Cultural variation in eye movements during scene perception

Hannah Faye Chua, Julie E. Boland, and Richard E. Nisbett*

Department of Psychology, University of Michigan, 530 Church Street, Ann Arbor, MI 48109-1043

In the past decade, cultural differences in perceptual judgment and memory have been observed: Westerners attend more to focal objects, whereas East Asians attend more to contextual information. However, the underlying mechanisms for the apparent differences in cognitive processing styles have not been known. In the present study, we examined the possibility that the cultural differences arise from culturally different viewing patterns when confronted with a naturalistic scene. We measured the eye movements of American and Chinese participants while they viewed photographs with a focal object on a complex background. In fact, the Americans fixated more on focal objects than did the Chinese, and the Americans tended to look at the focal object more quickly. In addition, the Chinese made more saccades to the background than did the Americans. Thus, it appears that differences in judgment and memory may have their origins in differences in what is actually attended as people view a scene.

attention | culture | memory | eye-tracking | visual cognition

A growing literature suggests that people from different cultures have differing cognitive processing styles (1, 2). Westerners, in particular North Americans, tend to be more analytic than East Asians. That is, North Americans attend to focal objects more than do East Asians, analyzing their attributes and assigning them to categories. In contrast, East Asians have been held to be more holistic than Westerners and are more likely to attend to contextual information and make judgments based on relationships and similarities.

Causal attributions for events reflect these differences in analytic vs. holistic thought. For example, Westerners tend to explain events in terms that refer primarily or entirely to salient objects (including people), whereas East Asians are more inclined to explain events in terms of contextual factors (3–5). There also are differences in performance on perceptual judgment and memory tasks (6–8). For example, Masuda and Nisbett (6) asked participants to report what they saw in underwater scenes. Americans emphasized focal objects, that is, large, brightly colored, rapidly moving objects. Japanese reported 60% more information about the background (e.g., rocks, color of water, small nonmoving objects) than did Americans. After viewing scenes containing a single animal against a realistic background, Japanese and American participants were asked to make old/new recognition judgments for animals in a new series of pictures. Sometimes the focal animal was shown against the original background; other times the focal animal was shown against a new background. Japanese and Americans were equally accurate in detecting the focal animal when it was presented in its original background. However, Americans were more accurate than East Asians when the animal was displayed against a new background. A plausible interpretation is that, compared with Americans, the Japanese encoded the scenes more holistically, binding information about the objects with the backgrounds, so that the unfamiliar new background adversely affected the retrieval of the familiar animal.

The difference in attending to objects vs. context also was shown in a perceptual judgment task, the Rod and Frame test (7). American and Chinese participants looked down a long box. At the end of the box was a rod whose orientation could be changed and a frame around the rod that could be moved independently of the rod. The participants' task was to judge when the rod was vertical. Chinese participants' judgments of verticality were more dependent on the context, in that their judgments were more influenced by the position of the frame than were those of American participants. In a change blindness study, Masuda and Nisbett asked American and Japanese participants to view a sequence of still photos and also to view animated vignettes of complex visual scenes (unpublished data). Changes in focal object information (e.g., color and shape of foregrounded objects) and contextual information (e.g., location of background details) were introduced during the sequence of presentations. Overall, the Japanese reported more changes in the contextual details than did the Americans, whereas the Americans reported more changes in the focal objects than did the Japanese. This finding has at least two possible explanations (see ref. 9). On one account, the Asian participants had more detailed mental representations of the backgrounds, whereas the Westerners had more detailed representations of the focal objects. On the other account, the mental representations did not differ with culture, but the two groups differed in their accuracy for detecting a deviation between their mental representation of the background/focal object and the current stimulus.

Clearly, there were systematic differences between the Americans' and the East Asians' performance in the causal perception, memory, and judgment studies. However, it is unclear whether the effects occur at the level of encoding, retrieval, mental comparison, or differences in reporting bias. To identify the stages in perceptual–cognitive processing at which the cultural differences might arise, consider what is known about scene perception: (*i*) Within 100 ms of first viewing a scene, people can often encode the gist of the scene, e.g., "picnic" or "building" (10). (*ii*) People then construct a mental model of the scene in working memory (11). The mental representation is not an exact rendering of the original scene and is usually incomplete in detail (12–13). (*iii*) Although the initial eye fixation may not be related to the configuration of the scene, the following fixations are to the most informative regions of the scene for the task at hand (14). The fixation positions are important because foveated regions are likely to be encoded in greater detail than peripheral regions (15). (*iv*) The mental representation of the scene is then transferred to and consolidated in long-term memory. (*v*) Successful retrieval from long-term memory relies on appropriate retrieval cues. (*vi*) During retrieval, the recalled information may be filtered by experimental demands and cultural expectations. Past studies (3–8) have failed to establish whether the effects are due to differences in perception, encoding, consolidation, recall, comparison judgments, or reporting bias.

To address this issue, we monitored eye movements of the American and the Chinese participants while they viewed scenes containing objects on relatively complex backgrounds. We chose

---

*To whom correspondence should be addressed. E-mail: nisbett@umich.edu.

Fig. 1. Sample pictures presented in the study. Thirty-six pictures with a single foregrounded object (animals or nonliving entities) on realistic backgrounds were presented to participants.

this measure because eye fixations reflect the allocation of attention in a fairly direct manner. Moreover, we have relatively little awareness of how our eyes move under normal viewing conditions. If differences in culture influence how participants actually view and encode the scenes, there will be differences in the pattern of saccades and fixations in the eye movements of the members of the two cultures. [Saccades are rapid, ballistic eye movements that shift gaze from one fixation to another (15).] In particular, we would expect Americans to spend more time looking at the focal objects and less time looking at the context than the Chinese participants. Furthermore, if the Chinese participants perceive the picture more holistically and bind contextual features with features of the focal object, they might make more total saccades when surveying the scene than the Americans. On the other hand, if no eye movement differences emerge between the two cultures, then previous findings of memory and judgment differences are likely due to what happens at later stages, e.g., during memory retrieval or during reporting.

Methods

Participants. Twenty-five European American graduate students (10 males, 15 females) and 27 international Chinese graduate students (14 males, 12 females, 1 data missing) at the University of Michigan participated in the study. The mean ages of Americans and Chinese were 24.3 and 25.4 years, respectively. All of the Chinese participants were born in China and had completed their undergraduate degrees there. Participants from the two cultures were matched on age and graduate fields of study. Participants were graduate students from engineering, life sciences, business programs, and, in a few cases, from the social sciences. Recruitment e-mails were sent to a Chinese student organization as well as to different graduate academic departments. Volunteers were each paid $14.00 for their participation in the study.

Materials. A collection of animals, nonliving things, and background scenes was obtained from the COREL image collection (Corel, Eden Prairie, MN), and a few were obtained from a previous study (6). The pictures were manipulated by using PHOTOSHOP software (Adobe Systems, San Jose, CA) to create 36 pictures of single, focal, foregrounded objects (animal or nonliving thing) with realistic complex backgrounds. The final set of pictures contained 20 foregrounded animals and 16 foregrounded nonliving entities, e.g., cars, planes, and boats (see Fig. 1 for examples of the pictures shown). The set was composed mostly of culturally neutral photos, plus some Western and Asian objects and backgrounds. This set of 36 pictures was used in the study phase, during which the eye movement data were collected.

For the recognition-memory task, the original 36 objects and backgrounds together with 36 new objects and backgrounds were manipulated to create a set of 72 pictures. Half of the original objects were presented with old backgrounds and the other half with new backgrounds. Similarly, half of the new objects were presented with old backgrounds and the other half with new backgrounds. This procedure resulted in four picture combinations: (*i*) 18 previously seen objects with original backgrounds, (*ii*) 18 previously seen objects with new backgrounds, (*iii*) 18 new objects with original backgrounds, and (*iv*) 18 new objects with new backgrounds. This set of 72 pictures was used in the object-recognition phase. All participants saw the same set and sequence of trials to make comparisons of performance comparable.

Procedure. *Study phase.* The participants sat on a chair and placed their chin on a chin rest to standardize the distance of the head from the computer monitor. The distance of the chin rest from the monitor was 52.8 cm. The size of the monitor was 37.4 cm.

At the start of the session, participants wore a 120-Hz head-mounted eye-movement tracker (ISCAN, Burlington, MA), and eye-tracking calibration was established before the presentation of stimuli. After this calibration, participants were given instructions on the screen. They were informed that they would be viewing several pictures, one at a time. Before each picture was presented, a blank screen with a cross sign (+) was to appear. Participants were told to make sure that they looked at that cross sign. Once the picture appeared, they could freely move their eyes to look at the picture. For each of the pictures, participants verbally said a number between 1 and 7, indicating the degree to which they liked the picture (1, don't like at all; 4, neutral; 7, like very much).† These instructions were followed by several screens showing a sample of how the task would proceed. Once ready, participants started the actual task of viewing the 36 pictures. Each picture was presented for 3 s. Afterward, participants engaged in several distracter tasks for about 10 min. Participants were moved to a different room and, for example, asked to do a backward-counting task, subtracting 7 starting from 100 until they reached zero.

Object-recognition phase. Participants were brought back to the computer room to complete a recognition-memory task. Participants were told that they would be viewing pictures. Their task was to judge as fast as they could whether they had seen an object before, that is, whether they had seen the particular animal, car, train, boat, etc. in the pictures during the study

---

†The Chinese participants gave higher liking ratings than did the Americans (*Ms*, 4.64 vs. 4.16; *P* < 0.005).
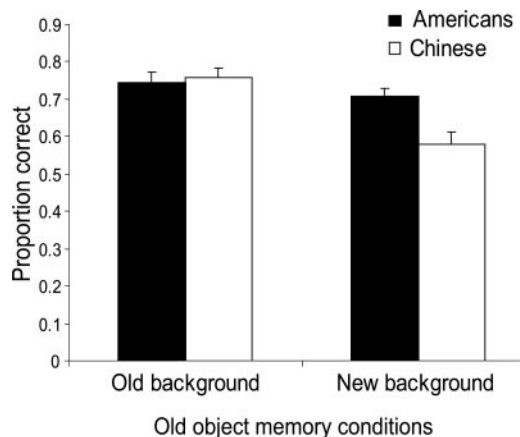
Chua *et al.*

Fig. 2. Mean accuracy rates from the object-recognition phase (22 Americans and 24 Chinese). Data shown refer to correct recognition of old objects, when the old objects were presented in old backgrounds, compared with when old objects were presented in new backgrounds. Object refers to the single foregrounded animal or nonliving entity on the picture; background refers to the rest of the realistic, complex spatial area on the visual scene.

phase. Participants pressed a key if they believed that they had seen the object before, and they pressed another key if they believed that it was new. If participants were unsure, they were told to make a guess. Participants then were shown a sample picture informing them which item in the picture was the object and that the rest of the visual scene was the background. Participants were informed that each picture would be shown only for a specified period. In the event that the picture had already left the screen, they could still input their response. Seventy-two pictures, including 36 original objects and 36 lure objects, were presented. The objects were presented with either an old or a new background. Each picture was again presented for 3 s, and a fixation screen was presented between the picture presentations.

Demographic questionnaire and debriefing. At the end of the study, participants engaged in an object-familiarity task. All 72 objects were presented against a white screen on a computer. Participants circled "yes" if they thought they had seen the object in real life or in pictorial information before coming to the study and "no" if they had not. This procedure was similar to that in a previous study (6). We repeated the analyses reported in this paper with familiarity as a covariate, and there were no changes in the statistical patterns. Participants also completed a demographic questionnaire asking information about their age, education, family history, and English language ability. Participants were debriefed and paid.

Data analysis. Six participants had a hit rate of <0.5 on the object-recognition task, averaged across conditions. These participants' data were excluded in all statistical analyses. One additional European American had poor eye-tracking data. These exclusions resulted in data for 21 European American and 24 international Chinese participants being included in the eye-tracking analyses.

Results

The results for the object-recognition task were consistent with previous findings (6), indicating that East Asians are less likely to correctly recognize old foregrounded objects when presented in new backgrounds [$F(1, 44) = 5.72$, $P = 0.02$] (Fig. 2). Thus, we have additional evidence for relatively holistic perception by East Asians: they appear to "bind" object with background in perception.

The eye-movement patterns of American and Chinese participants differed in several ways. As summarized in Fig. 3, the American participants looked at the foregrounded object sooner and longer than the Chinese, whereas the Chinese looked more at the background than did the Americans, confirming our predictions. Overall, both groups fixated the background more than the objects (Fig. 3A), probably because the background occupied a greater area of the visual scene [$F(1, 43) = 72.46$, $P < 0.001$]. The Chinese made more fixations during each picture presentation than the Americans [$F(1, 43) = 4.43$, $P < 0.05$], but this was entirely due to the fact that Chinese made more fixations on the background [$F(1, 43) = 9.50$, $P < 0.005$]. The Americans looked at foregrounded objects 118 ms sooner than did the Chinese [$t(43) = 2.41$, $P = 0.02$] (Fig. 3B). Participants from both cultures had longer fixations on the objects than on the backgrounds (Fig. 3C) [$F(1, 43) = 17.27$, $P < 0.001$], but this was far more true for the Americans than for the Chinese [$F(1, 43) = 5.97$, $P < 0.02$]. In short, the cultural difference in the memory study was reflected in the eye movements as well.[‡]

The cultural difference in eye-movement patterns emerged very early. At the onset of the picture slide, 32–35% of the time both the Americans and the Chinese happened to be looking at the object, but the first saccade increased that percentage by 42.8% for the Americans and only by 26.7% for the Chinese [$t(43) = 2.46$, $P < 0.02$].

To better understand the time course of cultural differences, we examined the fixation patterns across the 3-s duration of picture presentations. Fig. 4 shows that whereas the Americans were most likely to be looking at the object for about 600 ms of the first second, the Chinese exhibited a very different eye-movement pattern. For the first 300–400 ms, no cultural differences were observed; at picture onset, both Americans and Chinese fixated the backgrounds more than the focal objects [$F(1, 43) = 235.91$, $P < 0.001$]. By ≈420 ms after picture onset, the Americans were equally likely to be looking at the background and the focal object. At this point, there was an interaction of culture and fixation region, with only the Chinese fixating the backgrounds more than the objects [$F(1, 43) = 6.43$, $P < 0.02$]. Based on Fig. 4, the region during which the Americans attended preferentially to the object spanned 420–1,100 ms. Averaging the data across this interval, the Americans fixated the objects proportionately more than the backgrounds, whereas this was not at all true for the Chinese [$F(1, 43) = 7.31$, $P < 0.01$]. There was no time point at which the Chinese were fixating the objects significantly more than the backgrounds during the 3-s presentation. Averaging the data from 1,100 to 3,000 ms, the Chinese looked more at the backgrounds than at the objects, whereas this was much less true for the Americans [$F(1, 43) = 6.64$, $P < 0.02$]. Taken together with the summary data from Fig. 3, these findings provide clear evidence that cultural differences in eye-movement patterns mirror and probably underlie the cultural differences in judgment and memory tasks.

Discussion

The present findings demonstrate that eye movements can differ as a function of culture. Easterners and Westerners allocated attentional resources differently as they viewed the scenes. Apparently, Easterners and Westerners differ in attributing informativeness to foregrounded objects vs. backgrounds in the context of a generic "How much do you like this picture?" task. The Americans' propensity to fixate sooner and longer on the
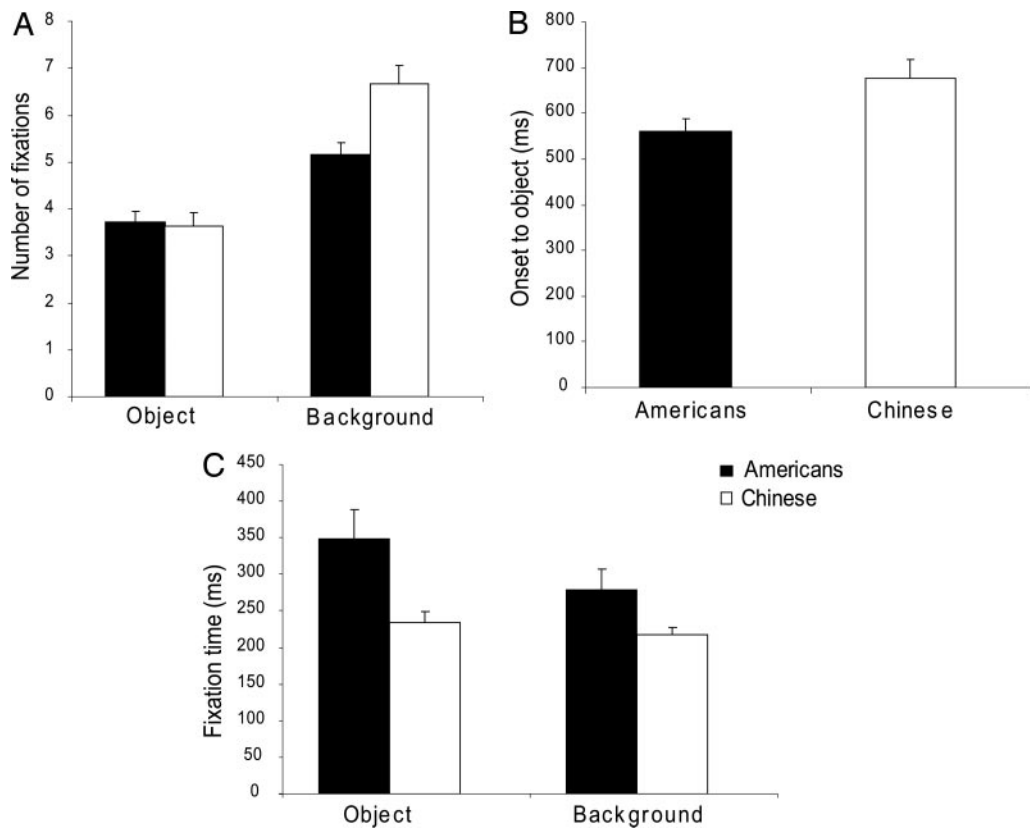
PSYCHOLOGY

Fig. 3. Eye movement data. (*A*) Number of fixations to object or background by culture (21 Americans and 24 Chinese). Each picture was presented for 3 s. (*B*) Onset time to object by culture. Time was measured from onset of each picture to first fixation to object, comparing Americans and Chinese. (*C*) Average fixation times to object and background as a function of culture. All figures represent mean scores over 36 trials and SEM.

foregrounded objects suggests that they encoded more visual details for the objects than did the Chinese. If so, this could explain the Americans' more accurate recognition of the objects, even against a new background. The Chinese pattern of more balanced fixations to the foreground object and background is consistent with previous reports of holistic processing of visual
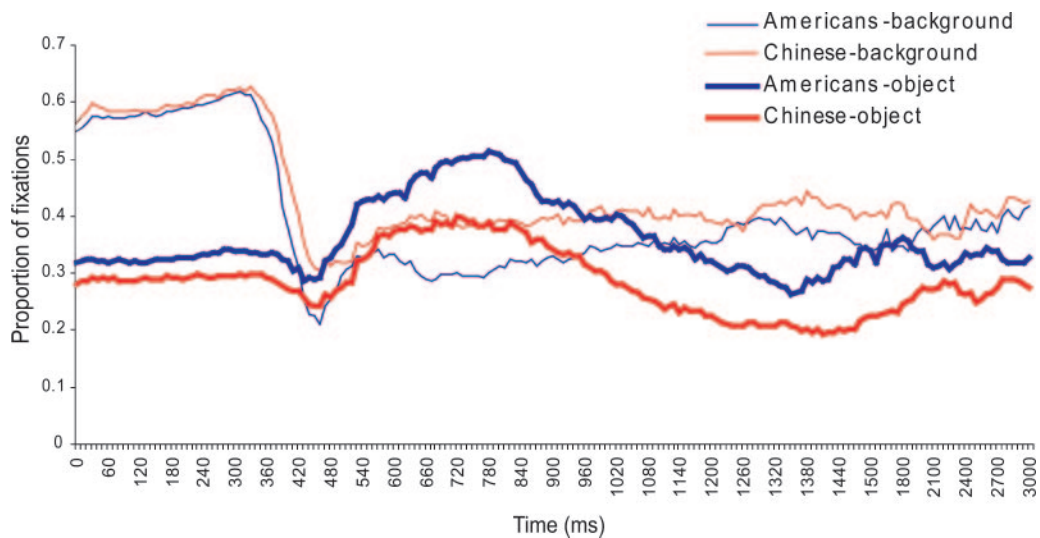

Fig. 4. Proportion of fixations to object or background, across the 3-s time course of a trial. Data points are sampled every 10 ms for 0–1,500 ms, and every 50 ms for 1,500–3,000 ms, averaging over all 36 trials. The sum of percentages at each time point may not total 100% because, at times, participants were in the process of making a saccade, thus they were in between fixations. The graph illustrates distinct eye tracking patterns of Americans and Chinese during the 3-s period. Cultural differences begin by 420 ms after onset, when an interaction of culture and region was observed, with the Chinese, but not the Americans continuing to fixate the background more than the focal object. Averaging the data from 420 to 1,100 ms, Americans were fixating focal objects at a greater proportion than backgrounds, compared with Chinese. Averaging the data from 1,100 to 3,000 ms, Chinese were fixating more often to the backgrounds and less to the objects, compared with Americans.

scenes (6–8). Thus, previous findings of cultural differences in visual memory are likely due to how people from Eastern and Western cultures view scenes and are not solely due to cultural norms or expectations for reporting knowledge about scenes.

Cultural differences in eye movements, memory for scenes, and perceptual and causal judgments could stem from several sources, including differences in experience, expertise, or socialization. It is common to consider such factors in high-level cognition, but because such factors can influence the allocation of attention, they influence lower level cognition as well. Our hypothesis is that differential attention to context and object are stressed through socialization practices, as demonstrated in studies on childrearing practices by East Asians and Americans (16, 17). The childrearing practices are, in turn, influenced by societal differences. East Asians live in relatively complex social networks with prescribed role relations (18, 19). Attention to

context is, therefore, important for effective functioning. In contrast, Westerners live in less constraining social worlds that stress independence and allow them to pay less attention to context.

The present results provide a useful warning in a world where opportunities to meet people from other cultural backgrounds continue to increase: people from different cultures may allocate attention differently, even within a shared environment. The result is that we see different aspects of the world, in different ways.

1. Nisbett, R. E., Peng, K., Choi, I. & Norenzayan, A. (2001) *Psychol. Rev.* **2,** 291–310.
2. Nisbett, R. E. & Masuda, T. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 11163–11170.
3. Choi, I. & Nisbett, R. E. (1998) *Pers. Soc. Psychol. Bull.* **24,** 949–960.
4. Morris, M.W. & Peng, K. (1994) *J. Pers. Soc. Psychol.* **67,** 949–971.
5. Chua, H. F., Leu, J. & Nisbett, R. E. (2005) *Pers. Soc. Psychol. Bull.* **31,** 10925–10934.
6. Masuda, T. & Nisbett, R. E. (2001) *J. Pers. Soc. Psychol.* **81,** 922–934.
7. Ji, L., Peng, K. & Nisbett, R. E. (2000) *J. Pers. Soc. Psychol.* **78,** 943–955.
8. Kitayama, S., Duffy, S., Kawamura, T. & Larsen, J. T. (2003) *Psychol. Sci.* **14,** 201–206.
9. Simons, D. J. & Rensink, R. A. (2005) *Trends Cognit. Sci.* **9,** 16–20.
10. Potter, M. C. (1976) *J. Exp. Psychol. Hum. Learn. Mem.* **2,** 509–522.
11. Enns, J. T. (2004) *The Thinking Eye, the Seeing Brain: Explorations in Visual Cognition* (Norton, New York).
12. Intraub, H. (1997) *Trends Cognit. Sci.* **1,** 217–212.
13. Potter, M. C., O'Connor, D. H. & Olivia, A. (2002) *J. Vision* **2,** 516.
14. Henderson, J. H. & Hollingworth, A. (1999) *Annu. Rev. Psychol.* **50,** 243–271.
15. Smith, E. E., Fredrickson, B., Loftus, G. & Nolen-Hoeksema, S. (2002) *Atkinson and Hilgard's Introduction to Psychology* (Wadsworth, Belmont, CA), 14th Ed.
16. Fernald, A. & Morikawa, H. (1993) *Child Dev.* **64,** 637–656.
17. Tardif, T., Gelman, S. A. & Xu, F. (1999) *Child Dev.* **70,** 620–635.
18. Markus, H. R. & Kitayama, S. (1991) *Psychol. Rev.* **98,** 224–253.
19. Nisbett, R. E. (2003) *The Geography of Thought: How Asians and Westerners Think Differently. . . And Why* (Free Press, New York).

PSYCHOLOGY