Voluntary control of illusory contour formation

William J Harrison [1,2,*] & Reuben Rideaux [1]

[1]Department of Psychology, University of Cambridge

[2]Queensland Brain Institute, The University of Queensland

*Corresponding author (willjharri@gmail.com)

ABSTRACT

The visual brain is tasked with segmenting visual input into a structured scene of coherent objects. Figures that produce the perception of illusory contours reveal minimal conditions necessary for the visual system to perform figure-ground segmentation, but it is unclear what role visual attention plays in such perceptual organisation. Here we tested whether illusory contours are formed under the guidance of voluntary attention by exploiting a novel variant of the classical Kanizsa figure. We used a psychophysical response classification technique to quantify spatial structures guiding perceptual decisions, and found that illusory contour formation was contingent on attended elements of the figure. However, the strength of these illusory contours was constrained by form implied by unattended figural elements. Our data thus reveal an interplay of top-down and bottom-up processes in figure-ground segmentation: while attention is not necessary for illusory contour formation, under some conditions it is sufficient.

MAIN TEXT

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 which requires interpreting an object's borders. Figures which produce illusory contours, such as the classic Kanizsa triangle[1], 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. In the present study, we investigated whether voluntary attention guides figure-ground segmentation of competing illusory contours.

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[2]. These “border-ownership” cells can distinguish figure from ground even when the monkey attends elsewhere in the display[3], and psychophysical adaptation aftereffects suggest such cells also exist in humans[4]. Further, neurophysiological work has revealed that V2 cells also process illusory edges[5], 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 pre-computed structure[6].

It is also known, however, that visual attention can modulate the perception of figure-ground relationships of luminance-defined stimuli. In particular, psychophysical work has shown that voluntary attention can alter perceived depth order[7], as in the case of Rubin’s face-vase illusion[8], and surface filling-in[9]. These findings raise the possibility that perceptual organisation may not be as automatic as previously hypothesized. Indeed, electrophysiological work has revealed that the activity of some border-ownership cells is modulated by visual attention[3]. Furthermore, visual attention has been shown to facilitate visual grouping according to Gestalt rules at both the neurophysiological[10] and behavioural[11] level. However, because these previous studies involve physically defined stimuli, it remains unclear whether visual attention simply modulates pre-attentively computed structure as suggested by neurophysiological work[3], 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 visual interpolation, here we combined a novel illusory figure with an attentionally demanding task, exploiting human observers’ propensity to use illusory edges when making perceptual decisions[12]. We developed a novel Kanizsa figure (**Fig. 1a**), in which "pacman" discs are arranged at the tips of an imaginary star. This figure includes multiple Gestalt cues that promote the segmentation of various forms. We predict that, because some of these cues suggest competing configurations,

selective attention can bias which figure elements are assigned to figure and which to ground. Although such a hypothesis is relatively uncontroversial, the critical question is whether grouping via selective attention promotes illusory contour formation in direct conflict with competing implied form. For example, while the black inducers of **Figure 1a** form part of an implied star, in isolation the black inducers imply an illusory triangle that competes with both the star form as well as the illusory triangle implied by the white inducers. We therefore assessed whether the figure perceived by observers is determined by which inducers are attended.

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 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 **Figure 1a**. By adding random visual noise to the target image on each trial (**Fig. 1b**), we could use reverse correlation to measure "classification images". An observer's classification image quantifies a correlation between each pixel in the image and the perceptual report revealing which spatial structures are used for perceptual decisions[12].

We generated hypotheses regarding how observers' voluntary attention may influence their perception of this figure. We used a support vector machine (SVM) classifier to perform the same experiment as observers after training it on one of three different protocols. 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 is analogous to that of an ideal observer[12], and reveals that only structure at the edges of the stimuli are 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 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, after testing on inducers only, found a classification image approximating a triangle (**Fig. 1d**). In the other case, we assumed that, although selective attention may guide the correct perceptual decision, the illusory form of a star may be determined pre-attentively according to the physical structure of the entire stimulus. In this case, we trained the classifier to

discriminate whether alternating tips of a star where relatively wide or narrow. When tested on the inducers, the resulting classification image reveals edges that are interpolated beyond the inducers, but that do not extend beyond the alternating star tips (**Fig. 1e**). These predictions provide qualitative comparisons for our empirical data.

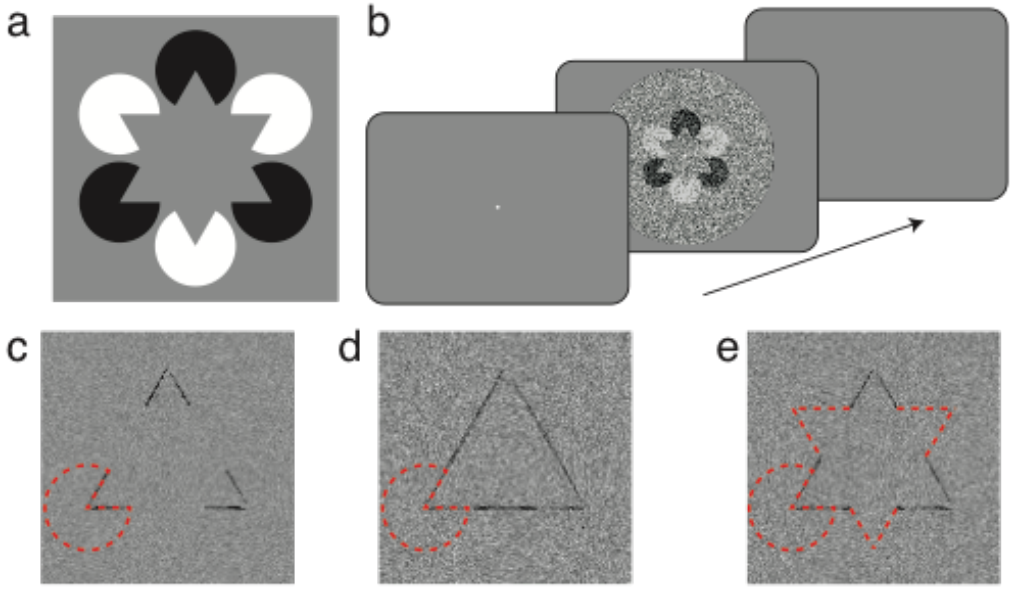


Figure 1. Novel illusory figure and design used to test the influence of attention on perceptual organisation. a) Our variant of the classic Kanizsa figure. "Pacman" inducers are arranged such that a star appears to occlude black and white discs. Whereas the ensemble of features may produce the appearance of a star, grouping features by polarity leads to competing illusory triangles. We test 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 250ms. The observer's task was to report whether the tips of the upright or 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 a SVM classifier perform the same task as observers after training it on three different protocols: (c) inducers alone or in combination with a (d) triangle or (e) star. Dashed red lines show the location of a pacman for reference, and in (e) the tips of the star that do not influence classification.

RESULTS

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 pacmen positioned at the tips of an imaginary star (**Fig. 1a**). Specifically, observers were instructed to report only the jaw size of inducers forming an upward (or downward) triangle within a

testing block. The uncued 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 Online Methods). We collapsed across inducer polarity by inverting the noise on trials in which the white inducers were cued, and across cue direction by flipping the noise on trials in which the downward facing illusory triangle was cued. The resulting images quantify the correlation between each stimulus pixel and the observer's report. So we could analyse a single axis of emergent spatial structure, we first averaged each observer's data with itself after rotating 120° and 240° such that correlations were averaged over the three sides of the triangle. Although this step involved bilinear interpolation of neighbouring pixels, no other averaging or smoothing was performed.

Classification images for three observers and their mean are shown in **Figure 2a** (see **Supp. Fig. 1a** for unrotated classification images). There are two obvious patterns that emerge. First, it is 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 sign of the correlation between the edges and tips of the triangle - noise pixels in these regions have the opposite influence on narrow/wide decisions, which is likely due to an illusory widening of the jaw centre which is not registered by the SVM (cf. **Fig. 1d**). Second, the edges clearly extend beyond the inducers, revealing observers' reports were influenced by illusory contours. However, it is also apparent that the spatial structure is non-uniform, with weaker correlations in the centre of the illusory edges than in the corners of the inducers. To test whether the illusory edge interpolation extended into the region of the implied competing figure, we performed Bayesian and Students' one-sampled t-tests of the pixel values along the edge of the triangle implied by the attended inducers (see red line in **Fig. 2a**). We selected only pixels that fell within the bounds of the competing implied triangle (see Online Methods and **Fig. 2b**), but nonetheless found that these 18 pixels were below zero for the naïve participant (mean and sem: $-3 \pm .9 \times 10^{-3}$, $BF_{10}=18.37$, $t(17)=3.59$, $p=0.002$), observer A2 (mean and sem: $-5 \pm .7 \times 10^{-3}$, $BF_{10}=8141.36$, $t(17)=6.94$, $p<0.001$), and the group (mean

and sem: -3 ± .4 x 10^{-3}, BF_{10}=16580, t(17) = 7.38, p<0.001), but not for A1 (mean and sem: -1 ± 1 x 10^{-3}, BF_{10}=0.42, t(17)=1.15, p=0.27). These data thus reveal illusory contour formation between attended visual elements even when the contour conflicts with equally plausible implied spatial structure.

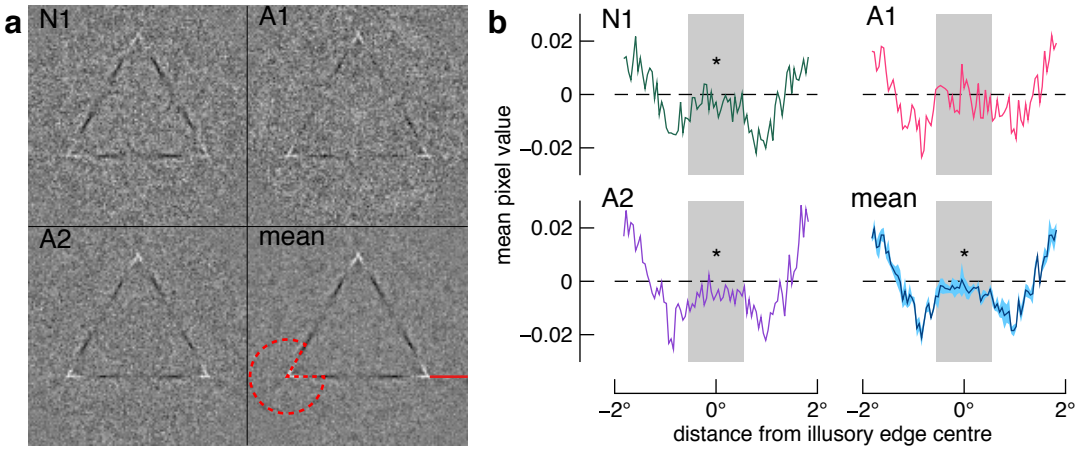


Figure 2. Classification image results. a) The individual and average classification images. Black and white pixels in this image are correlated with "narrow" and "wide" perceptual reports, respectively, after 9984 trials per participant. Data have not been smoothed, but were first averaged across triangle edges and cropped to be 122x122 pixels. In the mean image, a pacman outline is shown for reference, and a red line indicates the row of pixels aligned to the illusory edge, from which data in (b) are shown. b) Pixel values along the illusory edge. The grey shaded region corresponds to a conservative estimate of the extent of the gap that would appear if observers necessarily saw a star shape (e.g. Fig. 1e). The blue shaded region in the mean plot shows ±one standard error; asterisks indicate differences from zero ($BF_{10} > 10$ and $p < 0.05$). N1 is the naïve participant; A1 and A2 are authors.

We next tested the spatial specificity of illusory contour formation. For the two participants who showed a clear effect, we tested how spatially specific visual interpolation was by repeating the same analysis as above but for the row of pixels above and below the triangle boundary implied by the geometry of the attended inducers. Quite surprisingly we found that there was no evidence of illusory contour formation for either row of pixels for either participant, revealing the illusory contours were highly precisely aligned to the geometry of the triangle implied by the inducers. Consistent with this observation, psychophysical thresholds for identifying the relative inducer jaw size were reliably highly precise across testing sessions (see **Supp. Fig. 1a**).

Our data further address the extent to which the unattended figural elements were grouped. In our experiment, the uncued inducer jaw size was independent of the cued inducer jaw

size, and was thus uninformative of the correct report. Indeed, we found no evidence in the classification image that observers attended to these task-irrelevant cues. We modelled the possibility that these unattended cues were nonetheless grouped: 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 perceiving a whole star, but focussing their attention on only some regions of the figure. Because we designed our illusory figure to be geometrically invertible, the extent of the illusory star form is pronounced if we sum the classification image with a flipped version of itself (**Fig. 3a**). In **Figure 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 (**Supp. Fig. 2**). This result is strikingly similar to the SVM prediction, revealing that the changes in strength of edges of the illusory form are near-perfectly aligned with the geometry of the implied star. The reduction in pixel intensity along the triangle edge reported above (**Fig. 2b**) can thus likely be explained by the illusory star form constraining voluntary interpolation of the illusory triangle edge.

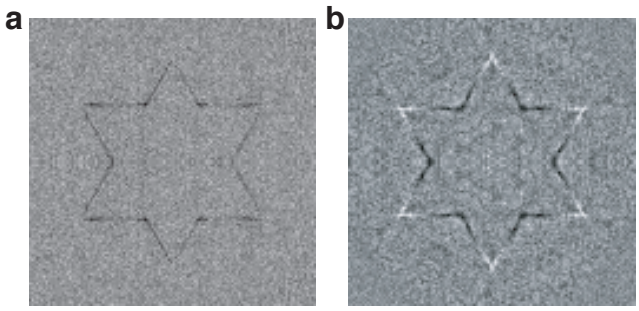


Figure 3. Pre-attentive grouping. a) Geometric form prediction of unattended grouping. The classification image derived from our SVM was summed with a flipped version of itself. Note that the inner corners of the star are well aligned due to the design of our original Kanizsa figure. b) Geometric form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

DISCUSSION

We used classification images to quantify which spatial structure guides perceptual decisions when attention is directed to only some elements of a scene. Our main findings are twofold. First, we found that voluntarily attending to a subset of Gestalt cues leads to visual interpolation between those cues, and the resulting illusory edges inform perceptual decisions. Second, the strength of visual interpolation is constrained by the grouping of unattended elements. Our study thus reveals that figure-ground segmentation of illusory

form is determined by both top-down (cognition-contingent) and bottom-up (stimulus-driven) processes.

Unlike previous studies that show visual attention modulates the appearance of physically defined surfaces (e.g., attending to different surfaces of the Necker cube), our study shows a rich interaction between attention and perception. The illusory edges of the triangle implied by the attended inducers directly conflict 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 figures and ground are illusory. Spatial structure is thus computed by neural regions whose operations can be contingent on the voluntary state of the observers. The precision of illusory contours were nonetheless tightly aligned to the geometry of luminance defined structure, indicating these inferential processes are also highly contingent on scene or task context. Indeed, observers' psychophysical thresholds for the inducer task reveal a correspondence between their precise objective psychophysical performance and subjective classification image.

We were able to quantify the influence of unattended 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 unattended inducers. Measuring perceived form in the absence of visual attention is notoriously difficult[8], which is perhaps one reason why many studies of figure-ground organisation rely on single-unit recordings. Whereas neurophysiological recordings have revealed the brain regions involved in perceptual organisation, 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 colour adaptation after-effects are biased according to alternating illusory contours at a similar location[13]. 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[14].

The influence of attention on figure-ground segmentation may be explained by feedback signals from the lateral occipital complex[15,16] that could act as early as V1[10], but also may involve modulating responses of border-ownership cells in V2[3]. Border-ownership cells indicate which side of a border is an object versus ground. Previous work showing the activity of border-ownership cells is modulated by visual attention[3] has been limited to luminance-defined borders. Our finding that information inferred by the visual system is influenced by voluntary attention suggests that attentional modulation of border-ownership may similarly apply to illusory contours[5]. Early psychophysical work suggested that illusory contours are perceived in the absence of attention[17,18], but 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 found attention can influence the appearance of existing surfaces[9]. In our study, visual attention had a causal role in forming the structure from which perceptual decisions were made.

ONLINE METHODS

Observers. Three healthy subjects, one naïve (N1) and two authors (A1 & A2 corresponding to authors RR and WH, 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. Stimuli were generated in MATLAB (The MathWorks, Inc., Matick, MA) using Psychophysics Toolbox extensions. Stimuli were presented on a calibrated ASUS LCD monitor (120Hz, 1920×1200). 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 500Hz using an EyeLink 1000 (SR Research Ltd., Ontario, Canada).

Stimuli and task. The stimulus was a modified version of the classic Kanisza triangle. Six pacman discs (radius = 1°) were arranged at the tips of an imaginary star centred on a

fixation spot. The six tips of the star were equally spaced, and the distance from the centre of the star to the centre of each pacman 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 grey background (77.5 cd/m^2). 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, while the others were black, and for half the trials this was reversed. Inducers had a Weber contrast of .75.

We added Gaussian noise to the stimulus on each trial to measure classification images. Noise was 250 x 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 x 500 pixels, such that each randomly drawn luminance value occupied 2 x 2 pixels (.05° x .05°). The amplitude of these luminance values was then scaled to have an effective contrast of 0.125 on the display background, and were then added to the Kanizsa figure. Finally, a circular aperture was applied to the noise to ensure the edges of the inducers were equally spaced from the noise edge (**Fig. 1b**).

The jaw size of 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 was consistent with a triangle that was fatter or thinner 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 fat or thin was pseudorandomly assigned across trials such that an equal number of all trial types were included in each block. The relative jaw size of attended inducers was independent of the unattended inducers; thus, the identity of the non-target 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 unlimited duration to report the jaw size using a button press. The

next trial would immediately follow a response. Throughout the experiment, eye tracking was used to ensure 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 them to maintain fixation during stimulus presentation, and then 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 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 minutes including a forced break. Each observer completed 16 blocks for a total of 9984 trials, which took a total of approximately five hours duration 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 was required before starting the experimental blocks.

Support vector machine models. Support vector machine (SVM) classifiers were trained and tested in MATLAB. We generated (3) hypotheses by training SVM classifiers on images of the inducers *i*) alone, or with either *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 (fat/thin) exemplars with no noise. To generate hypotheses in the form of classification images, we used each of the classifiers to perform the same task as the human observers in the experiment (trials = 9984), with the exception that here the inducers were always consistent with an equilateral triangle; thus, classification was exclusively influenced by the noise in the image.

Data and statistical analysis. The 9984 noise images for a participant were separated according to perceptual report (“narrow” or “wide”). To collapse across inducer polarity, we

reflected the distribution of noise on trials in which the cued inducers were white (i.e., we inverted the sign). We also collapsed across upright and inverted cue conditions by spatially flipping the noise on inverted trials. The noise values were then summed within each report type. The difference of these summed images is the raw classification image. 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 summarises all three edges. Note that this is a conservative estimate of the classification image and any spurious structure will 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 bottom right panel of **Fig. 2a**). To ensure that these pixels were not contaminated by averaging of nearest-neighbour pixels during rotation, described above, we excluded the three pixels closest to the inner corners of the star. We conducted a one-sample, two-tailed Bayesian and Students' T-Test on these pixel values using JASP software (JASP Team, 2017).

Data availability. The data that support the findings of this study are available from the corresponding author upon request.

REFERENCES

1. Kanizsa, G. *Sci. Am.* **234**, 48–52 (1976).
2. Zhou, H., Friedman, H.S. & Heydt, von der, R. *The Journal of Neuroscience* **20**, 6594–6611 (2000).
3. Qiu, F.T., Sugihara, T. & Heydt, von der, R. *Nature Neuroscience* **10**, 1492–1499 (2007).
4. Heydt, von der, R., Macuda, T. & Qiu, F.T. *J Opt Soc Am A Opt Image Sci Vis* **22**, 2222–2229 (2005).
5. Heydt, von der, R., Peterhans, E. & Baumgartner, G. *Science* **224**, 1260–1262 (1984).
6. Mihalas, S., Dong, Y., Heydt, von der, R. & Niebur, E. *Proceedings of the National Academy of Sciences* **108**, 7583–7588 (2011).
7. Driver, J. & Baylis, G.C. *Cognitive Psychology* **31**, 248–306 (1996).
8. Wagemans, J. et al. *Psychol Bull* **138**, 1172–1217 (2012).
9. Tse, P.U. *Vision Research* **45**, 1095–1098 (2005).
10. Wannig, A., Stanisor, L. & Roelfsema, P.R. *Nature Neuroscience* **14**, 1243–1244 (2011).
11. Houtkamp, R., Spekreijse, H. & Roelfsema, P.R. *Percept Psychophys* **65**, 1136–1144 (2003).
12. Gold, J.M., Murray, R.F., Bennett, P.J. & Sekuler, A.B. *Current Biology* **10**, 663–666 (2000).
13. van Lier, R., Vergeer, M. & Anstis, S. *Current Biology* **19**, R323–4 (2009).
14. Poort, J. et al. *Neuron* **75**, 143–156 (2012).
15. Murray, M.M. et al. *Journal of Neuroscience* **22**, 5055–5073 (2002).
16. Stanley, D.A. & Rubin, N. *Neuron* **37**, 323–331 (2003).
17. Davis, G. & Driver, J. *Nature* **371**, 791–793 (1994).
18. Mattingley, J.B., Davis, G. & Driver, J. *Science* **275**, 671–674 (1997).

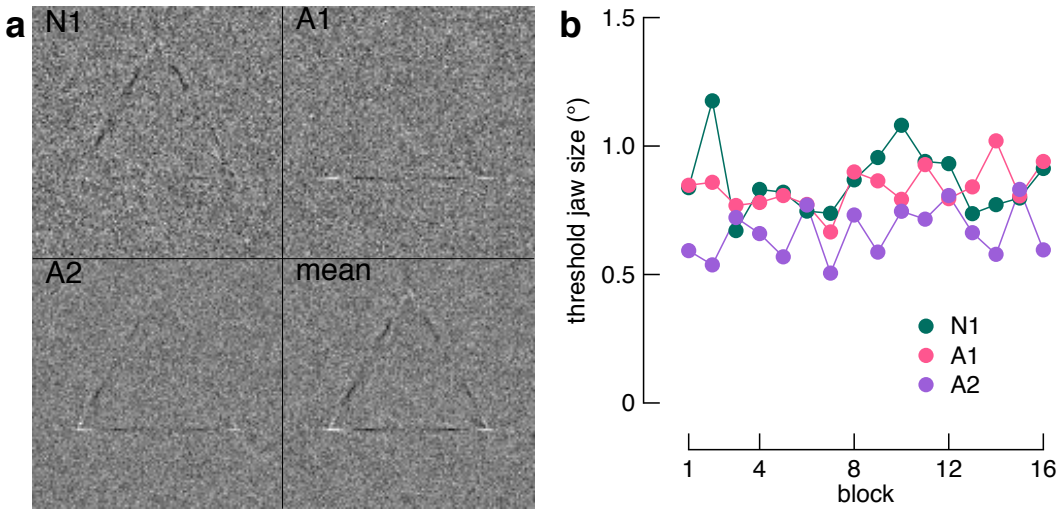
ACKNOWLEDGEMENTS

We are indebted to Peter Bex who developed the novel Kanizsa figure with us and provided helpful feedback on our study design and results. 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).

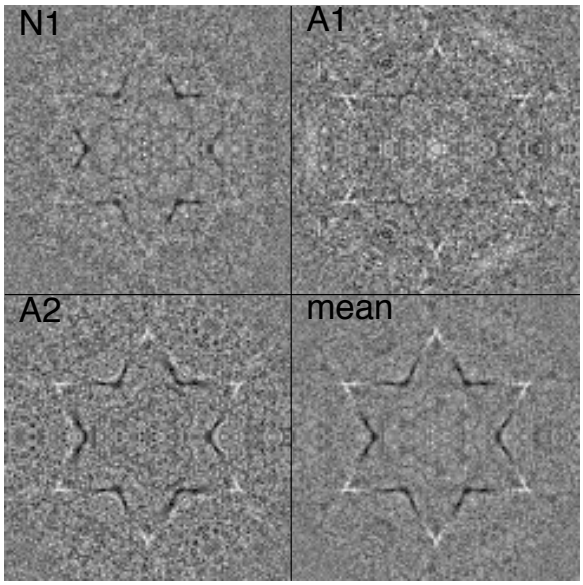
AUTHOR CONTRIBUTIONS

Both authors designed the experiment and collected the data. WJH analysed the experimental data, and RR performed the SVM analyses. Both authors contributed equally to the writing of the manuscript.

SUPPLEMENTARY FIGURES



Supplementary Figure 1. Raw classification images and psychophysical performance. a) Classification images without rotation of the image with itself. b) Threshold performance across blocks shown separately for each observer. Thresholds were the midpoint of a cumulative Gaussian fit to accuracy data for each session.



Supplementary Figure 2. Individual classification images revealing grouping of unattended structure. These images were created by summing each classification image with a flipped version of itself. Note that the emergent structure aligns to the geometry of the star implied by our Kanizsa figure (Fig. 1a).