



Voluntary control of illusory contour formation

William J. Harrison^{1,2}  · Reuben Rideaux¹

Published online: 13 February 2019
© The Psychonomic Society, Inc. 2019

Abstract

The extent to which visual inference is shaped by attentional goals is unclear. Voluntary attention may simply modulate the priority with which information is accessed by the higher cognitive functions involved in perceptual decision making. Alternatively, voluntary attention may influence fundamental visual processes, such as those involved in segmenting an incoming retinal signal into a structured scene of coherent objects, thereby determining perceptual organization. Here we tested whether the segmentation and integration of visual form can be determined by an observer's goals, by exploiting a novel variant of the classical Kanizsa figure. We generated predictions about the influence of attention with a machine classifier and tested these predictions with a psychophysical response classification technique. Despite seeing the same image on each trial, observers' perception of illusory spatial structure depended on their attentional goals. These attention-contingent illusory contours directly conflicted with other, equally plausible visual forms implied by the geometry of the stimulus, revealing that attentional selection can determine the perceived layout of a fragmented scene. Attentional goals, therefore, not only select precomputed features or regions of space for prioritized processing, but under certain conditions also greatly influence perceptual organization, and thus visual appearance.

Keywords Object-based attention · Cognitive and attentional control · Grouping · Segmentation

The clutter inherent to natural visual environments means that goal-relevant objects often partially occlude one another. A critical function of the human visual system is to group common parts of objects while segmenting them from distracting objects and background, a process that requires interpreting an object's borders. Figures that produce illusory contours, such as the classic Kanizsa triangle (Kanizsa, 1976), have provided many insights into this problem by revealing the inferential processes made in determining figure–ground relationships. These figures give rise to a vivid percept of a shape emerging from sparse information, and thus demonstrate the visual system's ability to interpolate structure from fragmented information, to perceive edges in the absence of luminance discontinuities, and to fill-in a shape's surface

properties (for a review, see Shapley, Rubin, & Ringach, 2004). In the present study, we exploit these figures to investigate whether voluntary attention influences perceptual organization.

Most objects can be differentiated from their backgrounds via a luminance-defined border. The visual system is tasked with allocating one side of the border to an occluding object, and the other side to the background. This computation can be performed by neurons in macaque visual area V2 whose receptive fields fall on the edge of an object (Zhou, Friedman, & von der Heydt, 2000). These “border-ownership” cells can distinguish figure from ground even when the monkey attends elsewhere in the display (Qiu, Sugihara, & von der Heydt, 2007), and psychophysical adaptation aftereffects suggest such cells also exist in humans (von der Heydt, Macuda, & Qiu, 2005). Furthermore, neurophysiological work has revealed that V2 cells also process illusory edges (von der Heydt, Peterhans, & Baumgartner, 1984), though it is unclear whether those cells possess the same properties as border-ownership cells. These findings have contributed to the claim that visual structure is computed automatically and relatively early in the visual system, and that visual attention is guided by this precomputed structure (Mihalas, Dong, von der Heydt, & Niebur, 2011).

It is also known, however, that visual attention can modulate the perception of figure–ground relationships of luminance-defined stimuli. Both voluntary and involuntary

Electronic supplementary material The online version of this article (<https://doi.org/10.3758/s13414-019-01678-8>) contains supplementary material, which is available to authorized users.

✉ William J. Harrison
willjharri@gmail.com

¹ Department of Psychology, University of Cambridge, Cambridge, UK

² Queensland Brain Institute, University of Queensland, St Lucia, Australia

forms of attentional allocation can impact higher level cognition (Posner, 2016) and basic perception (Carrasco, 2011). Allocating visual attention to an area of space, for example, prioritizes processing of stimuli presented at that location relative to other locations (Posner, 1980), and can alter apparent contrast (Carrasco, Ling, & Read, 2004). As early as 1832, Necker described his ability to alter the apparent depth of an engraved crystalline form, now referred to as a Necker cube, via an overt shift of attention. More recent psychophysical work has shown that voluntary attention can alter perceived depth order (Driver & Baylis, 1996) as in the case of Rubin's face–vase illusion (Rubin, 1915; Wagemans et al., 2012), surface transparency (Tse, 2005), speed (Anton-Erxleben, Henrich, & Treue, 2007; Turatto, Vescovi, & Valsecchi, 2007), contrast (Carrasco et al., 2004; Liu, Abrams, & Carrasco, 2009; Stormer, McDonald, & Hillyard, 2009), and spatial frequency (Abrams, Barbot, & Carrasco, 2010; Gobell & Carrasco, 2005). Furthermore, visual attention has been shown to facilitate visual grouping according to Gestalt rules at both the neurophysiological (Wannig, Stanisor, & Roelfsema, 2011) and behavioral (Barbot, Liu, Kimchi, & Carrasco, 2018; Houtkamp, Spekrijse, & Roelfsema, 2003) level. For instance, Barbot et al. found that the apparent perceptual organization of luminance defined multi-element arrays is either intensified or attenuated by the presence or absence of covert attention, respectively. These findings raise the possibility that, regardless of whether it is necessary, visual attention may play a determining role in visual appearance under certain conditions. However, because these previous studies involved physically defined stimuli, it remains unclear whether visual attention simply modulates pre-attentively computed structure as suggested by neurophysiological work (McMains & Kastner, 2011; Qiu et al., 2007), or whether structural computations depend on the state of attention. Rivalrous illusory figures are perfectly suited to address this issue: If attending to one illusory figure results in illusory contours that directly conflict with the form of another illusory figure, then structural computations must depend on attention.

To investigate the influence of voluntary attention on perceptual organization, here we combined a novel illusory figure with an attentionally demanding task, exploiting human observers' propensity to use illusory edges when making perceptual decisions (Gold, Murray, Bennett, & Sekuler, 2000). We developed a novel Kanizsa figure (Fig. 1a) in which "Pac-Man" discs are arranged at the tips of an imaginary star. This figure includes multiple Gestalt cues that promote the segmentation and integration of various forms not defined by the physics of the stimulus (Harrison, Ayeni, & Bex, 2019). We predicted that, because some of these cues suggested competing configurations, selective attention could bias which figure elements were assigned to figure and which to ground. Although such a hypothesis is relatively uncontroversial, the critical question was whether grouping via selective attention

promotes illusory contour formation in direct conflict with competing implied form. For example, whereas the black inducers in Fig. 1a form part of an implied star, in isolation the black inducers imply an illusory triangle that competes with both the star form and a second illusory triangle implied by the white inducers. The dependence of such perceptual organization on voluntary attentional selection thus could reveal the extent of top-down processing on visual appearance. We therefore assessed whether the apparent organization of the figure was determined by which inducers were attended.

Materials and method

Observers

Three healthy participants, one naïve (N1) and two authors (A1 & A2, corresponding to authors R.R. and W.J.H., respectively), gave their informed written consent to participate in the project, which was approved by the University of Cambridge Psychology Research Ethics Committee. All procedures were in accordance with approved guidelines. Simulations were run to determine an appropriate number of trials per participant to ensure sufficient statistical power, and our total sample is similar to those generally employed for classification images. All participants had normal vision.

Apparatus

The stimuli were generated in MATLAB (The MathWorks, Inc., Natick, MA) using Psychophysics Toolbox extensions (Brainard, 1997; Cornelissen, Peters, & Palmer, 2002; Pelli, 1997). Stimuli were presented on a calibrated ASUS LCD monitor (120 Hz, 1,920 × 1,200). The viewing distance was 57 cm, and participants' head position was stabilized using a head-and-chin rest (43 pixels per degree of visual angle). Eye movement was recorded at 500 Hz using an EyeLink 1000 (SR Research Ltd., Ontario, Canada).

Stimuli and task

The stimulus was a modified version of the classic Kanizsa triangle. Six Pac-Man discs (radius = 1°) were arranged at the tips of an imaginary star centered on a fixation spot. The six tips of the star were equally spaced, and the distance from the center of the star to the center of each Pac-Man was 2.1°. The fixation spot was a white circle (0.1° diameter) and a black cross hair (stroke width = 1 pixel). The stimulus was presented on a gray background (77.5 cd/m²). The polarity of the inducers with respect to the background alternated across star tips. For half the trials, the three inducers forming an upright triangle were white, and the other inducers were black, and for

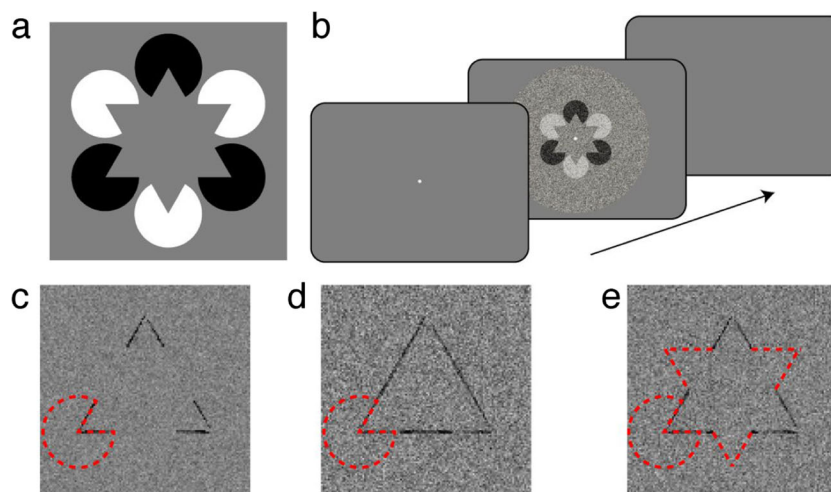


Fig. 1 Novel illusory figure and design used to test the influence of attention on perceptual organization. (a) Our variant of the classic Kanizsa figure. “Pac-Man” inducers are arranged such that a star appears to occlude black and white discs. Whereas the ensemble of features can produce the appearance of a star, grouping features by polarity leads to competing illusory triangles. We tested whether attending to one set of inducers (e.g., the white inducers) leads to interpolation of the illusory edge. (b) Example trial sequence. After an observer fixates a spot, the illusory figure with overlaid Gaussian noise is displayed for 250 ms. The observer’s task was to report whether the tips of the upright or the inverted

triangle were narrower or wider than an equilateral triangle. The target triangle was cued prior to, and held constant throughout, each testing block. The observer’s perceptual reports were then correlated with the noise on each trial to produce classification images. (c–e) Support vector machine (SVM) classifier images: We had an SVM classifier perform “narrow” versus “wide” triangle judgments after training it on three different protocols: (c) inducers, (d) a triangle, or (e) a star (see the Method section). Dashed red lines show the location of a Pac-Man, for reference, and in panel e they also show the tips of the star that do not influence classification

the other half of trials these associations were reversed. Inducers had a Weber contrast of .75.

We added Gaussian noise to the stimulus on each trial to measure classification images. The noise was 250×250 independently drawn luminance values with a mean of 0 and standard deviation of 1. Each noise image was scaled without interpolation to occupy 500×500 pixels, such that each randomly drawn luminance value occupied 2×2 pixels ($0.05^\circ \times 0.05^\circ$). The amplitude of these luminance values was then scaled to have an effective contrast of .125 on the display background, and the noise pixels were then added to the Kanizsa figure. Finally, a circular aperture was applied to the noise to ensure that the edges of the inducers were equally spaced from the noise edge (Fig. 1b).

The jaw size of the inducers was manipulated such that they were wider or narrower than an equilateral triangle, which would have exactly 60° of jaw angle for all inducers. The observer’s task was to indicate whether the jaws of the attended inducers were consistent with a triangle that was narrower or wider than an equilateral triangle. Prior to the first trial of a block, a message on the screen indicated which set of inducers framed the “target” triangle, and this was held constant within a block but alternated across blocks. The polarity of the target inducers and whether the triangles were narrow or wide were pseudorandomly assigned across trials, such that equal numbers of all trial types were included in each block. The relative jaw size of the attended inducers was independent of the unattended inducers; thus, the identity of the nontarget triangle was uncorrelated with the correct response.

Each trial began with the onset of the fixation spot and a check of fixation compliance for 250 ms. Following an additional random interval (0–500 ms, uniformly distributed), the stimulus was presented for 250 ms, after which only the background was presented while observers were given an unlimited duration to report the jaw size using a button press. The next trial would immediately follow a response. Throughout the experiment, eyetracking was used to ensure that observers did not break fixation during stimulus presentation. If gaze position strayed from fixation by more than 2° , the trial was aborted and a message was presented instructing participants to maintain fixation during stimulus presentation, after which the trial was repeated. Such breaks in fixation were extremely rare for all participants.

A three-down one-up staircase procedure was used to progress the difficulty of the task by varying the difference of the jaw size from 60° (i.e., from what would form an equilateral triangle). On each trial an additional angle was randomly added or subtracted to the standard 60° inducers. The initial difference was 2° . Following three correct responses, this difference would decrease by a step size of 0.5° , or it would increase by the same amount following a single error. When an incorrect response was followed by three correct responses (i.e., a reversal), the step size halved. If two incorrect responses were made in a row, the step size would double. If the step size fell below 0.05° , it would be reset to 0.2° . Blocks consisted of 624 trials, which took approximately 20 min, including a forced break. Each observer completed 16 blocks, for a total of 9,984 trials, which took approximately 5 h to

complete, spread over multiple days and testing sessions. To familiarize observers with the task, they underwent two training blocks of 624 trials each with no noise. They then were shown the stimulus with noise and completed as many trials as they felt were required before starting the experimental blocks.

Support vector machine models

Support vector machine (SVM) classifiers were trained and tested in MATLAB. We generated three hypotheses by training SVM classifiers on images of the (i) inducers, (ii) a triangle, or (iii) a star. We trained the classifiers using a quadratic kernel function and a least-squares method of hyperplane separation. The training images consisted of two exemplars (“narrow” and “wide”) with no noise (Supplementary Fig. 1). To generate hypotheses in the form of classification images, we used each of the classifiers to perform narrow/wide triangle judgments (trials = 9,984) with an equilateral triangle; thus, classification was exclusively influenced by the noise in the image.

Data and statistical analysis

The 9,984 noise images for a participant were separated according to perceptual report (“narrow” or “wide”). To collapse across inducer polarity, we inverted the sign of the noise on trials in which the cued inducers were white. We also collapsed across the upright and inverted cue conditions by spatially flipping the noise on inverted trials. To calculate which spatial locations influenced perceptual reports, we used a standard classification analysis in which each trial was classified according to the observer’s response with respect to the stimulus shown on that trial (Gold et al., 2000; Gold & Shubel, 2006; Mareschal, Dakin, & Bex, 2006; Neri & Heeger, 2002). Each stimulus was either narrow or wide (S_{narrow} or S_{wide}), and each response was either narrow or wide (R_{narrow} or R_{wide}), giving four trial types: (1) $S_{\text{narrow}}R_{\text{narrow}}$, (2) $S_{\text{wide}}R_{\text{narrow}}$, (3) $S_{\text{wide}}R_{\text{wide}}$, and (4) $S_{\text{narrow}}R_{\text{wide}}$. The classification images were generated by averaging and combining these response types according to the equation

$$CI = \left(\overline{S_{\text{narrow}}R_{\text{narrow}}} + \overline{S_{\text{wide}}R_{\text{narrow}}} \right) - \left(\overline{S_{\text{wide}}R_{\text{wide}}} + \overline{S_{\text{narrow}}R_{\text{wide}}} \right).$$

The resulting classification image showed the strength of correlation between each pixel’s location with the perceptual report made by the observer. Images were normalized to the “attend upright black inducers” condition, such that black pixels indicated locations where dark luminance noise was correlated with a “narrow” response and light luminance noise was correlated with a “wide” response. Conversely, white pixels indicated locations at which light luminance noise was correlated with a “narrow” response and dark luminance

noise was correlated with a “wide” response. The polarity and intensity of a given location thus provided information regarding that location’s contribution to perceptual decision making. To average across emergent triangle edges, we further summed the image with itself two times after rotating 120° and 240° using Matlab’s “imrotate” function using bilinear interpolation. This procedure results in a classification image that is invariant across edges, such that analysis of one edge summarized all three edges. Note that this was a conservative estimate of the classification image and that any spurious structure would only be diminished. To test for correlated pixels along the illusory edge of the classification image, we extracted 18 pixels along the bottom edge of the implied triangle but within the bounds of the implied star tip (see the bottom right panel of Fig. 2a). To ensure that these pixels were not contaminated by the averaging of nearest-neighbor pixels during rotation, described above, we excluded the three pixels closest to the inner corners of the star. We conducted one-sample, two-tailed Bayesian and Student’s *t* tests on these pixel values using the JASP software (JASP Team, 2017). The reported effect sizes are Cohen’s *d*.

We performed the model comparisons in Fig. 2c by first normalizing the noise of the mean classification image and each SVM prediction, such that the sum of the squared errors of each image equaled 1. We then subtracted the mean classification image from each prediction and found the sum of the squared error of the resulting difference. Finally, we normalized the difference scores to the model with the least error by subtracting from each distribution the mean of the distribution with the lowest error. This process was repeated for 200 repetitions of each SVM prediction. The mixture modeling (Fig. 3c) was performed similarly, but we further used Monte Carlo simulations to estimate the proportion of trials in which a triangle was perceived. In this case, each set of 200 simulated experiments included a proportion of triangle template trials, ranging from .33 (chance) to 1. We validated this model-fitting procedure by generating a simulated classification image with a known generative template or with proportional mixtures of templates, and then verified the model fitting’s returned results that approximated the ground truth. The Monte Carlo simulations were highly accurate for a range of simulated proportions, but slightly overestimated the contribution of the triangle template when the ground-truth contribution was close to .33, and conversely, slightly underestimated its contribution when the triangle was the only contributor.

Results

We used a response classification technique that allowed us to simultaneously assess where observers’ attention was allocated and whether such attentional allocation resulted in visual interpolation of illusory edges. At the beginning of each block

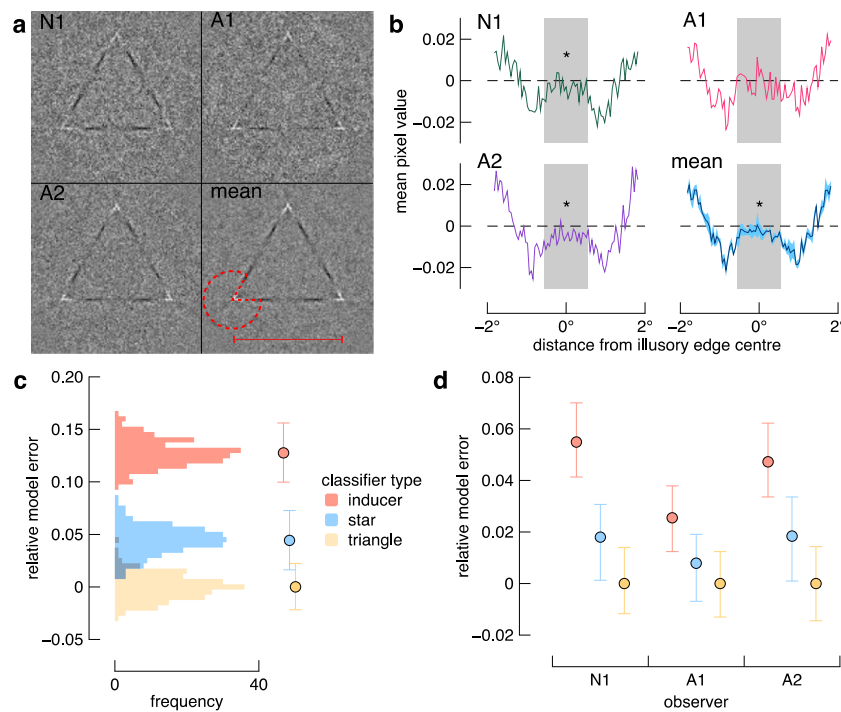


Fig. 2 Classification image results. (a) The individual and average classification images, normalized to the “attend upright black inducers” condition. Black pixels indicate locations where dark and light noise were correlated with “narrow” and “wide” judgments, respectively, and white pixels indicate the opposite relationship, after 9,984 trials per participant. The data have not been smoothed, but were first averaged across triangle edges and cropped to measure 122×122 pixels. In the mean image, a Pac-Man outline is shown for reference, and a red line indicates the spatial range of the implied triangle edge (from which the data in panel b are shown). (b) Pixel values along the illusory edge. The gray shaded region corresponds to a conservative estimate of the extent of the gap in the edge that would appear if observers necessarily saw a star shape (e.g., Fig. 1e).

The shaded interval around the line in the mean plot shows ± 1 standard error; asterisks indicate differences from zero ($BF_{10} > 10$ and $p < .05$; see the text). N1 is the naïve participant, and A1 and A2 are authors. (c) Comparison of SVM models for the averaged data. The distributions show comparisons of the mean classification image to the output of each SVM prediction, repeated 200 times. Data points and error bars represent the mean and 95% confidence intervals, respectively, for each SVM training regime. Model error has been normalized relative to the model with the least error, which is the model in which the SVM was trained to perceive a triangle within the attended inducers. (d) Comparison of the SVM models for each observer; the colors are as per panel c

of testing, observers were cued to report the relative jaw size of the inducers forming an upright (or inverted) triangle, corresponding to the white (or black) elements in Fig. 1a. By adding random visual noise to the target image on each trial

(Fig. 1b), 8we could use reverse correlation to measure “classification images.” An observer’s classification image quantified the correlation between each pixel in the image and the perceptual report, revealing which spatial locations were used

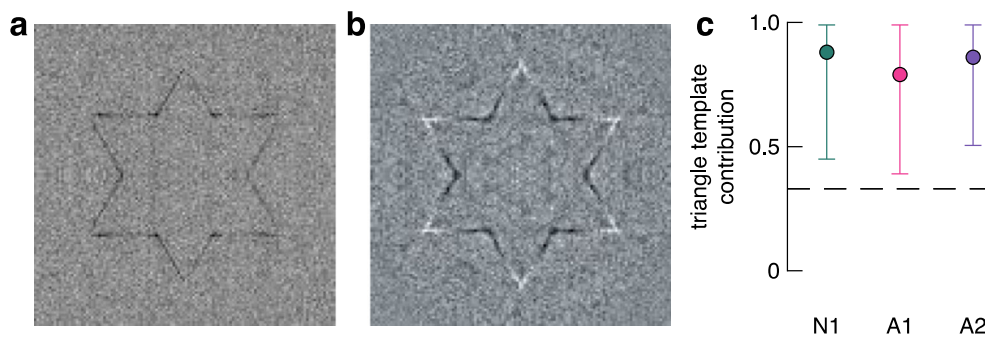


Fig. 3 Pre-attentive grouping. (a) Geometric form prediction of unattended grouping. The classification image derived from our SVM was summed with a flipped version of itself. Note that the inner corners of the star are well aligned, due to the design of our original Kanizsa figure. (b) Geometric form in the observers’ data. The mean classification image was summed with a flipped version of itself and reveals that the strength of the illusory edges is well aligned with the implied star. (c) Results of

mixture modeling used to explore the correspondence between the fluctuations of illusory edge strength and the implied figure geometry. The best-fitting model for each observer was one in which attention determined the perceptual outcome on 84% of trials. The dashed line indicates the proportion expected from a purely stochastic process. Error bars show 95% confidence intervals

for perceptual decisions (Abbey, Eckstein, & Bochud, 1999; Ahumada, 1996; Ahumada, Beard, & Ellis, 1998; Ahumada & Lovell, 1971; Beard & Ahumada, 1998; Gold et al., 2000).

We generated hypotheses regarding how observers' voluntary attention might influence their perception of this figure. We used an SVM classifier to judge small changes to a triangle image after training it on one of three different protocols (Supplementary Fig. 1). First, we generated a prediction of the hypothesis that observers can attend to the correct inducers but do not perceive illusory edges, by training a model to discriminate only the jaws of the inducers. This model was analogous to that of an ideal observer and revealed that only structure at the edges of the stimuli was used in generating a response (Fig. 1c). We next generated predictions of how illusory edges could be interpolated in this task. In one case, we assumed that illusory contours would be formed between attended inducers. We thus trained the classifier to discriminate whether a triangle's edges were bent outward or inward, and found a classification image approximating a triangle (Fig. 1d). In the other case, we assumed that, although selective attention might guide the correct perceptual decision, the illusory form of a star might be determined pre-attentively according to the physical structure of the entire stimulus. In this case, we trained the classifier to discriminate whether the alternating tips of a star—that is, the tips corresponding to a set of cued inducers—were relatively wide or narrow. The resulting classification image revealed edges that were interpolated beyond the inducers, but that they did not extend beyond the alternating star tips (Fig. 1e). These predictions not only provided qualitative comparisons for our empirical data, but they also allowed us to formally test which training regime produced a classification image that most closely resembled the human data.

To motivate observers to attend to only one possible configuration of the illusory figure, they were cued to report the relative jaw size (“narrow” or “wide”) of only a subset of Pac-Men positioned at the tips of an imaginary star (Fig. 1a). Each cued triangle was defined by three inducers, the jaw sizes of which were varied from 60° (an implied equilateral triangle) according to an adaptive staircase (see the Method section). In an early investigation into illusory contour perception, Ringach and Shapley (1996) had found that observers' perceptual thresholds were less than an angular degree with similarly sized stimuli. Specifically, observers were instructed to report only the jaw size of inducers forming an upward (or downward) triangle within a testing block. The noncued inducer jaws varied independently of the cued inducers and thus added no information regarding the correct response. To derive the spatial structure used for perceptual decisions, we added Gaussian noise to each trial and classified each noise image according to the observers' responses (Fig. 1b). To create the classification image for each observer, we summed all noise images for narrow reports and subtracted the sum of all noise images for wide reports (see the Method section). We

collapsed across inducer polarities by inverting the noise on trials in which the white inducers were cued, and across cue directions by flipping the noise on trials in which the downward-facing illusory triangle was cued. The resulting images quantified the correlation between each stimulus pixel and the observer's report. To analyze a single axis of emergent spatial structure, we first averaged each observer's data with themselves after rotating 120° and 240°, such that the correlations were averaged over the three sides of the triangle. Although this step involved bilinear interpolation of neighboring pixels, no other averaging or smoothing was performed, and this averaging is therefore most likely to have only reduced the strength of emergent illusory structure.

Classification images for the three observers and their mean are shown in Fig. 2a (see Supplementary Fig. 2a for the unrotated classification images). Images were normalized to the “attend upright black inducers” condition; black pixels indicated locations where dark and light noise were correlated with “narrow” and “wide” judgments, respectively, and white pixels indicated the opposite relationship. Two obvious patterns emerged. First, it was clear that observers based their reports on pixels within the jaws of the cued inducers, indicating that only some regions of the image—those aligned with the attended inducers—influenced perceptual decisions. Note the difference in the signs of the correlations between the edges and tips of the triangle—noise pixels in these regions had opposite influences on narrow/wide decisions, which was likely due to an illusory widening of the jaw center that was not registered by the SVM (cf. Fig. 1d). Second, the edges clearly extend beyond the red inducer outline shown in the mean image, revealing that observers' reports were influenced by illusory contours. However, it is also apparent that the spatial structure is nonuniform, with weaker correlations in the center of the illusory edges than in the corners of the inducers. We therefore quantitatively tested the extent of illusory contour formation.

To test whether the illusory edge interpolation extended into the region of the implied competing figure, we performed two analyses. First, we used Bayesian and Student's one-sample *t* tests to assess the pixel values along the edge of the triangle implied by the attended inducers (see the red line in Fig. 2a). We selected only pixels that fell within the bounds of the competing implied triangle (see the Method section and the gray-shaded regions in Fig. 2b) and found that these 18 pixels were below zero for the naïve participant [mean and *SEM*: $-3 \pm 0.9 \times 10^{-3}$; $BF_{10} = 18.365$, $t(17) = 3.585$, $p = .002$, $d = 0.845$], observer A2 [mean and *SEM*: $-5 \pm 0.7 \times 10^{-3}$; $BF_{10} = 8,141.356$, $t(17) = 6.944$, $p < .001$, $d = 1.637$], and the group [mean and *SEM*: $-3 \pm 0.4 \times 10^{-3}$; $BF_{10} = 16,580$, $t(17) = 7.38$, $p < .001$, $d = 1.738$], but not for A1 [mean and *SEM*: $-1 \pm 1 \times 10^{-3}$; $BF_{10} = 0.431$, $t(17) = 1.15$, $p = .266$, $d = 0.204$]. The lack of a difference in observer A1 may have been due to a difference in task-related strategy and/or increased lapse rate.

We next quantified the spatial structure content of the classification image by testing which prediction generated by the SVM was most similar to the human data (see Figs. 1c–1e). For each model, we generated 200 predictions, each with a unique distribution of noise, and computed the sum of the squared errors between the predictions and the mean classification image produced by the human observers (see the Method section). The resulting distributions of error, normalized to the best model, are shown in Fig. 2c and reveal that the model in which we trained the classifier to perceive a complete triangle is the best fit to the data (z test comparing the mean error for the triangle SVM vs. the distributions of error for the star and inducer SVMs: $ps < .0001$). The pattern of results was the same for all observers (Fig. 2d): For the inducer, star, and triangle templates, respectively, the mean standardized model errors (± 1 standard deviation) were: N1, 0.055 (0.007), 0.018 (0.008), 0 (0.007); A1, 0.026 (0.006), 0.008 (0.006), 0 (0.007); A2, 0.047 (0.008), 0.018 (0.008), 0 (0.007). Z tests comparing the mean error for the triangle SVM versus the distributions of error for the star and inducer SVMs were all significant (all $ps < .0001$). Taken together, these analyses reveal illusory contour formation between attended visual elements, and this interpolation occurred despite the contour conflicting with equally plausible implied spatial structures.

We next tested the spatial specificity of illusory contour formation. In the preceding analyses presented in Fig. 2b, we selectively tested only a single row of pixels aligned with the mouths of the inducers. For the two participants who showed a clear effect, we next tested how spatially specific the visual interpolation was, by repeating the same analysis but for the rows of pixels above and below the triangle boundary implied by the geometry of the attended inducers. Quite surprisingly, we found good evidence that there was an absence of illusory contour formation for the pixels below the implied triangle boundary (N1, $BF_{01} = 3.19$; A2, $BF_{01} = 3.31$), and equivocal evidence for the pixels above the implied triangle boundary (N1, $BF_{10} = 1.05$; A2, $BF_{01} = 1.83$). We therefore found evidence that only a single row of pixels extending between the inducer edges contributed to observers' perceptual decisions. These results thus revealed that the strength of the illusory contours was highly precisely aligned to the geometry of the triangle implied by the attended inducers. Consistent with this observation, psychophysical thresholds for identifying the relative inducer jaw size were reliably highly precise across testing sessions (see Supplementary Fig. 2b). Across sessions, the mean thresholds (± 1 standard error) for observers N1, A1, and A2 were $0.86^\circ \pm 0.03^\circ$, $0.84^\circ \pm 0.02^\circ$, and $0.66^\circ \pm 0.03^\circ$, respectively.

Our data further addressed the extent to which the noncued figural elements might have influenced the perceptual judgments. In our experiment, the noncued inducer jaw size was independent of the cued inducer jaw size, and was thus

uninformative as to the correct report. Indeed, we found no evidence in the classification image that observers' perceptual decisions were guided by these task-irrelevant cues. We modeled the possibility that these noncued elements were nonetheless grouped to form a star. In such a case in which a star was perceived, the task could still be performed accurately, were observers to base their reports on only the edges shared by the star and the triangle implied by the cued inducers. As expected, the SVM prediction of pre-attentive figure-ground segmentation shows gaps in the sides of the classification image triangle (Fig. 1e). Note that this model is equivalent to observers having perceived a whole star, but with a later-stage attentional signal focused on only some regions of the precomputed figure. Because we designed our illusory figure to be geometrically invertible, the extent of the illusory star form was pronounced if we summed the model's classification image with a flipped version of itself (Fig. 3a). In Fig. 3b, we show the result of performing this step with the observers' average classification image. Very similar patterns of results were found for all individual images (Supplementary Fig. 3).

Although flipping the classification image and producing a star-like figure might be somewhat trivial, more important is how the edges of the star were formed. We wished to test whether the extent of the illusory lines matched what would be expected were observers to have relied on a pre-attentively computed star form, rather than on two superimposed triangles determined by attentional selection. The critical aspect of the star-like figure shown in Fig. 3b is therefore whether or not the lines that form the star have terminators at the point where they intersect (i.e., the inner corners of the implied star). This is the case in Fig. 3a, because we were using the star model to generate the classification image. For the observers' data in Fig. 3b, on first glance the same appears to be the case: The emergent features appear to stop precisely at the point of intersection, suggesting that observers perceived a star but based their reports on only some parts of this figure. These qualitative results, however, are in contrast to the SVM analysis, presented above (Fig. 2c), in which we found that correlated noise in observers' classification images was best explained by observers having attended to the cued triangle. Thus, a remaining critical question, which we address below, was whether we could quantify the proportion of trials on which observers relied on different forms implied by the inducers.

In contrast to our initial quantitative analyses, the results of which suggested that visual attention determines which of two illusory triangles was perceived on a trial (Fig. 2), qualitative inspection of the star-like form shown in Fig. 3b suggests that observers may have based their reports on only some parts of a pre-attentively computed star. However, there are at least three possible explanations for the near-perfect alignment of the changes in illusory edge strength with the implied star figure (Fig. 3b). First, as discussed, a similar classification image would have been obtained had observers perceived a star on

every trial, a possibility that we discounted by our quantitative analysis of the illusory edge, described above (Fig. 2). Second, this qualitative result could be generated if trial-by-trial perceptual organization was stochastic, such that observers perceived each possible configuration approximately equally often across trials. Under this hypothesis, the resulting illusory contours shown in Fig. 2a are incidental rather than being determined by observers' attentional goals. The third possible explanation is that observers' voluntary allocation of attention determined the outcome on most, but not all, trials. To distinguish between the latter two possibilities, we used mixture modeling to quantify the proportion of trials on which observers' percept depended on the attentional instructions (see the Method section). A purely stochastic process would be implied were the proportion of trials accounted for by the triangle template no different from .33 (i.e., the apparent top-most surface was equally often a star, the cued triangle, or the noncued triangle; see Fig. 1c–1e). However, in the best-fitting model, the attention-contingent triangle template contributed to 84% of trials on average, which is much greater than would be expected by a stochastic process (Fig. 3c). At the individual level, the triangle contributions (and 95% confidence intervals) for N1, A1, and A2 were 88% (45%–100%), 79% (39%–99.5%), and 86% (50.5%–100%), respectively. This mixture modeling is thus consistent with observers' attentional goals determining the illusory contour interpolation on the vast majority of trials.

Discussion

We used classification images to address whether voluntary attention determines a scene's apparent visual structure. Using a psychophysical response classification paradigm, we tested which of three competing model predictions best described the influence of attention on illusory contour formation. Our results clearly showed that voluntary attention can guide the fundamental processes involved in the perceptual organization of illusory structure.

Unlike previous studies that had showed that visual attention modulates the appearance of physically defined surfaces (e.g., attending to different surfaces of the Necker cube; Necker, 1832), our study showed a rich interaction between attention and endogenously generated percepts. Classification images revealed the spatial location of the noise elements that influenced observers' responses, whereas interpreting subjective phenomenology was more difficult. However, our stimulus design ensured that the classification images revealed information about the perceived depth order of the image elements. The presence of lines in the classification image that extended between the inducers is clear evidence that at least two of the three observers based their judgments on the perception of a figure whose edges occluded the competing

(noncued) shape information. Given that the illusory edges of the triangle implied by the attended inducers directly conflicted with the regions of the competing implied figures (i.e., the star and inverted triangle), our finding that illusory edges were interpolated between attended inducers reveals that attention can determine depth order, even when the figures and ground are illusory. Spatial structure is thus computed by neural operations that are at least partially contingent on the voluntary state of the observer. The precision of illusory contours was nonetheless tightly aligned to the geometry of the luminance-defined structure, indicating that these inferential processes are also highly contingent on the scene or task context. Indeed, observers' psychophysical thresholds for the inducer task reveal a correspondence between their precise objective psychophysical performance and their subjective classification image.

We found clear interparticipant differences in the classification images. First, we found a clear effect of edge completion in our initial analysis of edge completion in only two of the three observers (Fig. 2). Such a difference across participants is not exceptional: Gold and Shubel (2006) also found classification image evidence of illusory edges in two out of three participants. Nonetheless, although the effect did not reach significance for one observer in our data, the same general direction of results was found in the classification image analysis (Fig. 2b), and the same results were found in the individual SVM model comparisons (Fig. 2d). A degree of homogeneity of our results across participants is also reflected by the fact that the group-average effect was significant. Importantly, the critical effect of a fully interpolated illusory edge was found in the naïve observer's data, and, across participants, we found relatively strong effect sizes of $d = 0.845$ (N1), $d = 1.637$ (A2), and $d = 0.204$ (A1), despite this not being significant for A1. The second interparticipant differences we found were in the raw classification images, which revealed varying degrees of completeness (Supplementary Fig. 2). For the two observers for whom the effect was significant, at least two edges of the triangle were clearly visible, and for the remaining observer one edge was clearly visible. We can think of at least three possible explanations for these individual differences (similar between-observer differences were reported by Gold et al., 2000, and Gold & Shubel, 2006). First, observers might have interpolated the edge of a single or a pair of unconnected lines between the cued inducers. Second, observers perceived a triangle, but only used part of this triangle to perform the task. Third, individual biases in attentional allocation may have differentially influenced interpolation of the different edges. Given the strength of the Kanizsa illusion—that is, the perception of a triangle—we think that the latter two explanations are more likely; however, we cannot definitively show this with the present data. The conclusion that attention influences illusory contour formation is equally valid under either of these explanations.

We were able to quantify the influence of noncued stimuli on perception by measuring a classification image across the entire stimulus. We found that changes in the strength of illusory contour formation between attended inducers were aligned with form implied by the noncued inducers. Our mixture modeling suggests that the non-cued stimuli influenced performance on approximately 16% of trials. Such a contribution of task-irrelevant features on perceptual decisions could be attributed to lapses in attentional allocation, or variability in the feed-forward processing of the incoming signal. Measuring perceived form in the absence of visual attention is notoriously difficult (Wagemans et al., 2012), which is perhaps one reason why many studies of figure–ground organization rely on single-unit recordings. Whereas neurophysiological recordings have revealed the brain regions involved in perceptual organization, they have left open the question of perceptual phenomena. Our data show that the influence of attention on perception is constrained by task-irrelevant information, providing yet further evidence that visual experience is the combination of both bottom-up and top-down processes. This conclusion sheds light on previous work in which competing color adaptation after-effects are biased according to alternating illusory contours at a similar location (van Lier, Vergeer, & Anstis, 2009). In these demonstrations, the onset of inducer elements likely attracts an observer's attention, resulting in perceptual completion processes specific to only the implied shape of attended elements. Surface filling-in would then follow the contours of the implied form (Poort et al., 2012). Indeed, other recent research from our lab has revealed that similar interactions may occur between attention and surface filling-in (Harrison et al., 2019).

The influence of attention on figure–ground segmentation may be explained by feedback signals from the lateral occipital complex (Murray et al., 2002; Stanley & Rubin, 2003) that could act as early as V1 (Wannig et al., 2011), but also might involve modulating the responses of border-ownership cells in V2 (Qiu et al., 2007). Border-ownership cells indicate which side of a border is object versus ground. Previous work showing that the activity of border-ownership cells is modulated by visual attention (Qiu et al., 2007) has been limited to luminance-defined borders. Our finding that the information inferred by the visual system is influenced by voluntary attention suggests that attentional modulation of border-ownership may apply similarly to illusory contours (von der Heydt et al., 1984). Early psychophysical work suggested that illusory contours are perceived in the absence of attention (Davis & Driver, 1994; Mattingley, Davis, & Driver, 1997), but these studies did not address the question of whether illusory contours can be formed *because* of voluntary attention, which we have shown here. Our findings are also distinct from other recent work that has shown that attention can influence the appearance of existing surfaces (Tse, 2005). In our study, visual attention had a causal role in forming the structure from

which perceptual decisions were made. We anticipate that our simple stimulus and task design might prove to be a useful neurophysiological assay to test further the neural substrates governing the interaction between voluntary attention and perceptual organization.

Author note We are indebted to Peter Bex, who developed the novel Kanizsa figure with us and provided helpful feedback on our study design and results. We also thank Tom Wallis for feedback on an earlier draft that led to the mixture modeling and for overall improvements in the manuscript. This research was supported by funding to W.J.H. from King's College Cambridge and the National Health and Medical Research Council of Australia (APP1091257), and by funding to R.R. (ECF-2017-573) from the Leverhulme Trust. Both authors designed the experiment and collected the data. W.J.H. analyzed the experimental data, R.R. performed the SVM analyses, and both authors performed the model comparisons. Both authors contributed equally to the writing of the manuscript. We declare we have no competing interests.

Data availability Data and analysis code are available via the Open Science Framework: <https://osf.io/42sp8/>.

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

References

- Abbey, C. K., Eckstein, M. P., & Bochud, F. O. (1999). Estimation of human-observer templates in two-alternative forced-choice experiments. In E. A. Krupinski (Ed.), *Proceedings of Medical Imaging 1999: Image perception and performance* (pp. 284–295). Bellingham: International Society for Optics and Photonics. <https://doi.org/10.1117/12.349653>
- Abrams, J., Barbot, A., & Carrasco, M. (2010). Voluntary attention increases perceived spatial frequency. *Attention, Perception, & Psychophysics*, 72, 1510–1521. <https://doi.org/10.3758/APP.72.6.1510>
- Ahumada, A., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, 49, 1751–1756. <https://doi.org/10.1121/1.1912577>
- Ahumada, A. J. (1996). Perceptual classification images from Vernier acuity masked by noise. *Perception*, 25(1 Suppl), 2. <https://doi.org/10.1068/v96i0501>
- Ahumada, A. J., Beard, B. L., & Ellis, S. R. (1998). *Response classification images in Vernier acuity*. Article presented at the 1998 Annual Meeting of the Association for Research in Vision and Ophthalmology, Fort Lauderdale.
- Anton-Erxleben, K., Henrich, C., & Treue, S. (2007). Attention changes perceived size of moving visual patterns. *Journal of Vision*, 7(11), 5. <https://doi.org/10.1167/7.11.5>
- Barbot, A., Liu, S., Kimchi, R., & Carrasco, M. (2018). Attention enhances apparent perceptual organization. *Psychonomic Bulletin and Review*, 25, 1824–1832. <https://doi.org/10.3758/s13423-017-1365-x>
- Beard, B. L., & Ahumada, A. J., Jr. (1998). Technique to extract relevant image features for visual tasks. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human vision and electronic imaging III* (pp. 79–85). Bellingham, WA: International Society for Optics and Photonics. 10.1117/12.320099
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436. <https://doi.org/10.1163/156856897X00357>
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51, 1484–1525. <https://doi.org/10.1016/j.visres.2011.04.012>

- Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, 7, 308–313. <https://doi.org/10.1038/nn1194>
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The EyeLink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34, 613–617. <https://doi.org/10.3758/BF03195489>
- Davis, G., & Driver, J. (1994). Parallel detection of Kanizsa subjective figures in the human visual system. *Nature*, 371, 791–793. <https://doi.org/10.1038/371791a0>
- Driver, J., & Baylis, G. C. (1996). Edge-assignment and figure-ground segmentation in short-term visual matching. *Cognitive Psychology*, 31, 248–306. <https://doi.org/10.1006/cogp.1996.0018>
- Gobell, J., & Carrasco, M. (2005). Attention alters the appearance of spatial frequency and gap size. *Psychological Science*, 16, 644–651. <https://doi.org/10.1111/j.1467-9280.2005.01588.x>
- Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology*, 10, 663–666. [https://doi.org/10.1016/S0960-9822\(00\)00523-6](https://doi.org/10.1016/S0960-9822(00)00523-6)
- Gold, J. M., & Shubel, E. (2006). The spatiotemporal properties of visual completion measured by response classification. *Journal of Vision*, 6(4), 5. <https://doi.org/10.1167/6.4.5>
- Harrison, W. J., Ayeni, A. J., & Bex, P. J. (2019). Attentional selection and illusory surface appearance. *Scientific Reports*. <https://doi.org/10.1038/s41598-018-37084-7>
- Houtkamp, R., Spekreijse, H., & Roelfsema, P. R. (2003). A gradual spread of attention during mental curve tracing. *Perception & Psychophysics*, 65, 1136–1144. <https://doi.org/10.3758/BF03194840>
- JASP Team. (2017). JASP [Computer software]. Retrieved from <https://jasp-stats.org/download/>
- Kanizsa, G. (1976). Subjective contours. *Scientific American*, 234, 48–52.
- Liu, T., Abrams, J., & Carrasco, M. (2009). Voluntary attention enhances contrast appearance. *Psychological Science*, 20, 354–362. <https://doi.org/10.1111/j.1467-9280.2009.02300.x>
- Mareschal, I., Dakin, S. C., & Bex, P. J. (2006). Dynamic properties of orientation discrimination assessed by using classification images. *Proceedings of the National Academy of Sciences*, 103, 5131–5136. <https://doi.org/10.1073/pnas.0507259103>
- Mattingley, J. B., Davis, G., & Driver, J. (1997). Preattentive filling-in of visual surfaces in parietal extinction. *Science*, 275, 671–674.
- McMains, S., & Kastner, S. (2011). Interactions of top-down and bottom-up mechanisms in human visual cortex. *Journal of Neuroscience*, 31, 587–597. <https://doi.org/10.1523/JNEUROSCI.3766-10.2011>
- Mihalas, S., Dong, Y., von der Heydt, R., & Niebur, E. (2011). Mechanisms of perceptual organization provide auto-zoom and auto-localization for attention to objects. *Proceedings of the National Academy of Sciences*, 108, 7583–7588. <https://doi.org/10.1073/pnas.1014655108>
- Murray, M. M., Wylie, G. R., Higgins, B. A., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). The spatiotemporal dynamics of illusory contour processing: Combined high-density electrical mapping, source analysis, and functional magnetic resonance imaging. *Journal of Neuroscience*, 22, 5055–5073.
- Necker, L. A. (1832). LXI. Observations on some remarkable optical phenomena seen in Switzerland; and on an optical phenomenon which occurs on viewing a figure of a crystal or geometrical solid. *London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1, 329–337. <https://doi.org/10.1080/14786443208647909>
- Neri, P., & Heeger, D. J. (2002). Spatiotemporal mechanisms for detecting and identifying image features in human vision. *Nature Neuroscience*, 5, 812–816. <https://doi.org/10.1038/nn886>
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442. <https://doi.org/10.1163/156856897X00366>
- Poort, J., Raudies, F., Wannig, A., Lamme, V. A. F., Neumann, H., & Roelfsema, P. R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron*, 75, 143–156. <https://doi.org/10.1016/j.neuron.2012.04.032>
- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25. <https://doi.org/10.1080/00335558008248231>
- Posner, M. I. (2016). Orienting of attention: Then and now. *Quarterly Journal of Experimental Psychology*, 69, 1864–1875. <https://doi.org/10.1080/17470218.2014.937446>
- Qiu, F. T., Sugihara, T., & von der Heydt, R. (2007). Figure-ground mechanisms provide structure for selective attention. *Nature Neuroscience*, 10, 1492–1499. <https://doi.org/10.1038/nn1989>
- Ringach, D. L., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research*, 36, 3037–3050.
- Rubin, E. (1915). *Synsoplevede Figurer: Studier i psykologisk Analyse/ Visuell wahrgenommene Figuren: Studien in psychologischer Analyse* [Visually perceived figures: Studies in psychological analysis]. Copenhagen, Denmark: Gyldendalske Boghandel.
- Shapley, R., Rubin, N., & Ringach, D. (2004). Visual segmentation and illusory contours. In L. M. Chalupa & J. S. Werner (Eds.), *The visual neurosciences* (Vol. 2, pp. 1119–1128). Cambridge: MIT Press. Retrieved from http://www.cns.nyu.edu/nava/MyPubs/Shapley-Rubin-Ringach_VisSeg_MITpress2004.pdf
- Stanley, D. A., & Rubin, N. (2003). fMRI activation in response to illusory contours and salient regions in the human Lateral Occipital Complex. *Neuron*, 37, 323–331. [https://doi.org/10.1016/S0896-6273\(02\)01148-0](https://doi.org/10.1016/S0896-6273(02)01148-0)
- Stormer, V. S., McDonald, J. J., & Hillyard, S. A. (2009). Cross-modal cueing of attention alters appearance and early cortical processing of visual stimuli. *Proceedings of the National Academy of Sciences*, 106, 22456–22461. <https://doi.org/10.1073/pnas.0907573106>
- Tse, P. U. (2005). Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Research*, 45, 1095–1098. <https://doi.org/10.1016/j.visres.2004.11.001>
- Turatto, M., Vescovi, M., & Valsecchi, M. (2007). Attention makes moving objects be perceived to move faster. *Vision Research*, 47, 166–178. <https://doi.org/10.1016/j.visres.2006.10.002>
- van Lier, R., Vergeer, M., & Anstis, S. (2009). Filling-in afterimage colors between the lines. *Current Biology*, 19, R323–R324. <https://doi.org/10.1016/j.cub.2009.03.010>
- von der Heydt, R., Macuda, T., & Qiu, F. T. (2005). Border-ownership-dependent tilt aftereffect. *Journal of the Optical Society of America A*, 22, 2222–2229. <https://doi.org/10.1364/JOSAA.22.002222>
- von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224, 1260–1262. <https://doi.org/10.1126/science.6539501>
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138, 1172–1217. <https://doi.org/10.1037/a0029333>
- Wannig, A., Stanisor, L., & Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience*, 14, 1243–1244. <https://doi.org/10.1038/nn.2910>
- Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, 20, 6594–6611. <https://doi.org/10.1523/JNEUROSCI.20-17-06594.2000>