The role of global cues in the perceptual grouping of natural shapes

James H. Elder

Centre for Vision Research, York University, Toronto, Canada

Timothy D. Oleskiw

Centre for Neural Science, New York University, New York, NY, USA

Ingo Fruend

Centre for Vision Research, York University, Toronto, Canada

Introduction

Real-world scenes can be highly cluttered, often causing an object to fragment into many pieces before it hits our retinae. Perceptual grouping is the process of reassociating these fragments to generate the percept of a whole object. Here we consider in particular the problem of grouping the fragments of the bounding contour based on their geometric relationships.

While the shape of these contour fragments can be arbitrary, approximating them as piecewise linear allows the problem to be modeled as the grouping of local orientation signals, coded by neurons in primary visual cortex (Hubel & Wiesel, 1968; Heitger, Rosenthaler, von der Heydt, Peterhans, & Kulber, 1992), to form a sequence or chain that describes the entire contour (Figure 1). Note the combinatorial nature of the problem: Given \( n \) oriented elements in the image, there are \( n!/(2(n-k)!) \) different sequences of length \( k \) to consider. This is far too many to evaluate for most contours in complex natural scenes—the brain must employ some prior knowledge or principles to limit the search.

The study of these perceptual organization principles has a long history stretching back to the early days of Gestalt psychology (Koffka, 1935; Wertheimer, 1938). Two key Gestalt principles relevant to contour grouping are proximity and good continuation. We will refer to these grouping cues as local cues because they can be defined for pairs of neighboring contour elements. In particular, proximity can be encoded in terms of the distance \( d_{ij} \) between elements \( i \) and \( j \), and good continuation can be coded in terms of the angles \( \alpha_{ij}, \beta_{ij} \) induced by a linear interpolation between the elements (Figure 2a).
What are the brain mechanisms underlying these Gestalt principles? Physiological experiments first in cat (Gilbert & Wiesel, 1989) and later in nonhuman primate (Li & Gilbert, 2002; Stettler, Das, Bennett, & Gilbert, 2002; Li, Piech, & Gilbert, 2006) have revealed a striking bias in the long-range horizontal connections between orientation columns in primary visual cortex that serves to connect neurons coding local orientation signals that can be smoothly interpolated, consistent with the classical Gestalt principles of proximity and good continuation. Psychophysical experiments suggest that these connections serve to facilitate geometrically consistent orientation responses in the human striate cortex (Cass & Spehar, 2005).

These physiological results have motivated more detailed computational models of the neural circuits in primary visual cortex responsible for contour grouping (e.g., Li, 1998; Yen & Finkel, 1998). In parallel, computational research has shown that these “cocircularity” cues can be helpful in guiding the extraction of contours from natural imagery (Parent & Zucker, 1989), and psychophysical studies revealed that the human visual system is highly sensitive to these local geometric relationships (Field, Hayes, & Hess, 1993). Finally, statistical models of these relationships rooted in empirical measurements of natural imagery have yielded accurate quantitative normative models for these local Gestalt cues that validate their inferential power (Figure 2b; Kruger, 1998; Geisler, Perry, Super, B. J., & Gallogly, 2001; Elder & Goldberg, 2002; Sigman, Cecchi, Gilbert, & Magnasco, 2007; Geisler & Perry, 2009). Together, these converging results have established a local “Association Field” model rooted in the circuitry of early visual cortex as the standard perceptual model for contour grouping.

Although these local Gestalt cues directly relate only pairs of contour elements, given certain assumptions they can be used to infer global contours spanning many elements (Elder & Zucker, 1996; Elder, Krupnik, & Johnston, 2003; Elder, 2013, 2015). To see this inference, it is helpful to introduce some notation. Let \( y_{ij} = \{d_{ij}, \alpha_{ij}, \beta_{ij}\} \) represent the local Gestalt cues relating elements \( i \) and \( j \). Now consider the hypothesis \( H \) that a particular sequence of elements \( i_1, \ldots, i_n \) forms a global contour, and let \( \neg H \) represent the negation of this hypothesis. The likelihood ratio \( L \) for the proposed grouping of elements can then be written as

\[
L = \frac{p(y_{i_1 i_2}, \ldots, y_{i_n i_1}|H)}{p(y_{i_1 i_2}, \ldots, y_{i_n i_1}|\neg H)}
\]

If we assume conditional independence between the cues relating pairs of elements, the likelihoods can be factored, so that

---

Figure 1. The problem of contour grouping. Neurons in primary visual cortex (V1) code the local orientation of contours in the image, but without explicit representation of how they are linked together. Our percept, on the other hand, is of coherent and complete objects, segmented from their background.

Figure 2. (a) Local geometric cues to contour grouping. Smoothness properties of objects give rise to the classical Gestalt cues of proximity and good continuation (Wertheimer, 1938). For a pair of local oriented elements these can be expressed through a variable \( d \) encoding the separation of the elements and two angles \( \{\alpha, \beta\} \) induced by a linear interpolation between the elements. (b) Posterior probability, derived from ecological statistics, that a local oriented element (shown in black) should be grouped directly with a second observed element, as a function of its proximity, direction, and relative orientation (Elder & Goldberg 2002). (c) Local Gestalt cues, together with a Markov independence assumption, lead to an efficient framework for extracting extended contours.
and Hess (1993), in which a sequence of oriented elements, the location and orientation of the next element depends only upon the location and orientation of the element immediately preceding it in the chain. This Markov property gives rise to an optimal substructure property that allows maximum probability contours to be identified in polynomial time despite the exponential size of the search space (Figure 2c).

Although this Markov assumption has been used by computer vision algorithms to extract global contours from complex natural images (Elder & Zucker, 1996), there are a number of reasons to be skeptical that it represents a complete model of human perceptual grouping. First, it has been noted that these algorithms can fail in some cases where the human visual system seems to have little trouble (Movahedi & Elder, 2013). Second, the distribution of intrinsic distances between high-curvature events on natural contours has been shown to be incompatible with the Markov assumption (Ren, Fowlkes, & Malik, 2008). Third, the Markov assumption does not accommodate important global topological constraints such as closure and avoidance of self-intersections (Elder & Zucker, 1996; Elder et al., 2003). Finally, we note that Equation 2 implies that the whole is exactly equal to the sum of its parts: not a particularly Gestalt assertion!

For all of these reasons, it seems likely that the standard model is incomplete, and that the human visual system is somehow capable of combining these important local cues with additional global cues for the perceptual grouping of contours. The goal of this paper is to psychophysically test this hypothesis.

Has it not already been tested? One would think so. Particularly in the last 25 years there have been a number of excellent quantitative psychophysical studies on the perceptual determinants of contour grouping. Most of these employ a method due to Field, Hayes and Hess (1993), in which a sequence of oriented elements following a curvilinear path is embedded in uniformly distributed and randomly oriented distractor elements. The observer’s task is to detect the path, and properties such as the spacing, orientation jitter, and photometric appearance of the elements can be varied to identify their influence on the underlying grouping mechanisms (e.g., Hess & Dakin, 1999; Geisler et al., 2001; Ledgeway, Hess, & Geisler, 2005; Watt, Ledgeway, & Dakin, 2008). However, in these studies the target contours are typically constructed to afford only local cues to grouping—the placement and orientation of each element on the path depends only upon the previous element. It is thus impossible to infer anything about global factors in the perceptual grouping of contours from these experiments.

In order to test whether global cues play a role in contour grouping, we need to incorporate global cues into the target contour. But exactly which cues should be incorporated? There are many possibilities, including closure, convexity, parallelism and symmetry (Zhu, 1999; Elder, 2013, 2015, 2018). One potential approach is to take each global cue in turn and test for its role in isolation. Although this reductionist approach has advantages, these stimuli would not be typical of the complex natural objects we see around us, which afford a number of these cues at the same time. We therefore elect to use natural shape stimuli, in particular animal shapes. We know from prior work that animals in natural scenes are detected efficiently (Thorpe, Fize, & Marlot, 1996), and that rapid animal detection is driven in large part by shape cues (Elder & Velisavljević, 2009). These animal shapes exhibit a diversity of global regularities to which the human visual system may have adapted to support grouping and recognition in highly cluttered (e.g., forest) environments. The key to isolating the effect of these global regularities will be the careful design of control stimuli that afford the same local cues as our animal stimuli but none of the global cues.

### Methods

#### Stimuli

The stimuli were based upon 264 blue-screened images of animal objects taken from the Hemera Photo-Objects dataset. We tried to select shapes that were recognizable as animals, as opposed to more generic shapes (e.g., snakes). The original images had maximum dimension (either height or width) of 1,024 pixels; the other dimension varied depending upon the shape of the animal. The alpha channel for each image signaled figure/ground for the blue-screened animal object. Contours were extracted by tracing points with nonzero gradient magnitude in the alpha channel using Moore neighborhood tracing (Pradhan, Kumar, Agarwal, Pradhan, & Ghose, 2010), and each contour was then represented as a high resolution polygon with this sequence of points as vertices. The number of vertices N ranged from 1,506–4,668 over shapes. See Figure 3 for examples.

From these high-resolution polygons, lower resolution polygons of length 2n were derived by selecting a random point on the original polygon as a first sample and then selecting 2n−1 additional samples at roughly equal arc-length intervals from this point. Alternate segments of these polygons were then labelled as...
“visible” and “gap” (Figure 4); rendering just the visible segments yields a roughly uniformly fragmented animal target shape of length \( n \) visible segments (Figure 5, top). In what follows we refer to the number of visible segments \( n \) as the target length; this was varied in order to assess how the influence of nonlocal factors in perceptual grouping and target detection depends upon level of detail.

Each of the gap segments in a polygon forms two angles with respect to the adjacent visible segments. Let the gap length and two angles identified with gap segment \( i \) be denoted \( d_i \), \( \alpha_i \), and \( \beta_i \), respectively (Figure 4). Note that the visible and gap segments varied slightly in length.

Our experiments assessed human ability to detect these animal shape targets in noise, compared with various forms of local metamer targets that have the same first-order statistical geometric properties between neighboring elements but lack the global regularities of natural shapes (Figure 5). We detail the method for constructing these metamer stimuli within the appropriate experiment-specific methodology sections below.

In these experiments, the shape targets subtended roughly 2.5° visual arc and were randomly oriented and positioned within a 20° diameter circular stimulus window. The task was made challenging by embedding the targets within a field of randomly positioned and oriented distractor segments. The lengths of the distractor segments were drawn randomly and uniformly from the lengths of the visible target segments when the target was present, and from a randomly selected target when the target was absent. Antialiased segments were drawn white on a midgray background. Figure 6 shows a sample target-present stimulus display.

**Procedure**

A yes/no design was employed. Each trial sequence consisted of a fixation display (0.5 s), stimulus display (160 ms in Experiments 1 and 2, 200 ms in Experiment 3), and then the fixation display again until response (Figure 6). A target shape appeared within the stimulus window with 0.5 probability. The observer pressed the left arrow key on a keyboard to indicate target absent, and the right arrow key to indicate target present. Feedback was provided in the form of a tone for incorrect responses. The number of distractor elements at criterion (75% correct) performance was estimated using an adaptive psychometric procedure (QUEST; Watson & Pelli, 1983). Each experiment was broken into three 100-trial blocks per condition; blocks were counterbalanced across conditions. QUEST was re-initialized for each block, yielding three independent threshold estimates for each condition.

### Experiment 1. Detection of animal and open metamer targets

**Methods**

**Stimuli**

For our first experiment, we used a simple procedure to construct metamer stimuli that have exactly the same first-order (local) geometric properties as our animal shapes, but none of the global regularities. The process is illustrated in Figure 7a. To construct a metamer, a fragmented animal shape was first selected, and lists of the lengths \( l_i \) of the visible segments and of the gap geometry \( (d_i, \alpha_i, \beta_i) \) were extracted, starting from a randomly selected visible segment on the animal shape.
These lists were then randomly and independently permuted. Finally, the fragmented metamer shape was constructed by sequentially rendering the sequence of visible segment lengths and gap geometries from the permuted lists. Since the individual gap geometries were copied faithfully, this transcription preserves the local pairwise grouping relationships between visible elements. However, since the order of these gap geometries is randomized, higher order nonlocal structure has been shuffled away. Figure 7b shows the result for an example animal shape.

Since each animal polygon is simple and closed, the sum of its turning angles \( \sum (x_i + \beta_i) \) is \( \pm 2\pi \). This property is commutative and so preserved in the metamer despite the shuffling. This regularity can be considered a “convexity bias” as it favors one sign of curvature over the other. As this convexity bias is a nonlocal property that could feasibly be a cue to perceptual grouping, we generated a second form of metamer without this bias by randomizing the sign of all of the turning angles. We refer to these two forms of metamer stimuli as \( M_1 \) and \( M_2 \), respectively.

Line segments were drawn in white on a midgray disk 20° in diameter, within a black rectangular screen.

We emphasize that although the metamer stimuli created for this experiment are specific to animal...
shapes, the same technique could be applied to any class of shape.

**Observers**

There were 10 observers in our first experiment. Each performed 3 blocks of 100 trials each for 3 target lengths (5, 10, 20) and 3 target conditions (animal, \(M_1\), \(M_2\)), for a total of 2,700 trials per observer. One of the observers was an author; the others were naïve to the goals of the experiment. All observers had normal or corrected-to-normal vision.

**Results**

Figure 8 shows the results. Mean noise thresholds over observers for the three conditions are shown on the left, and pairwise contrasts between conditions are shown on the right. We see that results for the two metamer conditions are nearly identical: The convexity bias induced by retaining the signs of the turning angles does not seem to impact target detection. However, there is a clear difference between these metamer conditions and the animal shape condition: Thresholds are significantly higher for the animal condition than for the metamer conditions at higher target lengths (Table 1). Since the information present in the pairwise relationships between successive elements is identical for animal and metamer stimuli, this result indicates that observers are making use of information in the animal targets beyond these local cues.

What is the nature of the nonlocal differences between animal and metamer stimuli responsible for the observed differences in detectability? There are many possibilities; we first seek to determine whether simple factors such as target size, elongation, eccentricity, self-intersections, or number of inflections could explain our results.

**Target size**

Whereas the animal and metamer stimuli consist of the same number of oriented elements, the metamer
stimuli tend to be less compact, and thus subtend a larger portion of the visual field. One way to quantify the size of the target stimulus is by the area of its convex hull. Figure 9a shows the results of this analysis. Observe that the metamer stimuli are considerably larger by this measure. To understand whether this difference could account for the observed differences in noise thresholds, we analyzed a subset of our data for animal and metamer (M1) stimuli at a target length of 20 elements, equalized according to the area of their convex hulls. Specifically, we sorted animal and M1 stimuli separately according to the area of their convex hull. We then sequentially selected animal stimuli in decreasing order of area and metamers in increasing order of area until the mean area of the two subsets matched (Figure 9b); this procedure resulted in area-matched subsets containing roughly 87% of the original dataset. Finally, we used QUEST to re-estimate noise thresholds from the original human data using only these area-matched stimulus subsets. The results for the animal and M1 targets at a target length of 20 elements are shown in Figure 9c. We find that results are not substantially affected by this equalization; comparing results for all target stimuli shown in Figure 8, thresholds are still significantly higher for animal targets relative to metamer targets: two-tailed matched sample t(9) = 4.4, p = 0.002. Thus it seems that the difference in the average area of the animal and metamer stimuli does not account for the observed difference in their detectability.

**Target elongation**

Since the metamers are not constrained to be closed, they tend to be more elongated than the animal stimuli. We measured the elongation of the stimuli by the length of the maximal inscribing chord of the convex hull. Figure 10a verifies that by this measure the metamer stimuli are more elongated than the animal stimuli. Applying the same equalizing technique we used for area now to elongation for 20-element stimuli yielded a matched subset comprising 63% of the original data. Again we find that results are essentially unchanged by this equalization (Figure 10b); thresholds are still significantly higher for animal targets relative to metamer targets: two-tailed matched sample t(9) = 4.0, p = 0.003. Thus it seems that the difference in the average elongation of the animal and metamer stimuli also does not account for the observed difference in their detectability.

**Target eccentricity**

We also assessed whether there were systematic differences in the eccentricity of the stimuli. Figure 11a shows the mean eccentricity of the target shape centroids as a function of target length. Note that on average, the animal stimuli tend to be more eccentric.
than the metamer stimuli. This is because targets were positioned randomly to lie entirely within the central 20° of the display. Since the animal stimuli tended to be more compact, they tended to assume more eccentric locations. Since more eccentric presentation would be expected to reduce performance, these differences are not likely to account for the superior detectability of animal targets.

**Self-intersections**

Although the animal targets are simple (nonintersecting) closed contours, our method for generating metamers does not prevent them from turning back on themselves, potentially generating multiple self-intersections (Figure 5). Is it possible that these self-intersections somehow obscure the target, lowering detectability? To assess this, we analyzed the subset of 20-element metamer stimuli that do not self-intersect. (This consists of 20.6% of $M_1$ stimuli and 30.5% of $M_2$ stimuli.) Results are substantially unchanged, and animal targets are still significantly more detectable than the metamers: Figure 11b, two-tailed matched sample $t(9) = 4.2, p = 0.002$ for Animal versus $M_1$ and $t(9) = 6.7, p = 0.00008$ for Animal versus $M_2$. Thus the absence of self-intersections does not appear to be responsible for the superior detectability of the animal targets.

![Figure 10. Experiment 1 analysis: target elongation. (a) Mean elongation of the animal and first-order metamer $M_1$ (with convexity bias) and $M_2$ (without convexity bias) stimuli. (b) Mean noise thresholds for elongation-equalized animal and $M_1$ metamer targets (with convexity bias) for target-length = 20 condition. Error bars indicate standard error of the mean.](image)

![Figure 11. Experiment 1 analysis: eccentricity and self-intersections. (a) Mean eccentricity of the centroid of the target animal and first-order metamer $M_1$ (with convexity bias) and $M_2$ (without convexity bias) stimuli. Error bars indicate standard error of the mean. (b) Mean noise thresholds for animal targets and the subset of open metamer targets that do not self-intersect. $M_1$: metamers with convexity bias; $M_2$: metamers without convexity bias. Error bars indicate standard error of the mean.](image)
Pettet, McKee, and Grzywacz (1998; see also Pettet, 1999) found that introducing larger turning angles or inflections in a contour lowers detectability. This raises the question: Could the enhanced detectability of closed animal shapes relative to open metamers be due to some difference in the average magnitude or sign of the turning angles? By construction, the magnitude of the turning angles for animals and metamers are identical, so certainly a difference in the average magnitude of the turning angles cannot explain our results. What about sign? While the signs of the turning angles for $M_1$ metamers match those for the animal contours, the signs of the turning angles were selected randomly for $M_2$ metamers, yet no effect on performance was observed. Thus it seems that differences in the signs of the turning angles also cannot explain our results.

It is important to note, however, that the number of inflections in the contour depends not only on the signs of the turning angles but also upon their sequencing. For the animal contours the signs of the turning angles at adjacent elements are correlated, and this could be expected to produce fewer inflections on average than for metamers, for which turning angles at adjacent elements are statistically independent.

Figure 12a shows the mean probability of an inflection occurring at any given element in our animal and metamer stimuli. We see that indeed for longer target lengths, the metamers contain slightly more inflections than the animal stimuli, as predicted. Could this explain why animals are more easily detected?

To answer this question, we analyzed the number of inflections in target contours that were successfully detected (hits) versus the number in target contours that were not detected (misses). Figure 12b shows the results. We see that there is no significant difference between the number of inflections in hit trials versus miss trials, for any of the stimulus conditions ($p > 0.05$ for matched-sample $t$ tests over our 10 observers for all three stimulus conditions at all three target lengths). We thus conclude that a difference in the number of inflections does not explain the observed difference in detectability of animal and metamer stimuli.

**Summary**

These control analyses appear to rule out explanations for the superior detectability of animal shapes over metamers based upon size, length, eccentricity, self-intersection, and number of inflections. However, one important difference between animal and metamer stimuli that we have not analyzed is closure. Closure has been suggested as potentially important both as a cue for contour grouping (Kovacs & Julesz, 1993) and as a bridge from one-dimensional to two-dimensional shape (Koffka, 1935; Elder & Zucker, 1993; Elder & Zucker, 1994; Garrigan, 2012), and thus could plausibly contribute to the detectability of our animal stimuli. Our next two experiments will explicitly assess the potential role of closure in the perceptual organization of fragmented shapes.

**Experiment 2. Detection as a function of closure**

**Methods**

Methods were identical to those of Experiment 1, except as noted below.

**Stimuli**

The goal of this experiment is to compare detection for partial fragmented animal shapes as a function of their closure, compared with metamer shapes matched...
Because the results of Experiment 1 suggest that the convexity bias that distinguishes our two types of metamer stimuli has very little impact on detectability, we elected to employ only the $M_1$ metamer stimulus for this comparison.

We selected a target length of $n = 40$ for the complete animal shapes, and measured detection performance for 50% closure ($n = 20$), 75% closure ($n = 30$), and 100% closure ($n = 40$), and for length-matched $M_1$ metamers. (For animal shapes with 100% closure [$n = 40$] we compare with metamer stimuli of length $n = 41$ to equate the number of pairwise connections between neighboring elements.)

Line segments were drawn in white on a midgray disk 20° in diameter, within a black rectangular screen.

Observers

Ten observers participated in Experiment 2. Five of these observers had already participated in Experiment 1 (including one author); the other five were new. All had normal or corrected-to-normal vision, and all but the author were naïve with respect to the goals of the experiment.

Results

Figure 13 shows the results. Noise thresholds are consistently higher for the partial animal targets relative to the partial metamer shapes, and this detectability advantage becomes statistically significant for higher target lengths (Table 2). These results suggest that the detectability advantage for animal targets is not dependent upon perfect closure: Seventy-five percent closure is sufficient to show a statistically significant advantage. This is consistent with prior work suggesting that perceptual closure, unlike mathematical closure, is a continuous property of contours, rather than a binary, all-or-nothing feature (Elder & Zucker, 1993, 1994). This experiment therefore leaves us with uncertainty about why the animal targets are easier to detect: Is it perceptual closure, or some other nonlocal property of these animal shapes? Our final experiment will address this question.

Experiment 3. Detection of animal, open metamer and closed metamer targets

Methods

Methods were identical to those of Experiment 1, except as noted below.

Stimuli

Construction of open metamers that match the local pairwise geometry of animal shapes was a simple matter, requiring only a shuffle of pairwise geometric properties. Creating simple (nonintersecting) and closed versions of these metamers is more complicated, as the constraints of nonintersection and closure introduce nonlocal dependencies between the placement of the oriented elements comprising the stimulus.

To solve this problem, we employed a very different approach. The process is somewhat involved: We summarize it here and provide additional detail in the appendix. We first used a novel method to construct from our high-resolution animal polygons, low-resolution 120-segment equilateral approximations. (The equilateral constraint simplifies the process, as the polygons can now be described entirely by the sequence...
of turning angles, without reference to the segment lengths.)

From the resulting 120-segment equilateral polygonal approximations of our animal shapes, we learned a nonparametric kernel density model of the turning angles. We then developed a method for creating novel equilateral polygons whose turning angles match this density model but which are otherwise maximum entropy. The method is seeded with a regular polygon (all angles equal). Proposals for new nonintersecting polygons are then formed using an isoperimetric constraint on inscribed quadrilaterals (Goldin, Delosme, & Bruckstein, 2010). Metropolis-Hastings sampling (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Hastings, 1970) is then used to fairly sample from these proposals, using the kernel density model for turning angles of animal shapes as the target density. The probability density of turning angles for the resulting closed metamer samples matches the density for the original animal shapes (Figure 14). We will use the label $M_3$ to refer to metamers generated from these shapes, to distinguish them from the open metamers used in Experiments 1 and 2.

We emphasize that although the closed metamer stimuli created for this experiment are specific to animal shapes, the same technique could be applied to any class of shape, given a sufficient number of exemplars to learn the statistical model for local turning angles.

Line segments were drawn in white on a black rectangular screen.

**Observers**

There were 13 observers. One of the observers was an author; the others were naïve to the goals of the experiment. All observers had normal or corrected-to-normal vision.

**Results**

Figure 15 shows the results of this experiment. We see that as the target length is increased to 20 elements and beyond, a dissociation between the three conditions emerges: Animal target are significantly more easily detected than either form of metamer, but at the same time the closed metamers are significantly more easily detected than open metamers (Table 3). These results indicate that closure is certainly one nonlocal cue that aids detection, but it is not the only nonlocal cue: There must be additional nonlocal regularities of the animal targets that play a role in the perceptual organization of shape.¹

As for open metamers, closed metamers have slightly more inflections than animal shapes, but the number of inflections did not seem to influence thresholds. This

![Figure 14. Experiment 3. The probability density of turning angles for our closed metamer shapes is a close match to the turning angle density for the original animal shapes.](image1)

![Figure 15. Experiment 3 results. A: Animal targets; $M_1$: first-order metamers (open, with convexity bias); $M_3$: first-order metamers (closed). Left: mean noise thresholds. Right: pairwise contrasts. Error bars indicate standard error of the mean.](image2)
was verified by a four-way ANOVA analysis of the number of inflections in the target for target-present trials with observer, target shape, target length, and response (hit or miss) as factors (Table 4). Whereas both target shape and target length are significant predictors of the number of inflections, observer and response are not.

### Discussion

With three experiments we have tested the hypothesis that nonlocal cues play a role in the perceptual grouping of contours. Rather than narrowing focus on a specific cue, we employed natural animal shapes that afford many nonlocal cues, on the assumption that the brain is adapted to take advantage of cues available in ecologically important stimuli. Critical to our method is the design of control stimuli that match the local cues afforded by our animal stimuli, but none of the global cues.

In our first experiment we used a simple shuffling trick to achieve this, which produced first-order metamer contours with exactly the same local geometry as animal contours. The results (Figure 8) clearly indicate that the animal shapes are easier to detect in oriented noise than the first-order controls. A series of control experiments ruled out trivial explanations based on differences in size, elongation, eccentricity, or self-intersections, pointing to the importance of more meaningful nonlocal regularities. A specific and somewhat controversial candidate is closure.

#### Closure

Closure has long been identified as an important perceptual bridge from one-dimensional to two-dimensional shape (Koffka, 1935). The empirical evidence for this claim derives in part from visual search experiments using outline figures in which observers must detect a unique 2D target shape embedded in a field of distractor shapes (Figure 16; Elder & Zucker, 1993, 1994). Search becomes progressively faster as the closure of the shapes is increased. There is also evidence that our ability to encode shape information for later recall is a strong function of closure (Garrigan, 2012).

However, the present experiments do not involve the discrimination of 2D shapes but rather the detection of a fragmented contour in random clutter. Using the Field, Hayes, and Hess methodology, Kovacs and Julesz (1993) found that a roughly circular closed contour was more detectable than an open contour of the same length, and argued from this that closure enhances the detectability of contours in clutter. However, the magnitude and sign of the turning angles used to form the open and closed contour stimuli employed in this experiment were not equated, making it unclear whether it was closure per se that was enhancing detection.

Pettet et al. (1998) addressed this question by fixing each turning angle to be exactly 30° in magnitude. For the closed contour, the sign of the turning angle was constant, whereas for the open contour it was randomly and uniformly selected to be either positive or negative. As a result, the open control contours typically had multiple inflections not present in the closed contours.

### Table 3. Experiment 3: Statistical analysis of noise threshold contrasts based on two-tailed matched-sample t tests over observers. Note: Animal targets contrasted with first-order metamers M1 (open, with convexity bias) and M3 (closed).

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Target length</th>
<th>Mean</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal vs. M1</td>
<td>20</td>
<td>68</td>
<td>2.7</td>
<td>12</td>
<td>0.019</td>
</tr>
<tr>
<td>Animal vs. M3</td>
<td>20</td>
<td>11</td>
<td>0.4</td>
<td>12</td>
<td>0.69</td>
</tr>
<tr>
<td>M3 vs. M1</td>
<td>20</td>
<td>57</td>
<td>2.4</td>
<td>12</td>
<td>0.033</td>
</tr>
<tr>
<td>Animal vs. M1</td>
<td>40</td>
<td>407</td>
<td>4.3</td>
<td>12</td>
<td>0.00096</td>
</tr>
<tr>
<td>Animal vs. M3</td>
<td>40</td>
<td>197</td>
<td>2.5</td>
<td>12</td>
<td>0.026</td>
</tr>
<tr>
<td>M3 vs. M1</td>
<td>40</td>
<td>210</td>
<td>4.0</td>
<td>12</td>
<td>0.0018</td>
</tr>
</tbody>
</table>

### Table 4. Experiment 3: Statistical analysis of noise threshold contrasts based on two-tailed matched-sample t tests over observers. Note: Animal targets contrasted with first-order metamers M1 (open, with convexity bias) and M3 (closed).

<table>
<thead>
<tr>
<th>Predictor</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observer</td>
<td>12</td>
<td>0.17</td>
<td>0.99</td>
</tr>
<tr>
<td>Target shape</td>
<td>2</td>
<td>96.32</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Target length</td>
<td>3</td>
<td>1174.89</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Response (hit or miss)</td>
<td>1</td>
<td>0.16</td>
<td>0.69</td>
</tr>
</tbody>
</table>
They found that the closed contours were more easily detected than the open contours, replicating the Kovacs & Julesz result. In order to isolate the underlying cause, they (see also Pettet, 1999) also compared the detectability of a circular closed contour with constant turning angles and no inflections to the detectability of a noncircular closed contour with no inflections but nonconstant turning angles (a “bean”), and to a noncircular closed contour with two inflections and nonconstant turning angles (a “moon”). They found that variation in turning angle (curvature) and the presence of inflections reduced detection performance. In a critical control, they showed that when closed and open contours were better matched for the presence of inflections, the advantage of closure disappeared. Tversky, Geisler, and Perry (2004) have also presented evidence that small enhancements in detectability found for closed contours may derive from probability summation over local configurations of elements rather than a global closure mechanism.

In summary, prior work suggests that, despite the clear role of closure in the formation of 2D shape percepts (Koffka, 1935; Elder & Zucker, 1993, 1994; Garrigan, 2012), its role as a global cue for detection of contours in clutter may be minimal, with evidence to the contrary attributed to local geometric covariates, in particular the magnitude and sign of turning angles.

Our novel metamer stimuli provide an opportunity to more carefully test for a role of closure in contour detection while controlling carefully for local geometry. In Experiment 2, we compared the detectability of fragments of our animal stimuli ranging from 50% closure to 100% closure against length-matched local metamer controls. The results show that the superior detectability of the animal shapes does not depend upon perfect closure: Seventy-five percent closure was sufficient to generate a statistically significant advantage. These results are consistent with prior work (Elder & Zucker, 1994) showing that perceptual closure is a graded phenomenon, increasing smoothly in influence as the degree of closure increases. However, they are not consistent with a more simplistic explanation of the effects of closure in terms of proximity and good continuation cues (Tversky et al., 2004), as these are the same for the partial animal shapes and the metamer stimuli.

Nevertheless, Experiment 2 leaves the factor of closure confounded with other potential nonlocal factors determining perceptual grouping. The relative improvement in detectability of the animal stimuli with increased target length could be due to closure but could also be due to other concomitant global shape cues that are progressively revealed. Experiment 3 serves to more clearly separate these two possibilities by introducing a new class of closed metamer stimuli. Local geometry (turning angles) of both open and closed metamers is by construction matched to each other and to the animal shapes. The closed metamers are also constrained to be closed but otherwise are maximum entropy. Thus the sole difference between closed and open metamers is closure. The finding that closed metamers are more detectable than open metamers is thus clear evidence that closure does play a role in the detection of contours in clutter.

Other global cues

In addition to closure, there are a number of other global shape attributes known to be perceptually important that might play a role in perceptual grouping. These include convexity (Jacobs, 1996; Liu, Jacobs, & Basri, 1999; Corcoran, Mooney, & Tilton, 2011) and symmetry and parallelism (Lowe, 1985; Mohan & Nevatia, 1992; Wagemans, 1995; Zisserman et al., 1995; Jepson, Richards, & Knill, 1996; Behrmann, Zemel, & Mozer, 1998; Jacobs, 2003; Feldman, 2007; Sasaki, 2007; Stahl & Wang, 2008; Machilsen, Pauwels, Wagemans, 2009). Wilder, Feldman, and Singh (2016) have also advanced a more general theory of shape detectability based on efficient coding. They suggest that shapes that are more likely to occur in our visual environment will be coded more concisely by the brain, and that these more compact encodings for common stimuli will lead to more efficient detection. Our results are at least partially consistent with this theory, as the animal shapes we use here are surely more probable ecological stimuli than the metamers. It also seems plausible that our closed metamers are more ecological than our open metamers.

Neural basis

What are the brain mechanisms underlying the integration of local and global cues for contour grouping? Neurons in early visual cortex have highly localized, oriented receptive fields well suited to coding the local edges of object boundaries (Hubel & Wiesel, 1968; Dobbins, Zucker, & Cynader, 1987; Heitger et al., 1992). The standard view is that perceptual grouping of these edges into contours relies upon long-range horizontal connections between orientation columns in primary visual cortex (Gilbert & Wiesel, 1989; Li & Gilbert, 2002; Stettler et al., 2002; Li et al., 2006).

The significance of the results reported here is that global cues beyond the local association field are seen to codetermine contour grouping. This evidence demands a revision to the standard model, a revision that allows for global shape properties to influence local grouping. These global shape and object proper-
ties are coded deeper in the ventral stream, in Area V4 and TE/TEO, for example (Pasupathy & Connor, 1999; Kourtzi & Kanwisher, 2000, 2001; Op de Beeck, Wagemans, & Vogels, 2001; Connor, Brincat, & Pasupathy, 2007; Op de Beeck, Torfs, & Wagemans, 2008) and the responses of neurons in these areas depend upon input from the early visual areas coding local grouping relationships. We consider two alternative neural architectures that might account for the influence of global shape properties on perceptual grouping.

**Feedforward architecture**

One possible account is that early visual cortex passes multiple, competing local grouping hypotheses forward to intermediate and higher visual areas, which through a combination of feedforward computation and lateral interactions bring to bear more global factors to select a unique solution that is globally consistent with the visual input. Under this model, no trace of this global computation would necessarily be evident in early visual cortex.

This kind of computational progression is consistent with some studies that examine the dynamics of neural tuning in intermediate and higher areas of the object pathway. Brincat and Connor (2004) studied the dynamics of contour shape selectivity in posterior inferotemporal cortex (PIT). They found that early PIT responses were selective for individual straight and curved contour fragments, but that tuning evolved toward nonlinear selectivity to multifragment configurations of contours over a time course of about 60 ms.

Yau, Pasupathy, Brincat, and Connor (2013) extended this analysis to neurons in Area V4. Analogously to PIT, they found that V4 neural response evolved over a time course of about 50 ms from selectivity to contour orientation to selectivity for curvature. They posited an intracortical facilitatory circuit between similarly tuned V4 neurons that would generate a sustained selectivity for curvature.

The temporal dynamics in V4 and PIT are consistent with a feed-forward progression of progressively more complex shape tuning. V4 neural selectivity for orientation peaks at roughly 100 ms poststimulus, and selectivity for curvature peaks at roughly 150 ms. In PIT, neural selectivity for individual straight and curved contour fragments peaks at roughly 150 ms poststimulus, while selectivity for multifragment configurations peaks at roughly 200 ms.

Orientation and turning angles (local curvature) are equated for our animal and metamer stimuli, suggesting that the greatest differentiation in neural response to our target shapes would be seen in the later phase of PIT response, when selectivity for multifragment configurations emerges. However, it is unclear whether the simple local facilitatory circuit proposed here would be sufficient to resolve local grouping ambiguities in cluttered displays like the ones we use here.

**Feedback architecture**

A second possibility is that higher visual areas receive grouping hypotheses from early visual areas and then provide feedback to these early visual areas to support hypotheses that are consistent with global evidence while suppressing hypotheses that are inconsistent with global evidence. This kind of model extends the function of V1 beyond a transient initial stage of local processing to include a more sustained role in representing high-resolution geometric detail (Lee, Mumford, Romero, & Lamme, 1998; Lee & Mumford, 2003; Roelfsema & de Lange, 2016). Anatomically, substantial feedback connections from higher to lower visual areas in ventral stream are known to exist (Ungerleider, 1995), and the feedback of global shape information to inform local grouping in early visual cortex could be one of their functions. Under this second model, one would expect to find a trace of this global influence in the later phase of response of the early visual neurons coding local contour geometry.

There is growing evidence for this feedback hypothesis. Perceptual learning and attention have been found to dynamically influence contour grouping mechanisms in early visual cortex (Gilbert & Sigman, 2007; Li, Piech, & Gilbert, 2008; Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013), and V1 neurons can adapt dynamically to different nonlocal shape properties depending on the immediate task (McManus, Li, & Gilbert, 2011; Gilbert & Li, 2013; Piech, Li, Reeke, & Gilbert, 2013; Ramalingam, McManus, Li, & Gilbert, 2013). Both of these effects are thought to be the result of feedback from higher shape-selective visual areas. In more recent physiological experiments involving simultaneous recording from neurons in macaque areas V1 and V4, selectivity for global contour information is first seen in V4, emerging in V1 roughly 40 ms later (Chen, Yan, Gong, Gilbert, Liang, & Li, 2014). These results point to a V1–V4 recurrent network underlying the global perceptual integration of contours.

Human brain studies of illusory contour perception using event-related potential (ERP; Halgren, Mendola, Chong, & Dale, 2003), magnetoencephalography (MEG; Yoshino, Kawamoto, Yoshida, Kobayashi, & Shigemura, 2006) and trans-cranial magnetic stimulation (TMS; Wokke, Vandenbroucke, Scholte, & Lamme, 2013) methodologies point to an analogous cooperative recurrent computation involving early and late visual areas in ventral stream. Similar feedback mechanisms have been proposed to account for the perceptual assignment of figure/ground relationships.
(Lamme, 1995; Zipser, Lamme, & Schiller, 1996; Lamme, Super, & Spekreijse, 1998; Peterson & Cacciamani, 2013; Cacciamani, Scalf, & Peterson, 2015; Poort, Self, van Vugt, Malkki, & Roelfsema, 2016) and border ownership (Craft, Schutze, Niebur, & von der Heydt, 2007).

Human behavioral studies also point to a recurrent circuit for perceptual grouping and shape perception. Most of these studies have been based on backward masking, where performance on a shape detection or discrimination task is disrupted by the subsequent presentation of a masking stimulus that is thought to interfere with feedback from higher shape areas to early visual cortex (Enns & Di Lollo, 1997, 2000; Habak, Wilkinson, & Wilson, 2006). Whereas these studies rely on a disruption of visual processing to infer feedback, Drewes, Goren, Zhu, and Elder (2016) used the animal and open metamer stimuli employed here to demonstrate that a behavioral facilitation effect can also be observed. Specifically, target shapes were presented twice at the same location but with variable delay, in dynamic oriented element noise. Performance was found to peak when the two target presentations were separated by a delay of roughly 60 ms, suggesting a fast recurrent circuit underlying the perceptual organization of shape.

These primate and human studies have all focused on the perceptual organization of shape. However, recent human fMRI evidence suggests that feedback modulation may also extend to more semantic object category information (Uyar, Shomstein, Greenberg, & Drewes, 2016) and border ownership (Craft, Schutze, Niebur, & von der Heydt, 2007).

In summary, recent anatomical, physiological and behavioral evidence suggests that cortical feedback of global shape information from higher visual areas to lower visual areas may play an important role in the perceptual grouping of contours, and our psychophysical results are consistent with this hypothesis. That said, we suspect that perceptual object formation is the outcome of a cooperative computation involving diverse feedforward, lateral, and feedback connections between a number of early and late visual areas. A computation that is progressive, generating quick approximate solutions that are refined over time, would have adaptive advantage in supporting rapid real-time vision, and such mechanisms could be probed through future behavioral and physiological experiments that jointly manipulate local and global grouping cues as well as stimulus dynamics.

**A role for parietal cortex?**

Our discussion so far has focused on the so-called object pathway in ventral stream, but posterior areas of parietal cortex are known to be involved in the processing of object configural information (Kravitz et al., 2013; Freud, Plaut & Behrmann, 2016), and could also be involved in the detection of contours in clutter. Kuai, Li, Yu, and Kourtzi (2017) have assessed to what degree the presence of a contour in a cluttered display can be decoded from visual areas in the object pathway but also from areas in parietal cortex. They employed a standard static form of the Field, Hayes, and Hess stimulus, but also an interesting dynamic form, in which the stimulus is revealed over time through a narrow slit. For smaller slit widths, the elements constituting the collinear contour appeared sequentially, moving either up or down the slit, depending on the orientation of the contour and the direction of motion (left or right). Under these conditions, not only do the elements not appear simultaneously, their relative geometry when they do appear is not colinear. As a consequence the stimulus is unlikely to strongly activate long-range horizontal connections in V1, which are thought to be primarily selective for simultaneously presented, colinear stimuli.

Kuai et al. found that even for narrow slit widths for which only one element of the contour was visible at a time, observers were still well above chance on the contour detection task. fMRI analysis revealed that contour presence could be decoded from higher dorsal visual areas, including V3B, and intraparietal sulcus (IPS) as well as LOC, but not from early visual areas (V1, V2) or from ventral visual areas (V3v, hV4). Most accurate decoding was from posterior IPS. In contrast, when the stimuli were presented simultaneously, without the vertical slit, decoding was possible from early as well as late visual areas. Interestingly, decoding of contour presence (but not contour orientation) for simultaneous stimuli was still possible from higher dorsal as well as ventral areas. These results suggest that higher areas of dorsal stream are also involved with contour integration over both time and space. Exactly how these visual areas collaborate with higher areas in ventral stream and early visual areas to support detection of contours in cluttered displays is not yet understood.

**Keywords:** perceptual organization, contours, shape

**Acknowledgments**

We thank the NSERC Discovery, York Vision: Science to Applications, York Research Chair, and York Bridging Fund programs for supporting this research.

Commercial relationships: none.
Corresponding author: James H. Elder.
Footnote

1 Thresholds are somewhat lower than for Experiments 1 and 2. This may be due to the fact that targets in Experiments 1 and 2 were slightly less eccentric on average than in Experiment 3, due to an overly conservative rule used to ensure targets remained within the 20° stimulus window. Also, in Experiments 1 and 2 the stimulus region was defined by a midgray disk within the black background of the screen, whereas in Experiment 3 the entire background was black. The absence of a clear boundary for the stimulus region may have created more spatial uncertainty for the observer.

References


Contour integration over time: Psychophysical and fMRI evidence. Cerebral Cortex, 27, 3042–3051.


Raftery, A. E., & Lewis, S. M. (1996). Implementing MCMC. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), Markov chain Monte Carlo in...


Appendix: Construction of closed metamer stimuli

From each of the high-resolution animal shape polygons, we generated a coarser, approximately equilateral polygonal stimulus with only 120 vertices.
Each of these simpler stimuli was formed by partitioning the sequence of high-resolution vertices representing a shape into a sequence of 120 disjoint, contiguous subsequences, and then representing the shape by vertices located at the 120 arc-length mean locations for each of these subsequences. To determine an approximately equilateral polygonal model, we searched the exponential space of possible partitions to find the partition minimizing the variance in side length, using an expansion moves algorithm (Boykov, Veksler, & Zabih, 2001). The resulting polygons were approximately equilateral, with a mean standard deviation in length of 2.0%.

**Maximum entropy sampling**

The creation of closed metamer stimuli depends upon a fair sampling from the space $S_N$ of all simple (nonintersecting) polygons of length $N$. To achieve this, we used a random walk over polygons inspired by the work of Goldin et al. (2010). In particular, we devised an isoperimetric procedure for incrementally changing the turning angles of a shape while not altering any of the segment lengths, so that the polygon remains equilateral. Further, the resulting shape must not self-intersect. Note that almost all angle updates will not satisfy these requirements. For example, changing only one turning angle will generally require the length of at least one segment to change.

What is the minimum number of turning angles that must be changed in order to preserve the equilateral property? Consider the triangle defined by a random triplet of points (generally not neighboring points) on the contour. Note that moving any of these three points relative to each other must entail a length change in at least one of the triangle sides, as the side lengths uniquely determine the triangle. But now consider a tetragon defined by a quartet of points on the contour. In this case, there is a one-dimensional family of tetragons with exactly the same sequence of side lengths. Let us express the original contour as a concatenation of four fragments, each completed by one of the four tetragon segments (Figure A1). Imagine continuously rotating through this one-dimensional family, keeping each contour segment rigidly attached to its associated tetragon segment, but allowing it to rotate relative to the adjacent fragments. Since the sides of the tetragon remain fixed in length, all of the sides of the associated contour fragments will also remain fixed in length; only four of the turning angles will change, and these four changes must sum to 0. In this way, an isoperimetric shape change preserving the equilateral nature of the contour has been induced by a change to just four of the turning angles. Note that perceptually the change in shape can be dramatic if the sampled quartet of points is widely spaced.

Higher order inscribing polygons could also be used. For example, if a quintet of points is sampled, the result will be a two-dimensional family of pentagons with exactly the same side lengths. Note, however, that in general a number of shape changes must be proposed before finding one that does not induce a self-intersection. We use the tetragon method because it involves the least computation, and is thus practical even when the proportion of valid proposals is small.

To be precise, given a nonintersecting polygon $P_n$, we construct a new polygon $P_{n+1}$ by first randomly selecting four vertices from $P_n$. These four points define a tetragon $T$ with edges of length $\ell_1, \ell_2, \ell_3, \ell_4$, corresponding to four fragments $p_1, p_2, p_3, p_4$ of the original polygon $P_n$. Note that this tetragon is only one member of a one-dimensional family $T$ of tetragons with identical edge length sequences but different turning angles. To produce a new shape, we sample uniformly from this one-dimensional family to generate a new tetragon $T_*$ and replace the edges of this tetragon with the four polygon fragments $p_1, p_2, p_3, p_4$ to form a new

![Figure A1. Constructing the next random shape in the sequence. (a) Random shape (solid) and associated tetragon (dashed) at iteration $i$. (b) Random isoperimetric deformation of the tetragon, preserving side lengths. (c) Deformed shape induced by reattaching four shape fragments to rotated tetragon segments.](image-url)
candidate polygon $P_\ast$. We then check $P_\ast$ for self-intersections: If none are found, we set $P_{n+1} = P_\ast$. Otherwise, we draw another uniform sample $T_\ast$ from $T$. If a valid nonintersecting polygon has not been generated after five draws, we draw another random tetragon $T$ and repeat this process. We initialize the sequence by setting all turning angles to the same value of $2\pi/120$, forming a regular (nonintersecting) 120-sided polygon.

The random walk constrained in this way is a Markov chain satisfying detailed balance for a uniform distribution on $S_N$, and it is irreducible. Therefore, the stationary distribution of the Markov chain is the uniform distribution on $S_N$ (Tierney, 1996).

Initial samples are colored by the initial conditions and are therefore not fair samples from the uniform distribution on $S_N$, necessitating a burn-in period prior to selecting samples for analysis or as psychophysical stimuli. In addition, correlation between successive samples in the chain increases the number of samples required for parameter estimation. To minimize correlation between shapes used for parameter estimation, we apply a thinning method (Gelman, 1996), selecting samples at regularly spaced intervals along the chain. Sample size, burn-in and thinning parameters are determined using a standard method (Raftery & Lewis, 1996): We used 38,415 samples. The ratio of between-chain variance to within-chain variance was less than 0.973 for all chains, reflecting reliable convergence of our Markov-chain Monte Carlo sampling method (Gelman, 1996).

Matching the turning angles of animal shapes

The foregoing procedure creates closed, nonintersecting polygons that are otherwise maximum entropy. To form the desired closed metamer control stimuli, we wish to add the additional constraint that the probability distribution over turning angles matches that of our animal shapes.

To this end, we used kernel density estimation with cross validated bandwidth (Wassermann, 2006) to describe the marginal distribution of the turning angles for our animal shapes and then performed Metropolis-Hastings sampling (Metropolis et al., 1953; Hastings, 1970) with this density model as the target distribution. Given this kernel density model, we can generate samples of the desired constrained maximum entropy shapes using a modification of the random walk model. In particular, instead of accepting every candidate shape that has no intersections, we accept candidate shapes conditionally (Zhu, 1999): If $f(P_\ast) \geq f(P_n)$ we always set $P_{n+1} = P_\ast$. If, however, $f(P_\ast) < f(P_n)$, we set $P_{n+1} = P_\ast$ with probability $f(P_\ast)/f(P_n)$. This is a special form of Metropolis-Hastings sampling (Metropolis et al., 1953; Hastings, 1970) and is guaranteed to sample fairly from the distribution associated with $f$ in the long run. Samples size, burn-in and thinning parameters were determined automatically (Raftery & Lewis, 1996).