Journal of Experimental Child Psychology 147 (2016) 22–38

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Cultural differences in visual object recognition in 3-year-old children

Megumi Kuwabara [a,*], Linda B. Smith [b]

[a] Child Development Program, California State University, Dominguez Hills, Carson, CA 90747, USA
[b] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:
Received 4 November 2015
Revised 5 February 2016

Keywords:
Object recognition
Culture
Perception
Feature
Configural
Development

ABSTRACT

Recent research indicates that culture penetrates fundamental processes of perception and cognition. Here, we provide evidence that these influences begin early and influence how preschool children recognize common objects. The three tasks ($N = 128$) examined the degree to which nonface object recognition by 3-year-olds was based on individual diagnostic features versus more configural and holistic processing. Task 1 used a 6-alternative forced choice task in which children were asked to find a named category in arrays of masked objects where only three diagnostic features were visible for each object. U.S. children outperformed age-matched Japanese children. Task 2 presented pictures of objects to children piece by piece. U.S. children recognized the objects given fewer pieces than Japanese children, and the likelihood of recognition increased for U.S. children, but not Japanese children, when the piece added was rated by both U.S. and Japanese adults as highly defining. Task 3 used a standard measure of configural progressing, asking the degree to which recognition of matching pictures was disrupted by the rotation of one picture. Japanese children's recognition was more disrupted by inversion than was that of U.S. children, indicating more configural processing by Japanese than U.S. children. The pattern suggests early cross-cultural differences in visual processing; findings that raise important questions about how visual experiences differ across cultures and about universal patterns of cognitive development.

---

* Corresponding author.
E-mail address: mkuwabara@csudh.edu (M. Kuwabara).

http://dx.doi.org/10.1016/j.jecp.2016.02.006

Introduction

Human visual object recognition is fast and robust. People can recognize a large number of instances of many different categories under varied and non-optimal conditions. By all accounts, this prowess depends on visual experiences with the categories (e.g., Kourtzi & DiCarlo, 2006; Peissig & Tarr, 2007); that is, the recognition of cars, cups, and dogs depends on one's experience with those categories (Gauthier, Skudlarski, Gore, & Anderson, 2000; Kovack-Lesh, McMurray, & Oakes, 2014; Malt & Majid, 2013). We ask whether the development of visual object recognition also depends on the culture in which one develops. The hypothesis is not that culture affects object recognition because of the kind or range of experienced instances but rather whether culture biases visual processing more generally, encouraging the processing of more local or global properties, and in so doing changes the information used and represented for recognizing objects. The idea that culture penetrates a core cognitive function such as visual object recognition is novel but is consistent with a growing set of findings showing pervasive cultural effects on visual processing.

The relevant cross-cultural studies have primarily focused on the processing of scenes (visual arrays composed of multiple objects) and have used a variety of measures, including recognition measures (Ishii, Tsukasaki, & Kitayama, 2009; Masuda & Nisbett, 2001, 2006), eye-tracking (Chua, Boland, & Nisbett, 2005; Kelly, Miellet, & Caldara, 2010; Masuda et al., 2008), and brain imaging (Goh et al., 2013; Han & Northoff, 2008; Hedden, Ketay, Aron, Markus, & Gabrieli, 2008; Masuda, Russell, Chen, Hioki, & Caplan, 2014). The findings show consistent differences in how Western adults (residing in North America and Europe) and Eastern adults (residing in China, Japan, and Korea) process visual information. In aggregate, the findings suggest that Western perceivers are more selective, more focused on local elements in scenes, and less affected by visual context than Eastern perceivers. In contrast, Eastern perceivers are more holistic and more sensitive to the relational structure among elements in a scene (Chua et al., 2005; Hedden et al., 2008; Kitayama, Duffy, Kawamura, & Larsen, 2003; Masuda & Nisbett, 2001, 2006; Masuda et al., 2008; Miyamoto, Yoshikawa, & Kitayama, 2011; Nisbett & Masuda, 2003; Nisbett & Miyamoto, 2005; Nisbett, Peng, Choi, & Norenzayan, 2001). These differences are not all-or-none, but they are pervasive across a variety of tasks and visual stimuli.

Critically, cultural differences like those found in adults are also found in children (Duffy, Toriyama, Itakura, & Kitayama, 2009; Imada, Carlson, & Itakura, 2013; Moriguchi, Evans, Hiraki, Itakura, & Lee, 2012; Senzaki, Masuda, & Nand, 2014), including children as young as 4 years (Kuwabara & Smith, 2012; Kuwabara, Son, & Smith, 2011). The demonstration of developmentally early cultural differences imposes constraints on explanations of their origins. The demonstration of these early differences in core psychological tasks such as visual search and selective attention (Kuwabara & Smith, 2012) also challenges our understanding of presumed universal properties of cognitive development. With these larger issues in mind, we ask the question: Are cultural differences in visual processing also evident in how young children recognize common objects?

As several reviews have lamented (Braddick & Atkinson, 2011; Nishimura, Scherf, & Behrmann, 2009; Smith, 2009), the development of visual object recognition—despite its centrality to many human competencies—is relatively understudied beyond the first year of life. This is so despite the fact that the literature also shows that developmental changes in visual object recognition extend well into adolescence (Bova et al., 2007; Jüttner, Wakui, Petters, Kaur, & Davidoff, 2013; Wakui et al., 2013). The evidence that we do have from young children derives primarily from studies of Western children. These findings suggest a developmental progression from recognition based more on local and piecemeal features to recognition based on the relational structure among the features and parts (Augustine, Jones, & Smith, 2015; Augustine, Smith, & Jones, 2011; Davidoff & Roberson, 2002; Diamond & Carey, 1986; Smith, 2009; Wakui et al., 2013). For young Western children, for example, one or two highly diagnostic features—cat ears and whiskers or wheels—may trump other category-incongruent visual information (Pereira & Smith, 2009; Rakison & Butterworth, 1998). We ask the question: Is this early reliance on local category-diagnostic features principally a fact about Western children or a fact about the development of object recognition more generally?

To answer this question, we used a converging evidence approach, as suggested by Garner (1974), and did so for two reasons. First, the contrast between more global, relational, and holistic processing versus local, piecemeal features, and analytic processing has multiple manifestations both in the cross-cultural studies cited above and in the larger literatures on perceptual processing in adults (Fink et al., 1996; Martin, 1979; Maurer, Le Grand, & Mondloch, 2002) and in children (Mondloch, Pathman, Maurer, Le Grand, & de Schonen, 2007; Poirel, Mellet, Houdé, & Pineau, 2008; Quinn, Tanaka, Lee, Pascalis, & Slater, 2013; Robbins, Shergill, Maurer, & Lewis, 2011). The very large literature on adult perceptual processing contrasting more holistic versus more local processing provides clear evidence for the distinction (Kimchi, 1992; Maurer et al., 2002; Nisbett et al., 2001; Palmeri & Gauthier, 2004) but also shows many complexities. In particular, the range of tasks used to measure these differences is unlikely to tap a single underlying mechanism (Bravo & Nakayama, 1992; Fink et al., 1997; Maurer et al., 2002; Richler, Palmeri, & Gauthier, 2012). Furthermore, the observed differences in this literature on more local versus more global visual processing may be more continuous than categorical (Kimchi, 1992; Richler, Mack, Palmeri, & Gauthier, 2011; Tsai, Meiran, & Lamy, 1995). Therefore, no single task may be sufficient to make strong conclusions about cultural differences. Second, culture may influence task performance in many ways, including how the task instructions are understood and the motivation to perform well. Therefore, no one demonstration task, no matter how well controlled or how seemingly culturally neutral or well motivated, will be sufficient for a strong conclusion about cultural differences. Accordingly, in the study that follows, we used three different tasks. All were focused on one manifestation of a distinction between local versus more holistic processing of category-defining features, namely that of piecemeal versus configural feature processing (Bushmakin & James, 2014; Folstein, Palmeri, & Gauthier, 2013). We operationally define more local processing as object recognition that is more based on piecemeal features, that is, features presented out of their larger context within the whole; we operationally define more holistic processing as recognition that is more dependent on the larger whole-object (relational/spatial) context within which those features are typically perceived. In sum, the three tasks were designed to assess, in different ways, recognition based on decontextualized features versus features in their spatial/relational context. The participating children were 3-year-olds, a full year younger than the youngest children tested to date for West–East cultural differences in visual processing.

Method

Participants

A total of 128 children between the ages of 28 and 43 months were recruited such that 64 pairs of children, consisting of one Japanese and one U.S. child, were formed in which the children within a pair were within 1 month of age (±1 month). The average age of the children was 36.8 months, and 34 of the Japanese children and 35 of the U.S. children were male. The Japanese children were residents of Yamanashi, Japan, and were monolingual speakers of Japanese; the U.S. children were residents of Monroe County, Indiana (in the U.S. Midwest), and were monolingual speakers of English. In both cultures, the children were from families of comparable economic and educational status, with at least one parent in each family having attained a college degree. Pairs of children were assigned to the three tasks—16 pairs to the Feature task, 16 pairs to the Puzzle task, and 32 pairs to the Conversion task—such that mean age and range of ages across the tasks were comparable; pairs of children were randomly assigned to the different order conditions within tasks. To limit possible carryover effects, each child participated in only one task. Age pairing was done solely to ensure comparability of tested children in the two cultures and across the three tasks; all statistical analyses took the conservative approach of assuming independent samples across the two cultures. Children were tested in their preschools in both countries.

Selection of stimulus categories and measures of cultural appropriateness

The stimuli in the three tasks were derived from photographs of common basic-level object categories. In all tasks, the categories were selected to include both animals and artifacts and to be common categories in the experiences of children in both cultures. The category names are normatively in the productive vocabulary of children in both cultures by 30 months of age (Fenson et al., 1993; Ogura & Watamaki, 1997). The photographs were open-use images found via Google image search selected by the two authors (one a Japanese national and the other a U.S. national) to be typical instances in both cultures, and then additional measures of the cultural appropriateness of the selected photographs were made as described under the specific tasks. The children in both countries were tested by the first author (fluent in English and Japanese) or assistants (native speakers in the testing language) directly supervised by the first author.

Feature task

The Feature task was designed to test children's ability to recognize objects by diagnostic local features when the larger object context within which those features are typically perceived was removed, an approach that has been used in both machine vision research (Wu, Osuntogun, Choudhury, Philipose, & Rehg, 2007; Wu & Rehg, 2012) and adult vision research (James, Huh, & Kim, 2010; James, Humphrey, Gati, Menon, & Goodale, 2000) to measure the role of local feature information in object recognition. The expectation was that U.S. preschool children would be able to use feature information alone to recognize the objects; the question was whether same-aged Japanese children would show the same skill. They should if visual object recognition for all children develops from recognition based on isolated features to recognition based on more relational representations of the features in the whole-object context and if the development of visual object recognition in the two cultures progresses at similar rates. In the task, children were asked to recognize instances of common object categories from three local regions showing only category-diagnostic features as shown in Fig. 1A. The 6 categories were airplane, dog, chair, duck, car, and hat; photographs of two unique instances for each category were used to create a test set of 12 images. Following the approach of James and colleagues (2000), these photographs were presented behind occluders as shown in Fig. 1A such that only three spatially separated regions were visible. The three regions for each photograph were selected by the authors to contain highly diagnostic information. The diagnosticity of the selected features for adults in the two cultures was tested by presenting adults living in Japan and adults living in the United States ($n = 7$ in each cultural group) with individual (one at a time) test objects that showed only the three features (as in Fig. 1A) and asking the adults to name the object. All adults produced the target name for all test photographs on first query, indicating that the three feature regions provide sufficient information for mature members of both cultures to recognize these objects. The entire set of stimuli used in the study consisted of (a) the Feature test trials (12 objects—2 unique instances × 6 categories—showing only the three diagnostic regions; (b) Untested Feature Foils trials (2 instances from 2 categories: donkey and cup) that also showed only three regions and were added to reduce the repetition of the tested category instances as foils; and (c) additional photographs of 6 objects (cat, orange, apple, spoon, flower, and bottle) that were used on Whole Picture trials designed to ensure that children understood the task. The Whole Picture trials were structured the same as the Feature test trials, but all objects were shown as complete pictures.

Feature test trials, Untested Feature Foils trials, and Whole Picture trials were arranged as pages in a picture book. Each trial (page in the book) consisted of a 6-alternative forced choice test; the experimenter named a category, and children's task was to point to the picture of the named target on a page showing 6 alternatives. Each page (8 × 11.5-inch size paper) showed a target and 5 foils selected such that 3 of them were randomly selected from the targets for other Feature test trials and Untested Feature Foils trials and 2 of them were from Whole Picture trials. Each occluding box (showing the 3 feature fragments) was 2 × 2 inches. Placement of images on the page varied from page to page to limit location response biases. The average distance between the images on a page was 1 inch and ranged from 1/8 to 2½ inches. The 21 pages were assembled into two pseudo-random orders, resulting in two books. For both books, the order was constrained such that the first page of the book presented a Whole Pictures trial and was followed by at least 2 Feature test trials; the remaining 4 Whole Pictures trials were randomly intermixed with the remaining 10 Feature test trials and 4 Untested Feature Foils trials. On each trial, children were asked to point to a named object: "Look, there is a car here. Where is the car?" in English or "Mite, kokoni kuruma ga aruyo. Kuruma dokoni arukana?" in Japanese. Children indicated their choice by pointing. No feedback was provided at any point. Children were tested while

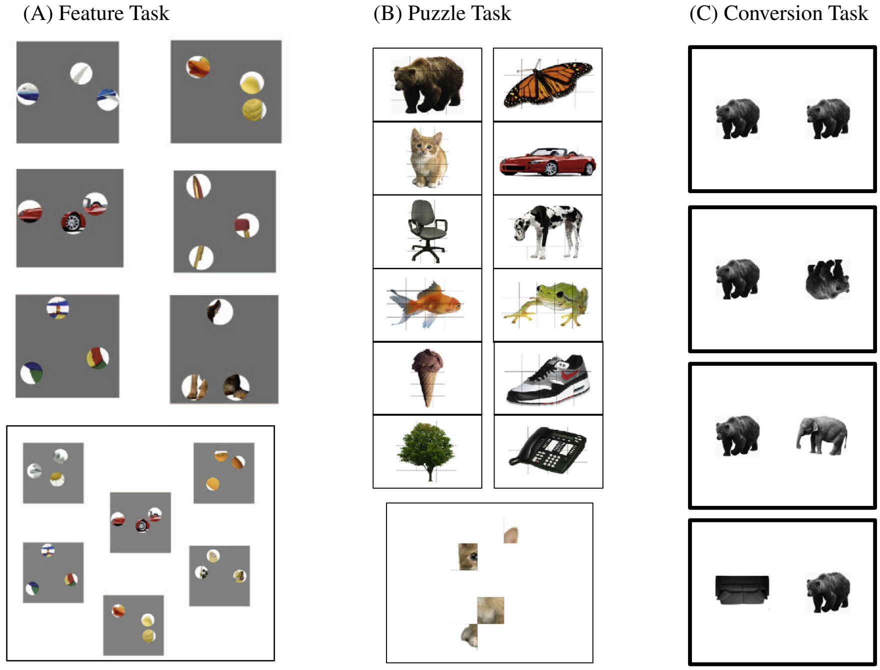

Fig. 1. Sample stimuli used in the three tasks. (A) Feature task: Example of six target categories (airplane, dog, chair, duck, car, and hat). Each picture of common objects was masked to show three diagnostic features of category. The bottom of the six pictures shows an example of a Feature test trial (6-alternative forced choice task). (B) Puzzle task: 12 puzzle pictures. Pictures of common objects were divided into pieces that were presented incrementally. The bottom of the 12 pictures shows an example of a trial (middle of "cat" trial). (C) Conversion task: Example of "bear" trial. Pairs of pictures either matched (Same trials) or did not match (Different trials), and either both pictures were presented upright or one was presented inverted (see text for details).

sitting at a table with the book on the table. The distance of an open book page to children's face was approximately 16 inches.

Puzzle task

The principal goal of the Puzzle task was to test the hypothesis that Japanese children required more visual information than the diagnostic features alone—that is, more information about the whole object—to recognize the object than did U.S. children. In addition, the task provided converging evidence on whether U.S. children were particularly reliant on *diagnostic* features. In the task, children were asked to guess a pictured object as it emerged piece by piece (see Fig. 1B). The pictured object was presented as a puzzle, beginning with a single piece in its proper spatial location, and children were asked to name the pictured object after each piece was added. The puzzle pieces were added randomly such that early in a trial there was no coherent spatial-relational information about the whole object or with respect to the individual pieces. However, as more pieces were added completing the whole, the holistic features—those in their larger spatial-relational context—should emerge. If U.S. children depend on piecemeal features, then they should need, on average, fewer pieces to recognize the object and the likelihood that they recognize an object should be dependent on the category diagnosticity of the information in the added piece on each trial. If Japanese children use more holistic information and require the features in the spatial-relational context of the whole, then they should require, on average, more pieces to recognize the object and be less dependent on the category diagnosticity of the information in a single added piece.

To test these predictions, the objects were 12 photographs of typical instances of 12 basic-level categories: bear, butterfly, kitten, car, chair, dog, fish, frog, ice cream cone, shoe, tree, and phone (see Fig. 1B). To check the recognizability of categories from the selected pictures in the two cultures, five 3-year-olds in Japan and five 3-year-olds in the United States (who did not participate in any experimental tasks) were tested in a 4-alternative forced choice recognition procedure in which they pointed to the named picture when asked (whole photograph). Each of the 12 pictures was tested, and the mean percentage correct was 100% in both cultural groups, indicating that young children in both cultures could recognize the whole pictures used to make the puzzles.

Test puzzles—one for each category—were made by dividing the image into a 3 × 5 or 4 × 4 grid (as appropriate to the shape of the pictured object). Categories differ naturally in the number and variety of diagnostic regions (Ullman, 2007). For example, every piece in the butterfly wing is highly similar to every other piece and potentially equally predictive of the category; however, some regions of a pictured kitten (e.g., the ears, the nose) are potentially more diagnostic than others (e.g., a piece showing only fur). To measure differences in diagnosticity of the puzzle pieces for the different puzzles, seven adults residing in each country were asked to pick the three most category-diagnostic pieces for each puzzle. The individual pieces in each puzzle were then classified as "highly diagnostic" in that culture if the piece was chosen by at least five of seven adults in that culture group as diagnostic. The diagnostic puzzle pieces selected for each picture by this method did not differ across cultures: Three puzzles (cat, dog, and frog) had two highly diagnostic pieces, seven puzzles had just one agreed-on diagnostic piece, and two puzzles (butterfly and ice cream) had no single puzzle piece reaching the "highly diagnostic" threshold.

During the task, each piece of the puzzle was shown on a computer screen in its appropriate location. The animation function in Microsoft PowerPoint was used to add pieces incrementally. Any pieces that were completely blank within the array were not used as incremental pieces but were present from the beginning (e.g., the white space between the legs of the frog); given this constraint, the number of incremental pieces for the 12 test puzzles varied from 13 to 16. The session began with two 6-piece warm-up puzzles (apple and pig pictures with easier bigger pieces) that were used to instruct children on the goals of the task. For each puzzle, two unique random orders were created that determined the sequence with which pieces were added to the puzzle. Half of the children in Japan and the United States were presented with each of the piece addition orders for a puzzle. These puzzles were also arranged in two different puzzle orders (crossed with the piece orders). On each trial, children were asked after a piece had been added: "What is this?" in English or "Kore wa nani kana?" in Japanese. Answers accepted as correct included the basic-level name ("cat"), variants of that name (e.g., "kitty," "kitten"), and subordinate or child variant names (e.g., "teddy" for bear). If children labeled the object incorrectly (e.g., said "cow" for the dog) or said "I don't know," the next piece was shown until they correctly named the object or the experimenter reached the final piece. The trials were presented using a 13-inch MacBook Pro, and the distance between the computer and children was approximately 16 inches.

Conversion task

The purpose of the Conversion-effect task was to measure configural processing, that is, children's dependence on the spatial and relational information among features. Children were asked to judge whether two pictures were the same picture or different pictures when both pictures were presented in the same upright orientation and when *one of them* was inverted. In adults, the different orientation of a matching picture leads to slower response times on same judgments and is interpreted as an indicator of the perceiver's dependence on the spatial relations among features, that is, the configuration rather than the elements (Bushmakin & James, 2014; Yin, 1969). This *conversion* effect is robust in adults for nonface objects as well as faces, in contrast to the more face-specific *inversion* effect (poorer performance when both to-be-compared faces are inverted) (see Bushmakin & James, 2014; Yin, 1969). Thus, whereas in the two prior tasks we *removed* the whole-object context that supports relational processing, in this task we *disrupted* that information by inverting the pictures. By hypothesis, this manipulation should affect Japanese children's recognition if they are processing the relational information of the features within the whole. However, if U.S. children are truly piecemeal feature processors, inversion might not alter recognition; it should not if U.S. children can effectively isolate the category-defining features within the inverted object.

The index of configural processing for the child version of the conversion task is more errors for children on Same trials when one picture is inverted relative to when both are upright. For these critical Same test trials, we used two copies of the 12 pictures from the Puzzle task. The nonmatching pairs for the Different trials were created by pairing one copy of each of the 12 photographs with nonmatching images selected from a set of 24 photographs from different object categories (see Appendix). These 24 object categories, used only on the Different trials, were selected to create "near category" and "far category" Different trials. The goal for these trials was to encourage deeper processing of the visual information by including some more difficult trials, the "near category" trials in which there was more overall similarity of the objects, but to also include enough "easy" Different trials to ensure that children were doing the task of making a same–different judgment. For the animal pictures, the "near" comparison objects were another animal (e.g., bear–elephant), and the corresponding "far" comparison objects consisted of an animal and an artifact (e.g., bear–sofa). There were no specific predictions for the Different trials; in adult research on the conversion effect (see Bushmakin & James, 2014), performance on Different trials is not interpreted as distinguishing more piecemeal versus more holistic processing. Fig. 1C shows examples of trial types.

Each child was tested on 6 of 12 categories. A total of 24 test trials were composed of 12 Same and 12 Different trials (2 Same and 2 Different trials for each category); among these trials, 6 Same and 6 Different trials showed both pictures in upright orientation, and 6 Same and 6 Different trials showed *one* picture inverted. The assignment of objects and trial types was counterbalanced across children and matched for same-aged pairs across cultures. For the 6 Different trials, 3 consisted of "near" comparisons and 3 consisted of "far" comparisons. On Upright–Inverted trials, the spatial location (left or right) of the inverted picture was counterbalanced across trials. Trials were presented on a 13-inch MacBook Pro, and the distance from the computer to perceivers' eyes was approximately 16 inches. On each trial, the two images on the screen were centered in their respective halves of the screen; the lateral separation between to-be-compared images depended on the overall shape of the image (more elongated horizontally vs. vertically) and was on average 2 inches. Each test slide (with two images) was shown for 1.5 s and transitioned to the blank white screen (using the Microsoft PowerPoint slide transition function). On the blank white screen, children were asked to judge whether the images they saw in the previous slide were the same or different.

The session began with 6 Instruction trials that used a cartoon face and an apple. On these trials, children were shown either two identical or two mismatching pictures presented with both pictures upright or with one picture upright and one inverted. Children were asked to judge whether they were the same picture or different pictures, with same referring to the picture itself and not its orientation. Children were asked, "Were the pictures you just saw the same or not the same?" in English or "Ima mita no onaji datta? Onaji ja nakatta?" in Japanese. Feedback was given on these Instruction trials to ensure that children understood that "same" referred to the picture and not its orientation. On these Instruction trials, children in both cultures sometimes (~45% of children in both cultures) incorrectly judged the first upright and inverted picture of the same object to be different on *the first upright–inverted* Instruction trial, but all children responded correctly on the remaining Instruction trials. The test trials immediately followed and used the same question on each trial as on the Instruction trials, but no feedback was given.

Results

Feature task

All children performed well on the Whole Picture trials; mean proportion correct for U.S. children was 1.00 ($SD = 0$) and for Japanese children was .99 ($SD = .05$). As shown in Fig. 2A, children from both cultures recognized the objects on Feature test trials better than chance (chance = .16), $t(15) = 11.20$, $p < .001$ for U.S. children and $t(15) = 7.27$, $p < .001$ for Japanese children. However, as shown in Fig. 2A, U.S. children ($M = .73$, $SD = .20$) recognized significantly more of the objects from piecemeal

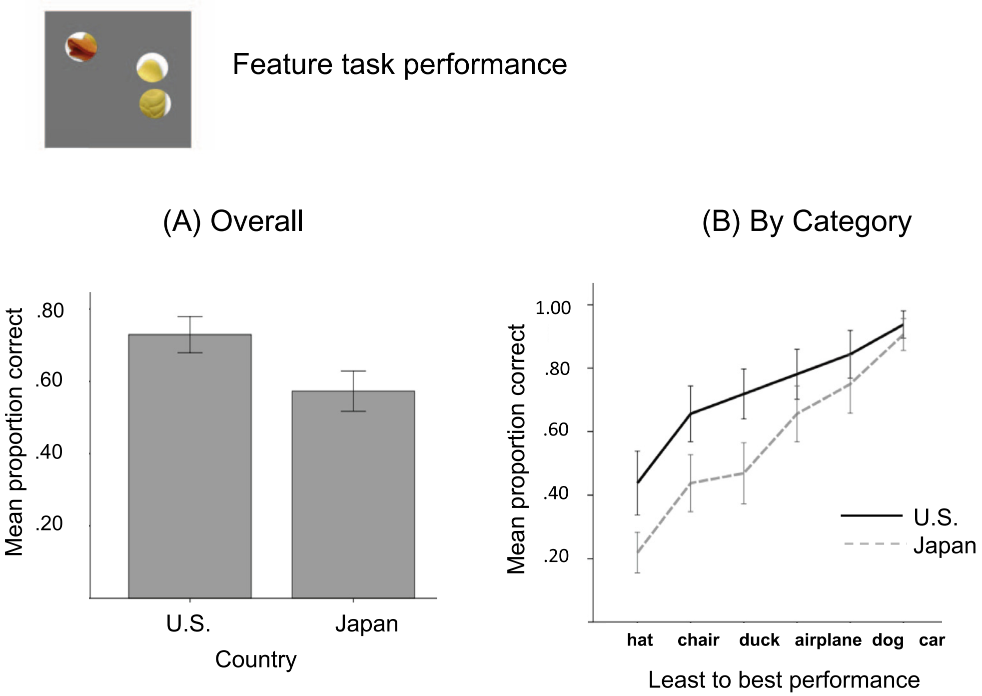

Fig. 2. Results of the Feature task. (A) The bar graph shows mean proportion correct (chance = .16) in recognizing the object from three features for children from the two cultures. (B) The line graph shows proportion correct for the individual categories as a function of culture. The error bars indicates ±1 standard error.

information than Japanese children ($M = .57$, $SD = .22$), $t(30) = 2.08$, $p < .05$, $d = 0.13$ (independent-sample *t*-test). The superior performance of U.S. children in this task characterized most of the categories (see Fig. 2B). Japanese children recognized cars and dogs from the diagnostic fragments as well as U.S. children did, a result suggesting that they too are learning about local features diagnostic of category membership. However, for most categories they were less likely than U.S. children to recognize the objects from the features. An examination of children's errors (points to foils) revealed no systematic patterns within or between culture groups. These results provide support for early cultural differences in object recognition. The occluder removed information about the whole and visually isolated the category-diagnostic features. U.S. children nonetheless recognized these common objects, whereas Japanese children were less likely to do so, a result that fits the hypothesis that Japanese children's visual object recognition is more dependent on the holistic information in the whole object than U.S. children.

Puzzle task

Only a very small number of the 12 puzzles were not recognized by children when the last piece was in place: .03 ($SD = .05$) for U.S. children and .12 ($SD = .14$) for Japanese children. Even though the number of unrecognized pictures is small in absolute magnitude in both groups, the difference is reliable, $t(30) = 2.47$, $p < .05$. Because preliminary evidence indicated that Japanese children recognized the picture when presented as whole as well as U.S. children, this difference may mean that randomly sequenced piece-by-piece presentation made perceiving the spatial coherence to recognize the whole object more difficult for young Japanese children. The trials on which children could not name the picture even after all pieces were presented were excluded from all subsequent analyses.

Fig. 3A shows the mean number of pieces children required to recognize the picture overall and for each puzzle (see also Fig. 3B). The order of each category was sorted by the least to best performance (lower to higher number of pieces) required for Japanese children (see Fig. 3B). An independent-sample *t*-test of the mean pieces to recognition yielded a significant difference between groups, $t(1, 30) = 3.71$, $p < .001$, $d = 0.32$. As predicted from the hypothesis that U.S. children are more piecemeal in their processing and Japanese children are more holistic; children from the United States needed fewer pieces ($M = 3.47$, $SD = 0.82$) than Japanese children to recognize the pictured object ($M = 4.92$, $SD = 1.33$). These group differences were not due to just one or a few categories being difficult for Japanese children; U.S. children recognized 9 of 12 of the categories earlier from the piecemeal information than Japanese children (as shown in Fig. 3B), whereas Japanese children did not recognize any puzzle picture with reliably fewer pieces than U.S. children ($t < 1.4$ for the pairwise comparisons of car, shoe, chair, and telephone).

By hypothesis, U.S. children rely more on the diagnostic feature information in an individual piece, whereas Japanese children rely more on the features in context and, thus, on the emerging picture as a whole. If this is correct, then different pieces—more versus less diagnostic—should have differential effects on U.S. children's recognition compared with Japanese children's recognition. Specifically, the effects of the diagnosticity of an added piece should be less for Japanese children. To assess this prediction, we asked whether highly diagnostic pieces presented early in the series (before more than

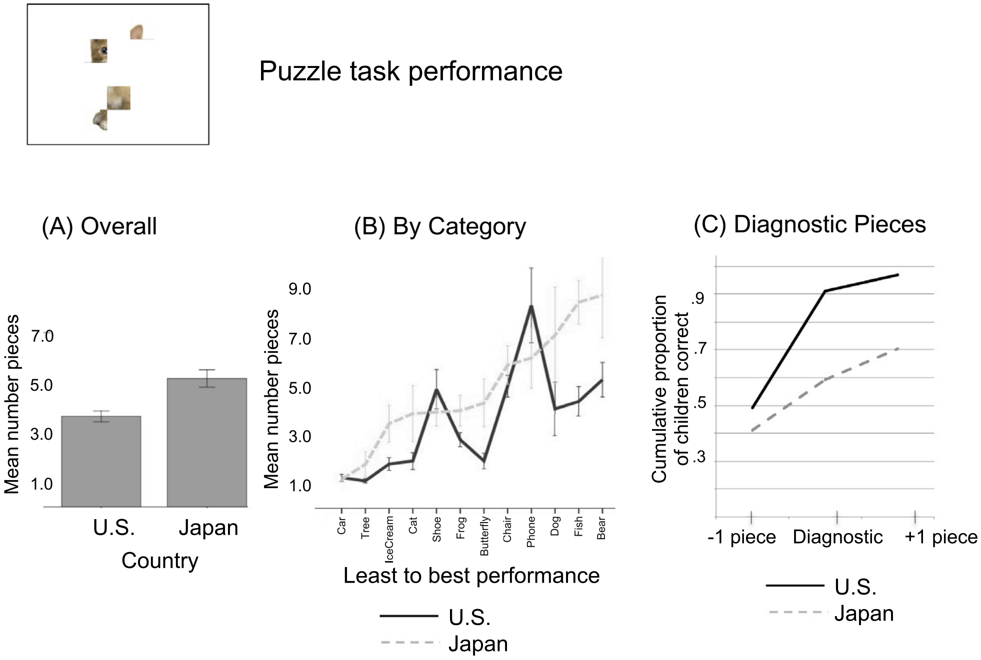

Fig. 3. Results of the Puzzle task. (A) The bar graph shows the mean number of pieces needed to recognize the pictures for children from the two cultures. (B) The line graph shows the mean number of pieces needed for recognition of each picture by children in the two cultures (ordered by performance of Japanese children). The error bars indicates ±1 standard error. (C) This figure shows the proportion of children in the two cultures who recognized the picture (when fewer than 50% of the pieces had been placed) when a highly diagnostic piece was added and when the piece just before and just after the diagnostic piece was added.

half of the pieces had been shown) had larger effects on U.S. children's recognition than on Japanese children's recognition. Fig. 3C shows that highly diagnostic pieces, when presented early in the series of pieces, had large effects on U.S. children's recognition but not on Japanese children's recognition. An independent-sample *t*-test with puzzle as the random variable was used to compare the gain in proportion of children who correctly named the object when the diagnostic piece was added relative to the just prior piece. The two cultural groups did not differ in number of children recognizing the puzzles at the piece just prior to the diagnostic piece, $t(89) = 0.80$, $p = .43$ ($M = .39$, $SD = .49$ for U.S. children and $M = .31$, $SD = .47$ for Japanese children), but did differ in the likelihood of naming the object correctly when the diagnostic piece was presented, $t(130) = 4.08$, $p < .001$ ($M = .81$, $SD = .39$ for U.S. children and $M = .49$, $SD = .50$ for Japanese children). To ensure that this effect was about adding a diagnostic piece (and not just adding any piece at that point in the series), we also calculated a baseline measure that was matched for timing of the added diagnostic piece but in which neither the just prior piece nor following piece was judged to be highly diagnostic; the two cultural groups did not differ in number of children recognizing the puzzle at one piece prior to the assigned piece, $t(102) = -0.16$, $p = .88$ ($M = .44$, $SD = .50$ for U.S. children and $M = .46$, $SD = .50$ for Japanese children), or at the assigned piece, $t(160) = -0.60$, $p = .55$ ($M = .46$, $SD = .50$ for U.S. children and $M = .51$, $SD = .50$ for Japanese children). Thus, the results from the Puzzle task provide converging evidence that U.S. children are better able to use local information and new evidence and that Japanese children require more information about the whole to recognize the object.

Conversion task

Three children who said "same" or "not same" on all test trials were excluded from the final analyses in this task (2 from the United States and 1 from Japan). The remaining children from both cultural groups performed equally well on the Different trials (76% correct for U.S. children and 80% correct for Japanese children). A 2 (Country) × 2 (Upright–Inverted) × 2 (Near–Far) analysis yielded no main effects or interactions ($F < 1.50$ for all cases). Thus, performance on the Different trials is not considered further.

The key measure of configural processing in this task is disruption of Same judgments by the inversion of one picture; we measured this conversion effect as a difference score in correct judgments on the Both Upright versus Upright–Inverted Same trials. Japanese children showed a larger conversion effect than U.S. children, $t(59) = 2.46$, $p < .05$, $d = 0.09$. For U.S. children the mean conversion effect was .13 ($SD = .24$), and for Japanese children this effect was .33 ($SD = .37$), a result suggesting that Japanese children are more reliant on the spatial configuration of visual information than U.S. children. Children from both cultures showed a reliable conversion effect (cf. to 0 = no conversion: $t(29) = 3.03$, $p < .01$ for U.S. children and $t(30) = 4.91$, $p < .001$ for Japanese children), a result suggesting that both groups of children are learning about and using the spatial relations among features to recognize objects, albeit to different degrees depending on culture.

Unexpectedly, however, U.S. children performed less accurately than Japanese children on same judgments in the Both Upright trials, the trials that should yield the best performance, $t(59) = 3.05$, $p < .001$ ($M = .75$, $SD = .30$ for U.S. children and $M = .93$, $SD = .13$ for Japanese children). To ensure that our conclusions about cultural effects on configural processing were not due to this difference, we recalculated the conversion effect for both groups of children, selecting only those children who performed at or above 75% in the Both Upright Same trials ($n = 19$ for U.S. children and $n = 28$ for Japanese children). For this selected sample, the conversion effect for U.S. children was .17 ($SD = .26$) and for Japanese children was .37 ($SD = .36$), $t(45) = 2.07$, $p < .05$, showing the same group effect as in the main analyses. The greater number of errors by U.S. children on the Both Upright Same trials may be a product of their strong reliance on piecemeal features to recognize objects. One problem with piecemeal features as opposed to configural features is that the comparison of separate images requires keeping track of spatially corresponding features across to-be-compared pictures in order to determine that they match. Comparing spatially aligned elements across arrays containing multiple elements has been shown to be quite difficult for (Western) preschool children (Vurpillot, 1968). A second problem with feature-based comparison when those features are presented in the context of the whole object is that the perceiver must allocate resources for perceptual analysis (e.g., Smith & Kemler, 1988) to extract those components. This leads to the testable hypothesis that for piecemeal feature processors—as young U.S. children are hypothesized to be—comparison with isolated features (as in the Feature task) would be easier than comparison with features in context.

Because inversion effects have been strongly linked to face processing (Valentine, 1988), we also examined the pattern of performances for animals versus artifacts. In the current study, and with children, the conversion effect was significantly stronger for artifacts ($M = .30$, $SD = .39$) than for animals ($M = .17$, $SD = .41$), $t(60) = 2.30$, $p < .05$ for both groups of children. In sum, the findings from this task are consistent with the overarching hypothesis under test that Japanese and U.S. children differ in the information used to visually recognize common objects, with Japanese children more dependent on holistic information and U.S. children more dependent on local information.

Cultural differences across tasks

The three tasks were selected to provide converging evidence for the hypothesis of cultural differences in object recognition in young children. To provide an overall assessment of cultural differences across the three tasks, we transformed each participant's score in each task to a *z* score determined across the combined set of scores for both cultural groups in that task such that a negative *z* score (deviation from the cross-culture mean) indicated more holistic recognition and a positive *z* score (deviation from the cross-culture mean) indicated more isolated feature recognition. Fig. 4 shows the overall frequency distribution of numbers of U.S. and Japanese children with *z* scores above and below the mean (with shaded areas indicating the contributions from the three different tasks to the overall distribution). The two distributions of the two populations of children clearly overlap, with most children

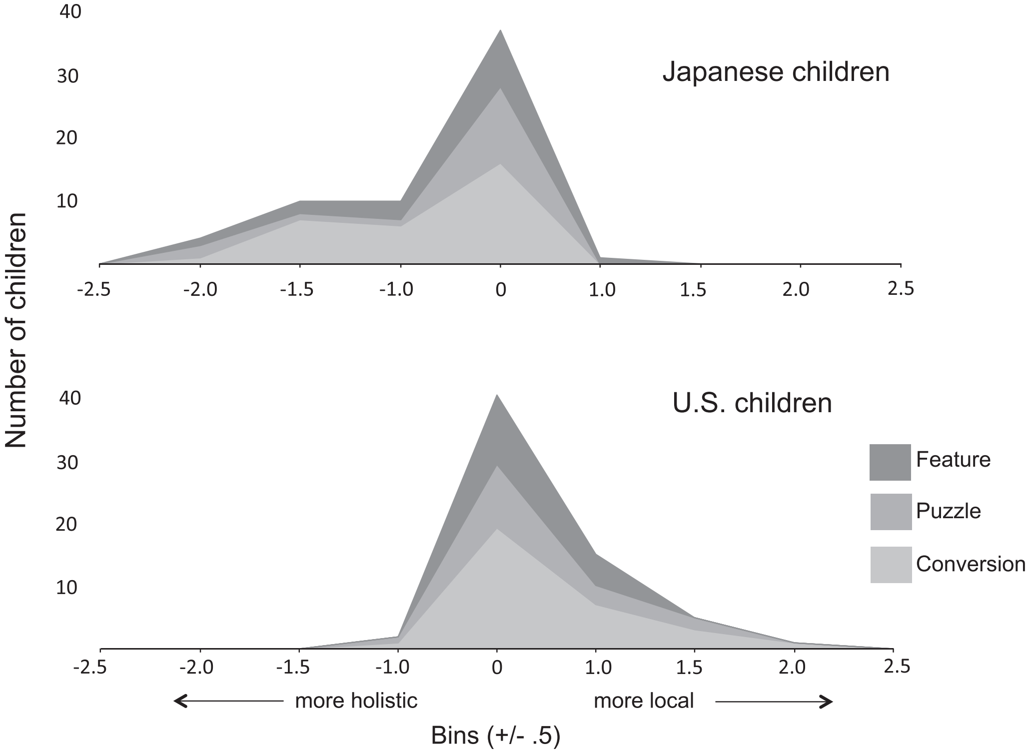

Fig. 4. Histograms of the distribution of individual children's performances in the three tasks, as measured by *z* scores with respect to the cross-cultural mean.

falling around the mean; however, the distributions skew in different directions. We used Pearson's (second) skewness coefficient to measure the skew in the three tasks separately, and in all three tasks the coefficient for Japanese children showed a positive (left-side) skew (0.91, 2.03, and 0.44 for the Feature, Puzzle, and Conversion tasks, respectively), whereas the coefficients for the U.S. children showed a negative (right-side) skew (−1.47, −3.13, and −1.79 for the Feature, Puzzle, and Conversion tasks, respectively). The fact that these differences in distribution across all three tasks—with their different task demands and instructions—shows the same pattern indicates that the cultural differences are general and shows the clear value of taking a converging evidence approach. The overlap of the distributions also clearly shows that the cultural differences are not all-or-none.

One question that emerges from examining these distributions of performance is whether the two groups of children are on the same developmental trajectory but one group is more advanced than the other: Are the more holistic Japanese children leading the way for all children, leading the way for Japanese children, or trailing in a universal developmental pattern? Are the more local U.S. children leading the way for all children, leading the way for U.S. children, or trailing in a universal developmental pattern? The current results do not provide answers to these questions. Individual performances within a culture (as measured by the *z* scores) were not correlated with age (United States: $r^2 = .001$; Japan: $r^2 = .04$); however, children in the current study did not vary greatly in their ages, which limits the interpretability of this result. Clearly, the next step is a developmental study of visual object recognition in the two cultures. A broad developmental study is critical both for answering the questions raised above and for understanding the developmental course of the cross-cultural differences observed in adults. Cultural differences could be exaggerated at early stages of development and decrease with age (see Nand, Masuda, Senzaki, & Ishii, 2014), or they could be the early beginnings of what will become stable larger differences. What the current results indicate is threefold. First, cultural differences in visual processing are evident by 3 years of age. Second, the differences are evident in object recognition and not just scene processing. Third, these differences are evident in a suite of tasks that contrasted object recognition via piecemeal features versus configural processing.

General discussion

The current findings show that cultural differences in visual processing begin early and are evident in how children recognize everyday object categories, a cognitive function that has been proposed to be immune from cultural effects (Gentner, 1978). The findings expand current knowledge about East–West cultural differences in visual processing, first by extending these differences to basic-level object recognition and second by showing that these differences are evident in children as young as 3 years. Past research on cultural differences in visual processing has focused on attention to local elements versus relations in scenes (multiple objects), but the current results suggest that the local versus global distinction applies to the processing of the features within objects as well. These findings—and the fact of early cultural differences in visual object recognition—pose challenges for current understanding of the experience-dependent nature of visual object recognition and, more broadly, for the role of culture in cognitive development.

Contemporary accounts of human visual object recognition have converged on the conclusion that there are multiple mechanisms and multiple routes through which visual objects are recognized (e.g., Alvarez, 2011; Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013; Peissig & Tarr, 2007). As adults, we readily recognize a dog from just the nose sticking out of the blanket, from the contour shape of its shadow, and from cartoon drawings that caricature the configural relations among eye, ears, and snout as well as potentially through other routes. The availability of multiple routes to object recognition may be the reason why human visual object recognition is as powerful as it is. We can recognize a dog as a dog—even under suboptimal conditions—because there are redundant pathways to the same functional end.

From this perspective, the observed cross-cultural differences in 3-year-old children's recognition of everyday basic-level categories may reflect culturally induced variations in the rates of development in these different routes to recognition rather than qualitatively different systems of object recognition. The specific pattern of results in the three tasks fits this conclusion. On the one hand, the differences between U.S. and Japanese children were systematic and evident across the three tasks, with U.S. children's recognition more based on individual diagnostic features and Japanese children's recognition more holistic. On the other hand, and again in each of the three different tasks, children in both cultures showed the ability to use the task-relevant information and the distributions of individuals' performances in the two cultures (Fig. 4) were clearly overlapping. Overall, the results suggest that the use of local and more holistic visual information to recognize objects is apparent in children in both cultures, albeit with different strengths. Past research with adults on the reliance on local versus configural properties in object recognition indicates that the likelihood and efficiency with which adults use one or the other form of information varies with kind of category (cars vs. faces; e.g., Maurer et al., 2002), expertise (less or more; e.g., Gauthier et al., 2000), and the task (e.g., Rousselet, Macé, & Fabre-Thorpe, 2003). These factors do not offer immediately obvious explanations of the differences observed in the current study, but they do point to the coexistence of multiple routes to object recognition that are biased as a function of task and experience.

Attempts to explain cross-cultural differences in adult scene processing have centered on two hypotheses. The first is that clutter in scenes encourages more configural processing of scene information, whereas less cluttered scenes encourage more selective focal attention. Thus, a history of experiencing more and less cluttered scenes could measurably alter visual processing in the ways observed across Western and Eastern cultures (Caparos et al., 2012; Davidoff, Fonteneau, & Goldstein, 2008; Miyamoto, Nisbett, & Masuda, 2006; Wang, Masuda, Ito, & Rashid, 2012; see also Berman, Jonides, & Kaplan, 2008). By this hypothesis, Japanese real-world and cultural representations are more cluttered, and U.S. scenes are less cluttered (Miyamoto et al., 2006; Wang, Masuda, Ito, & Rashid, 2012). The evidence to date for this hypothesis is not strong because it is not clear how pervasive the hypothesized cluttered versus noncluttered scene differences really are or whether they apply to the visual experiences of children prior to 3 years of age. Still, differences in the natural statistics of visual experiences provide a plausible—and testable—hypothesis of the origins of the observed differences in children's object recognition.

A second and widely discussed possibility (Kühnen & Oyserman, 2002; Markus & Kitayama, 1991; Nisbett et al., 2001; Uskul, Kitayama, & Nisbett, 2008; Varnum, Grossmann, Kitayama, & Nisbett, 2010) is that cultural differences in adult visual processing derive from differences in social organization (individualistic vs. collectivist) that influence the way in which visual information—and particularly social information—is sampled. Again, these cultural effects on visual information selection would need to start early to explain the observed cultural differences in this study as well as prior studies documenting differences in scene processing by children as young as 4 years (Kuwabara & Smith, 2012; Kuwabara et al., 2011). However, the idea that gaze patterns relevant to social interactions could influence visual processing more generally is plausible. Perceivers actively create their own visual input—about objects and scenes—with every eye movement. Thus, cultural variations in how visual information is collected also provide a tenable and testable account of the current results.

Because the relevant cultural differences emerge early in development, the relevant experiences will need to be understood in the context of cultural factors that affect children such as parenting, language, stories, and the visual artifacts presented to children. A large literature on cognitive development suggests that each of these may be relevant. Pervasive cultural differences in the social experiences of young children are well documented (Bornstein et al., 1992; Cole, Bruschi, & Tamang, 2002; Farver, Kim, & Lee, 1995; Friedlmeier & Trommsdorff, 1999). Moreover, within a single culture, parenting differences have been linked to the development of visual attention, including selective and sustained attention (e.g., Coll, 2005; Gartstein, Crawford, & Robertson, 2008; Stevens, Lauinger, & Neville, 2009). Language—how people talk about visual things—may also play a role because language provides cues as to the relevant information in a scene (Vales & Smith, 2015). Considerable evidence indicates systematic differences in the frequency of object names versus relational terms in Western and Eastern languages in general (e.g., Brown, 2008; Tardif, 2006) and in English and Japanese in particular (Fernald & Morikawa, 1993; Ogura, Dale, Yamashita, Murase, & Mahieu, 2006). These differences in language structure have been shown to be related to more systematic attention to objects in children learning English and to more systematic attention to relations among objects in children learning Japanese (Imai, Haryu, Okada, Lianjian, & Shigematsu, 2006; Yoshida, 2012). Finally, cultural artifacts—including the layouts of homes and the composition of pictures and picture books directed to children—could have direct effects on visual development.

These have not been systematically compared, but they provide a domain in which to directly test the hypothesis about child-relevant scenes and about the possible link between scene structure and social values of individualism versus collectivism. All of these components of children's developmental context could in principle influence the visual regularities in their world and/or how children visually explore their world. All of these potential systematic differences across cultures may also exist, less systematically, within a culture and may provide useful insights into individual differences in visual processing and visual attention within culture.

In conclusion, the cultural differences observed in the current study remind us that a complete developmental psychology—one that can disentangle the universals, the variations around those universals, and the potentially variable routes to cognitive maturity—requires studying cognitive processes, including those that seem culturally neutral, in different developmental environments.

Acknowledgments

We thank the children who participated in this study and the teachers at the day-care centers. We especially thank Fuji kindergarten, Aichijuku preschool, and Meisho kindergarten in Japan and Monroe County United Ministries, St. Mark's Nursery School and Kindergarten, Gan Shalom, Children's Village, Kid Angles, Penny Lane East, Carousel, and Prep School in the United States. We also thank the adults in each country who participated by providing judgments of the stimuli. This research was supported by R01HD 28675 to Linda B. Smith.

Appendix A

List of targets and corresponding "near category" and "far category" Different trials for the Conversion task.

Target	Near	Far
Chair	Music stand	Octopus
Cat	Fox	Clock
Ice-cream cone	Cup cake	Sea horse
Bear	Elephant	Couch
Butterfly	Dragonfly	Sun glass
Car	Motorcycle	Hippopotamus
Phone	Calculator	Turtle
Frog	Lizard	Sand box
Tree	Tent	Poodle
Dog	Cow	Table
Fish	Bird	Kettle
Shoe	Iron	Seal

References

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences, 15*, 122–131.

Augustine, E., Jones, S., & Smith, L. B. (2015). Relations among early object recognition skills: Objects and letters. *Journal of Cognition and Development, 16*, 221–235.

Augustine, E., Smith, L. B., & Jones, S. S. (2011). Parts and relations in young children's shape-based object recognition. *Journal of Cognition and Development, 12*, 556–572.

Berman, M. G., Jonides, J., & Kaplan, S. (2008). The cognitive benefits of interacting with nature. *Psychological Science, 19*, 1207–1212.

Bornstein, M. H., Tamis-LeMonda, C. S., Tal, J., Ludemann, P., Toda, S., Rahn, C. W., ... Vardi, D. (1992). Maternal responsiveness to infants in three societies: The United States, France, and Japan. *Child Development, 63*, 808–821.

Bova, S. M., Fazzi, E., Giovenzana, A., Montomoli, C., Signorini, S. G., Zoppello, M., et al (2007). The development of visual object recognition in school-age children. *Developmental Neuropsychology, 31*, 79–102.

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research, 51*, 1588–1609.

Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual-search tasks. *Perception & Psychophysics, 51*, 465–472.

Brown, P. (2008). Verb specificity and argument realization in Tzeltal child language. In M. Bowerman & P. Brown (Eds.), *Cross-linguistic perspectives on argument structure: Implications for learnability* (pp. 167–189). Mahwah, NJ: Lawrence Erlbaum.

Bushmakin, M. A., & James, T. W. (2014). The influence of feature conjunction on object inversion and conversion effects. *Perception, 43*, 31–42.

Caparos, S., Ahmed, L., Bremner, A., De Fockert, J. W., Linnell, K. J., & Davidoff, J. (2012). Exposure to an urban environment alters the local bias of a remote culture. *Cognition, 122*, 80–85.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America, 102*, 12629–12633.

Cole, P. M., Bruschi, C. J., & Tamang, B. L. (2002). Cultural differences in children's emotional reactions to difficult situations. *Child Development, 73*, 983–996.

Coll, C. G. (2005). Predicting individual differences in attention, memory, and planning in first graders from experiences at home, child care, and school. *Developmental Psychology, 41*, 99–114.

Davidoff, J., Fonteneau, E., & Goldstein, J. (2008). Cultural differences in perception: Observations from a remote culture. *Journal of Cognition and Culture, 8*, 189–209.

Davidoff, J., & Roberson, D. (2002). Development of animal recognition: A difference between parts and wholes. *Journal of Experimental Child Psychology, 81*, 217–234.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107–117.

Duffy, S., Toriyama, R., Itakura, S., & Kitayama, S. (2009). Development of cultural strategies of attention in North American and Japanese children. *Journal of Experimental Child Psychology, 102*, 351–359.

Farver, J. M., Kim, Y. K., & Lee, Y. (1995). Cultural differences in Korean- and Anglo-American preschoolers' social interaction and play behaviors. *Child Development, 66*, 1088–1099.

Fenson, L., Dale, P., Reznick, J. S., Bates, E., Hartung, J., Pethick, S., et al (1993). *MacArthur communicative development inventories*. San Diego: Singular Publishing.

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in Japanese and American mothers' speech to infants. *Child Development, 64*, 637–656.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees. *Nature, 382*, 626–628.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1997). Neural mechanisms involved in the processing of global and local aspects of hierarchically organized visual stimuli. *Brain, 120*, 1779–1791.

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex, 23*, 814–823.

Friedlmeier, W., & Trommsdorff, G. (1999). Emotion regulation in early childhood: A cross-cultural comparison between German and Japanese toddlers. *Journal of Cross-Cultural Psychology, 30*, 684–711.

Garner, W. R. (1974). *The processing of information and structure*.New York: Psychology Press.

Gartstein, M. A., Crawford, J., & Robertson, C. D. (2008). Early markers of language and attention: Mutual contributions and the impact of parent–infant interactions. *Child Psychiatry and Human Development, 39*, 9–26.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*, 191–197.

Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development, 49*, 988–998.

Goh, J. O. S., Hebrank, A. C., Sutton, B. P., Chee, M. W. L., Sim, S. K. Y., & Park, D. C. (2013). Culture-related differences in default network during visuo-spatial judgments. *Social Cognitive and Affective Neuroscience, 8*, 134–142.

Han, S. H., & Northoff, G. (2008). Culture-sensitive neural substrates of human cognition: A transcultural neuroimaging approach. *Nature Reviews Neuroscience, 9*, 646–654.

Hedden, T., Ketay, S., Aron, A., Markus, H. R., & Gabrieli, J. D. E. (2008). Cultural influences on neural substrates of attentional control. *Psychological Science, 19*, 12–17.

Imada, T., Carlson, S. M., & Itakura, S. (2013). East-West cultural differences in context sensitivity are evident in early childhood. *Developmental Science, 16*, 198–208.

Imai, M., Haryu, E., Okada, H., Lianjian, L., & Shigematsu, J. (2006). Revisiting the noun–verb debate: A cross-linguistic comparison of novel noun and verb learning in English-, Japanese-, and Chinese-speaking children. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 450–476). Oxford, UK: Oxford University Press.

Ishii, K., Tsukasaki, T., & Kitayama, S. (2009). Culture and visual perception: Does perceptual inference depend on culture? *Japanese Psychological Research, 51*, 103–109.

James, T. W., Huh, E., & Kim, S. (2010). Temporal and spatial integration of face, object, and scene features in occipito-temporal cortex. *Brain and Cognition, 74*, 112–122.

James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., & Goodale, M. A. (2000). The effect of visual object priming on brain activation before and after recognition. *Current Biology, 10*, 1017–1024.

Jüttner, M., Wakui, E., Petters, D., Kaur, S., & Davidoff, J. (2013). Developmental trajectories of part-based and configural object recognition in adolescence. *Developmental Psychology, 49*, 161–176.

Kelly, D. J., Miellet, S., & Caldara, R. (2010). Culture shapes eye movements of visually homogeneous objects. *Frontiers in Psychology, 1*. http://dx.doi.org/10.3389/fpsyg.2010.00006.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin, 112*, 24–38.

Kitayama, S., Duffy, S., Kawamura, T., & Larsen, J. T. (2003). Perceiving an object and its context in different cultures: A cultural look at new look. *Psychological Science, 14*, 201–206.

Kourtzi, Z., & DiCarlo, J. J. (2006). Learning and neural plasticity in visual object recognition. *Current Opinion in Neurobiology, 16*, 152–158.

Kovack-Lesh, K. A., McMurray, B., & Oakes, L. M. (2014). Four-month-old infants' visual investigation of cats and dogs: Relations with pet experience and attentional strategy. *Developmental Psychology, 50*, 402–413.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences, 17*, 26–49.

Kühnen, U., & Oyserman, D. (2002). Thinking about the self influences thinking in general: Cognitive consequences of salient self-concept. *Journal of Experimental Social Psychology, 38*, 492–499.

Kuwabara, M., & Smith, L. B. (2012). Cross-cultural differences in cognitive development: Attention to relations and objects. *Journal of Experimental Child Psychology, 113*, 20–35.

Kuwabara, M., Son, J. Y., & Smith, L. B. (2011). Attention to context: U.S. and Japanese children's emotional judgments. *Journal of Cognition and Development, 12*, 502–517.

Malt, B. C., & Majid, A. (2013). How thought is mapped into words. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*, 583–597.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224–253.

Martin, M. (1979). Local and global processing: The role of sparsity. *Memory & Cognition, 7*, 476–484.

Masuda, T., Ellsworth, P. C., Mesquita, B., Leu, J., Tanida, S., & De Veerdonk, E. V. (2008). Placing the face in context: Cultural differences in the perception of facial emotion. *Journal of Personality and Social Psychology, 94*, 365–381.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology, 81*, 922–934.

Masuda, T., & Nisbett, R. E. (2006). Culture and change blindness. *Cognitive Science, 30*, 381–399.

Masuda, T., Russell, M. J., Chen, Y. Y., Hioki, K., & Caplan, J. B. (2014). N400 incongruity effect in an episodic memory task reveals different strategies for handling irrelevant contextual information for Japanese than European Canadian. *Cognitive Neuroscience, 5*, 17–25.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6*, 255–260.

Miyamoto, Y., Nisbett, R. E., & Masuda, T. (2006). Culture and physical environment: Holistic versus analytic perceptual affordances. *Psychological Science, 17*, 113–119.

Miyamoto, Y., Yoshikawa, S., & Kitayama, S. (2011). Feature and configuration in face processing: Japanese are more configural than Americans. *Cognitive Science, 35*, 563–574.

Mondloch, C., Pathman, T., Maurer, D., Le Grand, R., & de Schonen, S. (2007). The composite face effect in six-year-old children: Evidence of adult-like holistic face processing. *Visual Cognition, 15*, 564–577.

Moriguchi, Y., Evans, A. D., Hiraki, K., Itakura, S., & Lee, K. (2012). Cultural differences in the development of cognitive shifting: East-West comparison. *Journal of Experimental Child Psychology, 111*, 156–163.

Nand, K., Masuda, T., Senzaki, S., & Ishii, K. (2014). Examining cultural drifts in artworks through development and history: Cultural comparisons between Japanese and Western landscape paintings and drawings. *Frontiers in Psychology: Cultural Psychology, 5*. http://dx.doi.org/10.3389/fpsyg.2014.01041.

Nisbett, R. E., & Masuda, T. (2003). Culture and point of view. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 11163–11170.

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Sciences, 9*, 467–473.

Nisbett, R. E., Peng, K. P., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*, 291–310.

Nishimura, M., Scherf, S., & Behrmann, M. (2009). Development of object recognition in humans. *F1000 Reports: Biology, 1*. http://dx.doi.org/10.3410/B1-56.

Ogura, T., Dale, P. S., Yamashita, Y., Murase, T., & Mahieu, A. (2006). The use of nouns and verbs by Japanese children and their caregivers in book-reading and toy-playing contexts. *Journal of Child Language, 33*, 1–29.

Ogura, T., & Watamaki, T. (1997). *Japanese communicative developmental inventories: User's guide and technical manual.*San Diego: Singular Publishing.

Palmeri, T., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews Neuroscience, 5*, 291–303.

Peissig, J. J., & Tarr, M. J. (2007). Visual object recognition: Do we know more now than we did 20 years ago? *Annual Review of Psychology, 58*, 75–96.

Pereira, A. F., & Smith, L. B. (2009). Developmental change in visual object recognition between 18 and 24 months of age. *Developmental Science, 12*, 67–80.

Poirel, N., Mellet, E., Houdé, O., & Pineau, A. (2008). First came the trees, then the forest: Developmental changes during childhood in the processing of visual local–global patterns according to the meaningfulness of the stimuli. *Developmental Psychology, 44*, 245–253.

Quinn, P. C., Tanaka, J. W., Lee, K., Pascalis, O., & Slater, A. M. (2013). Are faces special to infants? An investigation of configural and featural processing for the upper and lower regions of houses in 3- to 7-month-olds. *Visual Cognition, 21*, 23–37.

Rakison, D. H., & Butterworth, G. E. (1998). Infants' use of object parts in early categorization. *Developmental Psychology, 34*, 49–62.

Richler, J. J., Mack, M. L., Palmeri, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research, 51*, 333–342.

Richler, J. J., Palmeri, T. J., & Gauthier, I. (2012). Meanings, mechanisms, and measures of holistic processing. *Frontiers in Psychology, 3*. http://dx.doi.org/10.3389/fpsyg.2012.00553.

Robbins, R. A., Shergill, Y., Maurer, D., & Lewis, T. L. (2011). Development of sensitivity to spacing versus feature changes in pictures of houses: Evidence for slow development of a general spacing detection mechanism? *Journal of Experimental Child Psychology, 109*, 371–382.

Rousselet, G. A., Macé, M. J. M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision, 3*, 440–455.

Senzaki, S., Masuda, T., & Nand, K. (2014). Holistic versus analytic expressions in artworks: Cross-cultural differences and similarities in drawings and collages by Canadian and Japanese school-aged children. *Journal of Cross-Cultural Psychology, 45*, 1297–1316.

Smith, J. D., & Kemler, D. G. K. (1988). Is the more impulsive child a more holistic processor? A reconsideration. *Child Development, 59*, 719–727.

Smith, L. B. (2009). From fragments to geometric shape: Changes in visual object recognition between 18 and 24 months. *Current Directions in Psychological Science, 18*, 290–294.

Stevens, C., Lauinger, B., & Neville, H. (2009). Differences in the neural mechanisms of selective attention in children from different socioeconomic backgrounds: An event-related brain potential study. *Developmental Science, 12*, 634–646.

Tardif, T. (2006). But are they really verbs? Chinese words for action. In K. Hirsh-Pasek & R. M. Golinkoff (Eds.), *Action meets word: How children learn verbs* (pp. 477–498). Oxford, UK: Oxford University Press.

Tsai, Y., Meiran, N., & Lamy, D. (1995). Towards a resolution theory of visual attention. *Visual Cognition, 2*, 313–330.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences, 11*, 58–64.

Uskul, A. K., Kitayama, S., & Nisbett, R. E. (2008). Ecocultural basis of cognition: Farmers and fishermen are more holistic than herders. *Proceedings of the National Academy of Sciences, 105*, 8552–8556.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471–491.

Vales, C., & Smith, L. B. (2015). Words, shape, visual search, and visual working memory in 3-year-old children. *Developmental Science, 18*, 65–79.

Varnum, M. E., Grossmann, I., Kitayama, S., & Nisbett, R. E. (2010). The origin of cultural differences in cognition: Evidence for the social orientation hypothesis. *Current Directions in Psychological Science, 19*, 9–13.

Vurpillot, E. (1968). The development of scanning strategies and their relation to visual differentiation. *Journal of Experimental Child Psychology, 6*, 632–650.

Wakui, E., Jüttner, M., Petters, D., Kaur, S., Hummel, J. E., & Davidoff, J. (2013). Earlier development of analytical than holistic object recognition in adolescence. *PLoS One, 8*(4), 1–7.

Wang, H., Masuda, T., Ito, K., & Rashid, M. (2012). How much information? East Asian and North American cultural products and information search performance. *Personality and Social Psychology Bulletin, 38*, 1539–1551.

Wu, J., Osuntogun, A., Choudhury, T., Philipose, M., & Rehg, J. M. (2007). A scalable approach to activity recognition based on object use. In *Computer vision (ICCV): 2007 IEEE 11th international conference* (pp. 1–8). http://dx.doi.org/10.1109/ICCV.2007.4408865.

Wu, J., & Rehg, J. M. (2012). Object detection. In C. Zhang & Y. Ma (Eds.), *Ensemble machine learning: Methods and applications* (pp. 225–250). New York: Springer.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141–145.

Yoshida, H. (2012). A cross-linguistic study of sound symbolism in children's verb learning. *Journal of Cognition and Development, 13*, 232–265.