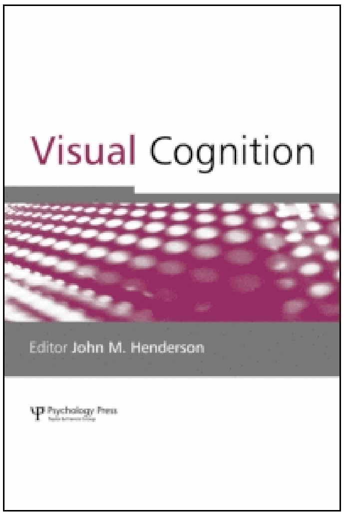

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683696

A cross-cultural study of the representation of shape: Sensitivity to generalized cone dimensions

Mark D. Lescroart [a]; Irving Biederman [a]; Xiaomin Yue [a]; Jules Davidoff [b]
[a] University of Southern California, Los Angeles, USA [b] Goldsmiths, University of London, UK

First Published on: 22 December 2008

To cite this Article Lescroart, Mark D., Biederman, Irving, Yue, Xiaomin and Davidoff, Jules(2008)'A cross-cultural study of the representation of shape: Sensitivity to generalized cone dimensions',Visual Cognition,

To link to this Article: DOI: 10.1080/13506280802507806

URL: http://dx.doi.org/10.1080/13506280802507806

VISUAL COGNITION, 0000, 00 (00), 1–17

Psychology Press
Taylor & Francis Group

A cross-cultural study of the representation of shape: Sensitivity to generalized cone dimensions

Mark D. Lescroart, Irving Biederman, and Xiaomin Yue
University of Southern California, Los Angeles, USA

Jules Davidoff
Goldsmiths, University of London, UK

Many of the phenomena underlying shape recognition can be derived from an assumption that the representation of simple parts can be understood in terms of independent dimensions of generalized cones, e.g., whether the axis of a cylinder is straight or curved or whether the sides are parallel or nonparallel. What enables this sensitivity? One explanation is that the representations derive from our immersion in a manufactured world of simple objects, e.g., a cylinder and a funnel, where these dimensions can be readily discerned independent of other stimulus variations. An alternative explanation is that genetic coding and/or early experience with extended contours—a characteristic of all naturally varying visual worlds—would be sufficient to develop the appropriate representations. The Himba, a seminomadic people in a remote region of Northwestern Namibia with little exposure to regular, simple artifacts, were virtually identical to western observers in representing generalized-cone dimensions of simple shapes independently. Thus immersion in a world of simple, manufactured shapes is not required for the development of a representation that specifies these dimensions independently.

Keywords: Shape perception; Multidimensional stimuli; Texture segregation; Cross-cultural research; Himba.

Any simple shape can be represented by a generalized cone (GC; Binford, 1971; Marr & Nishihara, 1978), which is the volume created by sweeping a cross section along an axis as, for example, when a circle is moved perpendicularly along a straight axis to produce a cylinder.

Please address all correspondence to Mark Lescroart, University of Southern California, 3641 Watt Way, Hedco Neuroscience Building, Los Angeles, CA 90089-2520, USA. E-mail: mark.lescroart@usc.edu

This research was supported by NSF 04-266415, 05-31177, and 06-17699 to IB and EU project SEDSU (012-984) to JD. Special thanks to Laurent Itti for his help with the classifier and Kemuu Jakurama, our guide.

http://www.psypress.com/viscog DOI: 10.1080/13506280802507806

Different volumes can be created through variations of independent GC dimensions, such as whether the axis is straight or curved, or whether the cross section remains constant in size. GCs assume central importance in parts-based accounts of object recognition (Biederman, 1987; Marr & Nishihara, 1978). It is one thing to show mathematically, as did Marr and Nishihara (1978), that any shape can be created by GCs; but are the dimensions that define GCs represented independently?

INDEPENDENT PROCESSING OF DIMENSIONS OF SHAPE

In both humans from the developed world and laboratory macaques, there is strong evidence that GC dimensions are encoded independently, in that selective attention to one dimension can be exercised without an effect of variations in another dimension. Such combinations of dimensions are said to be *separable* (Garner, 1974) or *analysable* (Shepard, 1964). Dimensions that cannot be treated independently, such as hue and saturation, are termed *integral* (Garner, 1974) or *nonanalysable* (Shepard, 1964), and selective attention to such combinations of dimensions is not efficient (Garner, 1974; Shepard, 1964).

With respect to the specific case of axis curvature and aspect ratio, Stankiewicz (2002) reported that University of Minnesota subjects could discriminate noisy variations in one of these GC dimensions, e.g., axis curvature, independently of the noise level on the other GC dimension, e.g., aspect ratio. Op de Beeck, Wagemans, and Vogels (2003) showed that the search slopes for a target that differed from the distractors in a value of a single dimension of either axis curvature or aspect ratio were less steep than when the distractors differed from the target in a conjunction of the values from both dimensions. For example, a low curvature target was more readily detected among high curvature distractors that varied in aspect ratio than when the target was defined as a low curvature–high aspect ratio shape and the distractor shapes were a mixture of high curvature–low aspect ratio shapes and low curvature–high aspect ratio shapes. A possible neural basis for this selectivity was discovered by Kayaert, Biederman, Op de Beeck, and Vogels (2005), who found that 95% of the variance of the firing of macaque IT cells to 2-D shapes could be accounted for by independent representation of GC dimensions.

POSSIBLE ROLE OF THE PRESENCE OF SIMPLE ARTIFACTS

However, in all of these studies the humans and the laboratory monkeys were raised in environments full of geometrically simple, manufactured

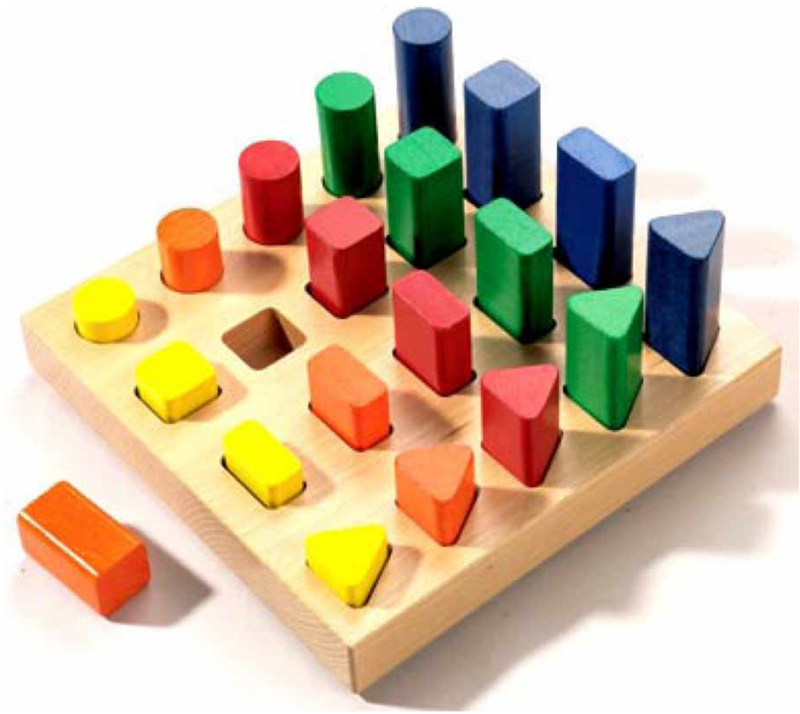

Figure 1. A "shape sorter" in which cross-section shape and aspect ratio (correlated with colour in this sorter) are varied independently (Haba Shape Sorter Board, from www.maukilo.com). To view this figure in colour, please see the online issue of the Journal.

objects, in which variation along single GC dimensions could be readily appreciated. For example, nails vary in aspect ratio and no other dimension, and pasta often varies in axis curvature and no other dimension. A popular toy for toddlers is a "shape sorter" (Figure 1) in which separate dimensions, cross-section shape and aspect ratio (correlated with colour for the toy in the image) are varied independently. It is possible that exposure to dimensional variation in such simple shapes facilitates the learning of independent shape dimensions. Supporting such a view are the reports of Schyns and Murphy (1994) and Schyns and Rodet (1997) suggesting that the features that we use for responding to our visual world are not fixed but flexible, reflecting our categorization needs. Consistent with the idea of encoding flexibility, Goldstone (1994) showed that humans can learn to perform fine judgements in one dimension of an integral combination of dimensions (brightness and saturation) without strongly affecting discrimination performance in the other.

Would individuals from a culture with only limited exposure to developed-world artifacts show the same independence of shape dimensions evidenced by the typical artifact-immersed laboratory subject? If dimensions are defined flexibly according to the needs of a culture, might the combination of dimensions that westerners represent separately not be so represented by people from a markedly different culture?

There is general belief that that the early cortical stages of the visual system have evolved (or have developed during infancy) in response to the statistics of the images that characterize the visual world (e.g., Baddeley & Hancock, 1991). For example, there is a 1/f relation between Fourier amplitude and spatial frequency in natural scenes, and intensity values for adjacent pixels in natural scenes are highly correlated. These spatial (or Fourier-like) statistics seem to characterize not only natural scenes, but artifactual environments as well (Switkes, Mayer, & Sloan, 1978; Tadmor & Tolhurst, 1994), and tuning properties strikingly similar to V1 cells' receptive fields can be derived from these statistics and a few simple assumptions (Olshausen & Field, 1996). Consequently one would expect little difference in the early coding stages of individuals either immersed or not immersed in a developed-world visual environment, and no one has proposed such differences.

What about later stages of processing? The sensitivity of human psychophysical performance and the tuning of macaque cells in more anterior visual areas cannot be predicted from the Fourier-like tuning properties evident at earlier stages of processing. Specifically, Fourier tuning does not make generalized cone dimensions explicit and thus cannot account for the tuning to GC dimensions evident in the tuning of IT cells (Kayaert et al., 2005) and human psychophysics (Op de Beeck et al., 2003; Stankiewicz, 2002). An example of non-Fourier tuning is the finding of Pasupathy and Connor (1999) that approximately 12% of the cells in V4 of the macaque respond to L-vertices at a particular orientation and angle (Pasupathy & Connor, 1999). These cells are unresponsive to either the angle bisector or the individual legs of the vertex, effects that would be expected from Fourier-like tuning. Another example is that macaque IT cells and human psychophysics demonstrate greater sensitivity to nonaccidental compared to metric variations in shape (Biederman & Bar, 1999; Kayaert, Biederman, & Vogels, 2003). The standard statistical analyses of Fourier components make none of this coding explicit. The origins of higher level perceptual categories are still unclear, and it is entirely possible that influences other than low-level Fourier-based image statistics (e.g., immersion in an environment filled with simple shapes, cultural emphasis, cognitive demands), could affect the human representation of shape.

A parallel can be drawn with speech perception. Although there is no evidence that the basic frequency-sensitivity tuning characteristics of early stages of audition differ from culture to culture, the particular set of phonemes that can be readily discriminated does vary with the particular language experienced in childhood. Children only retain the ability to hear the phonemic contrasts that convey semantic information.

THE HIMBA

The Himba are a seminomadic people living in a remote region of northwestern Namibia. Figure 2a and 2b show scenes typical of the Himba environment. In the more remote encampments, the Himba have little exposure to simple, modern artifacts and thus provide an opportunity to assess the effects of the presence of such artifacts (or lack thereof) on the representation of shape. We are not assuming that there is less variation in GC dimensions in the Himba's visual world. The issue is whether the exposure to simple shapes in which the dimensions are clearly contrasted, as in the shape sorter (Figure 1), facilitates the representation of the dimensions as *independent* dimensions.

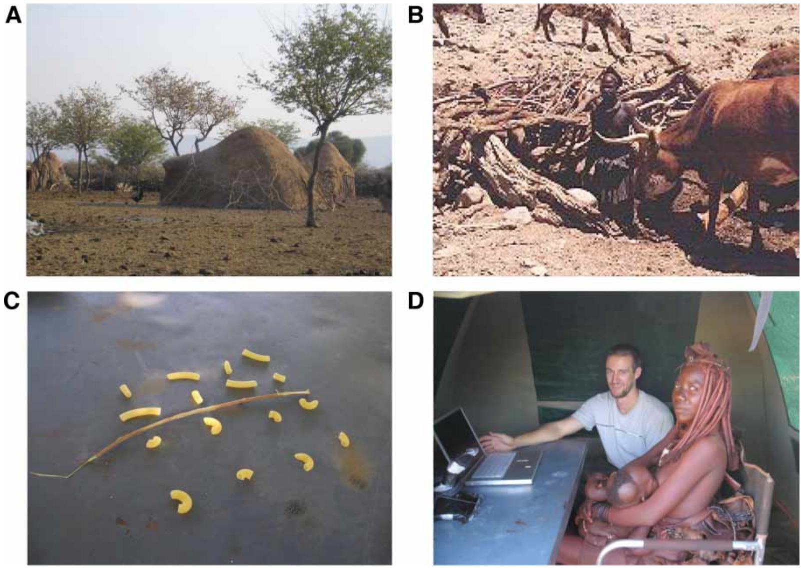

Figure 2. (a) Himba village showing dung and stick dwellings. (b) Watering hole. (c) Illustration of Himba training procedure with real macaroni and stick to indicate texture field boundary (slightly curved on top, highly curved on bottom), (d) Himba subject (and son) with experimenter (MDL). To view this figure in colour, please see the online issue of the Journal.

TEXTURE SEGREGATION

To examine independent GC coding we employed a texture segregation task as shown in Figure 3. Tasks such as the one illustrated in Figure 3a–c have been used to assess whether dimensions such as luminance and shape are independently coded (e.g., Bach, Schmitt, Quenzer, Meigen, & Fahle, 2000). The subject has to report whether the boundary between luminance and/or shape regions is vertical or horizontal. The boundary is always on either side of the middle row (if horizontal) or column (if vertical), so there is some uncertainty as to its location. In both Figure 3a and 3b the boundary is rapidly and effortlessly perceived. However, in Figure 3c, in

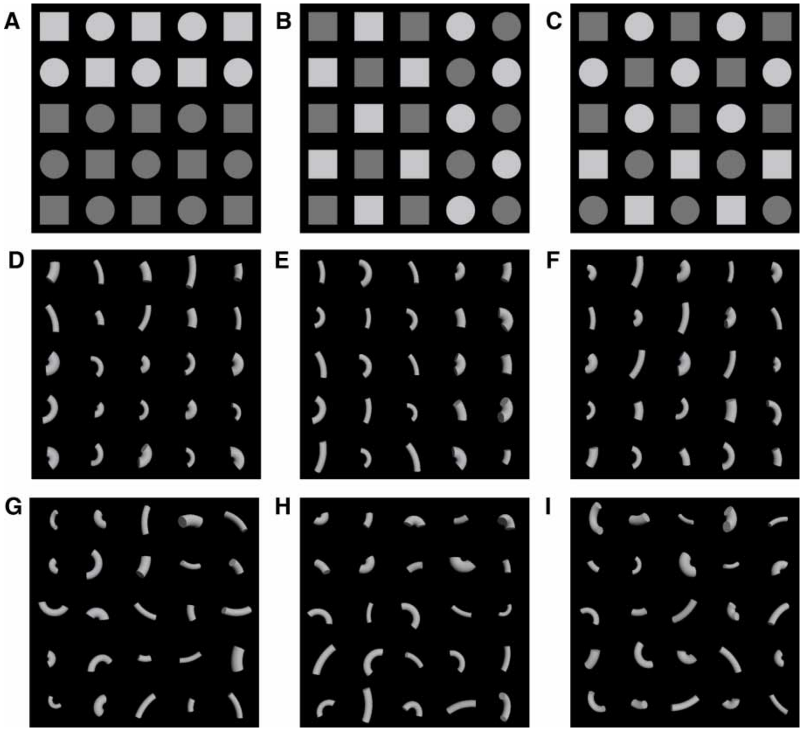

Figure 3. Illustration of texture segregation tasks. (a–c) Luminance and shape (not used in the experiment but shown here to illustrate conjunction costs in texture segregation). The boundary in (c) is horizontal, between Rows 3 and 4. (d–f) Examples of displays from the low-variability task. The boundary in (d) is defined by axis curvature (and is horizontal, between Rows 2 and 3); in (e) by aspect ratio (boundary is vertical, between Columns 3 and 4), and in (f) by axis curvature + aspect ratio (conjunction) (boundary is horizontal, between Rows 3 and 4). (g–i) Examples of displays from the high-variability task. In (g) the boundary is defined by axis curvature (boundary is vertical, between Columns 2 and 3), in (h) by aspect ratio (boundary is horizontal, between Rows 2 and 3), and in (i) by axis curvature + aspect ratio (conjunction) (boundary is vertical, between Columns 2 and 3).

which the texture fields are defined by a conjunction of luminance and shape, scrutiny is required. At first glance, the conjunction condition seems so different from the other two that the underlying similarity of the displays is not obvious. However, in all three panels, each texture field contains two of the four elements (darker and lighter circles and squares). So why should Figure 3c be more difficult? In Figure 3a and 3b, the elements on each side of the border differ in the values of one dimension, while the values of the other dimension vary across the whole display. If the relevant and irrelevant dimensions are separable (i.e., represented independently), then selective attention can be employed to respond only to the relevant dimension. In Figure 3c, both values of each dimension are on either side of the border so selective attention to one of the dimensions would not help segregate the fields. Because the border is defined by a conjunction of values, both dimensions must be processed and the task would be expected to be more difficult than the single dimension tasks in Figure 3a and 3b. It is only by virtue of a dimensionalized representation that conjunction tasks would be expected to be more difficult than the single dimension tasks.

It seems obvious that shape and luminance (Figure 3a–c) would be represented as independent dimensions. But what about different dimensions of shape itself? We used the texture-segregation tasks illustrated in Figure 3d–i to determine whether University of Southern California (USC) students, individuals immersed in the artifacts of the developed world, and the Himba would show independent representation of two generalized cone metric dimensions: Degree of axis curvature and aspect ratio. These dimensions are the best-studied examples of GC dimensions, and were used in the studies of Stankiewicz (2002), Op de Beeck et al. (2003), and Kayaert et al. (2005), among others. The dimensions also allowed effective variation of rotation in depth and the plane to eliminate a contribution of low-level cues of orientation and luminance. Also, concrete examples of shapes based on these dimensions (i.e., macaroni) were readily available for instructional purposes.

Why might the Himba, in contrast to people from the developed world, not represent shape dimensions independently? *Every* object that all people see will have some width and, to the extent that an axis can be ascribed to the object (or object part), some value of axis curvature. The issue under test, however, was not whether the individual attributes could be discriminated but whether, when varied in *combination*, variations in one attribute could be selected and the other ignored. As noted previously, developed-world environments provide frequent exposure to *simple* manufactured objects, or simple object parts, that vary in only a single dimension, such as aspect ratio, e.g., nails, pens, and soup cans, or only in axis curvature, such as coiled power cords and pasta, or where the dimensions are explicitly varied independently, as in shape sorters. We thus have more opportunity than the Himba for discrimination training on one of these dimensions, independently of the other. In addition, developed-world language and classroom instruction may allow us to express and selectively attend to these variations, whereas the Himba language, Otjiherero, provides a more limited vocabulary of shape terms (Viljoen & Kamupingene, 1983).

Might the Himba learn to encode dimensions independently through their exposure to, say, the aspect ratios of tree trunks or goats' legs? Possibly, but since many other attributes vary simultaneously in such examples, learning would be expected to be more difficult. Furthermore, if it were found that the Himba did differ from westerners in any aspect of their representation of shape, an obvious explanation would be based on the difference in visual environments.

If exposure to simple artifacts facilitated the learning of independent shape dimensions, we would expect that the Himba would process the single dimension tasks more like conjunctions, i.e., as integral combinations of dimensions, so there would be little or no advantage for the single dimension tasks.

METHOD

Logistics

The experimenter (MDL) flew to Windhoek, Namibia's capital, and undertook a two-day drive in an off-road capable vehicle to a township (Opuwo) at the edge of Himba territory, to meet the guide and obtain provisions. Because of ecotourism, which brings the Himba in ever greater contact with modern artifacts, it was necessary to go to even more remote regions, at least a full day or two's drive from Opuwo, to search for current encampments. These remote Himba still do have occasional interaction with traders bringing blankets, water jugs, and western clothes, and NGOs providing health services. They have no electronics of any kind, no western tools, no running water, and no furniture but rocks on which to sit and thick blankets for bedding. The experiment provided their first exposure to a computer. Upon encountering an encampment, the guide would approach the village chief and ask permission to camp on the outskirts of the compound and have members of the tribe participate in the experiment. Given that the guide could only facilitate one experiment at a time (two separate experiments were being run), with a number of subjects unable or unwilling to complete the experiment, a good "yield" would be one or two subjects per day.

Task

To test sensitivity to underlying shape dimensions we employed the texture segregation task illustrated in Figure 3d–i. The task was exactly the same as in Figure 3a–c, but instead of colour and shape, the texture elements varied on the metric dimensions of aspect ratio and axis curvature. As a result, our stimuli resembled macaroni noodles. The four different elements of each display were (informally): (1) Narrow, highly curved cylinders, (2) wide, highly curved cylinders, (3) narrow, slightly curved cylinders, and (4) wide, slightly curved cylinders. The radii of curvature for the slightly curved cylinders were 68 and 100 pixels for the narrow and wide elements, respectively; for the highly curved cylinders, the radii of curvature were 20 and 29 pixels, for the narrow and wide cylinders, respectively. Each narrow cylinder had an aspect ratio of 1:4, and each wide cylinder had an aspect ratio of 1.1:2. Each display was composed of 5×5 elements, divided into two regions, each with two types of cylinders. Subjects judged, as quickly and as accurately as possible, whether the boundary between the two regions was vertical or horizontal.

There were three possible ways to define the boundary: (1) By axis curvature (highly curved vs. slightly-curved), (2) by aspect ratio (wide vs. narrow), or (3) by a combination of the aspect ratio and axis curvature (narrow–highly curved and wide–slightly curved on one side vs. narrow–slightly curved and wide–highly curved on the other). In each of the first two conditions, subjects could perform the task based on only one GC dimension; in the third, the conjunction condition, they had to use information from two dimensions simultaneously. Each subject's sequence of trials was composed of all three conditions presented in pseudorandom order.

Stimuli

The texture field of 25 display elements (cylinders) spanned a square of 600×600 pixels on a 1024×768 pixel screen. The subjects sat approximately 0.66 m from the screen so the whole square subtended a visual angle of approximately 10.7°. The centres of the 25 display elements were evenly spaced but variations in size and planar and depth orientation produced some variability in the interelement distances. The average size of each display element at 0° orientation in depth subtended a visual angle of approximately 1.5°.

To deter participants from basing their decisions on low-level cues such as local or global orientation or pixel intensity values, we varied both the orientation and size of the elements. In the low- (high-)variability condition shown in Figure 3d–f (g–i), we randomly rotated each cylinder over 22.5° (360°) in the plane and, in depth, up to 22.5° (45°). All images were rendered in perspective projection, and after rendering, the size (in pixels) of each cylinder was randomly varied by 25% (33%) independent of the size variation from the depth rotation.

Stimulus presentation, response recording, and feedback were done on a Macintosh Powerbook G4 computer with a 15-inch screen. The stimulus presentation code was written using Psychtoolbox3 for Matlab (Brainard, 1997; Pelli, 1997).

Training

The complexity of the task required a thorough explanation and training procedure for both groups. For the Himba, the experimenter (with the help of the translator) first illustrated the task using actual macaroni noodles (Figure 2c). Subjects were asked to divide the array of noodles with a stick, keeping the shapes that were "the same" on the same side. Once this was grasped (usually after three or four trials), they then moved on to a practice sequence on the computer. Initially using a stick placed across the display, as in the training trials, they were taught to swipe the touchpad in the same (projected) orientation as that of the stick. After their response a line would appear on the display indicating the correct location of the border. Subjects were not required to distinguish between the two possible locations for a divide. Subjects would continue training until they correctly responded on seven out of eight consecutive trials, or until they completed 40 trials (at which point it was judged that they did not understand the task, and they were excluded from the experiment).

SUBJECTS

A total of 32 Himba (16 female, approximate mean age 25.1 years) and 9 USC subjects (7 female, mean age 20.5 years) participated in the experiment. (The Himba are uncertain as to their ages.) Himba were compensated with 0.5 kg of maize (corn meal) per hour tested. USC subjects received participation course credit or were compensated $8 for their time and effort.

Twelve Himba subjects (4 females, approx. mean age 21.7 years) and seven USC subjects (6 females, mean age 20.3 years) were included in the final analysis. Two of the Himba ran in both the low- and high-variability versions of the task, so a total of 14 Himba sessions were analysed. All USC subjects performed both versions of the task, so 14 USC subject sessions were analysed.

The data from one Himba subject was lost due to battery failure, and the data from 21 other USC and Himba subjects were excluded for a variety of reasons, including failure to meet training criterion on the low-variability task (4 Himba, 2 USC), failure to meet training criterion on the high-variability task (7 Himba), voluntarily quitting before half the trials were completed (4 Himba), different (pilot) testing conditions (3 Himba), and excessive westernization (1 Himba). *None of these individuals were excluded for failure to show a conjunction cost*, and the data from these subjects (only a few training trials in many cases) were in the same direction as the data from those who completed the experiment.

There are several reasons for the higher attrition rate among the Himba than among the westerners. Social customs required that some of the older Himba (three of the 19) be allowed to attempt the experimental task, even though two were incapable of performing above chance and the third older subject quit after less than half a session. Also, in accordance with USC Institutional Review Board requirements and to maintain harmonious relations with a village, all subjects were compensated with maize, whether they completed the experiment or not (which likely contributed to the higher drop-out rate). It was made very clear to subjects that they could quit at will, and given that the testing was repetitive and entirely outside their experience, it is a testament to their perseverance that only four chose to quit. The three Himba excluded for different testing conditions were run early in the investigation, while piloting appropriate testing procedures, i.e., whether to test inside a tent, in the dark and relative isolation, but often uncomfortable heat, or outside the tent, and deciding which response device to use (i.e., joystick or touchpad) and reasonable stimulus noise levels. One subject was excluded for excessive westernization, since during the day he was tested, it became obvious that many of the children in his village had been to a new nearby school, built since the guide's last trip to the village several years earlier.

For the first two weeks of the investigation, no runs of the high-variability version of the experiment were collected, due to a high degree of scepticism from a senior author who had worked with the Himba on several prior investigations and our guide as to whether the Himba would be able to perform the task at all. Thus the priority, in the second two weeks of training, was to run subjects in the high-variability version of the task, and subjects were advanced on to that task without first performing complete trials of the low-variability task (as there was a significant risk of losing subjects to goat-herding responsibilities). In those two weeks, seven subjects passed the training threshold for the low-variability task, but did not pass the threshold for the high-variability task.

Those 12 Himba who did complete the training/criterion phase required an average of 24 trials (mean of 10.8 minutes on the computer). Seven USC subjects successfully performed the same training procedure, although they were not trained with the real macaroni phase. The USC subjects required an average of 19 trials (5.1 minutes). It should be noted that the increased training time for the Himba included translation lags and familiarization with what was, for all of them, their first experience with a computer.

Himba were given 72 and western subjects 108 trials per condition. For 750 ms after the subject responded on every trial, a coloured line appeared over the actual position of the texture boundary as feedback. Green indicated a correct response, red incorrect. There was no question that the Himba understood the feedback: They showed obvious signs of displeasure at incorrect responses.

Conditions of testing were not completely comparable between Himba and USC subjects: None of the USC subjects breastfed their infants while performing the task (Figure 2d), nor did a noisy goat ever attempt to enter the USC testing room.

RESULTS

Reaction times and error rates were analysed in a mixed 2×3 repeated-measures analysis of variance, with factors tribe (Himba vs. USC, a between subjects, unequal Ns variable) and condition (axis curvature [single-dimension], aspect ratio [single dimension], and conjunction, a within-subjects variable). Separate ANOVAs were run for the high- and low-variability tasks although primary discussion will be on the high-variability tasks as these were better controlled for low-level features that could have produced an artifactual conjunction cost (as described in the Classifier Analysis section).

Given the differences in testing conditions noted earlier, as well as the complete unfamiliarity with the experience for the Himba and possible general ability differences, it was not surprising that error rates and RTs were higher for the Himba than the USC subjects, although only by 12.2% and 1.37 s. (Figure 4). These differences were significant; for error rates, $F(1, 10) = 5.06$, $p < .05$, and for RTs, $F(1, 10) = 5.31$, $p < .05$.

The primary interest of this investigation was whether the Himba would be able to selectively attend to a single dimension so that their RTs and error rates in the single dimension condition would be reduced compared to the conjunction condition. Both groups had reliably lower error rates and RTs when the boundary was defined by a single dimension (either aspect ratio or axis curvature) compared to when the boundary was defined by a conjunction of the two dimensions (Figure 4): For error rates, $F(2, 20) = 41.76$, $p < .001$, $\eta_p^2 = .81$; for RTs, $F(2, 20) = 56.14$, $p < .001$, $\eta_p^2 = .85$. In fact, the mean of the two single-dimension conditions, in both RTs and error rate,

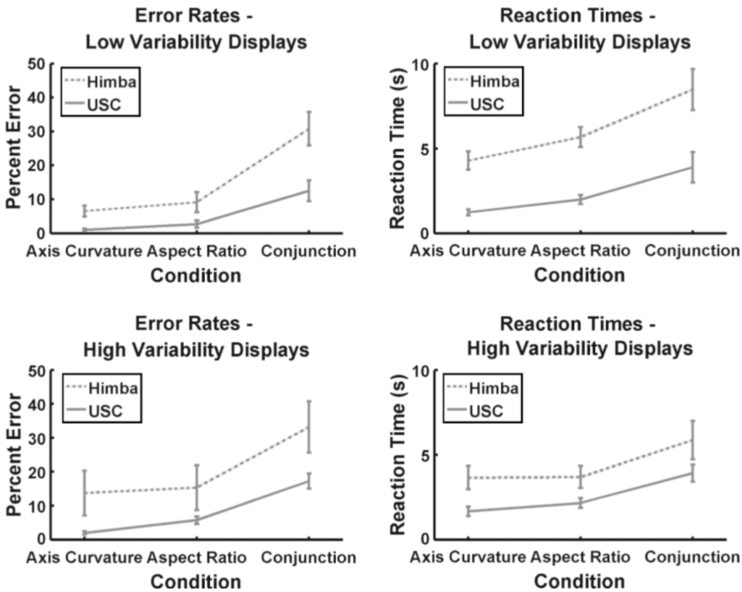

Figure 4. Mean percentage errors and correct reaction times for the Himba and USC subjects for the single and conjunction conditions in the low- and high-variability tasks.

was lower than the conjunction condition for every subject in the experiment. Moreover, the magnitude of the advantage of the single dimension conditions was comparable for both tribes, yielding nonsignificant interactions between tribe and conditions for error rates, $F(2, 20) = 1.24$, $p = .31$, $\eta_p^2 = .11$, and for RTs, $F(2, 20) = 0.612$, $p = .55$, $\eta_p^2 = .058$. A Tukey HSD test verified the greater difficulty of the conjunction condition compared to both single dimension conditions ($p < .01$ for both), but found no significant difference between the two single-dimension conditions.

Because only two of the Himba were run on both the high- and low-variability display conditions, whereas all of the western subjects performed both levels of the task, a single ANOVA encompassing all the data at both noise levels could not be run. The reported *F*-values are those for the high-variability displays. An ANOVA run on the low-variability displays (and the data from those subjects who quit only part-way through, or for whom we only have training data) gave the same picture: All subjects showed a conjunction cost. For the excluded subjects, the single-dimension conditions' mean error rate was 23.9% and the mean RT was 10.20 s. For the conjunction condition, the mean error rate was 42.8% and the mean RT was 15.90 s.

To investigate the possibility that the Himba quickly learned the dimensions of axis curvature and aspect ratio during the course of the experiment, we compared the first half of the trials to the second half to see if the conjunction costs increased over the course of the session. They did not. In fact the opposite was the case: The difference between the averaged single dimension conditions and the conjunction condition actually *diminished* from the first to the second half of the trials, being 21.2% for errors (RT = 2.59 s) in the first half and 16.2% (RT = 1.52 s) in the second half.

It could be the case that the Himba learned to separately encode the dimensions of aspect ratio and axis curvature within the first few minutes of task instruction. But if 10 minutes of training will produce an effect equal to a lifetime of increased exposure to simple shapes, then that, too, speaks to the primacy of GC dimensions in the neural representation of shape.

The greater difficulty of the conjunction condition is presumed to be a result of having to attend to two (rather than one) independent "high-level" shape dimensions, axis curvature and aspect ratio. However, if the border in the single dimension conditions could be defined by low-level, nonshape cues, either in orientation or average pixel intensity values, then the greater difficulty of the conjunction condition could be trivially explained by the unavailability of such cues in that condition. Since Fourier statistics, which encompass simple local features like orientation and pixel intensity, have been shown to be essentially identical in both natural and artificial environments (Tadmor & Tolhurst, 1994), it is essential to verify that the task could not be done using only such "preshape" information.

Classifier analysis

To test whether the low-level cues of luminance and orientation could be the source of the difference between the conditions, we created a classifier that performed the task based solely on orientation and intensity. The classifier used a subset of the Itti and Koch (2000) model—the feature channels that compute local orientation (four orientations at six scales) and intensity—to process each of the 5×5 experimental texture displays. The channels are based on generally accepted quantitative estimates of early visual filtering in both domains.

Three different decision schemes for the classifier were modelled: One in which the classifier chose the divide that gave the greatest difference in mean intensity or orientation, one in which it chose the divide that had the greatest difference in the variance in intensity or orientation, and one in which it combined the mean and variance in orientation or intensity of each side of each possible divide into a two-dimensional vector, and chose the divide that gave the greatest Euclidean distance between the vectors. These decision schemes represent simple ways of making use of low-level image information to do the task (i.e., "Does local orientation vary more on one side than another?" rather than "Does one side have greater axis curvature than the other?"). The model that used a vector consisting of both mean and variance performed slightly more accurately than the other two, so further discussion will refer to that model. As with humans, if the classifier correctly chose vertical, but chose the wrong vertical divide (i.e., between Columns 2 and 3 when the correct divide was between Columns 3 and 4), it was credited with a correct response. The classifier ran 100 trials of each condition.

For the high-variability displays, the orientation-based classifier showed no significant difference between its error rates on any of the three conditions (for all comparisons between conditions, bootstrapped $p > .05$). The intensity-based classifier produced as many errors for the axis curvature condition as the conjunction condition, with the aspect ratio condition associated with the fewest errors. The ordering of conditions for the classifier was thus inconsistent with the results shown in Figure 4. For the low-variability displays, the ordering of the conditions by the classifier did match the ordering of the human subjects, but the difference was smaller than that observed in the human subjects. The potential availability of a low-level, nonshape cue (primarily orientation) in the low-variability condition justifies our variation of orientation and size in the high-variability condition. Consequently, we conclude, especially for the high variability conditions, that neither low-level differences in pixel intensity nor differences in orientation could explain the ordering of conditions.

DISCUSSION

Every Himba and USC subject showed a significant behavioural cost when they had to perform a task based on a conjunction of generalized cone dimensions rather than on a single dimension. The displays for all three conditions contained exactly the same four elements, with two of the elements on either side of the border. The classifier ruled out an effect of luminance and orientation differences across the border. Consequently, nothing in the displays themselves would necessitate the greater difficulty of determining the boundary in the conjunction condition. *It is only by the coding of the display elements as independent dimensions* over which selective attention can be exercised that the advantage of the single dimension over the conjunction conditions can be understood. The experiment offers, to our knowledge, the most rigorous assessment of the effects—or lack thereof—of exposure to modern artifacts on the underlying dimensions of the representation of shape.

We attribute the decreased accuracy and longer reaction times of the Himba to the differential testing conditions already mentioned, as well as to their lack of experience with psychophysical testing (none of our subjects had ever seen a computer before, much less used one). Many Himba subjects, seemingly chagrined that they had missed more than they felt they should, told the experimenter (through the translator), "I'm not used to this." An additional factor could be the differences in general ability, known to affect performance on such tasks (Ree & Carretta, 1994). We also note that there was no question that not only could the Himba readily appreciate the shape of the images on the screen, they also appreciated that those shapes could be projections of real-world 3-D objects. One Himba went so far—jokingly—as to accuse the experimenter of wasting food by placing the macaroni noodles inside the computer, where he could not eat them!

The bottom line is that the Himba's pattern of responses did not differ from that of individuals living in what is, arguably, the most artifactual of environments (Los Angeles). The sensitivity of both the Himba and USC students to underlying dimensions of generalized cones suggests that such sensitivity does not require immersion in a regular, manufactured environment but, instead, is likely a consequence of non-Fourier statistics of shape, determined through genetics or early infancy, that characterize virtually any visual world. These constraints would presumably be incorporated into the tuning of later, shape-selective stages of the ventral pathway.

REFERENCES

Bach, M., Schmitt, C., Quenzer, T., Meigen, T., & Fahle, M. (2000). Summation of texture segregation across orientation and spatial frequency: Electrophysiological and psychophysical findings. *Vision Research*, *40*(26), 3559–3566.

Baddeley, R. J., & Hancock, P. J. (1991). A statistical analysis of natural images matches psychophysically derived orientation tuning curves. *Proceedings of the Royal Society B: Biological Sciences, 246*(1317), 219–223.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147.

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Research*, *39*(17), 2885–2899.

Binford, T. O. (1971). *Visual perception by computer*. Paper presented at the Proceedings of the IEEE conference on Systems and Control, Miami, FL.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Lawrence Erlbaum Associates, Inc.

Goldstone, R. (1994). Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General*, *123*(2), 178–200.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506.

Kayaert, G., Biederman, I., Op de Beeck, H. P., & Vogels, R. (2005). Tuning for shape dimensions in macaque inferior temporal cortex. *European Journal of Neuroscience*, *22*(1), 212–224.

Kayaert, G., Biederman, I., & Vogels, R. (2003). Shape tuning in macaque inferior temporal cortex. *Journal of Neuroscience*, *23*(7), 3016–3027.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *200*(1140), 269–294.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607–609.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2003). The effect of category learning on the representation of shape: dimensions can be biased but not differentiated. *Journal of Experimental Psychology: General*, *132*(4), 491–511.

Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, *82*(5), 2490–2502.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Ree, M. J., & Carretta, T. R. (1994). The correlation of general cognitive ability and psychomotor tracking tests. *International Journal of Selection and Assessment*, *2*, 209–216.

Schyns, P. G., & Murphy, G. L. (1994). The ontogeny of part representation in object concepts. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 31, pp. 305–349). San Diego, CA: Academic Press.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 681–696.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, *1*, 54–87.

Stankiewicz, B. J. (2002). Empirical evidence for independent dimensions in the visual representation of three-dimensional shape. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(4), 913–932.

Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research*, *18*(10), 1393–1399.

Tadmor, Y., & Tolhurst, D. J. (1994). Discrimination of changes in the second-order statistics of natural and synthetic images. *Vision Research*, *34*(4), 541–554.

Viljoen, J. J., & Kamupingene, T. K. (1983). *Otjiherero Woordeboek, Dictionary, Embo Romambo*. Windhoek, Namibia: Gamsberg.

Manuscript received February 2008
Manuscript accepted September 2008
First published online month/year