

Running head: CULTURAL DIFFERENCES IN PERCEPTUAL REORGANIZATION

Cultural differences in photo-triggered perceptual reorganization

Jennifer Yoon¹, Nathan Witthoft², Jonathan Winawer¹, Michael C. Frank², Daniel Everett⁴, &
Edward Gibson³

¹Department of Psychology, New York University

²Department of Psychology, Stanford University

³Brain and Cognitive Sciences, MIT

⁴Department of Sociology, Bentley University

Author note

This research was supported by the European Commission project, Characterizing Human Language by Structural Complexity. We thank the Pirahã for their participation, and Keren Madora, Katie O'Neil, and Sonia Poltaratski for their help in data collection and analysis. These data were reported to the Cognitive Science Society in Yoon et al. (2011).

Correspondence should be addressed to Jennifer Yoon, Department of Psychology, New York University, New York, NY 10003, USA. Email: davie.yoon@nyu.edu. Phone: 650-455-9023.

ABSTRACT

Visual illusions and other perceptual phenomena can be used as tools to uncover the otherwise hidden constructive processes that give rise to perception. Although many perceptual processes are assumed to be universal, variable susceptibility to certain illusions and perceptual effects across populations suggests a role for factors that vary culturally. One striking phenomenon is the deficient recognition of two-tone images—photos that have been subjected to a simple transformation—in young children under conditions that trigger automatic recognition in adults. Here we show a similar lack of photo-triggered perceptual reorganization in the Pirahã, a hunter-gatherer tribe with limited exposure to modern visual media, suggesting such recognition is experience- and culture-specific.

INTRODUCTION

A core principle of vision science is that perception is not simply a passive reflection of the external world, but a process of constructive interpretation of inherently ambiguous input. Consider a shadow projected onto a wall. The same silhouette can be created by different objects of different sizes at different distances from the viewer. Images projected onto the retina have the same inherent ambiguity, and a wide range of perceptual judgments ranging from lightness (Adelson, 1993), to color, to depth, to shape and identity, are the result of “unconscious inferences” by the visual system (Helmholtz, 1878). Such inferences are often presumed to be automatic and culturally universal (Gregory, 2005; Kohler, 1929; Spelke, 1990).

The interpretative processes that give rise to a coherent percept or “gestalt” often occur effortlessly and without awareness, but they can be made explicit by examining images that are not correctly interpreted upon initial viewing, such as the famous two-tone depiction of the Dalmatian in the snow (Gregory, 1970) or the two-tone ocelot in Figure 1 (right column, second row). People often fail to recognize the two-tone image; when shown the corresponding photograph, however, they find the two-tone often transforms suddenly into a coherent percept. Observers viewing the ocelot in the two-tone will often make figure-ground errors, incorrectly assigning some background regions to the figure, some figure regions to the background. Reconfiguring figure-ground assignments after viewing the photograph is to “reorganize” one’s initial grouping to achieve a different perceptual state (Kovacs & Eisenberg, 2004). If the viewer ultimately recognizes the previously unrecognized image, perception reorganization is said to have been successful.¹

¹ Following the Gestalt school, we use the terms “perceptual organization” and “perceptual reorganization” to emphasize the process by which local image features are appropriately integrated (‘grouped’) or segregated in order to arrive at a meaningful interpretation of the

Many of the principles underlying perceptual organization are thought to be universal, based on demonstrations of sensitivity to these principles even in very young infants and remote cultures (Pica, Jackson, Blake, Troje, 2011; Spelke, 1990). There is also a body of evidence reporting variable susceptibility to certain illusions and other perceptual phenomena across different populations. These results suggest an important role for culturally variable factors, including experience with artifacts such as photographs (Segall, Campbell, & Herskovits, 1966) and digital clocks (Whitaker & McGraw, 2000), culture-specific processing biases (de Fockert, Davidoff, Fagot, Parron, & Goldstein, 2007), and exposure to urban versus rural vistas (Leibowitz, Brislin, Perlmutter, & Hennessy, 1969; for review of older work, see Jones and Hagan, 1980).

Culturally invariant mechanisms of development such as the physiological maturation of the visual system predict differences in perception between children and adults. But children may also become more strongly enculturated into the practices of perceptual inference and interpretation accepted in their particular community over time, similarly predicting differences in how children and adults perceive the world (Vygotsky, 1978). One particularly striking phenomenon in perceptual development is the deficient recognition of two-tone images in young children under conditions that trigger automatic and effortless recognition in adults (Kovacs & Eisenberg, 2004; Yoon, Winawer, Witthoft, & Markman, 2007). When faced with images like Figure 1 (even ones containing familiar creatures), children—like adults—often struggle to recognize the animal. But unlike the adults, children have very significant difficulty recognizing the animal even when the two-tone image is placed side by side with the original picture.

One issue that is left unaddressed by these developmental data is whether the rapid

image—a “gestalt” (Kohler, 1929).

perceptual reorganization reported by adults is a necessary consequence of having a mature visual system, or whether it is the result of knowledge and experience acquired in a specific cultural context. To address this question, we turned to a population with very different visual experience than the participants in typical perception experiments: the Pirahã. The Pirahã are a hunter-gatherer tribe inhabiting a remote region of the Amazon. They are of particular interest for our current study because of their limited contact with modern visual media and their sparse material culture (Everett, 2005). Like young children in a modern industrial culture, Pirahã adults have little experience or knowledge of the visual transformation that links a photo and two-tone image. But unlike children, they possess both physiologically mature visual systems and a lifetime of experience with complex visual tasks such as hunting and fishing.

We tested Pirahã adults and English-speaking controls on their ability to recognize two-tone images given the corresponding photographs as cues (Figure 1). We predicted that, like children and U.S adults, the Pirahã would have difficulty recognizing two-tone images uncued, that is without seeing the accompanying photo. If expertise in interpreting symbolic visual materials is a key factor in photo-cued two-tone reorganization, then the Pirahã—like children but unlike U.S. adults—would have trouble recognizing the cued image even in the presence of the photo.

METHODS

Participants included adult members of the Pirahã tribe ($n = 9$, mean estimated age = 30y) and as controls tested with the same stimuli, Stanford University students, faculty, and staff ($n = 8$, mean age = 26y). An additional control task with additional stimuli was tested on Stanford students ($n = 10$, mean age = 19y). The visual acuity of the Pirahã population was tested by DE and others some years earlier as part of a basic screen for medical services; the population was

on the whole normal, with no cataracts and a small incidence of nearsightedness.

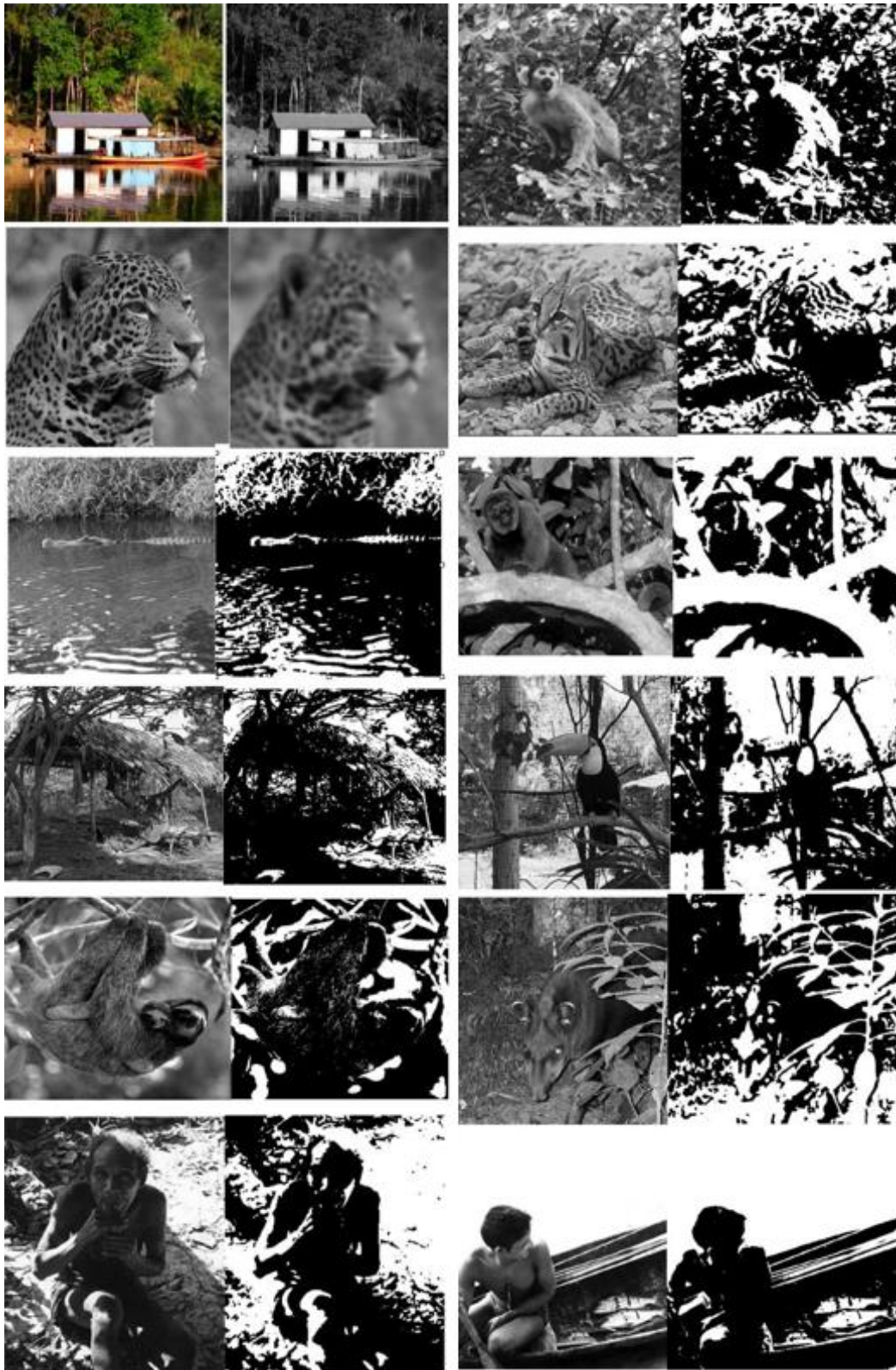


Figure 1. All stimuli used in the experiment. Left column from top to bottom: houseboat, jaguar, alligator, woman in hut, sloths, older man. Right column from top to bottom: squirrel monkey ocelot, howler monkey, toucan, tapir, fisherman. Cue items are shown to the left of test items. Houseboat and jaguar are warm-up items with simpler transformations. For full size stimuli, see online supplemental materials.

Ten two-tone images were created in Photoshop by blurring and posterizing (reducing the number of distinct gray scale values, in this case to two: black and white) grayscale photographs of animals and individuals found in the Pirahã participants' everyday environment (Figure 1). The amount of blur and the black/white threshold points were set independently for each photograph based on a repeated trial and error procedure until we were satisfied with the subjective impressions that the two-tone was (a) hard to recognize without first seeing the photograph from which it was derived ("uncued") and (b) easy to see after seeing the photograph ("cued"). This stimulus creation and selection were guided by the perceptual judgment of the experimenters. Images were printed onto 12x12cm cards.

Two other image pairs were tested which did not include two-tones and for which the correspondence was easier to see (Figure 2, top left two image pairs). These served as warm-up items and to ensure participants understood the task.

Each trial proceeded in three stages. In stage 1, participants were shown a two-tone image and asked to indicate their recognition by pointing to the location of the eye or Pirahã person in the picture (Figure 2). Responses were marked by placing a sticker at the indicated locations. Trials in which the target was not initially identified were considered "candidate reorganization trials." These trials were of particular interest as they provided a test of whether an initially unrecognized two-tone image could be successfully reinterpreted after seeing the corresponding photo. These trials proceeded to stages 2 and 3. In stage 2, participants were shown the corresponding photograph alone and asked to point to the location of the eye or Pirahã person. In

stage 3, the two-tone image and photograph were shown side-by-side. The experimenter then pointed back and forth between the two images using the Pirahã word for “same” to convey the correspondence between photo and two-tone. After this instruction, the subject was again asked to point to the location of the eyes or person in the two-tone image.

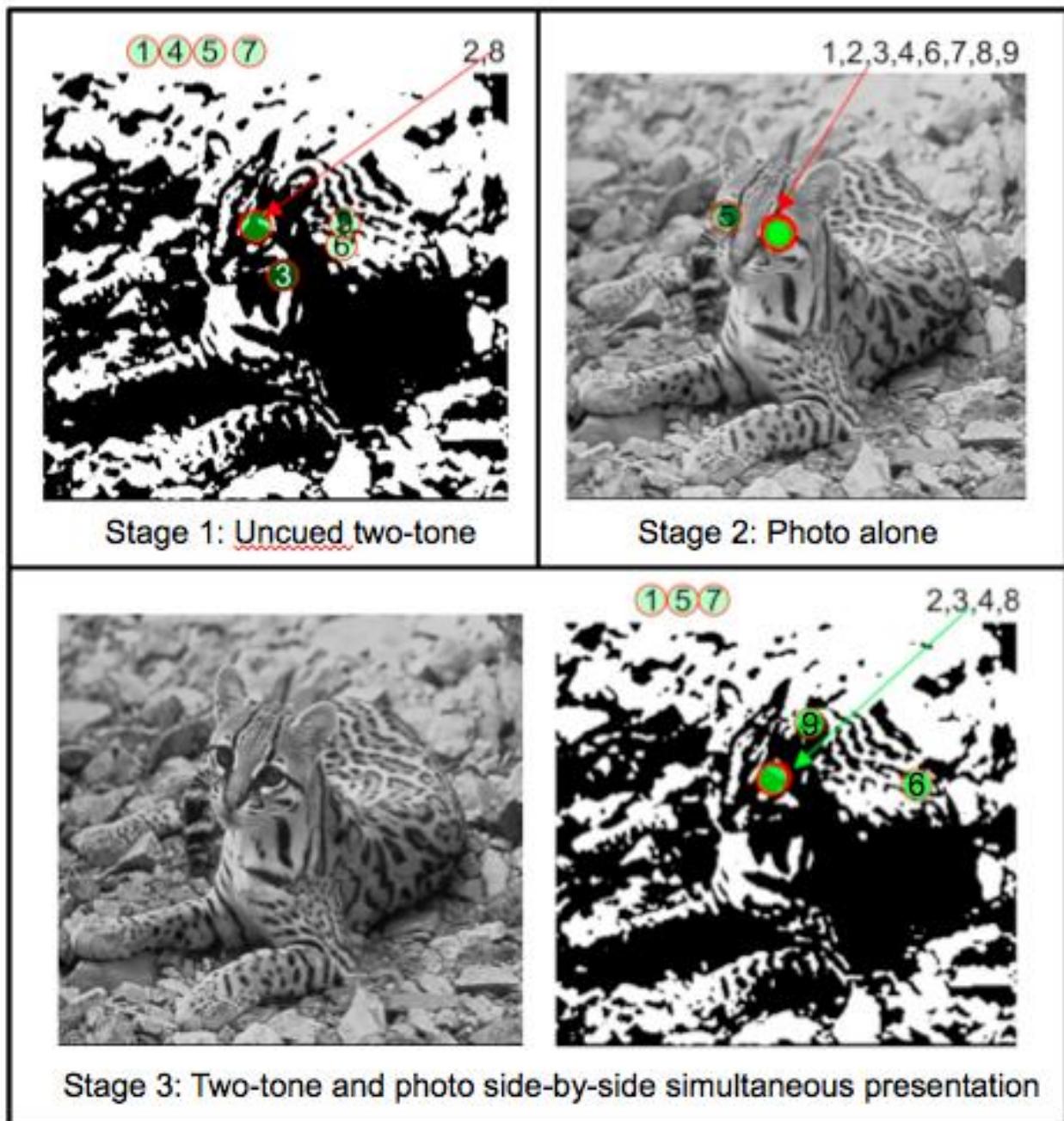


Figure 2. Upper left: an example two tone-stimulus from the Pirahã study. Subjects were first

presented the two-tone alone and asked to point to the location of an eye or person in the image. Red circles mark where Pirahã participants indicated an eye, and numbers indicate individual participants. Circles outside the image show responses of the form “there are no eyes here.” Only two participants (2,8) correctly pointed to an eye in this image during Stage 1. Upper right: Performance of Pirahã participants on the original photo, which was presented alone after the two-tone image was removed from view. All participants correctly pointed to one of the two eyes. Bottom row: performance of Pirahã participants on the two-tone image during Stage 3, when it was shown side-by-side with the photo. Two Pirahã participants succeeded uncued (2 & 8), two more succeeded with the photo present, indicating reorganization (3 & 4), and five did not show evidence of photo-triggered perceptual reorganization (1, 5, 6, 7, 9).

Viewing distance for Pirahã participants ranged from about 1.5 to 3ft and was not precisely controlled. Variability in this range is unlikely to affect recognition. In a separate control study to test for the possibility that close viewing interfered with perceptual reorganization, U.S. adults viewed two-tones from a much closer viewing distance than seen in any participants (9 in.) and performed at ceiling on candidate reorganization trials (100%). In addition, U.S. preschoolers, a similarly low reorganization population, viewed two-tones from distances of 2 and 4 feet with no difference in performance (Yoon, 2012).

We additionally tested Stanford students on an alignment manipulation task. This task controlled for the possibility that U.S. participants’ performance on the task was not due to recognizing the two-tone images, but merely to locating the point on the two-tone card in the same location as the corresponding point in the photograph. This study was identical to the main study, except that the images were cropped by 10% on two adjacent sides (e.g., top and left), chosen at random, with the constraint that the corresponding two-tone and photo were not cropped on the same two sides. Thus the eye or head was in a different location on the printed card in the photo and in the two-tone. If US participants were solving the task by pointing to the same location on the cards rather than by identifying the image features in the two-tone image, they would not have successfully located the eye in the two-tone image in this experiment.

RESULTS

Pirahã participants and U.S. control participants on the same task successfully indicated the target locations (either eye or person) on the non-two-tone practice images without the corresponding photo cue (controls 100%, Pirahã 88.9%), showing participants understood the task (Figure 3, white bars). Controls located the targets successfully in uncued two-tone images on 72.5% of trials. Initial recognition in Pirahã participants was less frequent (22.5% of trials). Controls identified the targets in the corresponding, untransformed photos 100% of the time and the Pirahã 90.3% of the time (Figure 3, black bars). All Pirahã participants correctly indicated the target on at least 7 of the 10 photos. Data from trials where the Pirahã did not correctly recognize the photo were excluded from subsequent analysis.

The trials of primary interest were candidate reorganization trials: trials on which participants did not initially locate the target in the two-tone (incorrect Stage 1), but did locate it in the photo cue image when it was presented alone (correct Stage 2). We assessed performance on only these trials (% correct Stage 3) by calculating the percentage of two-tones recognized after viewing the photo cue and dividing by the total number of candidate reorganization trials. U.S. control participants consistently showed cue-driven perceptual reorganization, always (100%) correctly indicating the eye or the Pirahã person on previously unrecognized two-tones. In contrast, Pirahã participants succeeded on candidate reorganization trials only 31.6% of the time. Two Pirahã participants never demonstrated perceptual reorganization, and the highest rate of reorganization for any Pirahã individual was 60%.

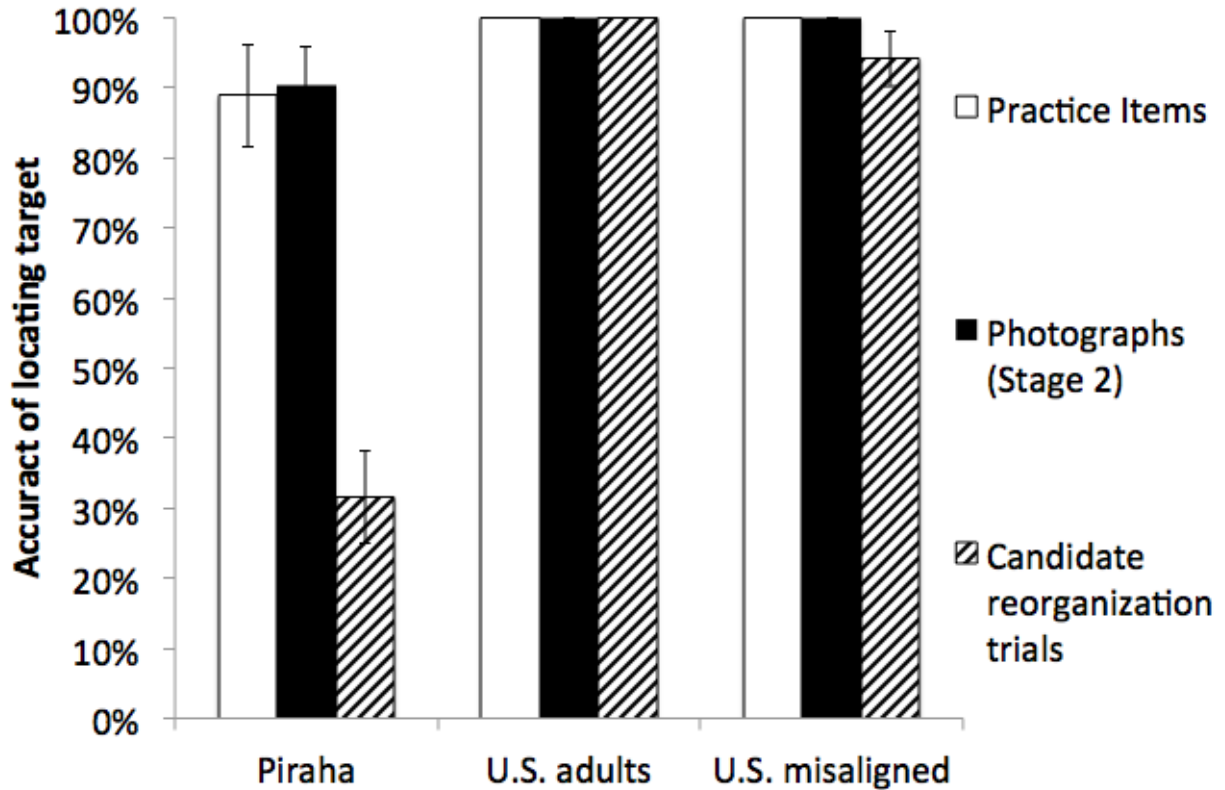


Figure 3. Summary of results from the Pirahã and the two U.S. control groups. Bars show participants' accuracy on photographs, practice items, and candidate reorganization trials (those trials on which the two-tone image was not recognized uncued). Error bars show the standard error of the mean.

Control participants in the misaligned condition—like the controls in the main experiment but unlike Pirahã participants—showed near perfect performance on candidate reorganization trials (94.2%). This result would be expected if control participants experienced reorganization, and their performance did not depend solely on a spatial alignment strategy to localize features.

DISCUSSION

We tested whether Pirahã participants showed perceptual reorganization of two-tone

images when they were viewed side-by-side with the original (cue) photograph from which they were generated. Although U.S. control participants performed at ceiling, successfully identifying the target location in every previously unrecognized image, the Pirahã found this task extremely challenging. The relative lack of cue-driven perceptual reorganization in the Pirahã is especially striking in contrast to the reported ease, vividness, and automaticity of reorganization in the control group, which persisted in the face of spatial misalignment. But why is reorganization so much less frequent in the Pirahã group? We begin by reviewing and rejecting hypotheses relating to task interpretation, stimulus familiarity, and task difficulty, and then discuss possible conceptual or experiential sources of differences in perceptual reorganization.

First, our data are inconsistent with an account driven purely by differences in task interpretation between the Pirahã and the U.S. controls. Good performance on the practice trials and on the photos themselves demonstrates that the Pirahã understood the general task instructions. The experimenters made their best effort, both verbally and gesturally, to explicitly indicate that the photo and two-tone images corresponded and were “the same.” Even if Pirahã participants did not know how to interpret the experimenter’s instructions initially, they would presumably understand that the image pairs corresponded after a successful spontaneous reorganization. However, success on one candidate reorganization trial did not lead to greater accuracy on subsequent trials (41% accuracy including only trials after first successful reorganization vs 32% across all trials, two sample t-test, $p=0.4$), suggesting that the experience of reorganization did not change the overall difficulty of the task. Pirahã participants did sometimes respond “no eyes/person” when asked to find these targets on the two-tone images. However, they gave this response only 37 of 90 trials, and 27 of the 37 occurred after they had already made an accurate response by locating eyes or people in previous two-tones. We thus

interpret “no eye/person” responses as indicating lack of recognition as opposed to a different interpretation of the task instructions.

Second, since the photographs depicted people and animals in their environment, it is also unlikely that the result is due to a lack of familiarity with the pictured items. In fact, the particular animals and people depicted in the photographs were more familiar to the Pirahã than the controls (one of the pictures contained a person known to members of the tribe, and several of the Pirahã participants spontaneously produced his name on seeing the photo).

Two-tone image recognition can fail at multiple stages by different mechanisms. A third possibility therefore is that poorer performance among Pirahã participants in the candidate reorganization trials was not due to a difficulty in using a photograph to reinterpret a two-tone (presumably a more top-down process), but rather a difficulty in recognizing the *uncued* two-tones. Recognition could have failed at a coarser level than photo and two-tone comparison—perhaps during early bottom-up stages of processing like basic perceptual organization, due to overall greater difficulty. Indeed, the Pirahã participants had significantly lower accuracy at two-tone recognition in the un-cued condition, indicating that two-tones are more difficult for them to see. To address whether this also explains why they were less likely to perceptually re-organize after seeing the photo, we compared items of similar difficulty in U.S. and Pirahã groups. We define difficult two-tones behaviorally, as those that are rarely recognized uncued; similarly, easy two-tones are those that are recognized frequently even when they are uncued. In pilot testing with a larger set of stimuli (40) on U.S. adults ($n=9$), successful reorganization occurred consistently regardless of variation in uncued two-tone “difficulty.” Even for the 5 stimuli with the poorest uncued recognition (46% success), participants successfully re-organized on nearly all trials (93%), far more often than the Pirahã in the main experiment reported here. In contrast,

considering only the two stimuli with the highest uncued recognition (78% success), the Pirahã who did not initially recognize these stimuli re-organized only infrequently (25%, 1 of 4 trials). Assuming a binomial distribution based on the reorganization probability of English speaking participants (93%), the likelihood of the Pirahã data—reorganizing only once or fewer out of 4 trials—is less than one percent. Thus, we do not believe that task difficulty or related issues explain our findings. How then should we interpret the striking failure to reorganize the two-tones in the Pirahã participants?

Our data indicate that a mature visual system is insufficient to guarantee good performance on this task. It is possible that the similarly low rates of photo-triggered reorganization of two-tone images in young U.S. children (Kovacs & Eisenberg, 2004; Yoon, Winawer, Witthoft, & Markman, 2007) and Pirahã adults are unrelated. But the existence of reduced reorganization in an adult population opens up the possibility that developmental failures in perceptual reorganization in young U.S. children may also be explained by a mechanism distinct from visual system maturation. Thus, one possibility is that the results reported here, together with the previous studies on young children (DeLoache, 1997; Kovacs & Eisenberg, 2004; Yoon et al., 2007; Yoon, 2012), reflect a role of perceptual literacy in photo-triggered perceptual reorganization: the very act of bringing our knowledge and experience to bear on perception in the way required for reinterpretation of a two-tone image may be the result of training and experience that is culture-specific. This perceptual literacy likely involves growing skill in “imposing one’s imagined structure” as in the reversal of ambiguous figures (Rock, Hall & Davis, 1994), though in this case the structure is derived from one image (photo) and imposed onto another (two-tone).

In their lack of expertise with visual symbols, young U.S. children are similar to the adult

Pirahã, whose material culture does not include a writing system, maps, or representational artwork that would entrain such a skill (Everett, 2005). Enculturation with visual symbolic materials may provide the training required for observers to navigate the dual nature of the two-tone and photo as (1) objects in and of themselves, as well as (2) representations of one another that are mutually informative. In Deloache's research, U.S. children were asked to find a doll hidden in a target room based on the location demonstrated in a symbolic representation of that room. She found surprising and robust failure to use a smaller scale model room to decide where to find a doll hidden in a larger target room, despite accurate memory for the configuration of the model room. The cognitive challenge was to inhibit a prepotent interpretation of the symbolic representation as an entity in and of itself (representation 1), and instead rely on a representation of the correspondence between target and symbol (representation 2). Older more successful children could be impaired if encouraged to play with the objects in the scale model room, strengthening their representation of the room as an entity in and of itself, overriding the representation that mapped the scale model (visual symbolic artifact) to the target room. Conversely, younger children were more successful when the scale model room was placed behind glass, a small change that could help inhibit their representation of the scale model as an entity in and of itself (Deloache, 2000).

Simplifying the challenge of dual representation (inhibiting representation 1 and strengthening representation 2) has been shown to aid inexperienced members of a culture (young U.S. children) in using symbolically corresponding visual representations, even before they acquire expertise in 'reading' visual symbols such as writing and maps (Deloache, 1997). One memorable method involved convincing children that the scale model room was physically the same entity as the target room due to transformation by "shrinking machine" – thus inhibiting

representation 1, removing an interfering representation of the scale model room as a distinct entity. It may be that similar manipulations could enable Pirahã adults to reliably experience photo-cued perceptual reorganization in the absence of a lifelong enculturation with visual symbolic materials. Further research should address the possibility that cultures that provide training in how to ‘read’ visual symbols such as writing and maps can influence the practices of perceptual inference and interpretation required for successful perceptual reorganization.

REFERENCES

- Adelson, E.H. (1993). Perceptual organization and the judgment of brightness. *Science*, 262, 2042-2044.
- de Fockert, J., Davidoff, J., Fagot, J., Parron, C., & Goldstein, J. (2007). More accurate size contrast judgments in the Ebbinghaus Illusion by a remote culture. *Journal of Experimental Psychology: Human Perception and Performance*, 33(3), 738-742.
- Deloache, J. S. (1997). The credible shrinking room: Very young children's performance with symbolic and nonsymbolic relations. *Psychological Science*, 8(4), 308-313.
- Deloache, J.S. (2000). Dual representation and young children's use of scale models. *Child Development*, 71(2), 329-338.
- Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology*, 46(4), 621-634.
- Gregory, R. L. (1970). *The Intelligent Eye*. London: Weidenfeld & Nicolson.
- Gregory, R. L. (2005). The Medawar Lecture 2001: Knowledge for vision: vision for knowledge. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1458), 1231-1251.

- Helmholtz, H. (1878). *Selected Writings of Hermann Helmholtz* (R. Kahl, trans.). Middletown, CT: Wesleyan University Press.
- Jones, R. K., & Hagen, M. A. (1980). A perspective on cross-cultural picture perception. In M. A. Hagen (Ed.), *The Perception of Pictures Volume 2* (pp.193-226). New York: Academic Press.
- Kohler, W. (1929). *Gestalt Psychology*. Oxford: Liveright
- Kovacs, I., & Eisenberg, M. (2004). Human Development of Binocular Rivalry. In D. Alais & R. Blake (Eds.), *Binocular Rivalry* (pp.101-116). Cambridge, MA: MIT Press.
- Leibowitz, H., Brislin, R., Perlmutter, L., & Hennessy, R. (1969). Ponzo perspective illusion as a manifestation of space perception. *Science*, 166, 1174-1176.
- Pelli, D. (1999). Close encounters — An artist shows that size affects shape. *Science*, 285, 844-846.
- Pica, P., Jackson, S., Blake, R., & Troje, N. F. (2011). Comparing biological motion perception in two distinct human societies. *PloS one*, 6(12), e28391.
- Rock, I., Hall, S., & Davis, J. (1994). Why do ambiguous figures reverse? *Acta Psychologica*, 87(1), 33-59.
- Segall, M. H., Campbell, D. T., & Herskovit, M. J. (1966). *The Influence of Culture on Visual Perception*. Indianapolis: Bobbs-Merrill.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press. Ed. Michael Cole.
- Whitaker, D., & McGraw, P. V. (2000). Long-term visual experience recalibrates human orientation perception. *Nature Neuroscience*, 3(1), 13.
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian

blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, 104(19), 7780-7785.

Yoon, J. M. D., Winawer, J., Witthoft, N., & Markman, E. M. (2007). Striking deficiency in top-down perceptual reorganization of two-tone images in preschool children. *Proceedings of the 6th IEEE International Conference on Development and Learning*, 181-186.

Yoon, J.M.D. (2012). *Vision and Revision: Cue-Triggered Perceptual Reorganization of Two-Tone Images in U.S. Preschool Children and Adults*. (Unpublished doctoral dissertation). Stanford University, Stanford, CA.