vol. 167, no. 5 the american naturalist may 2006

E-Article

The Structures of Letters and Symbols throughout Human History Are Selected to Match Those Found in Objects in Natural Scenes

Mark A. Changizi,[1,*] Qiong Zhang,[2,†] Hao Ye,[2,‡] and Shinsuke Shimojo[3,§]

1. Sloan-Swartz Center for Theoretical Neurobiology, California Institute of Technology, Pasadena, California 91125;
2. Division of Biology, Computation and Neural Systems, California Institute of Technology, Pasadena, California 91125;
3. Japan Science and Technology Program–Exploratory Research for Advanced Technology Shimojo Implicit Brain Function Project, California Institute of Technology, Pasadena, California 91124

Submitted April 11, 2005; Accepted December 19, 2005; Electronically published March 23, 2006

abstract: Are there empirical regularities in the shapes of letters and other human visual signs, and if so, what are the selection pressures underlying these regularities? To examine this, we determined a wide variety of topologically distinct contour configurations and examined the relative frequency of these configuration types across writing systems, Chinese writing, and nonlinguistic symbols. Our first result is that these three classes of human visual sign possess a similar signature in their configuration distribution, suggesting that there are underlying principles governing the shapes of human visual signs. Second, we provide evidence that the shapes of visual signs are selected to be easily seen at the expense of the motor system. Finally, we provide evidence to support an ecological hypothesis that visual signs have been culturally selected to match the kinds of conglomeration of contours found in natural scenes because that is what we have evolved to be good at visually processing.

Keywords: natural scenes, letter shape, visual signs, object junctions, ecological vision, evolution of writing.

Despite the great deal of variation in character shape in writing systems over human history (Fairbank 1968; Nakanishi 1980; Ifrah 1985; Sampson 1985; Coulmas 1991; Robinson 1995; Daniels and Bright 1996; Ager 1998; Helfman 2000), might there be underlying similarities? In this article, we will address three questions concerning the shapes of human visual signs. First, we will ask whether there are any empirical regularities governing the shapes of human visual signs. In an effort to answer this, we will identify a wide array of topologically distinct contour configurations and will measure the relative frequency of these configurations as they occur in letters across 100 writing systems over human history, Chinese characters, and nonlinguistic symbols. We will demonstrate that there are strong correlations among the configuration distributions of these three classes of visual sign, suggesting that the configuration distribution for human visual signs tends to possess a characteristic signature. Second, we will show that this signature correlates highly with the configuration distribution found in trademark symbols (signs that are selected primarily for visual recognition, not for the motor system) and with measures of visual stimulus complexity but that the visual sign signature correlates poorly with the configuration distribution found in scribbles and shorthand (drawings that are selected primarily for motor execution, not for visual recognition) and with measures of motor complexity. We will conclude from this that visual sign shapes are selected for optimization of visual recognition, not motor execution. Finally, we will test an ecological hypothesis that the shapes of visual signs have been selected to resemble the conglomerations of contours found in natural scenes, thereby tapping into our already-existing object recognition mechanisms, somewhat akin to sensory exploitation hypotheses in animal signaling (Endler 1992; Guilford and Dawkins 1993; Ryan 1998). We will provide evidence that the more common configuration types found in natural scenes tend to be the more common ones found in human visual signs.

A Topological Notion of Shape: Configuration Types

Characters in writing systems are configurations of strokes, or segments, and for the purpose of quantifying their

* Corresponding author; e-mail: changizi@caltech.edu.

† E-mail: qiong@its.caltech.edu.

‡ E-mail: hye@cs.caltech.edu.

§ E-mail: sshimojo@its.caltech.edu.

Am. Nat. 2006. Vol. 167, pp. E117–E139. © 2006 by The University of Chicago. 0003-0147/2006/16705-41010$15.00. All rights reserved.

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

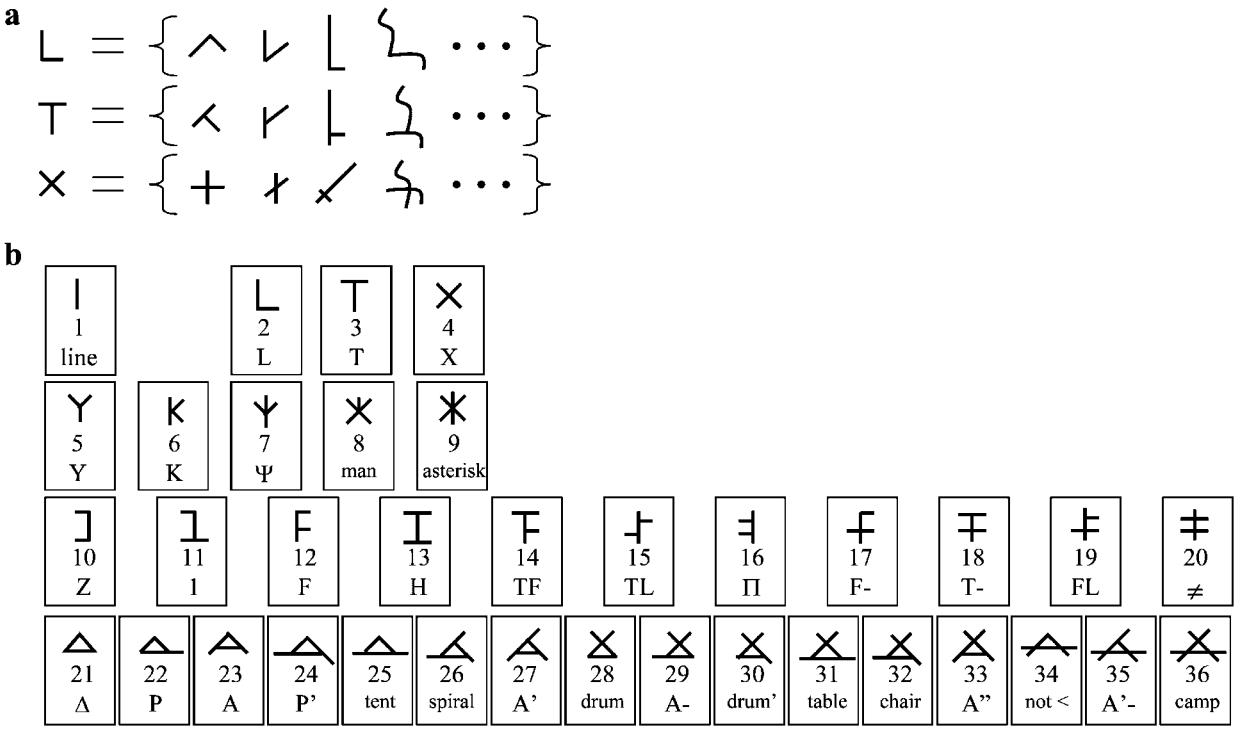


Figure 1: *a*, Illustration of the kinds of variation allowed within a configuration type. Each configuration shown stands for a large class of configurations of the same type, where the following features can vary independently: overall orientation of the configuration, relative orientation of the segments, relative lengths of the segments, and shapes of the segments. Example cases are shown for the three configuration types with a length of two. *b*, Catalog of all 36 distinct configuration types with three or fewer segments, with ID number and proper name. The catalog allows only configuration types where for any two distinct segments x and y, x can intersect y at most once and no segment can intersect itself.

structure, we developed the notion of a configuration type that captures the fundamental topological structure of a configuration. We have confined our catalog of configuration types to three or fewer segments because we found (Changizi and Shimojo 2005) that across 115 nonlogographic writing systems over human history, varying in number of characters from about 10 to 200, the average number of strokes per character is approximately three and does not appear to vary as a function of writing system size. Also, allowing configuration types with more than three segments leads to an explosion of possibilities, severely hindering analysis. Figure 1 shows all 36 distinct configuration types possible using three or fewer straight segments (along with their name and identification number), where each of the shown configuration types represents a large class of configurations with the same topology and where the relative orientation (between 0° and 180°), lengths, and shapes of the segments can vary independently, as can the overall orientation of the configuration. For example, L, T, and X represent the three topologically distinct configuration types built with exactly two segments, and each defines a class of type-identical configurations (see fig. 1*a*).

The configuration distribution for a set of visual signs is the frequency distribution of configuration types across the signs in that set (normalized so that the frequencies sum to 1). For the configuration distributions we measure here, we are not interested in whether the distributions are significantly different (they usually are), but instead we are interested to test whether, more weakly, x is a more common configuration type in distribution B ⇔ x is a more common configuration type in distribution C. Accordingly, unless otherwise stated, frequency ranks will be used when quantitatively comparing two configuration distributions, and the ranks are computed so that rank 1 is the highest frequency, tied frequencies receive identical rank, and after a tie, the subsequent rank is the next higher rank; for example, 50%, 10%, 10%, and 3% would have ranks 1, 2, 2, and 3. Note that we have designed our methodology in such a way that the frequency for each three-segment configuration type is independent of every other type (see "Configuration Measurement and Tests of

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Observer Veridicality" in the appendix), and the same is true among the two-segment configuration types (namely, L, T, and X). However, the two-segment configuration types are not independent of the three-segment configuration types. For this reason, our quantitative comparisons of distinct configuration distributions will concentrate primarily on the 32 three-segment configuration types. Finally, we note that our configuration distributions were measured by trained human observers; see "Configuration Measurement and Tests of Observer Veridicality" and figures A1 and A2 in the appendix for validations of their objectivity.

The above notion of shape is topological and therefore ignores geometrical aspects of shape. For the purposes of this article, however, this topological notion of shape is appropriate for the following reasons. First, many visual signs can undergo significant geometrical shape variation without losing their identity, but their topological properties are more highly constrained. For example, each time you draw a T, the specific geometrical features differ, but the topology remains identical. And consider the tremendous number of fonts, where it is primarily the geometrical properties that vary. Second, our notion of shape also has to be applied to the conglomerations of contours found in natural scenes, and the geometrical properties of such conglomerations will change substantially with the observer's viewpoint (i.e., accidental properties), whereas the topological properties will be viewpoint invariant (nonaccidental) to a much greater extent. Topologically defined shape—our configuration types—therefore provides a robust notion of shape for both visual signs and conglomerations of contours found in natural scenes, allowing us to compare these two realms.

Furthermore, there are theoretical and empirical reasons to believe that topological shape is more important than geometrical shape for object recognition in the visual system. Theoretically, it has long been noticed within the computational vision literature that recognizing object junctions—which are topologically defined conglomerations of contours—will be crucial to recognizing objects (Guzman 1969; Clowes 1971; Huffman 1971; Turner 1974; Waltz 1975; Chakravarty 1979; Kanade 1980; Barrow and Tenenbaum 1981; Binford 1981). And empirically, there is evidence that the visual system represents topological structures as intermediate levels in the computation of objects (Biederman 1987; Biederman and Cooper 1991; Biederman and Gerhardstein 1993). Therefore, in addition to allowing us to compare human visual signs with natural scenes, the topological notion of shape we use here may have some psychological reality.

Human Visual Signs Possess a Characteristic Signature Configuration Distribution

With our notion of shape—configuration types—in hand, what shapes do we in fact find among human visual signs? We considered three principal classes of visual sign. The first is a set of 96 writing systems, including numerals, abjads (characters for consonants but not for vowels), abugidas (characters for consonants but diacritical marks for vowels), alphabets (characters for consonants and vowels), and syllabaries (characters for syllables) from five major taxa—Ancient Near Eastern, European, Middle Eastern, South Asian, and Southeast Asian—as well as invented writing systems (for details, see "Writing Systems and Shorthand" in the appendix). This set included only nonlogographic writing systems, that is, where characters do not stand for whole words, and thus we refer to it as nonlogographic writing systems. The configuration types were determined for 1,442 characters across these writing systems. The second class of visual sign we considered is Chinese characters, from which we measured 4,759 configurations (for details, see "Chinese, Nonlinguistic Symbols, Trademarks, and Children Scribbles" in the appendix). The third principal class of visual sign is nonlinguistic symbols, for example, music or traffic symbols, from which we measured 3,538 configurations (for details, see "Chinese, Nonlinguistic Symbols, Trademarks, and Children Scribbles" in the appendix).

The configuration distributions for these three classes of visual sign are displayed in figure 2*a*. There are significant differences among these distributions, and the study of these differences is not our purpose here. Rather, we are interested in examining the degree to which they are similar and, in particular, in testing the extent to which more common in distribution B $\Leftrightarrow$ more common in distribution C. The correlation of the ranks of the 32 independent three-segment configuration types are as follows: between writing and Chinese, $R = 0.69$ ($P < .00001$); between writing and symbols, $R = 0.83$ ($P < .000001$); and between Chinese and symbols, $R = 0.81$ ($P < .000001$). The rank orders among the three two-segment configurations—L, T, and X—are identical for all three kinds of visual sign. We averaged these three distributions to obtain a summary distribution for visual signs, shown in figure 2*b*.

To see that the visual sign distribution is not some kind of mathematical tautology, it is useful to consider what the configuration distribution would be for images of tilings. For example, an image of a honeycomb, or hexagonal tiling, would have a distribution consisting of only Ys. A triangular tiling would, alternatively, possess only asterisks. An image of a brick wall would, for example, possess just Ts and TLs.

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

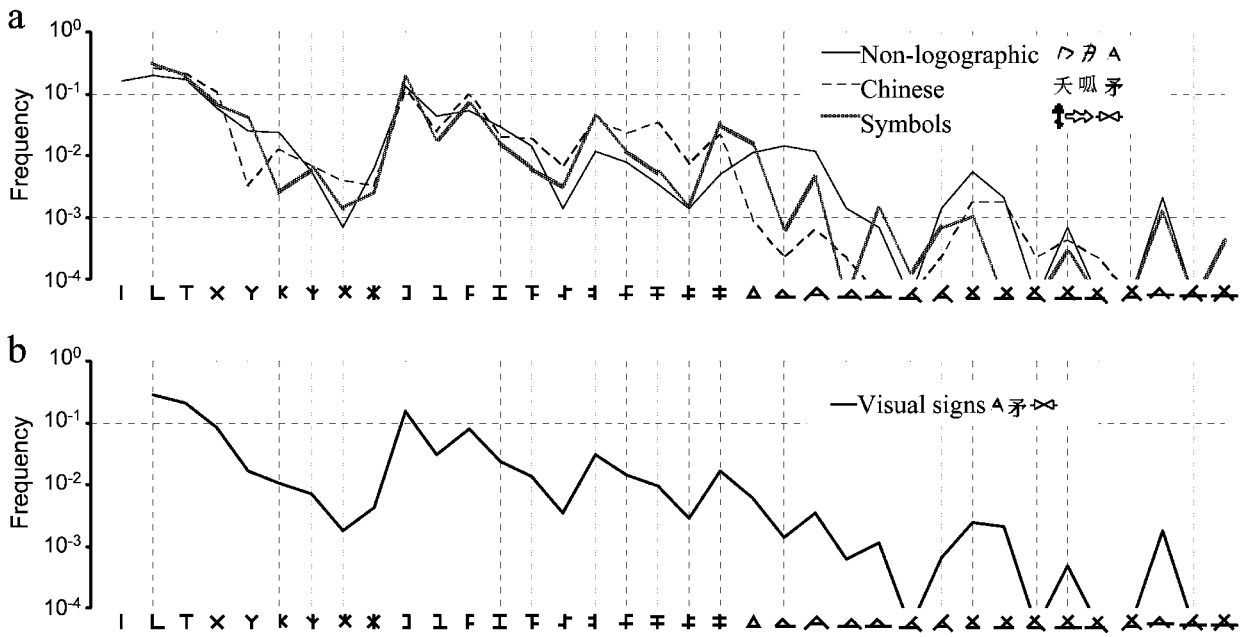


Figure 2: *a*, Configuration distribution across visual signs, including nonlogographic writing systems (*solid line*), Chinese (*dashed line*), and nonlinguistic symbols (*hatched line*). The correlation between writing and Chinese is $R = 0.69$ ($P < .00001$), between writing and symbols is $R = 0.83$ ($P < .000001$), and between Chinese and symbols is $R = 0.81$ ($P < .000001$). *b*, Summary distribution for visual signs, which is the average of the three plots in *a*. For comparison's sake, we will display this as a dotted line in several upcoming plots.

To show that the configuration distribution for visual signs is not a consequence of random lines on a page, visual images were fabricated via randomly placing strokes on the page, as described in "Configuration Measurement and Tests of Observer Veridicality" in the appendix. The measured configuration distribution for random lines is shown alongside the visual sign distribution in figure 3*a*, and one can see that the distributions are drastically different, having no correlation between them ($R = 0.1$, $P > .35$) for the ranks of the 32 independent three-segment configuration types. In particular, the random line distribution is largest for configuration types X, ≠, and "camp"—very complex configuration types—a distribution unlike the signature distribution found for visual signs.

We saw that the visual sign distribution is not due to random lines on the page, but perhaps a better notion of random would be where one randomly draws on the page (rather than randomly throws strokes onto the page). To show that the configuration distribution for visual signs is not a tautological consequence of writing (i.e., where even randomly drawing leads to it), we measured the distribution (726 configurations) from 17 scribble drawings from 26- to 32-month-old children (fig. 3*b*). For details, see "Chinese, Nonlinguistic Symbols, Trademarks, and Children Scribbles" in the appendix. Among the 32 independent three-segment configuration types, the scribble distribution ranks do not significantly correlate with those of visual signs ($R = 0.30$, $P > .05$). Because it is possible, in principle, for correlations between distributions to be absent and yet for the distributions not to be significantly different from one another, we computed the 95% confidence interval around the visual sign distribution assuming 726 configurations were sampled from it, and this is shown figure 3*b* (*gray*); one can see that the scribble distribution deviates considerably from the visual sign distribution.

Human Visual Signs Are Selected for Vision at the Expense of Motor

In the previous section, we have seen that there are strong similarities in the configuration distributions found across diverse classes of human visual sign and that these regularities are not some kind of tautology and not a consequence of random lines or random motor activity. There are two obvious potential selection pressures shaping visual signs, corresponding to reading and writing: visual signs must have shapes that we are able to recognize with our visual systems and that we can produce (very often) by hand. Although it is almost surely the case that both reading and writing contribute to the shapes of visual signs,

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

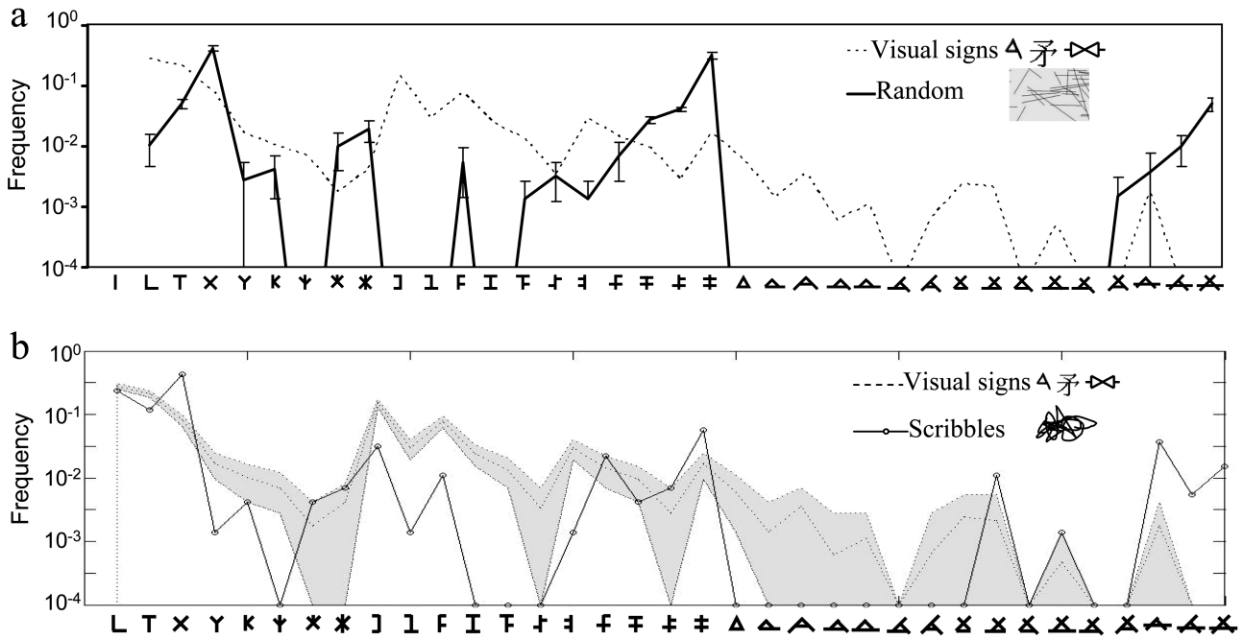


Figure 3: *a*, Configuration distribution measured from images with random lines (*solid line*), with standard error bars shown (as seen in fig. A2), but here shown alongside the average distribution across all visual signs (*dotted line*; i.e., the average of nonlogographic writing, Chinese, and symbols). There is no significant correlation between the ranks among the three-segment configuration types ($R = 0.1$, $P > .35$), and one can see that the configuration distribution for random line images differs significantly from that of visual signs (i.e., nearly always falls outside of the error bars). *b*, Configuration distribution for children scribbles (*solid line*) shown alongside the average distribution across all visual signs (*dotted line at center of gray band*). There is no significant correlation among the ranks for the 32 independent three-segment configuration types ($R = 0.3$, $P > .05$). Because only 726 configurations were sampled from scribbles, in principle, the distributions could be not significantly different, despite the lack of correlation. To check this, the gray areas show the 95% confidence interval regions around the visual signs distribution, given that 726 configurations were sampled from scribbles. That is, if 726 configurations were to be sampled using the visual sign configuration distribution, then the gray region shows the 95% confidence interval for each configuration type. Sixteen of the 35 configuration type frequencies for scribbles fall outside of the 95% confidence interval for visual signs. Among the 20 configuration types that actually occur within the scribble distribution, 12 fall outside the 95% confidence interval for visual signs.

here we ask whether one of these may be the principal driver of the configuration distribution (i.e., of topological shape). We test the extent to which visual signs appear to be selected for motor and then test the extent to which they appear to have been selected for vision.

Not Selected for Motor: Shorthand and Motor Complexity

To test whether the visual sign distribution is a result of selection pressure to optimize motor complexity of drawing, we carried out two analyses. First, because shorthand characters are selected for writing at the expense of reading, we measured the configuration distribution for shorthand (121 configurations from six shorthands; see "Writing Systems and Shorthand" in the appendix), as shown in figure 4*a*. The shorthand distribution possesses no correlation with the ranks from three-segment configuration types from visual signs ($R = 0.35$, $P > .05$) or from two-segment configuration types ($R = 0.33$, $P > .20$). Because of the small number of configurations sampled for shorthand, there is an increased possibility that, although there is no correlation, there may be no significant differences between the distributions. We computed 95% confidence intervals around the nonlogographic writing system distribution (fig. 4*a*, *gray*), and one can see that they differ considerably; for example, for the two-segment configuration types (L, T, and X), the shorthand distribution varies more than two orders of magnitude, whereas the nonlogographic writing system distribution does not vary much.

Next, we determined the minimum number of hand sweeps (number of pen strokes plus the number of pen lifts) required to draw each configuration type (fig. 4*b*, *left*). For example, F and A each require four hand sweeps (three pen strokes and one pen lift), and H and "spiral" each require five hand sweeps (three pen strokes and two pen lifts). There is no significant correlation ($R = 0.22$,

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

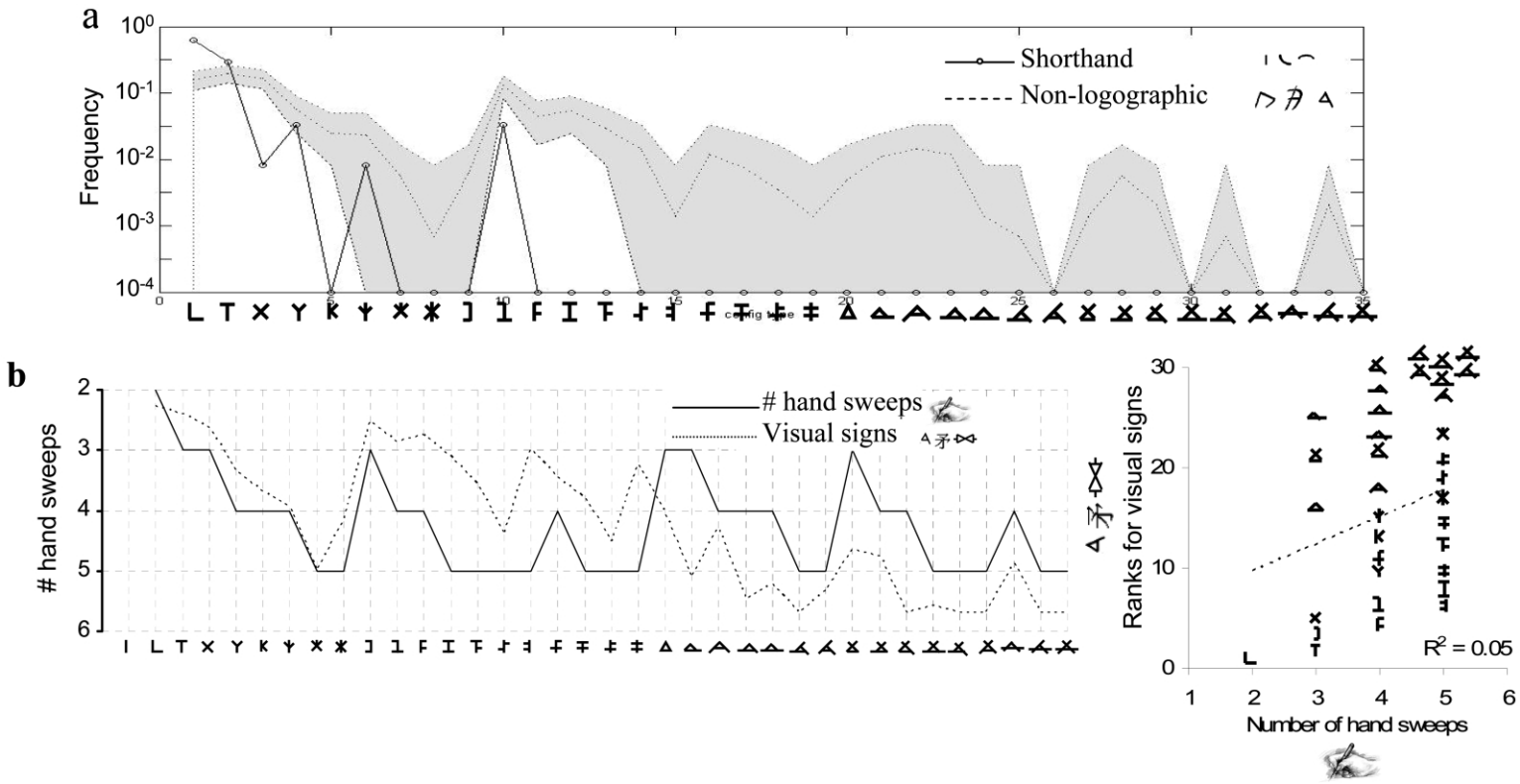


Figure 4: *a*, Configuration distribution for shorthand shown alongside that of the nonlogographic writing systems (*dotted line at center of gray band*). There is no significant correlation among the ranks of the three-segment configuration types between visual signs and shorthand ($R = 0.35, P > .05$) or among the two-segment configuration types ($R = 0.33, P \approx .50$). Because only 121 configurations were sampled (from six shorthand writing systems), in principle, the distributions could be not significantly different, despite the lack of significant correlation. To check this, the 95% confidence interval around the nonlogographic writing distribution is shown, assuming that 121 configurations were chosen from the distribution. Eight of the 36 configuration type frequencies for shorthand fall outside the 95% confidence interval for nonlogographic writing. Among the six configuration types that actually occur in shorthand, four fall outside the 95% confidence interval for visual signs. *b*, Minimum number of hand sweeps required to draw each configuration type. For comparison, the plot of the frequency ranks for visual signs (nonlogographic writing, Chinese, and symbols) is shown by the dotted line (rescaled to fit the range and area under the plot of the hand sweep plot). The plot on the right shows the frequency ranks for visual signs as a function of the number of hand sweeps; the linear regression line shown, and correlation, is for the three-segment configuration types only (not L, T, and X). There is no significant correlation ($R = 0.22, P > .2$).

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

$P > .2$) among the three-segment configuration types between the number of hand sweeps and the frequency ranks for visual signs (fig. 4*b*, *right*).

These two results—the lack of correlation between shorthand and visual signs and the lack of correlation between motor complexity and visual signs—suggest that visual sign topological shape is not strongly selected for the motor system.

Selected for Vision: Trademarks and Visual Stimulus Complexity

To test whether the visual sign distribution is a result of selection pressure for visual recognition, we carried out two analyses. Trademark symbols of businesses are selected to be easily recognized by the visual system, and because they are typically not drawn by hand, they should be under little selective pressure for motor optimization. The configuration distribution (consisting of 1,403 configurations in all; see "Chinese, Nonlinguistic Symbols, Trademarks, and Children Scribbles" in the appendix) across trademarks (fig. 5*a*) closely matches (among the ranks of the three-segment configuration types) that for visual signs ($R = 0.89$, $P < .000001$), indicating that visual signs may generally be selected primarily for viewing.

Next, we asked how well the configuration distribution for visual signs correlates with human judgments of visual stimulus complexity. Human judgments of visual stimulus complexity have been found to correlate highly with the number of angles in the stimulus (Hochberg and McAlister 1953; Attneave 1957; Arnoult 1960), and figure 5*b* shows how this relates to the distribution for visual signs. Most of the signature peaks and troughs found in visual signs are absent (fig. 5*b*, *left*), and there are clear nonrandom trends for the residual around the regression line (fig. 5*b*, *right*). Nevertheless, among the three-segment configuration types, there is a highly significant and strong correlation ($R = 0.69$, $P < .00001$) between the number of angles and the frequency ranks for visual signs.

These two results—the strong correlation between trademarks and visual signs and the strong correlation between visual stimulus complexity and visual signs—suggest that visual signs are strongly selected for the visual system. In combination with the previous parts of this section, we can conclude that visual sign configuration types appear to be selected for vision at the expense of motor.

Configuration Distribution Signature Is Found in Natural Scenes

Thus far, we have seen that there are regularities in the kinds of topological shape found across human visual signs and that these regularities appear to be selected for vision at the expense of motor. We consider an ecological and visual explanation for why visual signs are shaped the way they are. The hypothesis states that visual sign topological shapes have been selected (by cultural evolution or by trial-and-error design) so that more common configuration types among visual signs are the more common configuration types among natural scenes, thereby exploiting what humans have evolved to be good at visually processing. Note that the hypothesis does not predict that the probability distributions for visual signs and natural scenes do not significantly differ. Rather, much more weakly, the prediction is that more common in visual signs ⇔ more common in natural scenes, and thus we minimally expect the ranks to correlate. This hypothesis has direct analogies to sensory exploitation hypotheses in sexual selection theories of animal signaling (Endler 1992; Guilford and Dawkins 1993; Ryan 1998). Note, counterintuitively, that if the hypothesis is true, then if human visual signs were placed within natural scenes, the visual signs would actually be very difficult to detect. However, human visual signs have been selected to be read and distinguished from bare sand, plain soil, paper, papyrus, walls, and so on, not distinguished from natural scene backgrounds.

In the first subsection of this section, we measure the configuration distribution from images of scenes and demonstrate that its ranks correlate with that of the signature distribution found among human visual signs. It is reasonable to have some initial reservations about sampling natural images. First, there is some reason to worry about how representative of our ecology any given set of images might be. Second, and more fundamental, it is not entirely clear what the relevant ecology is for the hypothesis. Is it the ancestral ecology, perhaps the savanna? Or is it the environment in which the various visual signs were invented or evolved? We reasonably should expect that it is some combination of the two because what our visual systems are competent to recognize depends on both our evolutionary history and the surroundings in which we were raised. But how do we possibly weight one of these ecological settings over the other in our data collection?

These are difficult questions to answer, but the configuration distribution for natural scenes appears to be highly robust across very different environmental settings. This is primarily because the notion of shape we have used here is topological—whereas the distribution of geometrical shapes may well vary considerably across ecological settings, the distribution of topological shapes is much more invariant. Informally, nearly any environment with opaque, macroscopic objects strewn about (and thereby partially occluding one another) will possess strong correlations with this signature configuration distribution. Because of the robust notion of shape and the robust con-

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

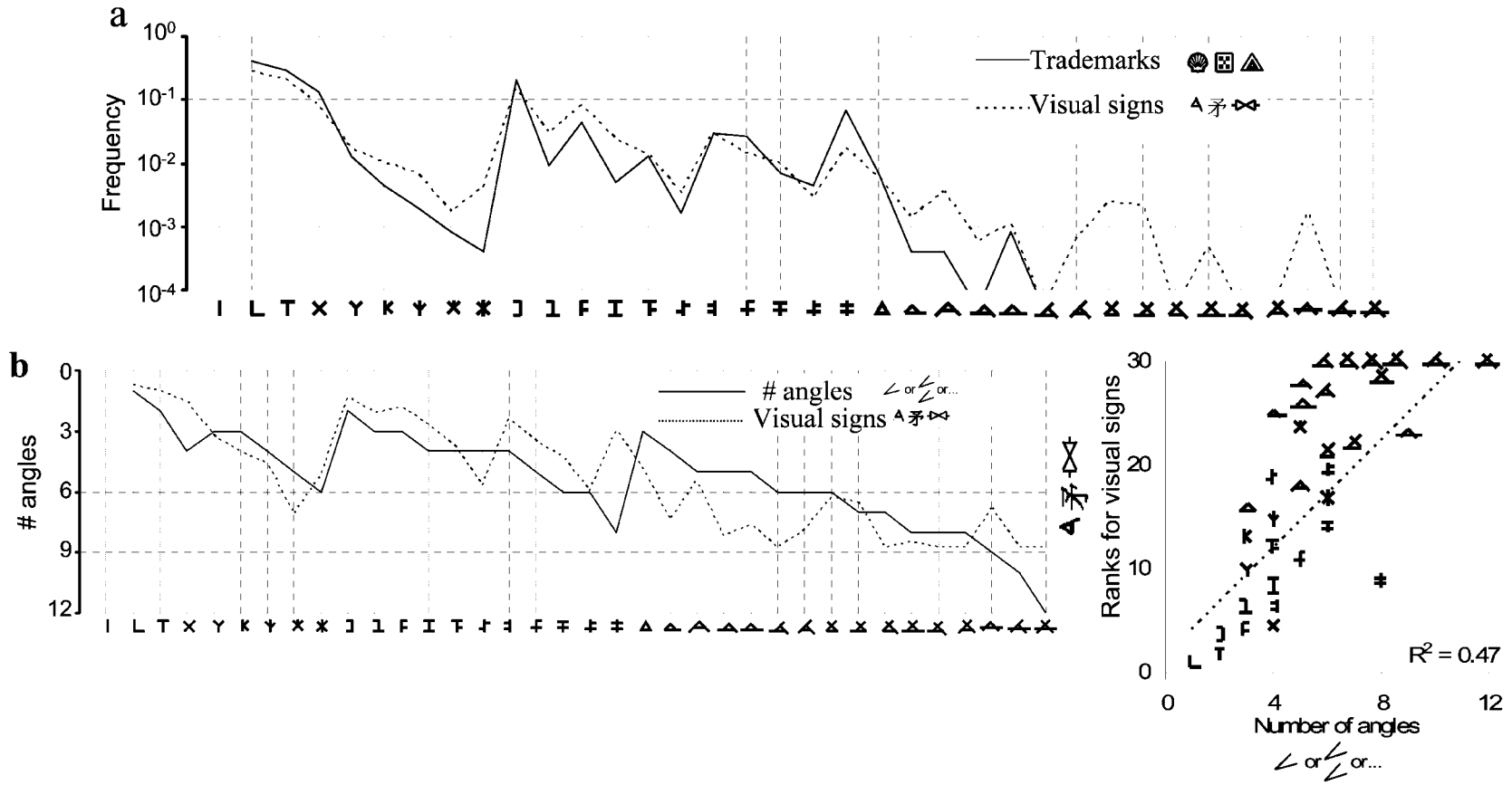


Figure 5: Two arguments that the visual sign configuration distribution is selected for vision. *a*, Configuration distribution across trademark symbols (*solid line*) shown alongside that of the average across all visual signs (*dotted line*). The distributions are highly correlated among the ranks for their 32 independent three-segment configuration types ($R = 0.89$, $P < .000001$). *b*, Number of angles in each configuration. The plot of the frequency ranks for visual signs is again shown as a dotted line for comparison. On the right is the plot of the frequency ranks for visual signs versus the number of angles; the linear regression and correlation shown are for the three-segment configuration types. The distributions are highly significantly correlated ($R = 0.69$, $P < .00001$).

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

figuration distribution to which it leads in natural scenes, we largely circumvent these difficult conceptual issues about how to properly sample from natural images, issues we probably could not circumvent were we to be using a geometrical notion of shape.

In an effort to better understand why natural scenes possess such a robust configuration distribution, in the second subsection of this section, we take an a priori approach to understanding the relative frequency of the configuration types. We show that with some very weak empirical assumptions concerning natural scenes, we can derive which configuration types will be expected to be more probable than others. In addition to helping elucidate why diverse classes of natural scene have similar configuration distributions, this a priori approach allows us to predict the relative frequency ordering of configuration types found among visual signs, and we will see that human visual signs conform nearly perfectly to these predictions.

Configuration Distribution Measured from Natural Scenes Is Similar to That for Visual Signs

We measured configuration distributions from three classes of natural image (for descriptions of the images, see "Images of Natural Scenes" in the appendix, and for descriptions of the measurements and validation by trained observers, see "Configuration Measurement and Tests of Observer Veridicality" in the appendix). The first class of images we refer to as "ancestral" and consisted of 27 photographs of savannas and tribal life (see fig. A3 in the appendix); 535 three-segment configurations were sampled. The second class of images we refer to as "National Geographic" and consisted of 40 miscellaneous photographs of rural and small-town life taken from the *National Geographic* Web site (see fig. A4 in the appendix); 471 three-segment configurations were sampled. The third (and final) class of images we call "CGI buildings" and consisted of 40 computer-generated realistic images of buildings (see fig. A5 in the appendix); 799 three-segment configurations were sampled.

The distributions for these three kinds of environment are shown in figure 6*a*, and one can see that they correlate very highly with one another, despite the great difference in the three classes of image. Among the three-segment configuration types, the correlation of the ranks between ancestral and National Geographic images is $R = 0.90$ ($P < .000001$), between ancestral and CGI buildings is $R = 0.87$ ($P < .000001$), and between National Geographic and CGI buildings is $R = 0.87$ ($P < .000001$). Not only do the three kinds of image have configuration distribution ranks that strongly correlate, but also plotting the frequencies directly against one another leads to best-fit power laws of approximately $y = x$. In particular, the best-fit power laws are $y = 1.054x^{1.083}$ for ancestral versus National Geographic images, $y = 0.935x^{0.924}$ for CGI buildings versus ancestral images, and $y = 0.931x^{0.992}$ for CGI buildings versus National Geographic images.

These results strongly suggest that the configuration distribution for natural scenes does not vary much as a function of the kind of natural scene. As mentioned earlier, we believe this is in part because the notion of shape we have chosen to employ here is topological and largely insensitive to large changes in the kind of environment from which one samples pictures. Informally, we expect that if there is intelligent alien life, then so long as they live among macroscopic opaque objects strewn about, they will probably have a configuration distribution that is somewhat similar to those in figure 6*a* (and our ecological hypothesis would predict that their visual sign ranks will be similar to ours).

Recall that our main ecological hypothesis predicts that the configuration distribution ranks for human visual signs will match that found in natural scenes. We averaged the three natural scene configuration distributions from figure 6*a*, and this average is shown in figure 6*b* (*solid line*, *left*). Also plotted there (*dotted line*) is the summary configuration distribution for visual signs from figure 2. One can immediately see that there is a strong similarity between the ecological configuration distribution and the visual sign distribution, and most of the signature peaks and valleys for visual signs are found in the ecological distribution. Figure 6*b* (*right*) shows that the ranks of the three-segment configurations for the natural images and the visual signs correlate highly with each other ($R = 0.88$, $P < .000001$). The frequencies (in logarithmic space) are also highly correlated with each other ($R = 0.89$, $P < .000001$). The rank correlation among three-segment configuration types of the summary ecological distribution with the three different kinds of visual sign alone are also very strong: nonlogographic writing systems ($R = 0.76$), Chinese characters ($R = 0.84$), and nonlinguistic symbols ($R^2 = 0.88$). To summarize, figure 6*a* shows us that the configuration distribution found in natural scenes is highly robust, being largely invariant across disparate environments, and figure 6*b* demonstrates that visual signs and natural images possess a similar configuration distribution, consistent with our ecological hypothesis.

A Priori Approaches to Measuring the Configuration Distribution in Natural Scenes

Here we more closely examine the configuration types in natural scenes and attempt to understand why the configuration distributions across such diverse classes of image have such similar distributions. We will be able to, with

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

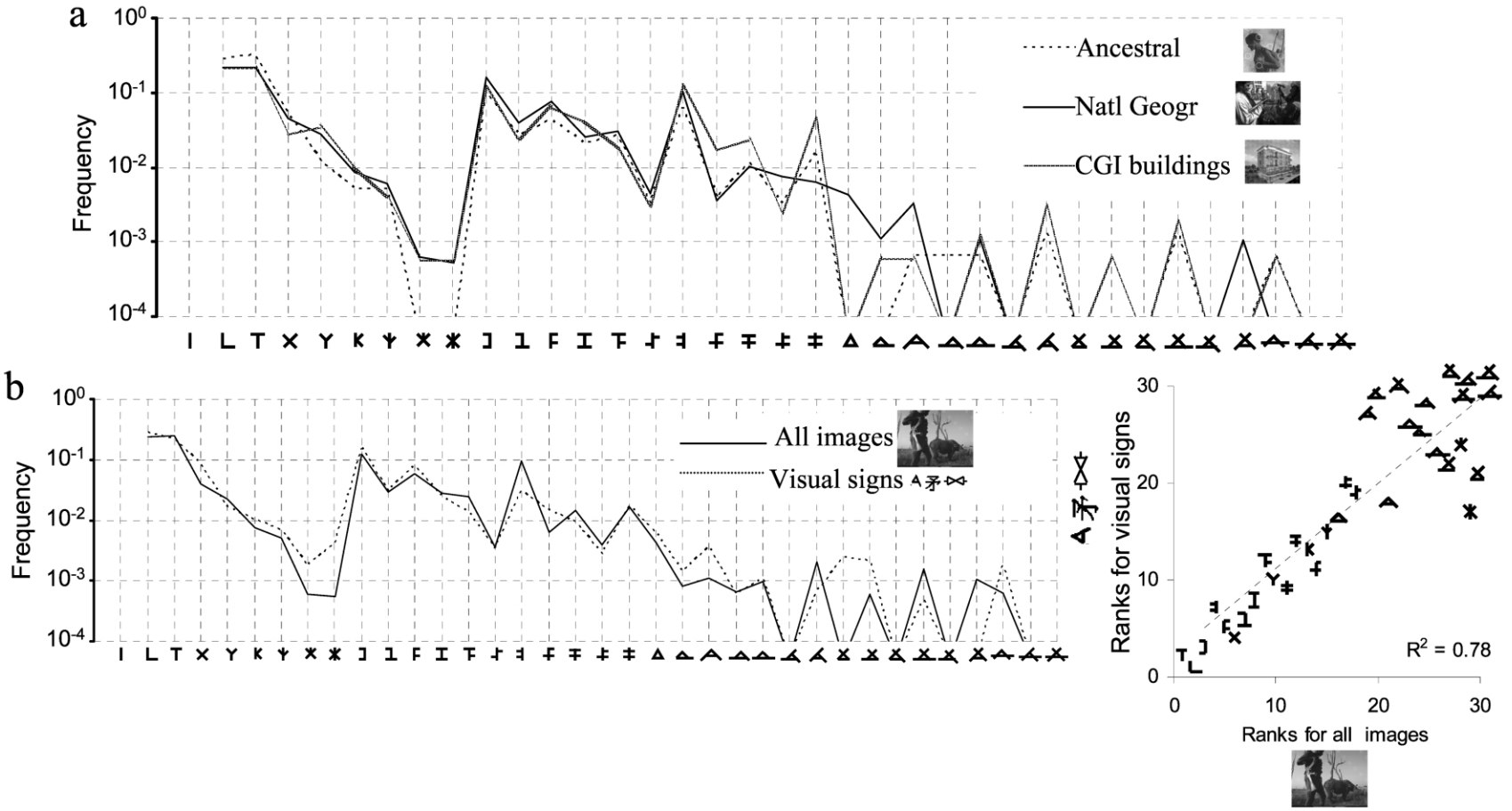


Figure 6: *a*, Configuration distribution for three kinds of image: ancestral, miscellaneous National Geographic, and computer-generated images of buildings. One can see that these diverse classes of natural images possess similar configuration distributions. *b*, Distribution for visual signs is shown along with the average distribution from the three kinds of natural image from *a*, and one can see a very close match in the peaks and valleys. The plot on the right shows the frequency ranks for visual signs versus the summary ecological frequency ranks. The correlation displayed is for the three-segment configuration types only.

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

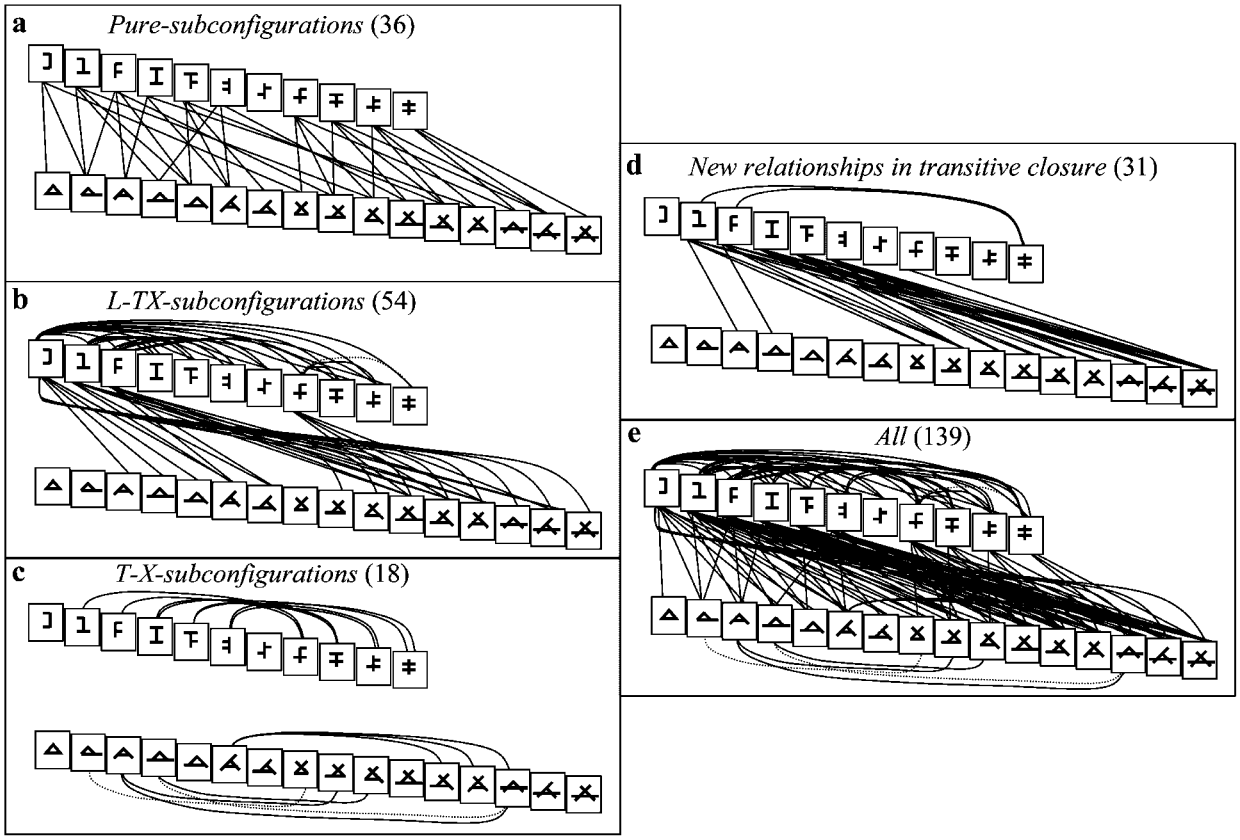


Figure 7: *a*, All 36 pure subconfiguration relationships, where when a connection exists between two configuration types, the one higher on the page is a subconfiguration of the other. *b*, All 54 L-TX subconfiguration relationships. See "L-TX Subconfigurations" in the appendix. *c*, All 18 T-X subconfiguration relationships are shown. See "T-X Subconfigurations" in the appendix. *d*, Thirty-one subconfiguration relationships resulting by taking the transitive closure of the preceding relationships. *e*, All of the subconfiguration relationships are shown in a single network with 139 relationships. Dotted edges in any of the above graphs (there are a total of three) signify the predictions that are mistaken for the set of visual signs.

either no or weak empirical assumptions, derive a partial order over the configuration types; that is, for many pairs of configuration types, we will derive that one configuration type will almost always be more frequent than the other.

We begin with an ecological relationship so strong that it must hold across all possible ecological settings: α is a pure subconfiguration of β ($\alpha >_{\text{psub}} \beta$), just in case α is one of the configuration types resulting from removing one or more junction from β. When a junction is removed from a configuration, we mean only that the junction is removed but not the segments; one should imagine erasing just the ink at the junction itself. The configuration type P (i.e., if the arms of F met at their tips) would not be a pure subconfiguration of A because removing part of a junction is not allowed. If $\alpha >_{\text{psub}} \beta$, then every time β occurs (i.e., is projected onto the retina), by necessity, so do all its pure subconfigurations—and not vice versa—and thus it follows that α is ecologically more probable than β. There are 36 pure subconfiguration relationships among the 32 configuration types with exactly three segments (fig. 7*a*). These 36 pure subconfiguration relationships amount to 36 predictions concerning the relative frequency of the respective pairs of configuration types in visual signs. For example, configuration type F is a pure subconfiguration of A, and thus the prediction is that F configuration types should be more frequent across visual signs than A configuration types. To ensure that these pure subconfiguration predictions are not tautologically true for visual signs, our measurements of configurations excluded all configurations of a length of three if they were pure subconfigurations of another configuration of a length three (see "Configuration Measurement and Tests of Observer Veridicality" in the appendix). Surprisingly, all 36 of these pure subconfiguration predictions are, in fact, true across visual signs.

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Other strong ecological relationships can be derived with the help of some defensible empirical assumptions concerning the relative probability of L, T, and X junctions. Within the computational vision and visual perception literature (Guzman 1969; Clowes 1971; Huffman 1971; Turner 1974; Waltz 1975; Chakravarty 1979; Kanade 1980; Barrow and Tenenbaum 1981; Binford 1981; Cutting and Massironi 1998; Kellman et al. 2001), L junctions are typically the result of contiguous contours, T junctions are the result of partial occlusions, and X junctions are the result of object surface adjacencies (such as stacks or tiling of objects; see Waltz 1975) or partial transparency (Adelson and Anandan 1990; Anderson 1997). As one may see in figure 6*a*, consistent with intuition, Ls and Ts are commonplace, having a frequency of the same order of magnitude, whereas Xs are relatively rare. From these empirical results, it is possible to derive two other kinds of subconfiguration (fig. 7*b*, 7*c*), which are defined and justified in "L-TX Subconfigurations" and "T-X Subconfigurations" in the appendix. Unlike the pure subconfiguration relationships discussed earlier—which must be true across all ecological settings—these two new subconfiguration relationships are only probabilistically expected. The transitive closure on all three kinds of subconfiguration leads to 39 new subconfiguration relationships (fig. 7*d*).

In all, there are a total of 139 LTX subconfiguration relationships, that is, strong ecological relationships likely to be invariant across a wide variety of human ecological settings, shown in figure 7*e*. These 139 subconfiguration relationships make predictions about only the configuration types with more than one junction (i.e., configuration types 10–36) and amount to 40% of all 351 potential relationships. All but four (2.88%) of these 139 relationships are, in fact, true of the average of our ecological distributions (fig. 6*a*). More important, among these 139 predictions, we find near-perfect (three mistaken predictions, or 2.16%) agreement with the relative orderings across visual signs, consistent with the ecological hypothesis. The three mistaken predictions are $\langle \text{F-} >_{\text{sub}} \neq \rangle$, $\langle \text{P} >_{\text{sub}} \text{drum} \rangle$, and $\langle \text{tent} >_{\text{sub}} \text{not} < \rangle$, shown in figure 6 (*three dotted edges*). Among nonlogographic writing systems, there is just one mistaken prediction ($\langle \text{tent} >_{\text{sub}} \text{not} < \rangle$), among Chinese characters there are three mistakes ($\langle \text{F-} >_{\text{sub}} \neq \rangle$, $\langle \text{P} >_{\text{sub}} \text{drum} \rangle$, and $\langle \text{tent} >_{\text{sub}} \text{not} < \rangle$), and for symbols, there are five mistakes ($\langle \text{F-} >_{\text{sub}} \neq \rangle$, $\langle \text{P} >_{\text{sub}} \text{drum} \rangle$, along with $\langle 1 >_{\text{sub}} \Pi \rangle$, $\langle 1 >_{\text{sub}} \neq \rangle$, and $\langle \text{H} >_{\text{sub}} \neq \rangle$). In contrast, 60 of these 139 relationships (43.2%) are not met by the relative orderings for children scribbles.

Discussion

The first result of our article is that there do appear to be regularities in the (topological) shapes of human visual signs (see fig. 2). We saw that the configuration distribution is very similar across approximately 100 nonlogographic writing systems, Chinese, and nonlinguistic symbols. This configuration distribution signature is not a result of any simple tilings, random line images, or scribbles (see fig. 3). We should reiterate that in this article, by "shape," we specifically mean "configuration type," as shown in figure 1. This topological notion of shape has an advantage over more geometrically based notions, where the specific stroke/contour shape and orientation would matter: typically, any given human visual sign can undergo significant variability in its geometrical structure without losing its identity, but typically, its topology cannot vary. Nevertheless, why visual signs have the geometrical shapes they do is an interesting question, and this research does not touch on this.

Our second result is that the shapes of visual signs appear to be selected primarily for vision at the expense of motor. We concluded this by examining shorthand versus trademarks, where the former is selected for writing at the expense of vision and the latter is selected for vision at the expense of the motor system. More generally, we saw that shorthand does not, but trademarks do, possess the signature configuration distribution found in visual signs (see figs. 4*a*, 5*a*). We also made this "vision at the expense of motor" conclusion by comparing the motor complexity versus the visual complexity of the 36 configuration types. We found that motor complexity does not correlate with the frequency ranks of configuration types, whereas visual complexity strongly correlates (see figs. 4*b*, 5*b*). Although the visual sign signature may be primarily selected for vision, we believe there are certainly selection pressures from writing. For example, consider the "drum" configuration type (28 in fig. 1). Figure 4*b* shows that it requires less motor complexity compared with its neighbors in the plot, and figure 5*b* shows that it possesses similar visual complexity compared with its neighbors. But note that in visual signs, as shown in figure 2, "drum" tends to be common compared with many of its neighbors; "drum" is also nonexistent in natural scenes, as shown in figure 6*a*, and nonexistent in trademarks, as shown in figure 5*a*, which are selected for vision entirely, with little or no motor contribution. This suggests that "drum" may occur to the extent it does in visual signs primarily because it is easy to write, not because it is easy to see (and not because it occurs in natural scenes). Also, we note that our notion of motor complexity is somewhat natural when writing by hand but would not be as appropriate for other methods of producing visual signs, such as chiseling into stone or incising into clay. In the future, we plan on acquiring configuration distributions for visual signs grouped on the basis of the production mechanism and measuring the differences between the distributions; we expect that al-

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

though the geometrical shapes would vary considerably as a function of the kind of implement, the topological shapes would vary considerably less, being primarily selected for vision (as we see, e.g., in trademarks, which tend not to be drawn by hand and yet possess the signature configuration distribution).

Our third and final result was that the signature configuration distribution for human visual signs closely matches that of natural scenes, in that their ranks are highly correlated. This serves to provide evidence for the ecological hypothesis that because we have evolved to be competent at processing the configuration types found in natural scenes, there has been cultural selection pressure for human visual signs to disproportionately possess the naturally common configuration types. We tested this in two distinct manners. First, we obtained measurements of the configuration distribution from three classes of image. We saw that despite these three classes of image being radically different, the configuration distributions were similar (see fig. 6*a*). More important, we demonstrated that the configuration distribution for these natural images matches (at the level of rank correlation) very closely the signature distribution for human visual signs (see fig. 6*b*). Second, we presented a theoretical framework for making a priori and semi–a priori claims about the relative probability of configuration types in natural scenes. This theoretical framework allowed us to make 139 predictions concerning the relative order of the probability of configuration types, as shown in figure 7. We saw that the configuration types found in human visual signs closely conform to these 139 predictions, with only several mistaken predictions, or about 2%.

We foresee at least the following three future directions for this research. The first is methodological, and it concerns the development and validation of software for automatically measuring configuration distributions from images. (Care must be taken because any such software will rely on image segmentation software, and it is well known that the resulting segmentation of an image into contours depends crucially on the choice of algorithm, of which there are dozens of choices.) Although we have tests of the objectivity of our human observers (see "Configuration Measurement and Tests of Observer Veridicality" in the appendix), validated software would radically increase the size of our data sets. Second, in this article, we have focused on the similarities between different classes of human visual sign. However, there are differences as well, as can be seen in figure 2*a*, and it is of interest in the future to quantify these differences and to test whether the differences can also be found in differences in the respective ecologies where the visual signs originated and developed. Finally, the ecological hypothesis of visual sign shape suggests a new theory of visual stimulus complexity. Rather than visual stimulus complexity being the result of specific stimulus features (like the number of angles) or of abstract mathematical properties, we propose that visual stimulus complexity is determined by its ecological probability: more improbable configuration types are expected to be the more visually complex ones (i.e., they are complex because they are improbable, not the other way around). More specifically, in an optimal code for configuration types, the length of the code will be proportional to $-\log(p)$, where p is the ecological probability of that configuration type (something fig. 6*a* possesses).

Acknowledgment

Support for this research was given by National Institutes of Health grant 5F32EY015370-02.

APPENDIX

Supplementary Information on Methods and Theory

Configuration Measurement and Tests of Observer Veridicality

All configurations measured—whether from visual signs or images—were carried out by trained observers using their subjective judgment. The primary author (M.A.C.) obtained the configuration data from writing systems, Chinese, and shorthand, and two undergraduate coauthors (Q.Z. and H.Y.) worked full-time for one summer measuring configurations from symbols (Q.Z.), trademarks (Q.Z.), scribbles (Q.Z.), and natural images (both Q.Z. and H.Y.). Each undergraduate also subsequently took measurements from a set of methods-testing images created by M.A.C., described later. The primary author trained the undergraduates to recognize the variety of configuration types; training consisted of measuring data from one of the data sets for several days under the watch of the primary author, who would check whether eligible configurations were missed and give feedback.

For nonlogographic writing systems and shorthand (see "Writing Systems and Shorthand"), configurations were measured only from characters with three or fewer strokes, and for each such character, its configuration type was determined. Nearly every such character has one unambiguous configuration type and is easily measured. For Chinese writing, symbols, trademarks, and scribbles (see "Chinese, Nonlinguistic Symbols, Trademarks, and Children Scribbles"), each drawing typically possesses more than three segments, and all subconfigurations of lengths of two or three were measured from the drawing. For example, the symbol ⇔ possesses two Ls, four Ts, four Fs, and two Hs, and the symbol ✡ possesses six Ls, six Xs, 12 F-'s, and six not<'s. In our analysis, we did not count

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

configurations of a length of three if they were subconfigurations of some other configuration of a length of three. For example, Z is a subconfiguration of Δ, and so if a Δ occurred in a character, the three Zs that are subconfigurations of Δ would not be included. And in ✡, F- is a subconfiguration of not<, and the F-'s would not be counted in this case. This was done so that the predictions we make concerning pure subconfigurations later are not tautological (see main text) and so that the set of three-segment configuration types possesses frequency distributions that are independent. Measuring configurations from Chinese, symbols, trademarks, and scribbles required some practice because a seemingly simple sign like the double arrow ⇔ possesses 12 eligible subconfigurations. However, once one is trained to spot the many subconfigurations of a larger sign, there are almost always no ambiguities, and the configurations are readily measured.

Measuring configurations from natural images was more difficult, and the task of measurement was simplified by placing sampling circles at regular intervals over the image; having one set of natural images be, in actuality, computer-generated images (of commercial buildings), which are easier to measure from than actual pictures; and having the two trained observers specialize their efforts on two different kinds of data collection. In particular, one undergraduate observer (H.Y.) was given the task of concentrating just on the single-junction configuration types, namely, L, T, X, Y, K, Ψ, "man," and "asterisk," and the other undergraduate observer (Q.Z.) was given the task of identifying only the configuration types with exactly three segments, namely, Y, K, Ψ, "man," "asterisk," Z, 1, F, H, …, "camp." Each judged both the contours and the configurations and was instructed to judge on the basis of the local stimulus properties. Single-junction configurations were measured differently from multiple-junction configurations. Single-junction configurations were measured by taking all single-junction configurations found within a sampling circle.

Multiple-junction configurations, however, are not pointlike like a single junction and can extend across the image—accordingly, a more subtle measuring technique is required. Multiple-junction configurations were measured by finding within a sampling circle the contour nearest to the center (if there was any contour within the sampling circle at al), and recording all configurations in the image (possibly outside of the sampling circle) for which that contour is a part. The observer made a subjective determination of which contour was nearest to the center of a sampling circle. For example, suppose that the top horizontal contour in the double arrow ⇔ lies within a sampling circle and is the nearest to the center of the circle. The observer would record all of the three-segment configurations that this contour is a part of. In this case, the top horizontal contour is a part of two Fs and one H. This is the case no matter the size of the double arrow, for example, whether the entire double arrow is small and lies within a sampling circle or is as large as the entire image and the top horizontal contour happens to cross only through the center of the sampling circle. If the exact same configuration (not configuration type) was found via two different sampling circles—which could happen if one contour of a configuration enters one sampling circle and another contour of the configuration enters another sampling circle—then the configuration was not counted again.

As discussed in the legend of figure A1, it is possible to compare the measurements made by these two different observers and their two distinct measuring techniques over the eight single-junction configuration types. Figure A1 shows that, for the relative frequencies among single junctions, the two observers—and the two measuring techniques—highly agree with each other. Because of the two distinct measuring techniques for single and multiple junctions, we normalized the distributions by virtue of the fact that both observers had an overlapping task of measuring the three-segment, single-junction configuration types (Y, K, Ψ, "man," "asterisk").

In order to test the objectivity of our trained observers' measurement of configurations, a variety of random line images were created. First, an image type was created (using Matlab) called "Random Ls," which consisted of 100 Ls placed on the page, as seen in figure A2*a*. Another image type was created called "Random Ts," which used the same parameters as in "Random Ls" but where now the vertex of each L was severed and one segment was slid randomly along the other segment, making a random T. A third type of image created was called simply "Random," which used the same parameters again but where now the vertex of each L was severed altogether, leaving just two randomly placed lines. The idea behind the three image types is that the last kind—Random—would consist of only randomly placed lines and the other two would, although possessing many random crossings, possess a significant contribution from Ls and Ts, respectively, for image type Random Ls and Random Ts. Five images of each type were created, and the two trained observers measured from them using the same technique they had used for natural images. For Random, 240 single-junction configurations and 276 multiple-junction configurations were measured. For Random Ls, 272 single-junction configurations and 412 multiple-junction configurations were measured. And for Random Ts, 273 single-junction configurations and 342 multiple-junction configurations were measured. The single-junction and multiple-junction configurations were sampled using distinct methods, but the multiple-junction

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

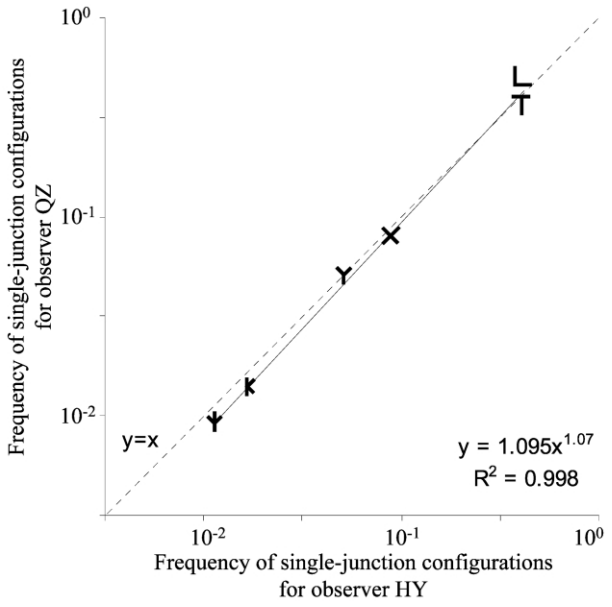


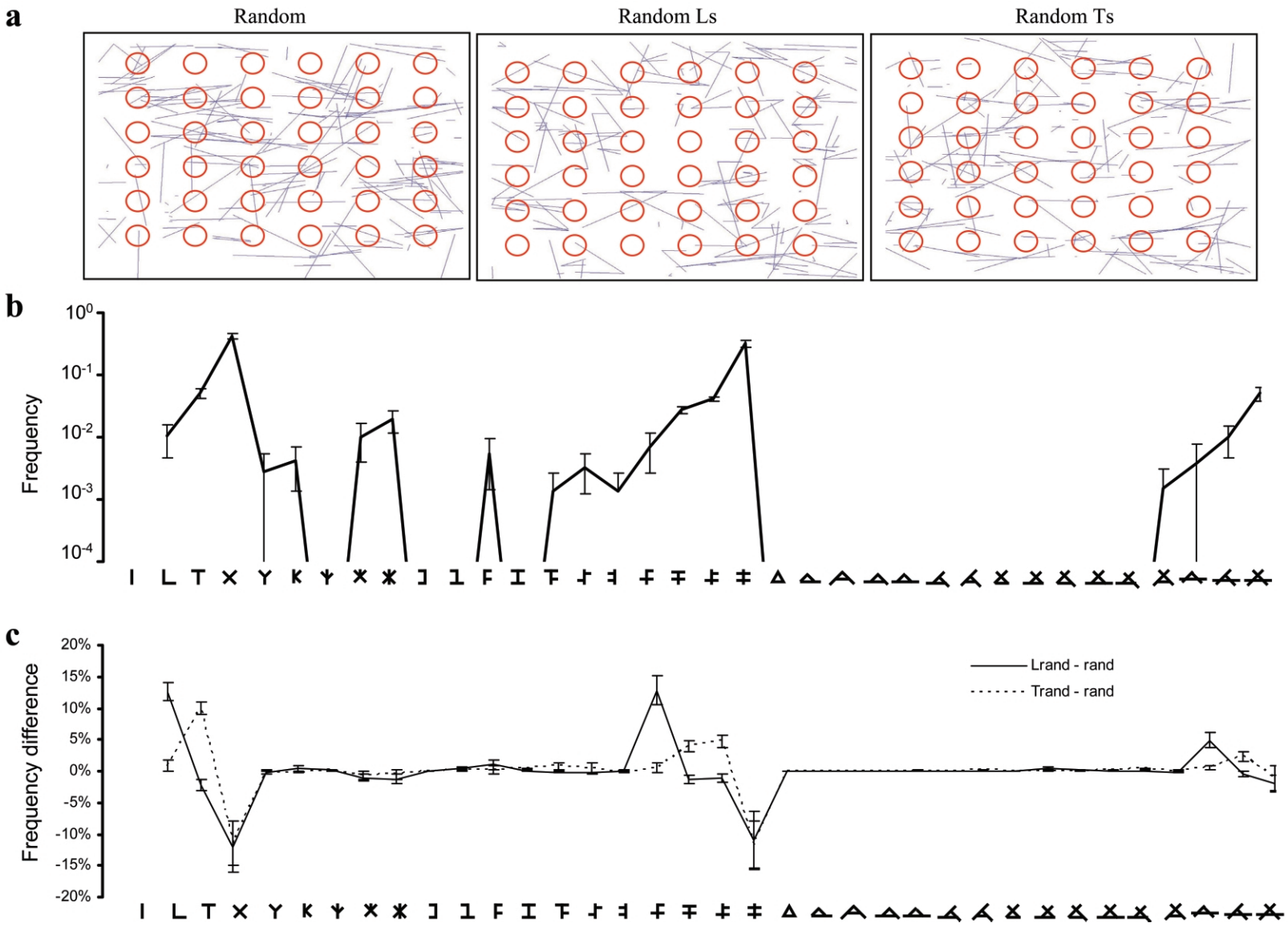
Figure A1: Log-log plot of the frequency of the single-junction configuration types for one observer versus that of the other observer, where one observer (H.Y.) concentrated on single-junction configuration types and used one measuring technique and another observer (G.Z.) concentrated on three-segment configuration types and used a distinct measuring technique. Note that the three-segment configuration types include the following single-junction configuration types: Y, K, Ψ, "man," and "asterisk." Furthermore, each junction in any of the configuration types with two or three junctions (i.e., configuration types 10–36) is either an L, T, or X, and thus from their distribution, it is possible to estimate the distribution for Ls, Ts, and Xs. The latter distribution was normalized so that the ratio of X to Y was the same as that measured by the other observer. The eight single-junction configuration types were therefore measured (at least implicitly) via both measuring techniques and can be directly compared to one another, which is what this plot shows. The best-fit (by least squares) solid line corresponds to a relationship of $y = 1.1x^{1.07}$, very close to $y = x$ (*dashed line*), showing that the single-junction configuration type distribution does not vary much between the two observers and the two distinct measurement techniques. Of the eight single-junction configuration types, for two of them—"man" and "asterisk"—one observer (G.Z.) found no occurrences, whereas the other observer (H.Y.) found them with frequencies 0.118% and 0.102%. These are not shown in the log-log plot because one cannot take the logarithm of 0. Plotting these data on an unlogged plot (not shown) where all eight single-junction configuration types can be plotted leads to a best-fit equation $y = 1.0303x - 0.0047$ ($R = 0.99$) or, again, very nearly $y = x$.

observer also measured single-junction configurations via the multiple-junction technique and the single-junction and multiple-junction measurements could thereby be relativized to each other. The two observers were not informed about the manner in which these three types of image were created. These random images allow tests of our methods.

First, consider just the Random image type, which consists of 200 random lines on the page. It is possible to derive approximately what the configuration distribution looks like for random line images, and once we determine this, we can check to see whether our trained observers measured the mathematically expected distribution. The measured distribution is shown in figure A2*b*. For random line images, the most probable junction type is X. Ts are comparatively rare because for two lines to make a T requires a coincidental alignment, where one segment abuts against the side of another segment. Ls are even more improbable because for two lines to make an L requires a very coincidental alignment, where one segment's endpoint happens to touch another segment's endpoint. We therefore expect the following to be true about the configuration distribution for random line images: Among the two-segment configuration types (namely, L, T, and X), X should be by far the most probable configuration type, followed by T. This can be seen to be true for the measured distribution in figure A2*b*. Among the two-junction configuration types (namely, Z to ≠), ≠ should be by far the most probable, followed by configuration types with one X and one T (namely, T- and FL), and indeed this can be seen to be true for the measured distribution. Finally, among the three-junction configuration types (namely, Δ to "camp"), "camp" should be by far the most probable, followed by configuration types with two Xs and one T (namely, A-′), and, again, this can be seen in the measured distribution in figure A2*b*. This plot provides evidence, then, that our trained observers are capable of measuring distributions significantly different from that found in visual signs and natural scenes (as we will see later) and that our measurements are driven by the objective, underlying distribution. Furthermore, the configuration distribution for random line images is skewed toward more complex configuration types, and the fact that our observers are capable of measuring this provides evidence that our observers are not biased toward simpler configuration types.

To further test our observers' measurements, we may check to see that their distributions for the Random L and Random T image types are as expected. Recall that these image types consist of randomly placed Ls and Ts, respectively. However, each is similar to the Random case, in that they will possess many X junctions because of random crossovers. Compared with the Random case, however, we expect Random L image types to have significantly more configuration types with L, and we expect Random T image types to have significantly more configuration types with T. Figure A2*c* shows two plots. The first is Lrand–rand, which shows the difference between the Random L configuration distribution and the Random distribution. The second is Trand–rand, which shows the difference between the Random T configuration distri-

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).



This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

bution and the Random distribution. Standard error bars are shown on both. If the observers are objectively measuring the configuration distribution of the Random L and Random T images, then we expect that the two plots in figure A2*c* should not significantly differ from one another except in the following two cases. First, if a configuration type has just one L and no T, then the Lrand–rand plot should be significantly above the Trand–rand plot. Second, if a configuration type has just one T and no L, then the Trand–rand plot should be significantly above the Lrand–rand plot. This is, indeed, exactly the case, as can be seen by examination of the plot here. (Note that we do not expect configuration types with, say, two Ls and no Ts, or vice versa, to be more probable because the Random L images have negligible probability of having configurations with two Ls.) We also note that the two plots in figure A2*c* look nearly identical except when they differ in regard to having one L or T. And, except for the expected reduction in the number of Xs, $\neq$'s, and "camp"s, there is no other loss of more complicated configuration types in the Random L or Random T image types, suggesting that the presence of the simpler L or T junctions does not bias the observers from noticing the presence of the more complicated configuration types (e.g., "asterisk," which is found with frequencies of 0.66%, 0.68%, and 1.48% for Random, Random L, and Random T, respectively).

Finally, we reiterate that one set of our natural images—the CGI buildings—was computer generated. This was originally so as to make data collection simpler for our observers because computer-generated images tend to have contours that are cleaner and easier to locally identify. Our thinking was that although data collection would be simpler, these computer-generated images may nevertheless possess a sample of configurations that is representative of truly natural (urban) images. In regard to methods, having measurements from both computer-generated images and truly natural images is useful because the configuration distribution for the computer-generated images matches very well the distribution for the photographs of natural scenes; for example, the correlation between CGI buildings and the National Geographic images is $R = 0.92$ ($\mathrm{df} = 20$, $P < .000001$), and best-fit power law is $y = 0.931x^{0.992}$ (i.e., very nearly $y = x$). Therefore, any extra difficulty that our observers encountered in measuring from photographs of natural scenes—compared with computer-generated images—did not appear to lead to differences in the resulting distribution.

Writing Systems and Shorthand

Configuration types were taken from all characters with three or fewer strokes across the 115 writing systems from Ager (1998) and Daniels and Bright (1996) used in a companion study (Changizi and Shimojo 2005) that included numerals, abjads, abugidas, alphabets, and syllabaries from five major taxa—Ancient Near Eastern, European, Middle Eastern, South Asian, and Southeast Asian—as well as invented writing systems. For our sample of writing systems, the number of writing systems per taxon correlates highly ($R = 0.90$, $\mathrm{df} = 3$, $P < .05$) with the number of sections devoted to the taxon of Daniels and Bright (1996), who wrote one of the most comprehensive books on writing systems. Only the following 96 systems possessed at least one eligible character (i.e., having one of the configuration types used in this study), and there were 1,442 eligible characters in all: Ahom, Albanian (Elbasan), Ancient Berber (Horizontal), Arabic, Arabic numerals, Aramaic, Armenian (Eastern), Asomtavruli, Avestan, Bassa, Batak, Bengali numerals, Brahmi, Buhid, Burmese, Burmese numerals, Carrier, Celtiberian, Cherokee, Chinese numerals, Cypriot, Cyrillic, Dehong, Dehong numerals, Dhives akuru, Enochian, Ethiopic, Etruscan (archaic), Faliscan, Fraser, Gothic, Greek, Gujarati, Gujarati numerals, Gurmukhi, Hanuno'o, Hebrew, Hungarian Runes, Hungarian Runes numerals, Iberian (Northern), Iberian (Southern), Kharosthi, Kharosthi numerals, Korean (Hangeul), Kpelle, Latin (ancient), Latin (all caps), Latin (modern), Lepcha (Rong), Limbu, Linear B, Marsiliana, Meroitic, Messapic, Middle Adriatic (South Picene), Middle Persian (Pahlavi), Mkhedruli (Georgian), Mongolian, Mongolian numerals,

Figure A2: *a*, Example random line images. *Left*, random consists of 200 random lines. *Middle*, random Ls consists of 100 random L junctions. *Right*, random Ts consists of 100 random T junctions. The first two are created using parameters identical to those in the Random L image, except that in the Random T images, each L is broken and turned into a T, with the stem randomly positioned along the length of the top of the T, and in the Random images, each L is broken altogether, placing the two segments of the L randomly on the page. Circles are the sampling circles used by the trained observers to measure configurations from these image types. *b*, Configuration distribution for configurations measured from images with random lines (i.e., Random images). The distribution is the average of five distributions from five random line images, and the error bars show standard error. The random line configuration distribution is as expected because X junctions are more probable than T junctions and T junctions more probable than L junctions. This figure serves to demonstrate that our observers are capable of measuring the expected distribution. *c*, Difference between the Random L image configuration distribution and the Random configuration distribution (*solid line*; the latter distribution shown in *b*). This is called Lrand–rand. Difference between the Random T image configuration distribution and the Random configuration distribution (*dotted line*). This is called Trand–rand. Standard error bars are shown. This plot helps to demonstrate that our measurements are modulated as expected as we modulate the ground truth configuration types.

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Figure A3: Ancestral images (photographs) used in our measurements of the configuration distribution in natural scenes, taken from the public domain (including many from the Web page of the late Professor Norbert Schoenauer, http://www.arch.mcgill.ca/prof/schoenauer/arch528/lect01/n01.htm, and also from http://www.plantzafrica.com/vegetation).

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Figure A4: Rural images (photographs) used in our measurements of the configuration distribution in natural scenes, taken from http://www.nationalgeographic.com.

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Figure A5: Urban images (CGI) used in our measurements of the configuration distribution in natural scenes, taken from http://www.nemetschek.net/gallery/.

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

Nabataean, Ndjuka, New Tai Lue, New Tai Lue numerals, Nikhilipi, Nikhilipi numerals, N'Ko, N'Ko numerals, North Picene, Old Church Slavonic, Old Permic, Oscan, Pahawh Hmong, Pahawh Hmong numerals, Parthian, Phags-pa, Phoenician, Pollard Miao, Psalter, Redjang, Runic (Danish Futhark), Runic (Elder Futhark), Sabaean, Samaritan, Santali, Santali numerals, Sil'oti Nagri, Somali, Somali numerals, South Arabian, Syriac, Tagalog, Tagbanwa, Thaana, Tifinagh, Umbrian, and Varang Kshiti. We also measured configurations from 121 letters from five shorthands (Daniels and Bright 1996): Tironian, Characterie, Willis's Stenographie, Pitman, and Gregg. All of these writing systems are deemed nonlogographic because the characters do not stand for whole words; Chinese is largely logographic, for example. (Numerals are an exception because numerals stand for abstract numbers. Because of the close phylogenetic connection between the above writing systems and many numeral systems, the numeral systems are placed within this nonlogographic data set.)

Chinese, Nonlinguistic Symbols, Trademarks, and Children Scribbles

Chinese characters, nonlinguistic symbols, corporate trademarks, and scribbles typically possess more than three contours, and we measured all of the eligible subconfigurations found within each of these signs. For the data set from Chinese, the first (simplified) character on each of the first 340 pages of Manser's (2003) *Pocket Oxford Chinese Dictionary* was sampled, leading to 4,759 configurations (2,694 two-segment configurations and 2,065 three-segment configurations). For symbols, 3,538 three-segment configurations were measured from 1,107 symbols in Dreyfuss's (1972) *Symbol Sourcebook*. The symbols in the book are divided into the following 27 categories, followed by the number of symbols we sampled from that category: introduction/basic symbols (uncategorized), 85; accommodations and travel, 22; agriculture, 141; architecture, 108; astronomy, 5; biology, 33; business, 99; chemistry, 4; communications, 75; engineering, 93; folklore, 84; geography, 8; geology, 12; handling of goods, 4; home economics, 26; manufacturing, 54; mathematics, 11; medicine, 48; meteorology, 21; music, 13; photography, 7; physics, 13; recreation, 8; religion, 4; safety, 40; traffic, 36; and vehicle controls, 53. In total, 863 two-segment configurations were separately sampled from the book, along with Y configurations with which the two- and three-segment configuration frequencies were relativized to one another. For trademarks, 1,085 three-segment configuration types were measured from 126 symbols in Capitman's (1976) *American Trademark Designs*. We separately sampled 318 two-segment configurations from the book, along with Y configurations with which the two- and three-segment configuration frequencies were relativized to one another. For symbols and trademarks, to qualify for sampling, a symbol must have distinct contours, not be an obvious representation of a three-dimensional or real-world object, and not be a duplicate of a previously sampled symbol. For children scribbles, 577 single-junction configurations and 149 multiple-junction configurations were measured from 17 scribble drawings by 26–32-month-old children (Kellogg 1969, pp. 27–30).

Images of Natural Scenes

Three sets of natural images were used: ancestral, National Geographic, and CGI buildings. The first class of images we refer to as "ancestral" and consisted of 27 photographs of savannas and tribal life (see fig. A3); 535 three-segment configurations were sampled. The second class of images we refer to as "National Geographic" and consisted of 40 miscellaneous photographs of rural and small-town life taken from the *National Geographic* Web site (see fig. A4); 471 three-segment configurations (and 822 two-segment configurations) were sampled. The third (and final) class of images we call "CGI buildings" and consisted of 40 computer-generated, highly realistic images of exteriors and interior views of commercial buildings (see fig. A5); 799 three-segment configurations (and 3,213 two-segment configurations) were sampled. See "Configuration Measurement and Tests of Observer Veridicality" for more information on how configurations were measured.

L-TX Subconfigurations

The idea behind L-TX subconfigurations is that adding a topology-preserving L junction (e.g., adding an L to L to make Z is topology preserving but not adding an L to Z to make Δ) to a configuration type decreases the ecological probability less than adding a T or X junction. To see why, consider first the relationship between L and T junctions, which appear to have ecological probabilities of similar order of magnitude (fig. 6*a*). Consider starting with an L junction and adding to it a topology-preserving L junction; the resulting configuration type is Z. Instead, consider starting with that same L junction but adding to it a T junction; there are two different possible resulting configuration types, 1 and F, depending on whether the stem or the top of the T is appended to one of the arms of the L. Because the ecological probability of L is of the same order of magnitude as that of T (fig. 6*a*), the ecological probability of Z is of the same order of magnitude as the ecological probability of ⟨1 or F⟩. Therefore, the ecological probability of Z will be greater than either of them alone. Because of this combinatorial difference in the way Ls and Ts append to configuration types and because Ls and Ts

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

are approximately equally ecologically probable, the general conclusion is that adding a topology-preserving L junction will decrease the ecological probability of a configuration type less than adding a T junction.

Now consider the relationship between L and X junctions. As mentioned earlier, L junctions are ecologically more probable than X junctions (fig. 6*a*). Furthermore, each is combinatorially similar (unlike Ts), in that for any arm of a configuration type on which an L junction can be appended, there is also only one topologically distinct way to append an X junction. It follows that adding a topology-preserving L junction will decrease the ecological probability of a configuration type less than adding an X junction.

This motivates the following definition: α is an L-TX subconfiguration ($>_{\text{L-TXsub}}$) of β, just in case there is a configuration type γ such that $\gamma >_{\text{psub}} \beta$ via the loss of N T or X junctions; $\gamma >_{\text{psub}} \alpha$ via the loss of at least one topology-preserving L junction; and the number of segments in α is no greater than that in β. By virtue of the above arguments, if $\alpha >_{\text{L-TXsub}} \beta$, then we (statistically) expect that α is ecologically more probable than β.

There are 54 L-TX subconfiguration relationships (fig. 7*b*), and 50 of these are upheld within the average of the three ecological distributions from figure 6*a*, the mistaken predictions being $\langle 1 >_{\text{L-TXsub}} \Pi\rangle$, $\langle \text{F-} >_{\text{L-TXsub}} \Pi\rangle$, $\langle \text{F-} >_{\text{L-TXsub}} \text{T-}\rangle$, and $\langle \text{F-} >_{\text{L-TXsub}} \neq\rangle$. In regard to visual sign structure, these 54 L-TX subconfiguration relationships make 54 predictions concerning the relative probability of configuration types across visual signs. All but one ($\langle \text{F-} >_{\text{L-TXsub}} \neq\rangle$) of these 54 L-TX subconfiguration predictions are true across visual signs.

T-X Subconfigurations

The idea behind T-X subconfigurations is that adding a partial-occlusion T junction to a configuration type decreases the ecological probability less than adding an X junction. To understand why, similar to the L versus T combination, consider starting with an L junction and adding to it an X junction; the resulting configuration type is F-. But as mentioned earlier, starting with that same L junction but adding to it a T junction results in two different possible resulting configuration types, 1 and F. If X and T junctions were approximately equiprobable, then F- would be more probable than either 1 or F. However, because partial-occlusion T junctions are more than twice as ecologically probable as X junctions, adding a partial-occlusion T junction to a configuration type should decrease the ecological probability less than adding an X junction.

T junctions are not always the result of partial occlusions, however; in our image data set, T junctions are due to partial occlusions half (49.4%) of the time. One difficulty we are left with, then, is how to know when an added T junction is probably a partial-occlusion junction. There are, in fact, at least three scenarios in which an added T is probably not a partial-occlusion junction.

First, T junctions can be due to stacks and tilings (e.g., abut two cubes against one another), and for this reason, the probability that a T junction is due to partial occlusion is expected to fall if there is also an X junction in the configuration (X junctions being due to stacks and tilings 4.12 times more often than partial transparency in our set of images). For our image data, the probability that a T is due to partial occlusion is 0.627 if the T is not adjacent to an X junction, but the probability falls to 0.146 if it is adjacent to an X junction. For this reason, we expect that adding a T junction to configuration type γ decreases the probability less than adding an X junction but only when γ does not possess any X junctions.

Second, consider appending a T junction to a partial-occlusion T junction. There are four possible resulting configuration types: H, TF, Π, and TL. However, for only the first three of these cases is the second T junction possibly the result of partial occlusion. TL (15 in fig. 1) has been known for more than 30 years to be a simple example of an "impossible" object subpicture (Huffman 1971). It is not literally impossible because it can result from certain surface adjacencies, such as where one cube is placed on top of another and slid horizontally so that the cubes are not aligned. However, at most one of the Ts in TL can be due to partial occlusion. We note that, consistent with this, there is a deep trough at TL for visual signs and for the ecological measurements.

Finally, even when an appended T junction can consistently be interpreted as a partial-occlusion junction, there are cases where this is relatively improbable, as in when there are occlusion loops, such as in configuration type "spiral," where contour c_1 occludes contour c_2, which occludes c_3, which in turn occludes c_1, or in configuration type P′, where contour c_1 occludes c_2, which is part of a contiguous contour with c_3, which occludes c_1.

With these three scenarios in mind, we say that a configuration possessing a T junction and three or fewer segments is occlusion probable if it possesses no X junctions, it is not configuration type TL, and it is neither configuration type P′ or "spiral." This discussion motivates the following definition: α is a T-X subconfiguration ($>_{\text{T-Xsub}}$) of β, just in case there is a configuration type γ such that $\gamma >_{\text{psub}} \beta$ via the loss of N X junctions, $\gamma >_{\text{psub}} \alpha$ via the loss of at least N T junctions, and α is occlusion probable. By virtue of the above arguments, if $\alpha >_{\text{T-Xsub}} \beta$, then we (statistically) expect that α is ecologically more probable than β.

There are 18 T-X subconfiguration relationships (fig.

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).

7*c*), and all of these are found within the summary ecological distribution. In regard to our ecological hypothesis for visual sign structure, these 18 T-X subconfiguration relationships make 18 predictions concerning the relative probability of configuration types across visual signs. All but two ($\langle P >_{\text{T-Xsub}}$ drum$\rangle$ and $\langle$tent$ >_{\text{T-Xsub}}$ not$<\rangle$) of these 18 T-X subconfiguration predictions are true across visual signs.

Literature Cited

Adelson, E. H., and P. Anandan. 1990. Ordinal characteristics of transparency. Paper presented at the AAAI-90 Workshop on Qualitative Vision, July 29, 1990, Boston, MA.

Ager, S. 1998. Omniglot: a guide to writing systems. http://www.omniglot.com.

Anderson, B. L. 1997. A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions. Perception 26:419–453.

Arnoult, M. D. 1960. Prediction of perceptual responses from structural characteristics of the stimulus. Perceptual and Motor Skills 11:261–268.

Attneave, F. 1957. Physical determinants of the judged complexity of shapes. Journal of Experimental Psychology 53:221–227.

Barrow, H. G., and J. M. Tenenbaum. 1981. Interpreting line drawings as three-dimensional surfaces. Artificial Intelligence 17:75–116.

Biederman, I. 1987. Recognition-by-components: a theory of human image understanding. Psychological Review 94:115–147.

Biederman, I., and E. E. Cooper. 1991. Priming contour-deleted images: evidence for intermediate representations in visual object recognition. Cognitive Psychology 23:393–419.

Biederman, I., and P. C. Gerhardstein. 1993. Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. Journal of Experimental Psychology: Human Perception and Performance 19:1162–1182.

Binford, T. O. 1981. Inferring surfaces from images. Artificial Intelligence 17:205–244.

Capitman, B. B. 1976. American trademark designs. Dover, New York.

Chakravarty, I. 1979. A generalized line and junction labeling scheme with applications to scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 1:202–205.

Changizi, M., and S. Shimojo. 2005. Character complexity and redundancy in writing systems over human history. Proceedings of the Royal Society of London B 272:267–275.

Clowes, M. B. 1971. On seeing things. Artificial Intelligence 2:79–116.

Coulmas, F. 1991. The writing systems of the world. Blackwell, Oxford.

Cutting, J. E., and M. Massironi. 1998. Pictures and their special status in perceptual and cognitive inquiry. Pages 137–168 *in* J. Hochberg, ed. Perception and cognition at century's end: history, philosophy, and theory. Academic Press, San Diego, CA.

Daniels, P. T., and B. Bright. 1996. The world's writing systems. Oxford University Press, New York.

Dreyfuss, H. 1972. Symbol sourcebook. Wiley, New York.

Endler, J. A. 1992. Signals, signal conditions, and the direction of evolution. American Naturalist 139(suppl.):S125–S153.

Fairbank, A. 1968. A book of scripts. Penguin, Baltimore.

Guilford, T., and M. S. Dawkins. 1993. Receiver psychology and the design of animal signals. Trends in Neuroscience 16:430–437.

Guzman, A. 1969. Decomposition of a visual scene into three-dimensional bodies. Pages 243–276 *in* A. Grasselli, ed. Automatic interpretation and classification of images. Academic Press, New York.

Helfman, E. S. 2000. Signs and symbols around the world. Authors' Guild Backinprint.com, Lincoln, NE.

Hochberg, J., and E. McAlister. 1953. A quantitative approach to figural "goodness." Journal of Experimental Psychology 46:361–364.

Huffman, D. A. 1971. Impossible objects as nonsense sentences. Pages 295–323 *in* B. Meltzer and D. Michie, eds. Machine intelligence. Vol. 6. Elsevier, New York.

Ifrah, G. 1985. From one to zero: a universal history of numbers. Viking, New York.

Kanade, T. 1980. A theory of origami world. Artificial Intelligence 13:279–311.

Kellman, P. J., S. E. Guttman, and T. D. Wickens. 2001. Geometric and neural models of object perception. Pages 183–245 *in* T. F. Shipley and P. J. Kellman, eds. From fragments to objects: segmentation and grouping in vision. Elsevier, Oxford.

Kellogg, R. 1969. Analyzing children's art. Mayfield, Palo Alto, CA.

Manser, M. H., Z. Yuan, W. Liangbi, R. Yongchang, W. Jingrong, M. Ping, R. Xiaoping, and S. Qinan. 2003. Pocket Oxford Chinese dictionary. Oxford University Press, Oxford.

Nakanishi, A. 1980. Writing systems of the world. Tuttle, Rutland, VT.

Robinson, A. 1995. The story of writing. Thames & Hudson, London.

Ryan, M. J. 1998. Sexual selection, receiver biases, and the evolution of sex differences. Science 281:1999–2003.

Sampson, G. 1985. Writing systems. Stanford University Press, Stanford, CA.

Turner, K. J. 1974. Computer perception of curved objects using a television camera. PhD thesis. University of Edinburgh.

Waltz, D. 1975. Understanding line drawings of scenes with shadows. Pages 19–91 *in* P. H. Winston, ed. The psychology of computer vision. McGraw-Hill, New York.

Associate Editor: Michael J. Ryan
Editor: Jonathan B. Losos

This content downloaded from 071.183.025.092 on January 29, 2016 07:46:28 AM
All use subject to University of Chicago Press Terms and Conditions (http://www.journals.uchicago.edu/t-and-c).