



Visual noise consisting of X-junctions has only a minimal adverse effect on object recognition

Eshed Margalit^{1,2} · Sarah B. Herald^{1,3} · Emily X. Meschke¹ · Isabel Irawan¹ · Rafael Maarek⁴ · Irving Biederman^{1,5}

© The Psychonomic Society, Inc. 2019

Abstract

In 1968, Guzman showed that the myriad of surfaces composing a highly complex and novel assemblage of volumes can readily be assigned to their appropriate volumes in terms of the constraints offered by the vertices of coterminating edges. Of particular importance was the L-vertex, produced by the cotermination of two contours, which provides strong evidence for the termination of a 2-D surface. An X-junction, formed by the crossing of two contours without a change of direction at the crossing, played no role in the segmentation of a scene. If the potency of noise elements to affect recognition performance reflects their relevancy to the segmentation of scenes, as was suggested by Guzman, gaps in an object's contours bounded by irrelevant X-junctions would be expected to have little or no adverse effect on shape-based object recognition, whereas gaps bounded by L-junctions would be expected to have a strong deleterious effect when they disrupt the smooth continuation of contours. Guzman's roles for the various vertices and junctions have never been put to systematic test with respect to human object recognition. By adding identical noise contours to line drawings of objects that produced either L-vertices or X-junctions, these shape features could be compared with respect to their disruption of object recognition. Guzman's insights that irrelevant L-vertices should be highly disruptive and irrelevant X-vertices would have only a minimal deleterious effect were confirmed.

Keywords Perceptual organization · Shape perception · Vertices · Nonaccidental properties

An impressive feat of scene perception is that people can view an unfamiliar complex assemblage of partially occluded volumes, such as those shown in Fig. 1, and segment the mass readily and rapidly into individual volumes by assigning each surface to its appropriate volume. As is evident from this

example, this achievement of perceptual organization can be executed monocularly from a line drawing alone, without the need for adjacent volumes to differ in surface properties, such as color or texture, to facilitate their segmentation.

We will use the term *vertex* to refer to the configuration created by the *cotermination* of two or three contours.¹ The basis for this segmentation, solely using monocular shape cues, was described by Adolfo Guzman in his seminal 1968 dissertation, in which he showed that vertices formed by the cotermination of two or three contours were sufficient to account for the appropriate assignment of surfaces to volumes. (The volumes in Guzman's work were polyhedra, volumes with flat polygonal faces, straight edges and sharp corners.) The term *junction* will refer to the meeting or crossing of two or more contours, without the cotermination of any pair. Figure 2 shows the most common (and important) of Guzman's set of vertices and junctions. The dashed lines

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13414-019-01840-2>) contains supplementary material, which is available to authorized users.

✉ Irving Biederman
bieder@usc.edu

¹ Program in Neuroscience, University of Southern California, Los Angeles, CA, USA

² Neuroscience Graduate Program, Stanford University, Stanford, CA, USA

³ Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

⁴ Department of Biomedical Engineering, University of Southern California, Los Angeles, CA, USA

⁵ Department of Psychology, University of Southern California, Los Angeles, CA, USA

¹ We reserve the term *vertex* for the pattern formed from the *cotermination* of two or three contours. We use the term *junction* to refer to the pattern formed by two or more contours meeting without terminating at a common point. Typically, junctions will be formed by the crossing of contours without a change of direction at their crossing, as in an X-junction. Another type of junction is the T-junction, where one contour terminates on another contour, but not at its termination.

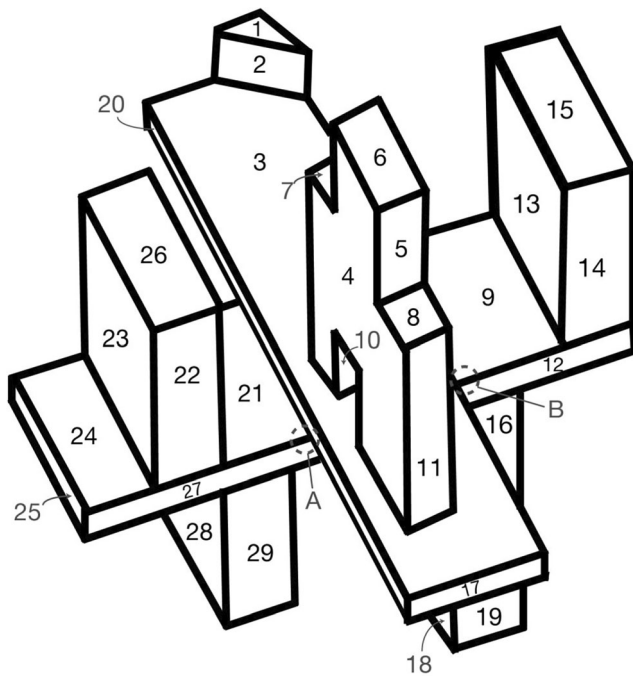


Fig. 1 Guzman's "bridge." Observers can readily assign surfaces to their appropriate volumes in novel scenes (redrawn from Guzman, 1968, with permission). The L-vertices, such as those at the top right of surface #15 and the bottom right of surface #29, are of great importance for segmenting scenes, in that they signal the termination of a surface—that is, where one surface is partially occluding the background

indicate which surfaces are to be grouped to the same volume. The pair of T-junctions in Fig. 2D do not depict cotermination. Instead, the stem of each of the two Ts terminates at the bounding contours of a surface that serves to occlude the smooth continuation of the collinear stems of the Ts. Consequently, the collinear stems of the Ts provide evidence that their contours should be grouped behind the occluding surface. In Fig. 1, matching T-junctions promote appropriate groupings of surface #9 with #21 and #12 with #27, and the "fork" (or Y) vertex promotes the grouping of surfaces #13, #14, and #15 to the same, single volume.

Common vertices and their grouping constraints

The grouping constraints from vertices were formally derived for polyhedra by Clowes (1971) and Huffman (1971), and extended by Waltz (1975) to include shadows, the classification of edges as concave, convex, obscuring (i.e., a depth discontinuity), or a crack, and the disambiguation caused by "accidental alignments." Here we describe common vertices and the grouping constraints that follow.

The L-vertex is the pattern created from the cotermination of two contours and is likely the most common of the vertices, as it is characteristic of both 2-D and 3-D shapes. It is of the greatest importance for segmenting scenes, in that it provides strong evidence for the termination of a surface—that is, where one surface is partially occluding the background, as at the upper right L at surface #15 and the lower right L of surface #29 of Fig. 1.

Another important junction for segmentation is the T-junction, where a contour terminates on another contour (rather than coterminating on the end of a contour as in the L or Y vertices). The nonterminating contour is a partially occluding surface (or surfaces) along the shaft of the terminating contour, as illustrated where the contour at the edge (an orientation discontinuity) between #9 and #12 terminates on surface #3, defining a T-vertex at that point (Fig. 1). Similarly, the contour at the edge between #21 and #27 terminates at #20, also defining a T-junction in Fig. 1 (Fig. 1). That the stems of these two T-junctions are collinear promotes the grouping of surface #9 with surface #21 and surface #12 with #27, allowing the volumes defined by surfaces #9 and #21 to be grouped together and interpreted as being partially occluded by the volume whose upper surface is #3.

Unlike L, fork, or arrow vertices, the two contours defining X- and T-junctions do not have to be at the same locus in depth. Consequently, a rotation in depth of the viewpoint can produce drastic changes in the position of the junction, or even its complete disappearance. The dependence of these junction types on viewpoint suggests that they are *not*

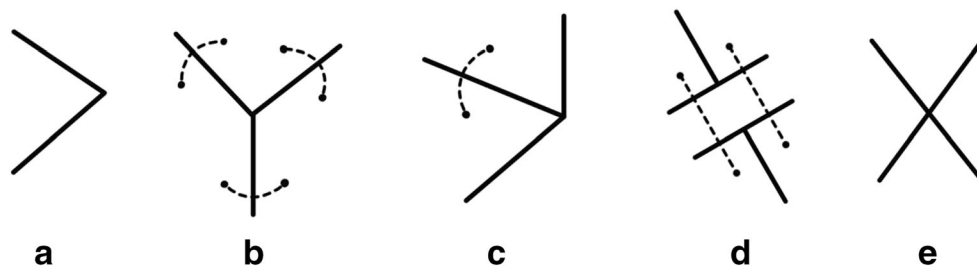


Fig. 2 A subset of Guzman's vertices. (A) An L-vertex. (B) A fork vertex. (C) An arrow vertex. (D) Matched T-junctions. (E) An X-junction. The dashed links indicate what surfaces are to be grouped to the same volume. If a surface is not to be grouped with another surface to a

common volume, there is no dashed link. The two surfaces of the L (inside vs. outside the vertex) are, therefore, not to be grouped as part of the same volume. (Redrawn from Guzman, 1968, with permission.)

nonaccidental shape properties (Biederman, 1987; Lowe, 1985), unlike the L, fork, and arrow vertices, which are defined by their cotermination at a common point in depth.

Accidental alignments Accidental alignments can create local ambiguities, as in the lower right L-vertex of surface #2 in Fig. 1 coterminating with the occluding edge of surface #3, creating an accidental Y vertex (Fig. 2) that would imply (incorrectly) that surfaces #2 and #3 and the background belong to the same volume. The accidental collinearity of the contour between #21 and #22 with the contour between #28 and #29, which creates matching T-junctions (Fig. 2) with surface #27, leads to the incorrect (and ambiguous) local grouping of surfaces #21 and #29.

In contrast to the importance of the L-vertex for the perceptual organization of shape is the X-junction, formed when two contours cross without a change of direction at the crossing point, as is shown in Condition OX in Figs. 3 and 6 below. Guzman noted that the X-junction was of no import for segmenting scenes.

Does the addition of L-vertices interfere with object recognition more than the addition of X-junctions?

We tested an implication of Guzman's work on shape segmentation that derives from his observation that X-junctions are irrelevant for segmenting scenes. Specifically, we assessed whether the addition of an irrelevant (noise) contour that produced L-vertices (which are relevant to segmentation by signaling, in this case, inappropriate terminations of a surface) would be more disruptive to object recognition than when the same irrelevant contours produced X-junctions (which are not relevant). Guzman's scheme achieved a grouping of surfaces to volumes and the segmentation (separation) of individual volumes from each other. He did not claim that such a representation was sufficient for recognition, but by some accounts (e.g., Biederman, 1987), the segmentation of an arrangement of surfaces to separate volumes constitutes a key bottom-up

stage leading to the recognition of objects, where the individual volumes constitute the simple parts of the object. Such accounts require that the *relations* between the volumes also be defined in order to yield a structural description (Lescroart & Biederman, 2012), consisting of an object's simple parts and the relations between these parts.

In the present experiment, we had observers view briefly presented masked line drawings of objects that they were to name as quickly and as accurately as possible. The contours of some of the images had midsegment gaps that, in the contour-deleted (CD) condition (Fig. 3), could readily be bridged by the Gestalt routine for "smooth continuation." Segments were added to other image variants so that they coterminated with the end points, thus bridging the gap with L-vertices, as shown in the contour-deleted with L-vertices (CDL) condition (Fig. 3). The L-vertices, in signaling the termination of a surface, would thus suppress the smooth continuation that would otherwise allow grouping across the gap. Other images had segments that crossed the contours midsegment, producing X-vertices, as in the original with X-junctions (OX) and the contour-deleted with X-junctions (CDX) stimulus conditions shown in Fig. 3.

Would the insertion of L-vertices that inappropriately signaled the termination of an object's surface, as shown in example CDL in Fig. 3, interfere with the recognition of that object, as compared to the CD or CDX condition? Would the insertion of contours that produced inappropriate X-junctions have only a minimal effect on the speed or accuracy of recognition, as illustrated in OX and CDX in comparison to O and CD, respectively, in Fig. 3, beyond what might be expected from the addition of the irrelevant noise segments themselves?

Method

Stimuli

The stimuli were 56 line drawings of common objects presented on an iMac 27-in. screen (2,560 × 1,440 pixel resolution)

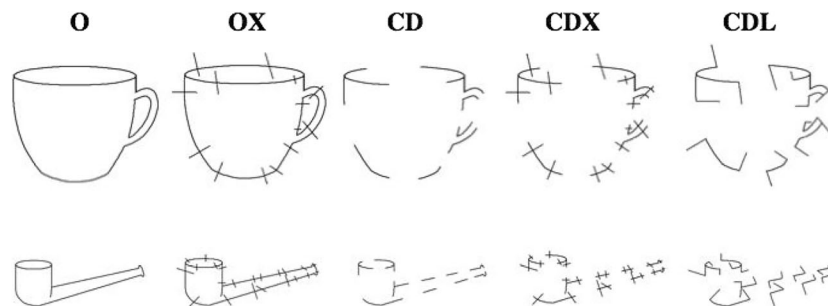


Fig. 3 An example of the stimuli in each of the five experimental conditions for two objects. O = original intact image, OX = original image with X-junctions, CD = contour-deleted image, CDX = contour-

deleted image with X-junctions near the gaps, CDL = contour-deleted image with the gaps bridged by L-vertices

using Psychtoolbox-3 (Kleiner, Brainard, & Pelli, 2007). Drawings of animals were excluded, as their eyes provided critical information about the object out of proportion to the extent of their contours. All drawings were sized to 600×600 pixels and subtended a visual angle of approximately 13.1° at arm's length. The contour-deleted (CD) condition (Fig. 3) was produced by deleting approximately 50% of the contours at their midpoint, to produce gaps that left the original vertices of the image intact. The contour-deleted L-vertex condition (CDL) was produced by adding the end points of line segments of half the extent of the deletions to each end of the contour at the gaps at approximately 90° to the endpoints to create the L shape as illustrated in Fig. 3, along with the other stimulus conditions. The two added segments on each side of the gap thus added approximately the same number of pixels as the deleted segment. If the addition of a segment for the L-vertex at 90° were to create another vertex or junction, then the segments were rotated until the added segment only produced an L-vertex. The contour-deleted X-vertex condition (CDX) was produced by moving the segments added in the CDL condition away from the gap and centering it over the original contour such that it produced an X-vertex. The OX stimuli were created by simply inserting the segments added in the CDX condition to the original images so that the added segment intersected the line from the existing drawing, creating an X-vertex. This ensured that the segments added to the CDL, CDX, and OX stimuli had identical numbers of pixels.

Subjects

Forty-eight University of Southern California students (15 males, 33 females; mean age = 19.9 years, range 18–25) participated in the experiment for credit in psychology courses or for monetary compensation. All subjects had normal or corrected-to-normal vision. The work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). All subjects gave informed consent in accordance with the procedures approved by University of Southern California's University Park Institutional Review Board.

Method and procedure

Subjects viewed line drawings of objects in a dimly lit room with a lamp facing away from the computer. Each object was presented for a maximum of 5 s, followed by a mask, but trials were advanced when a response was detected. All trials were separated by an interstimulus interval of 1,000 ms. The mask was a random appearing assemblage of straight lines. Subjects were instructed to name each object as quickly and accurately as possible as soon as the image appeared on the screen.

Reaction times (RTs) were measured by a microphone that stopped the computer clock started by the stimulus presentation. The response threshold for the microphone was recalibrated for each subject on the basis of a set of sample trials that did not include stimuli from the main experiment. Error rates were determined by the experimenter, postsession, on the basis of a series of criteria for each object—for example, the cup was considered correct if it was named as “mug” or could reasonably designate the basic-level class of the stimulus. Subjects were told to avoid prevocalizations, such as “um,” or from making any extraneous noise that would exceed the microphone threshold and thus would be falsely recorded as a response. Following microphone calibration, subjects performed six practice trials with simple geometric shapes (circle, square, etc.) to familiarize themselves with the procedure. If the subject did not satisfactorily perform the practice trials—by, for example, answering too quietly or uttering prevocalizations—the subject was instructed as to the desired response amplitude and performed the practice trials again.

Each subject viewed the 56 line drawings in a given condition only once, with the restriction that the first presentation of a given object was not the intact original (O) version of the object, which might have diminished the effects of the other conditions. The order of the images was shuffled and counterbalanced across subjects so that the average trial number, across subjects, of each object in each condition (save for the O condition) was the same: 140. Due to an error in the script, five of the 56 objects were viewed in only four of the five conditions. Thus, each subject was only presented with 276 of a possible 280 trials. This error constituted only 1.4% of possible trials, and an analysis that excluded those objects departed only negligibly from the data, so the objects were retained in the final analysis.

Trials in which subjects made extraneous noise that was incorrectly recorded as their answer for that stimulus, and that therefore did not represent either the subject's reaction time or their intended response, were not included in the analyses. The average number of usable trials was 271.4 (of a possible 276). In our analysis of reaction times, we further restricted our analysis to trials in which the subject made the correct response. The average number of trials used in this analysis was 256 of a possible 276 (92.8%).

Data analysis

The correct reaction times (RTs) and error rates were analyzed separately using generalized linear mixed-effect models (GLMMs). This type of linear regression allowed better satisfaction of normality assumptions than a typical log transformation of reaction time (Lo & Andrews, 2015), and also allowed the inclusion of subjects and stimuli as random effects (Judd, Westfall, & Kenny, 2012). The RT data were modeled using as

an inverse Gaussian distribution with the identity link function. Only correct trials were used in the RT analyses. For the accuracy analysis, the correct and incorrect responses were modeled as a binomial distribution using the logit link function. The odds ratio is reported as the effect size for the error rate results. The Holm–Bonferroni method was used to correct for multiple comparisons separately for RT and accuracy models.

To model the effect of seeing each object multiple times, albeit under different conditions, the number of times a stimulus had been shown, including the current trial, was counted and labeled the “repetition” for that trial. For example, the first time seeing a cup would have repetition value of 1, and the next viewing of the cup would have a repetition value of 2. Because one expects repetition effects to eventually saturate, repetition was modeled as $(1 - x^{-1.5})$, where x is the number of times the stimulus had been seen. Thus, the repetition values of (1, 2, 3, 4, 5) were transformed into (0, .65, .81, .88, .91). Various exponentiations were evaluated, and anywhere between $x^{-1.25}$ and x^{-2} produced nearly equal model fits, in terms of Akaike information criterion and R^2 . For both the RT and accuracy analyses, the transformed repetition value and the condition (e.g., CD) were included as fixed effects, and the subject ID and base object were included as random effects with random intercepts.

All analyses were done using R (R Core Team, 2018) and RStudio (RStudio Team, 2016). The following R packages were used: lme4 for the GLMM analysis (Bates, Mächler, Bolker, & Walker, 2015), sjPlot for viewing the GLMM output and obtaining confidence intervals (Lüdtke, 2018), multcomp for the corrected pairwise comparisons between each condition (Hothorn, Bretz, & Westfall, 2008), ggplot2 for plotting the data (Wickham, 2016), and tidyverse for managing the data (Wickham, 2017).

Results

Adding irrelevant segments forming X-junctions to the contours of an intact line drawing (OX – O) resulted in a small (28 ms), but reliable ($p = .01$) cost to RTs (Fig. 4). Deleting the contours midsegment (CD – O) resulted in a similar cost of 31 ms ($p < .001$). The cost of adding X-vertices and deleting midsegment contours (CDX – O) was 59 ms ($p < .001$). This was identical to the additive costs of the X-vertices (OX – O) and the contour deletion (CD – O) conditions alone ($28 + 31 = 59$ ms), suggesting separate linear contributions of each effect. (The strict additivity did not arise as a consequence of a calculation error, in that fractional millisecond values of the factors departed slightly from strict additivity.) Critically, when the segments identical to those forming X-vertices were shifted to form L-vertices at the gaps of the contour-deleted image (CDL – O), the increase in RT relative to

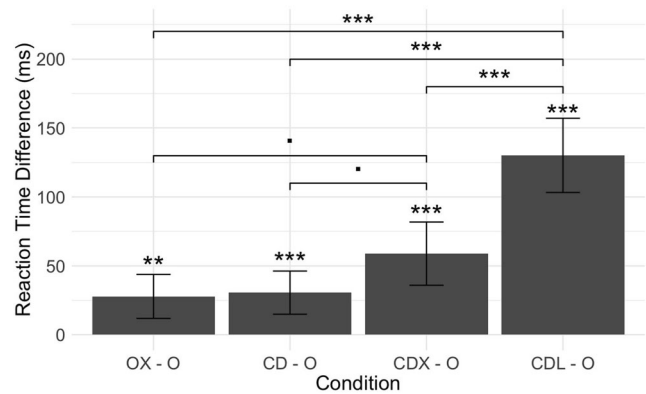


Fig. 4 Effects of the four conditions of image manipulations relative to the original, intact image (Condition O) on correct reaction times (RTs). Error bars indicate the 95% confidence intervals. •, $p < .1$; **, $p < .01$; ***, $p < .001$, Holm–Bonferroni corrected

the intact image grew to 130 ms ($p < .001$), more than double the cost of the CDX condition. This was accompanied by a doubling of the odds of making an error, relative to the intact images (log odds = .6865, $p < .001$). Only the CDL condition produced a significant change in error rates (Fig. 5).

Repetitions

Across the five presentations of the same base object, repetition effects decreased RTs by 189, 236, 256, and 265 ms, and decreased the odds of making an error by .52, .44, .41, and .40 for the second, third, fourth, and fifth viewings of the stimulus, respectively.

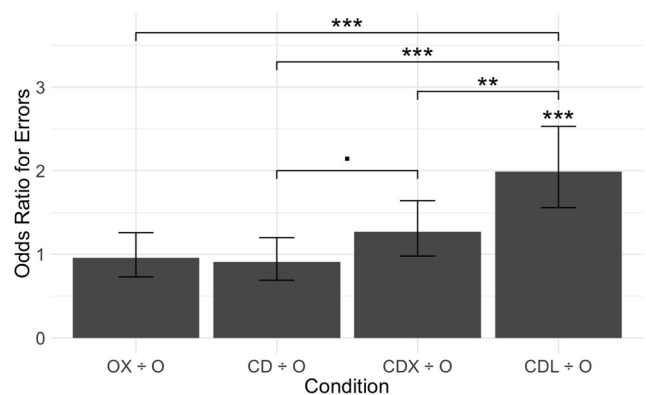


Fig. 5 Effects of the four conditions of image manipulations, relative to the original, intact image (Condition O), on the odds ratio for errors. Error bars indicate the 95% confidence intervals. In the present context, an odds ratio of 1 means that there is no difference in error rates between the two conditions—for example, the probability of an error for condition OX is equal to that of condition O. An odds ratio of 2 means that the condition—that is, condition CDL—has twice the likelihood of an error as in condition O. •, $p < .1$; **, $p < .01$; ***, $p < .001$, Holm–Bonferroni corrected

Discussion

Guzman and other early workers in computer vision demonstrated that sufficient constraints emanate from the vertices and junctions of novel, complex rectilinear objects to assign the surfaces to their appropriate volumes. This segmentation could be achieved solely on the basis of monocular shape, without any need to appeal to surface properties such as color or texture or to binocular vision or motion. The present study provides some evidence that the visual system gives high priority to vertices that make a strong contribution to segmentation, such as the L-vertex, and largely ignores those junctions, such as the X-junction, that are irrelevant to segmentation.

Adding extraneous contours to form L-vertices to suppress the smooth continuation of an extended contour of an object produced sizeable costs to that object's recognition speed and accuracy, whereas adding the same segments so that they formed X-vertices with the object's contours incurred only a minimal cost. Even this modest deleterious effect in the CDX condition may have been exaggerated in the present stimuli, as the added segments to produce X-vertices not infrequently crossed into a nearby contour in dense regions of the object, suggested a near-accidental L-vertex or occluding relevant contour, as can be witnessed with the pipe in the CDX condition of Fig. 3.

The lack of potency of X-vertices to serve as visual noise has been confirmed by author I.B. in another task, that of detecting targets in RSVP sequences. Subjects (members of his class) viewed RSVP sequences of line drawings of common objects with each image presented for 84 ms. Prior to each sequence, a target was specified by name. On some trials, the preceding and following frames consisted of a dense array of a spaghetti of contours that, at the frame rates, appeared to be superimposed over the object images, forming X-junctions somewhat similar to those illustrated in Fig. 6. There was no loss of identifiability of these images as compared to sequences without any noise.



Fig. 6 An illustration that the addition of a large array of uninterpretable contours forming only X-vertices has only a minimal effect on an object's identification

X- and T-junctions

Vessel, Biederman, Subramaniam, and Greene (2016), reported an experiment with a similar task and design to the present experiment. Their subjects named contour-deleted objects in which the gaps in the contours could be bounded by T-junctions or L-vertices. As would be expected from Guzman's analyses, naming RTs and error rates were higher for the gaps bounded by L-vertices—which suppressed the smooth continuation of the contours—than for the T-junctions, where the gaps were now interpreted as occluding surfaces behind which the contours could be grouped. T- and X-vertices thus are similar with respect to their lack of potency, relative to L-vertices, for suppressing smooth continuation of contours.

What determines the potency of the L-vertex for signaling the termination of a surface?

The line drawings of objects in the Vessel et al. (2016) experiment were depicted as either white or black contours on a gray background. Those authors demonstrated a critical condition for the L-vertex to retain its potency to signal the termination of a surface: the two legs of the vertex defining the L must have the same direction of contrast (both darker or both lighter than the background). If one leg of the L was darker and the other leg was lighter than the background, then RTs and error rates for images with gaps bounded by Ls were equivalent to when the gaps were bounded by Ts. Both segments of the L-vertices in the present stimuli had the same direction of contrast in that both legs were always darker—that is, black—than the white background. It would appear that the visual system has incorporated that statistical regularity as a necessary condition for the L-vertex to signal the termination of a surface. The consistency of directions of contrast in L-vertices is, statistically, a strong characteristic of L-vertices, in general, which can be readily confirmed by looking at such vertices in the reader's environment.

Why is partial occlusion typically not disruptive to perceptual recognition?

A more pervasive phenomenon than the recognition of contour-deleted objects is likely relevant to the minimal interference of Xs and Ts relative to Ls, in the present experiment and that of Vessel et al. (2016). We often view objects and scenes through partially occluding surfaces, such as light foliage or lace drapery. These occluding surfaces rarely result in any noticeable decrement in the perception of the scene. An explanation from the present perspective (derived from Guzman and Lowe) is that when such lightly occluding surfaces are randomly projected onto a scene, it would be rare or, more specifically, an *accident*, for contours of the occluding

surface to *coterminate* with the contours of the scene to produce an L-vertex. That is, the cotermination of contours (that define an L) is a nonaccidental property (Lowe, 1985).

Accidental vertices

As we noted previously, vertices produced by cotermination, such as Ls, forks, and arrows, are instances of nonaccidental properties. Accidental alignments can mimic these vertices and create local ambiguities, as in the accidental collinearity of the contour in Fig. 1 between surfaces #21 and #22 with the contour between #28 and #29 creating matching T-junctions with surface #27, leading to the incorrect local grouping of surfaces #29 with #21 and #28 with #22. The ambiguous groupings produced by accidental alignments may itself be taken as evidence for the ubiquitous role played by these junctions and vertices in the perceptual organization of shape.

Neural correlates of shape-based object representations

What might be the locus of Guzman's bottom-up account of shape segmentation, as viewed from the perspective of current research in the neuroscience of shape-based object recognition? Guzman's account made no appeal to prior familiarity with the object. Recent studies of the neural correlates of object recognition provide strong support for this aspect of Guzman's scheme. The lateral occipital complex (LOC) is a cortical region, consisting of the lateral occipital cortex and the posterior fusiform gyrus, that has been shown to be critical for object recognition, in that its bilateral lesioning renders an individual unable to recognize shapes of any kind—objects, faces, or print—while sparing the perception of color, texture, and motion (James, Culham, Humphrey, Milner, & Goodale, 2003). LOC can be localized with fMRI as the cortical region that shows a greater BOLD response to intact objects than to their scrambled versions, resembling texture (Malach et al., 1995). Margalit et al. (2016; Margalit, Biederman, Tjan, & Shah, 2017) compared the magnitude of the LOC BOLD response to familiar objects, depicted as an arrangement of geons, to when the same set of geons were rearranged so that they appeared as a novel assemblage. The magnitude of the BOLD response in LOC was completely unaffected by whether or not the object was familiar. Thus, the exclusion of prior familiarity in Guzman's predictions is consistent with the activity of object-selective regions in the brain.

The work in computer vision in the 1970s and 1980s that proved to be so inspiring to both computational and biological vision had, as its goal, a model of explicit shape-based object recognition grounded in projective geometry. Despite the great insights gained from this work, the study of geometric-based explicit shape representation appears to have fallen by the wayside, in favor of learning by deep networks which

currently give greater weight to surface properties (Geirhos, 2019). It would be of interest to see whether such networks, without being explicitly trained to employ the vertices for segmentation, nonetheless reflect the constraints offered by the vertices. In the present experiment we focused on only one class of these insights, but the clear results and the interpretations offered by the seminal work in computer vision suggest that this work still has much to offer to those studying the biology and psychophysics of visual shape recognition.

Open Practice Statement The data have been deposited in the Open Science Framework, osf.io.

Funding This research was funded by the Dornsife Research Fund.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48. doi:<https://doi.org/10.18637/jss.v067.i01>
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94, 115–147. doi:<https://doi.org/10.1037/0033-295X.94.2.115>
- Clowes, M. (1971). On seeing things. *Artificial Intelligence* 2, 79–116.
- Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2019). *ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness*. arXiv preprint. <https://arxiv.org/abs/1811.12231>
- Guzman, A. (1968). Computer recognition of three-dimensional objects in a visual scene (Ph.D. dissertation). Dept. of Electrical Engineering, MIT.
- Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, 50, 346–363.
- Huffman, D. (1971). Impossible objects as nonsense sentences. In B. Meltzer & D. Michie (Eds.), *Machine intelligence 6*. Edinburgh, Scotland: Edinburgh University Press.
- James, T. W., Culham, J., Humphrey, K. G., Milner, D. A., & Goodale, M. A. (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: An fMRI study. *Brain*, 126, 2463–2475. doi:<https://doi.org/10.1093/brain/awg248>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103, 54–69. doi:<https://doi.org/10.1037/a0028347>
- Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, 36(ECVP Abstract Suppl.), 14
- Lescroart, M. D., & Biederman, I. (2012). Cortical representation of medial axis structure. *Cerebral Cortex*, 23, 629–637. doi:<https://doi.org/10.1093/cercor/bhs046>
- Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, 6, 1171. doi:<https://doi.org/10.3389/fpsyg.2015.01171>
- Lowe, D. G. (1985). *Perceptual organization and visual recognition*. Boston, MA: Kluwer.

- Lüdtke, D. (2018). sjPlot: Data visualization for statistics in social science. Retrieved from <https://CRAN.R-project.org/package=sjPlot>
- Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., & Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92, 8135–8139. doi:<https://doi.org/10.1073/pnas.92.18.8135>
- Margalit, E., Biederman, I., Tjan, B. S., & Shah, M. P. (2017). What is actually affected by the scrambling of objects when localizing the lateral occipital complex? *Journal of Cognitive Neuroscience*, 29, 1595–1604. doi:https://doi.org/10.3389/10.1162/jocn_a_01144
- Margalit, E., Shah, M. P., Tjan, B. S., Biederman, I., Keller, B., & Brenner, R. (2016). The Lateral Occipital Complex shows no net response to object familiarity. *Journal of Vision*, 16, 2298–2306. doi:<https://doi.org/10.1167/16.11.3>
- R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- R Studio Team. (2016). RStudio: Integrated development for R. Boston, MA: RStudio, Inc. Retrieved from <http://www.rstudio.com/>
- Vessel, E., Biederman, I., Subramaniam, S., & Greene, M. R. (2016). Effective signaling of surface boundaries by L-vertices reflect the consistency of their contrast in natural images. *Journal of Vision*, 16(9), 15:1–10. doi:<https://doi.org/10.1167/16.9.15>
- Waltz, D. (1975). Understanding line drawings of scenes with shadows. In P. Winston (Ed.), *The psychology of computer vision* (pp. 19–91). New York, NY: McGraw-Hill. <http://dspace.mit.edu/handle/1721.1/41058>
- Wickham, H. (2016). ggplot2: Elegant graphics for data analysis. New York, NY: Springer.
- Wickham, H. (2017). tidyverse: Easily install and load the “Tidyverse” (R package version 1.2.1). Retrieved from <https://www.tidyverse.org/>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.