Shihui Han,[a,b,*] Yi Jiang,[a,1] Glyn W. Humphreys,[c] Tiangang Zhou,[d] and Peng Cai[d]

[a]Department of Psychology, Peking University, Beijing, PR China
[b]Learning and Cognition Laboratory, Capital Normal University, Beijing, PR China
[c]Behavioral Brain Sciences, School of Psychology, University of Birmingham, Birmingham, UK
[d]Key Laboratory of Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing, PR China

Received 4 August 2004; revised 23 September 2004; accepted 29 September 2004
Available online 18 November 2004

Keywords: fMRI; Neural substrates; Real world; Virtual visual worlds

Virtual reality is increasingly used for training in a wide range of contexts. For example, virtual human agents simulated using cartoons have been used to help students learn to perform physical, procedural tasks (Rickel and Johnson, 1999). Animated agents in virtual environments have also been used for training skills that require a high level of flexible, interpersonal interactions such as psychotherapy (Beutler and Harwood, 2004). However, whether the human brain differentially perceives and interacts with agents in the real and virtual worlds has been poorly understood. Recent functional magnetic resonance imaging (fMRI) studies have shown that, when we deal with actions assumed to come from real human agents, specific brain regions, such as the medial prefrontal cortex (MPFC), show stronger activation compared with when we assume the actions come from animated agents simulated by computers (Gallagher et al., 2002; Ramnani and Miall, 2004), suggesting that specific neural substrates may be involved in discrimination between human and non-human agents.

The current study assessed whether, when we simply perceive human agents in the real world, different brain regions are engaged compared with when we perceive agents in virtual worlds. To investigate this, we used fMRI to measure brain activations when participants observed movie and cartoon clips, which presented brief sequences of actions involving humans in real-life situations (movie clips) or actions involving either human or non-human agents in virtual worlds (cartoon clips). Movies present real images (photographs of a physical environment) whereas cartoons present virtual images (a simulation on physical principles of that environment). Brain activity when watching the clips was compared with random order static images from the movie and cartoon clips to control for any differences in low level visual feature processing. Relative to the baseline with static random images, movie or cartoon clips presented continuous and coherent visual events that induced explanatory predictions of behaviour. We aimed to identify if there are neural substrates differentiating the perception of human agents in the real visual world (in movie clips) from the perception of human or non-human characters in a virtual world (in cartoons).

Subjects

Twelve adults (6 male; 21–41 years of age, mean 25.5) with no neurological or psychiatric history participated in this study. All

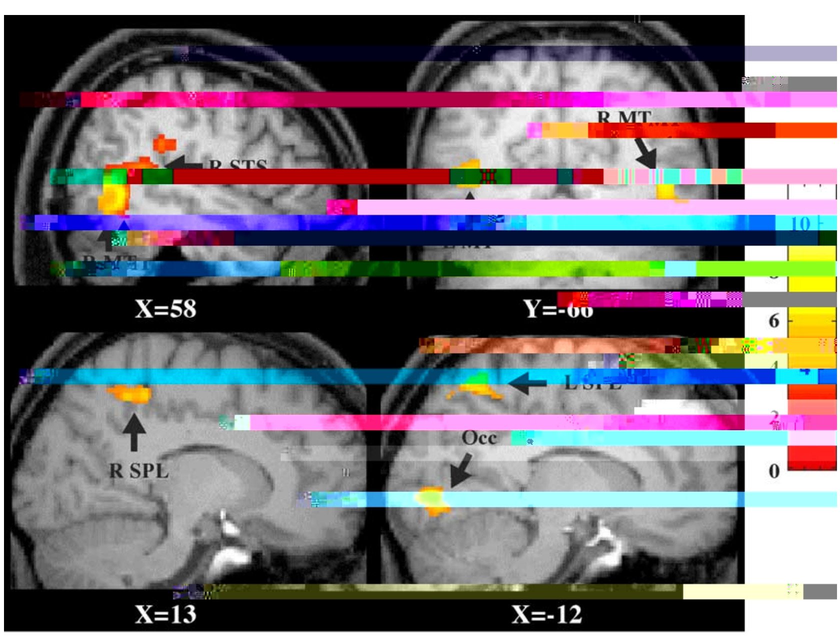
1053-8119/$ - see front matter
doi:10.1016/j.neuroimage.2004.09.046

participants were right-handed, had normal or corrected-to-normal vision, and were not colour blind. Informed consent was obtained from all participants prior to scanning. This study was approved by the Academic Committee of Department of Psychology, Peking University.

Stimuli and procedure

The stimuli were presented through a LCD projector onto a rear-projection screen located at a subject's head. The screen was viewed with an angled mirror positioned on the head-coil. The stimuli were movie and cartoon clips without accompanying sound and static images extracted from them. As illustrated in Fig. 1, one set of movies (Movie I, Fig. 1a) and one set of cartoons (Cartoon II, Fig. 1c) depicted human beings in real visual scenes. There were two clips of Movie I. One clip showed human activities at a subway station (meeting and walking away). Another clip showed students' activities in a classroom (raising hand and talking). There were two clips of Cartoon II, which also showed human activities at a subway station and students' activities in a classroom, similar to those of clips of Movie I. There were two clips of Cartoon I (Fig. 1b) depicting artificial characters in virtual visual scenes (e.g., robots or machine dinosaurs walking and fighting with transformed shapes) and two clips of Movie II (Fig. 1d) depicting animals in real visual scenes (e.g., gorillas walking and playing in jungles). Both the movies and cartoons were made by showing 29 frames of images per second. At a viewing distance of 70 cm, Movie I, Movie II, Cartoon I, and Cartoon II subtended visual angles of $28 \times 16^\circ$ (width × height), $27 \times 17^\circ$, $27 \times 20^\circ$, and $30 \times 16^\circ$, respectively. The sizes of static images were the same as the corresponding movies and cartoons. Four scans of 420 s were first obtained from each subject. Each scan consisted of six 1-min epochs, alternating pseudo-randomly between movie clips, cartoon clips, and the corresponding static images across subjects. There was a 10-s black screen before each epoch to set up a baseline for each epoch of stimuli. Thirty images were extracted at every 2 s from the corresponding movie and cartoon clips and were presented (each with a duration of 2 s) in a random order during the epochs of static images. Fig. 2

the corresponding static images were upright in two scans whereas



R MT
R SPL
Occ
X=58
X=13
X=-12
10
6
0

such as motion, shape, and colour were identical in the upright and inverted stimuli, we would expect effects associated with low-level features to be the same in Conditions C and A. The contrast of inverted movies–random static images showed bilateral activation of MT and the posterior STS (centred at −50, −66, 7, $Z = 4.90$, $P < 0.001$, corrected, and 55, −62, 8, $Z = 4.62$, $P < 0.001$, corrected), the occipital cortex (centred at −22, −93, 8, $Z = 4.11$, $P < 0.001$, corrected), and the left fusiform gyrus (centred at and 39, −60, −13, $Z = 4.57$, $P < 0.02$, corrected, see Fig. 5). Now there was also bilateral activation of the SPL (centred at −10, −55, 60, $Z = 4.11$, $P < 0.01$, corrected, and 28, −61, 49, $Z = 3.92$, $P < 0.01$, corrected), similar to that associated with upright cartoons. However, no activation in the MPFC or the cerebellum was observed in Condition C, indicating that activations observed in these areas in Condition A could not be attributed to specific low-level feature variations in upright movies. The increased activation of the fusiform gyrus for inverted movies may be due to increased attention to the visual features of the scenes or the difficulty of recognising inverted faces, so that extra neural activation is required (though see Kanwisher et al., 1998).

In Condition D, we assessed whether perception of human agents in a virtual visual world activated similar brain regions associated with the perception of human agents in the real visual world. Subjects viewed cartoon clips of human agents in visual scenes similar to those in the movies (Fig. 1c, Cartoon II). Performance was again contrasted relative to static images extracted from the clips. The contrast of human cartoons–random static images showed bilateral activation in MT and the posterior STS (centred at −51, −66, 7, $Z = 4.66$, $P < 0.001$, corrected, and 55, −60, 8, $Z = 5.23$, $P < 0.001$, corrected), the STS (centred at −63, −43, 15, $Z = 4.85$, $P < 0.001$, corrected, and 64, −38, 19, $Z = 5.00$, $P < 0.001$, corrected), and the SPL (centred at −4, −50, 52, $Z = 3.18$, $P < 0.04$, corrected, and 20, −49, 61, $Z = 4.35$, $P < 0.03$, corrected, see Fig. 6), similar to those observed for the cartoon clips of artificial characters. However, there was no activation in the MPFC or the cerebellum. The results indicate that SPL is recruited when coherent actions are viewed in cartoons, regardless of whether the cartoons depict artificial characters or human agents; in contrast, the MPFC and the cerebellum were not automatically engaged in perception of virtual reality. A further region-of-interest analysis confirmed that there was stronger MPFC activation associated with movies clips of human beings relative to cartoon clips of human or non-human agents ($t = 6.00$ and 4.75, respectively, $P < 0.001$), whereas no differential activation in the MPFC was observed between cartoon clips of human and non-human agents.

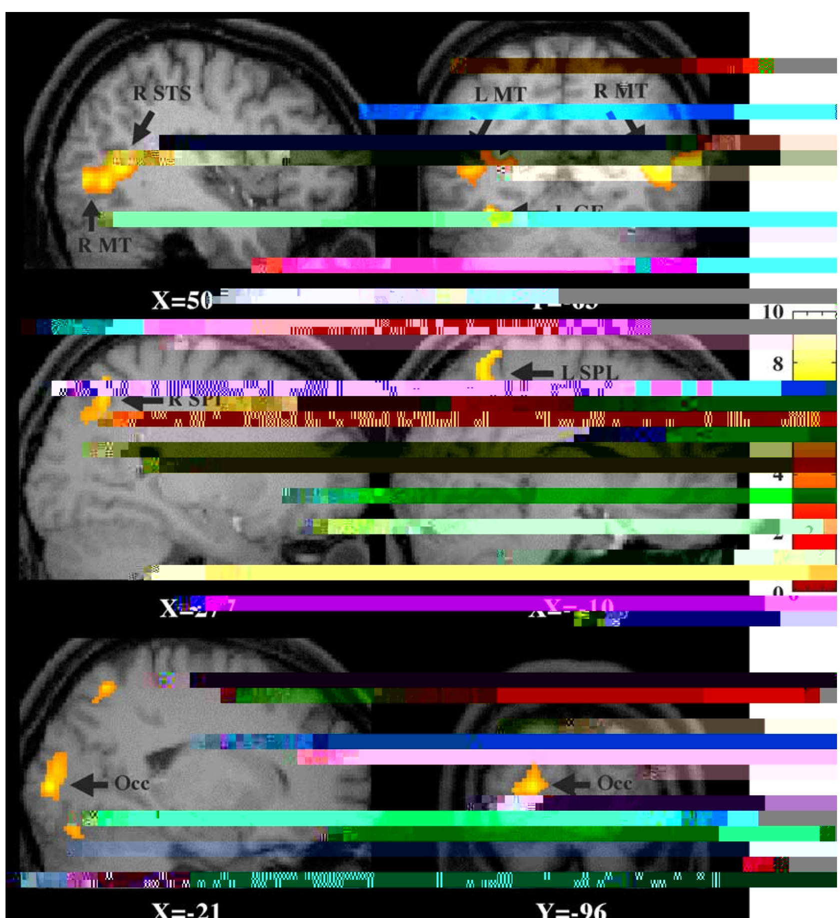


Fig. 5. Brain activations associated with viewing inverted movies of humans. Activations were observed in bilateral MT and the posterior STS, the left occipital

To assess if the MPFC and cerebellum activities are specifically involved in the perception of other humans but not animals in the real world, in Condition E, we had the same group of subjects watch upright movie clips that depicted gorillas walking and playing in jungles or the corresponding static images presented in a random order (Fig. 1d). The contrast between the movie clips of gorillas and static images showed bilateral activation in MT and the posterior STS (centred at −51, −66, 3, Z = 5.70, P < 0.001, corrected, and 51, −73, 2, Z = 5.39, P < 0.001, corrected), the medial occipital cortex (centred at −2, −91, −2, Z = 5.74, P < 0.001, corrected), and the SPL (centred at −14, −53, 61, Z = 4.31, P<

to their static image baselines. The medial occipital cortex and MT are likely engaged by the processing of low-level visual features of the moving images, such as changes in the shape, colour (Livingstone and Hubel, 1998), and motion direction (Tootell et al., 1995) of the stimuli, and the posterior STS is involved in the processing of biological motion (Puce et al., 1998) even when inverted images were perceived. The fMRI results suggest that the processing of low-level visual features of the real and virtual visual worlds is mediated by similar neural mechanisms in the posterior brain areas.

However, distinct neural substrates at higher level brain structures are involved when we view agents in coherent events in the real and virtual visual worlds. Most important, there were distinct brain regions activated by movie clips of human beings, on the one hand, and movie clips of animals and cartoon clips of human and non-human agents, on the other. The MPFC and the left cerebellum showed enhanced activity when adults watched movies depicting human activities in everyday-life situations. Our preliminary separate analysis of the fMRI data associated with the movie clips showed MPFC activation for both clips regardless of the difference in contents between them. Nevertheless, none of the clips of Cartoon II showed MPFC activation even though Cartoon II showed similar scenes as those of Movie I. Thus, it appears that the way to present the visual scenes rather than the content is critical for the MPFC and cerebellum activations. The MPFC and cerebellum activations could not result from summed amounts of motions in the movies because motion patterns were the same in the upright and inverted movies whereas the MPFC and cerebellum showed enhanced activities only to the upright movies. The movie clips of gorillas were also different from cartoon clips in motion patterns but did not induce the MPFC and cerebellum activations. It is unlikely that the effects are simply due to the perception of social relations between the characters (Iacoboni et al., 2004) or to the presence of coherent actions in the movie condition but not in its baseline (with random, static images). There were social relations between the individuals and coherent actions in the cartoons of people, similar to those in the movie clips, but the MPFC did not show enhanced activity. The effects cannot be attributed to simulation of human agents' actions with a third person perspective, which activated different brain regions (Ruby and Decety, 2001). It is also unlikely that the effects were caused by the stimuli being differentially interesting. We asked the participants to rate the movie and cartoon clips for interest, and found no differences in the rated interest for the critical conditions: movies of humans, cartoons of non-humans, cartoons of human characters, and movies of animals [the mean ratings were 3, 3.25, 2.88, 4, respectively, on a 5-point scale where 5 = very interesting, $F(1,11) = 0.25$, $P > 0.5$]. The cerebellum activity could be argued to arise from enhanced working memory when viewing a familiar scene (Desmond and Fiez, 1998), but there is no reason to expect this only for the movies of people compared with the matched cartoons of humans.

Given that the MPFC and cerebellum activations were eliminated when viewing animals in real visual scenes, it may be proposed that these areas are possibly engaged in 'mentalising' about other humans or forming a theory of mind of other humans (Frith and Frith, 1999; Premack and Woodruff, 1978; Wellman, 1990), consistent with previous neuroimaging findings (Brunet et al., 2000; Fletcher et al., 1995; Gallagher et al., 2000), and brain lesion results (Stone et al., 1998; Stuss et al., 2001). It is possible that aspects of mental state reasoning are engaged only when adults view other people in real-life situations (in movies), whereas this might not occur when viewing both human and non-human agents in virtual scenes (in cartoons). The MPFC and cerebellar regions were also not differentially activated when movies of people were inverted, presumably because the difficulty of encoding the people and objects in such scenes (Farah et al., 1995; Rock, 1986) impaired mental state reasoning. Prior studies have shown that the MPFC and associated regions can be activated when participants are required to carry out mental state reasoning on static cartoon pictures (Gallagher et al., 2000) and even moving shapes (Casteli et al., 2000), suggesting that 'mentalising' associated with the MPFC can be conducted on agents in cartoons (and with non-human agents) when explicitly required. However, the contrast in our data between activation states when viewing humans in everyday behaviours and when viewing the other stimuli indicates that the MPFC is not spontaneously engaged in mental state reasoning unless we observe other humans in the real visual world.

Watching characters in cartoons and animals in movies generated distinct patterns of activation bilaterally in the SPLs. Previous neuroimaging studies have shown the SPLs are engaged in viewing static images with social cues. It is possible that mental state reasoning from text stories involved the posterior part of the brain whereas mental state reasoning when viewing other people's behaviour in the real visual world depends mainly upon the involvement of the MPFC. This, however, needs to be clarified in thework.

The present study extends in important ways the prior brain imaging studies contrasting effects of actions assigned to humans or to computers (Gallagher et al., 2002; Ramnani and Miall, 2004). We show engagement of a neural loop including the MPFC and the cerebellum even when we simply view humans in real scenes, but not when active agents are seen in virtual environments. Hence, the perception of real and virtual visual worlds can be mediated with distinct neural substrates. The distinct neural and functional processes engaged in perceiving agents in real and virtual worlds may contribute to the limited efficacy of virtual visual environments for training (Seidel and Chatelier, 1999). Our findings suggest that human brains possibly when we interact with
real people in everyday life, in a way different from when we view or interact with artificial characters or static social stimuli. It may be further speculated that the brain may not attempt to model the behaviour of cartoon characters as it does with real people, which possibly constrains social impact of cartoons on adult viewers.

This work was supported by National Natural Science Foundation of China (Project 30225026 and 30328016), the Ministry of Science and Technology of China (Project 2002CCA01000), the Ministry of Education of China (02170), the Medical Research Council (UK),and the Royal Society (UK).

Beutler, L.E., Harwood, T.M., 2004. Virtual reality in psychotherapy training. J. Clin. Psychol. 60, 317–330.

Brunet, E., Sarfati, Y., Hardy-Baylé, M., Decety, J., 2000. A PET investigation of the attribution of intentions with a nonverbal task. NeuroImage 11, 157–166.

Casteli, F., Happé, F., Frith, U., Frith, C., 2000. Movement and mind:a unctional imaging study of perception and interpretationof complex intentional movement patterns. NeuroImage 12, 314–325.

Desmond, J.E., Fiez, J.A., 1998. Neuroimaging studies of the cerebellum: language, learning and memory. Trends Cogn. Sci. 2, 355–362.