain areas were subsequently subjected to one-sample *t*-tests. The results of the