Cultural differences in the lateral occipital complex while viewing incongruent scenes

Lucas J. Jenkins,1 Yung-Jui Yang,2 Joshua Goh,3 Ying-Yi Hong,2,4 and Denise C. Park5

1Department of Psychology, University of California, Davis, CA 95616, 2Department of Psychology, 3Beckman Institute, University of Illinois, Urbana-Champaign, IL 61801, USA, 4Nanyang Business School, Nanyang Technological University, Singapore 639798, and 5School of Behavioral and Brain Sciences, University of Texas, Dallas, TX 75235, USA

Converging behavioral and neuroimaging evidence indicates that culture influences the processing of complex visual scenes. Whereas Westerners focus on central objects and tend to ignore context, East Asians process scenes more holistically, attending to the context in which objects are embedded. We investigated cultural differences in contextual processing by manipulating the congruence of visual scenes presented in an fMR-adaptation paradigm. We hypothesized that East Asians would show greater adaptation to incongruent scenes, consistent with their tendency to process contextual relationships more extensively than Westerners. Sixteen Americans and 16 native Chinese were scanned while viewing sets of pictures consisting of a focal object superimposed upon a background scene. In half of the pictures objects were paired with congruent backgrounds, and in the other half objects were paired with incongruent backgrounds. We found that within both the right and left lateral occipital complexes, Chinese participants showed significantly greater adaptation to incongruent scenes than to congruent scenes relative to American participants. These results suggest that Chinese were more sensitive to contextual incongruity than were Americans and that they reacted to incongruent object/background pairings by focusing greater attention on the object.

Keywords: culture; scene perception; context; incongruence; lateral occipital cortex

INTRODUCTION

A growing behavioral literature suggests that culture can influence the way in which we perceive the visual world. Westerners tend to engage in an analytical style of processing marked by a focus on salient objects independent of the context in which they are embedded. In contrast, East Asians process visual information in a more holistic fashion, attending to the relationship between object and context (Li et al., 2000; Nisbett et al., 2001; Norenzayan et al., 2002; Kitayama et al., 2003; Nisbett and Masuda, 2003; Nisbett and Miyamoto, 2005; Masuda et al., 2008). For example, Westerners fixate sooner and longer on the focal objects in a scene, whereas East Asians fixate more often on the background (Chua et al., 2005). In a change detection paradigm, East Asians were more likely than Westerners to notice changes to the background, but Westerners were more likely than East Asians to notice changes to objects (Masuda and Nisbett, 2006). East Asians recognize previously seen objects better when they are presented with their original backgrounds than when they are presented with novel backgrounds, whereas Westerners’ object recognition is less affected by background manipulation (Masuda and Nisbett, 2001; Chua et al., 2005).

These cultural biases in scene perception are supported by neuroimaging evidence as well (Gutchess et al., 2006; Goh et al., 2007; Han and Northoff, 2008; Hedden et al., 2008). For example, Goh et al. (2007) investigated cultural differences in object and background processing using an fMR-adaptation (fMR-A) paradigm. In fMR-A experiments, regions of the brain that show a decrease in blood oxygen level dependent (BOLD) response associated with the repeated presentation of a particular stimulus are assumed to be involved in processing that stimulus (Grill-Spector and Malach, 2001; Grill-Spector et al., 2006). Goh et al. (2007) presented Chinese and American participants with sets of four pictures in which either the central object or the background varied from one picture to the next. During trials in which the object changed and the background remained the same, both groups showed BOLD signal adaptation in a bilateral region corresponding to the parahippocampal place area (PPA), suggesting that this region is involved in background processing (Epstein and Kanwisher, 1998). During trials in which the background changed while the object remained the same, both groups showed adaptation in the lateral occipital complex (LOC) bilaterally, suggesting a role in object processing (Malach et al., 1995; Grill-Spector et al., 2001). Critically, Americans showed significantly more adaptation than did Chinese in the right LOC, although the cultural effect was evident only in elderly participants. This finding is consistent with a Western cultural bias towards the processing of objects, and suggests a possible interactive effect with age.
In the present fMRI experiment, we modified the adaptation paradigm used by Goh et al. (2007) (see also Goh et al., 2004; Chee et al., 2006) by manipulating the congruence of the picture quartets presented. Young Chinese and American participants passively viewed sets of novel or repeated scenes in which the central object was either congruent with the background (e.g. a deer in the woods) or incongruent with the background (e.g. a television in the desert) (Figure 1). As in Goh et al. (2007), we expected to see BOLD signal adaptation for repeated vs novel scenes in both the LOC and PPA, reflecting the roles of those regions in object and background processing, respectively. Moreover, we expected the magnitude of adaptation in these regions to vary as a function of culture and scene congruence, consistent with previous work demonstrating attentional modulation of adaptation effects in the ventral visual processing stream (Eger et al., 2004; Murray and Wojciulik, 2004; Yi and Chun, 2005, Yi et al., 2006; Chee et al., 2006; Chee and Tan, 2007). We hypothesized that the Chinese participants would show greater adaptation to incongruent than to congruent scenes, reflecting a cultural bias to attend to and elaborate upon contextual relationships. In contrast, Americans were expected to be less sensitive to contextual incongruence, focusing on the objects independently of the context in which they were embedded. We hypothesized that American participants would show less of a difference between incongruent and congruent adaptation, or none at all.

**Methods**

**Participants**

Sixteen Chinese (age: $M = 26.3$, SD = 4.1; eight females) and 16 age- and gender-matched American participants (age: $M = 26.1$, SD = 3.0; eight females) were recruited in compliance with the human subjects protocols of the University of Illinois. Chinese participants spoke Mandarin as their primary language, and none had lived in the United States for longer than 1.5 years. American participants were native English speakers of non-Asian heritage. All participants were right-handed and had normal or corrected-to-normal vision.

**Procedure**

During the main fMRI experiment, participants passively viewed sets of full-color scenes consisting of an object superimposed on either a congruent or an incongruent background (Figure 1). They were instructed to attend to the scenes but were given no other information about the task paradigm. Incongruent object/background combinations were designed to appear improbable (e.g. a bear in a supermarket) but not physically impossible (e.g. a bear in a teacup). Scenes were presented in sets of four and belonged to four different experimental conditions: novel-congruent, consisting of four novel scenes with congruent object and background; repeated-congruent, consisting of a single repeated scene with congruent object and background; novel-incongruent, consisting of four novel incongruent scenes; repeated-incongruent, consisting of a single repeated scene with congruent object and background.

![Fig. 1](https://example.com/fig1.png) Picture stimuli and presentation sequence. The four conditions were novel-congruent: four novel congruent scenes; repeated-congruent: one repeated congruent scene; novel-incongruent: four novel incongruent scenes; repeated-incongruent: one repeated incongruent scene. Scene duration was 1.5 s, separated by an interval of 250 ms and a mean intertrial interval of 9 s.
novel-incongruent, consisting of four novel scenes with incongruent object and background; and repeated-incongruent, consisting of a single repeated scene with incongruent object and background. Each scene within a quartet was presented for 1.5 s and separated from the next by an interval of 250 ms, for a total presentation time of 6.75 s. Quartets were separated by a variable intertrial interval of 5.25, 8.25 or 11.25 s. A total of 80 quartets were presented in random order across four 6 min 22 s runs.

Prior to the four event-related runs, participants were scanned during two 5 min localizer runs in which they passively viewed alternating 15 s blocks of objects, scenes and phase-scrambled control images. The purpose of these scans was to identify object- and scene-selective regions of interest in the ventral visual cortex for time course analysis.

After completing the scanning session, participants were taken to a separate testing room, where they rated each of the 200 scenes they had just seen on a 5-point scale, ranging from very normal to very abnormal. This was done in order to rule out the possibility that one cultural group would perceive certain pictures to be inherently more incongruent than the other. No between-group differences were observed for 157 of the 200 pictures. Differences for the remaining pictures ($P<0.05$, uncorrected) were due in each case to the Americans’ greater tendency to utilize the endpoints of the scale—rating congruent pictures more ‘normal’ than Chinese participants and rating incongruent pictures more ‘abnormal’. Although this response bias is typical in cross-cultural studies (Chun et al., 1974; Chen et al., 1995), we note that even if this finding represents a valid psychological difference and not merely a measurement artifact, a tendency in American participants to see a greater contrast between congruent and incongruent scenes would only serve to weaken the predicted effect.

All written forms and instructions were provided in Mandarin translation for the benefit of the Chinese participants, and a Mandarin-speaking experimenter was present during the entire session to ensure that oral instructions were communicated clearly.

**Image acquisition**

One-hundred eleven functional scans were acquired in each of the four runs of the experimental task, along with 95 in each of the localizer runs, using a gradient-echo EPI sequence with the following parameters: TR = 3 s, FOV = 19.2 $\times$ 19.2 cm, matrix size = 64 $\times$ 64. Thirty-six 3 mm axial slices (0.3 mm gap) were acquired parallel to the AC-PC line. Coplanar T2 and high-resolution MPRAGE anatomical images were also acquired to aid image coregistration. Stimuli were projected on a screen at the back of the magnet and visible to participants through a mirror mounted on the RF coil.

**Data preprocessing and analysis**

All fMRI preprocessing and statistical analyses were conducted using SPM5 (http://www.fil.ion.ucl.ac.uk/spm/) customized with in-house scripts. Functional images were slice time corrected, realigned to the mean image to correct for motion artifacts, and normalized to a common template (Montreal Neurological Institute) using subject-specific drop-out masks to reduce image distortion. The normalized images were then resampled to a 3 mm isomorphic voxel resolution, and smoothed with a Gaussian kernel of 8 mm FWHM.

We first analyzed the localizer data, using a general linear model (GLM) including separate regressors for object, scene and scrambled image blocks. We identified for each participant the coordinates in both left and right hemispheres maximally responsive to object stimuli (object–scene contrast) and scene stimuli (scene–object contrast) and constructed $3 \times 3 \times 3$ voxel regions of interest around those points. In order to ensure that object-sensitive ROI corresponded anatomically to the LOC (Malach et al., 1995), they were selected from an anatomical mask that included the inferior occipital gyrus, the fusiform gyrus and the posterior inferior temporal gyrus. Likewise, scene-sensitive ROI were constrained to fall in or near the PPA (Epstein and Kanwisher, 1998) by selecting them from a region of the posterior parahippocampal gyrus bounded by $\{-35 < x < 15$, left; $15 < x < 35$, right; $-60 < y < -35$; $-15 < z < 2\}$.

Activity during the experimental task was modeled using a GLM with eight finite impulse response regressors for each of the four conditions (novel-congruent, repeated-congruent, novel-incongruent, repeated-incongruent), estimating BOLD response for eight scans, or 21 s, after stimulus onset. Time courses for each condition were extracted from the bilateral LOC and PPA regions of interest (Supplementary Figures 3 and 4). Rather than conduct statistical tests over the entire time course, we restricted our analyses to the point of peak BOLD activity and the point directly following (peak + 1).\(^1\) Peaks were defined as the point of maximum BOLD response in any task condition, averaged across participants. In the LOC regions of interest, these were time points 4 and 5 (9 and 12 s after stimulus onset). In the PPA regions of interest, they were points 3 and 4 (6 and 9 s after stimulus onset). Adaptation magnitude was calculated at these time points for both the congruent and incongruent conditions by subtracting repeated activity from novel activity (Figure 2a and b; Supplementary Tables 1 and 2). One participant from the Chinese group was discarded because of adaptation values that were over 3 SD from the overall mean.

\(^1\)Although adaptation studies frequently focus on activity at the hemodynamic peak, it was unclear whether this was appropriate for our paradigm. Our stimulus offset occurred at 6.75 s. Given a normal hemodynamic lag of $\sim 6$ s, we might reasonably expect to see the greatest effect at our 12 s time point. However, peak LOC activity was observed at 9 s. Therefore, we analyzed BOLD signal magnitude at both peak and peak + 1 in order to capture late-occurring adaptation effects.
Following our hypothesis that Chinese participants would show greater sensitivity to the congruence manipulation than would Americans, a 2 (congruence) × 2 (culture) mixed ANOVA was conducted on the adaptation values at each time point. Time points that showed a significant congruence × culture interaction were subjected to additional planned comparisons to further explore the effect.

RESULTS

Object processing: LOC

We began by analyzing the BOLD signal time courses extracted from the bilateral LOC regions of interest (Supplementary Figure 3). As expected, typical adaptation effects were observed in both groups such that BOLD signal response was significantly lower to Repeated than to Novel scenes at both of the a priori timepoints (2.15 < t < 6.39, 0.001 < P < 0.02). In addition, adaptation values in both the left and the right LOC showed a significant congruence × culture interaction at the peak +1 timepoint (Figure 2a and b) (left, F_{1,29} = 13.77, P = 0.001; right, F_{1,29} = 15.29, P = 0.001). Planned comparisons showed that in the left LOC, this effect was due to significantly greater adaptation to incongruent than to congruent scenes in Chinese participants (t_{14} = 4.59, P < 0.001) but not in American participants (t_{15} = 0.96, n.s.). A similar pattern was observed in the right LOC, but whereas the Chinese participants showed greater adaptation to incongruent than to congruent scenes (t_{14} = 3.44, P = 0.004), American participants showed significantly less adaptation to incongruent than to congruent scenes (t_{15} = 2.37, P = 0.03).

Background processing: PPA

Although the bilateral PPA regions of interest, like the LOC regions of interest, showed the typical BOLD signal attenuation for repeated vs novel scenes at both a priori time points (3.86 < t < 8.86, 0.001 < P < 0.002; Supplementary Figure 4), adaptation in the bilateral PPA showed no evidence of a congruence × culture interaction. However, a main effect of congruence was found at both a priori time points in the right PPA, with adaptation to congruent scenes significantly greater than adaptation to incongruent scenes (6 s, F_{1,29} = 9.84, P = 0.004; 9 s, F_{1,29} = 4.63, P = 0.04).

DISCUSSION

Consistent with behavioral research suggesting that people from East Asian cultures attend to and process visual context more extensively than do Westerners, we report evidence for greater object processing in Chinese participants when the objects appear in incongruent contexts than when they appear in congruent contexts. Object processing in our paradigm was indexed by BOLD signal adaptation in the LOC bilaterally, which was significantly greater in response to repeated incongruent than to repeated congruent scenes. The same effect was not demonstrated in Americans. We interpret these results to suggest that Chinese participants, sensitive to the context in which an object is embedded, focus greater attention on that object when the context is semantically inconsistent. American participants, in contrast, are less likely to bias attention to an object that violates normal semantic relationships.

Our results correspond with a number of eye-tracking studies that report longer fixation on objects that are inconsistent with a scene than on consistent objects (Loftus and Mackworth, 1978; Friedman, 1979; De Graef et al., 1990; Henderson et al., 1999; for review, see Henderson and Hollingworth, 1999). For example, Henderson et al. (1999) presented participants with line drawings of complex scenes containing either a congruent or an incongruent target object in preparation for a recognition test. They found that incongruent objects were fixated more often than congruent objects during the course of viewing and that the durations of the initial fixations were significantly longer for incongruent objects. The results of these studies—all
conducted with Western participants—indicate that the cultural differences observed in our fMRI experiment may simply be a matter of degree and should not be interpreted as proof for qualitatively different cognitive processes. Westerners clearly do attend to incongruent objects to some extent in similar task paradigms. Discrepancies between the fMRI and eye-tracking data are likely due to general methodological differences or to specific differences in the experimental tasks (e.g., active vs. passive viewing, line drawings vs. photographs). We therefore interpret with caution the finding that American participants showed less adaptation to incongruent than to congruent scenes in the right LOC. Although it does suggest that the Americans were sensitive in some fashion to the congruence manipulation, the direction of the effect was unexpected and difficult to explain within the context of the experimental design. Future work integrating eye-tracking and fMRI measures may address this issue more directly.

It is not clear why the cultural differences observed were limited to object processing in the LOC and did not extend to our scene-sensitive PPA regions of interest. Both groups exhibited less adaptation to incongruent than to congruent scenes in the right PPA, which suggests that improbable objects actually diverted attention away from the scene background. This is consistent with the eye-tracking literature mentioned above, as well as with work by Chee et al. (2006), who showed using the same paradigm that instructing participants to attend to the focal object in a scene resulted in decreased PPA adaptation. Since gist information can be extracted very rapidly from the background of a scene—as quickly as 80 ms by one estimate (Davenport and Potter, 2004; see also Biederman et al., 1982; Bar, 2004), it is possible that any cultural differences were far too subtle to be detected with the limited temporal resolution of fMRI.

To our knowledge, only two previous neuroimaging studies have investigated cultural differences in visual scene processing. Gutches et al. (2006) scanned Chinese and American participants while viewing complex scenes and found greater activation for Americans relative to Chinese in a network of brain regions related to object processing, including the bilateral middle temporal gyrus, right superior temporal gyrus and left superior parietal. As already mentioned, Goh et al. (2007) reported greater object-related adaptation in the LOC for American elderly participants than for Chinese elderly participants while viewing repeated scenes, suggesting that aging is associated with greater reduction in object processing for Chinese than for Americans. Both studies provide evidence for a greater focus on object processing among Westerners. The results of the present study complement and extend these earlier findings by demonstrating that object processing among East Asians is instead dependent on the relationship between the object and the context in which it is embedded. This constitutes evidence for cultural differences at a relatively high level of visual processing, involving access to and utilization of stored semantic representations. It reflects findings from the behavioral literature that East Asians tend to group objects based on their functional or contextual relationships rather than on category membership (Chiu, 1972; Ji et al., 2004), and suggests that the same biases may influence the way in which they attend to the visual world. Taken together, these studies show that prolonged experience within a given cultural environment may influence the neural processes underpinning visual perception.

SUPPLEMENTARY DATA
Supplementary data are available at SCAN online.

FUNDING
National Institute on Aging Grant RO1 AG015047 to Denise Park.

REFERENCES


