



Research article

Demonstrating Invariant Encoding of Shapes Using A Matching Judgment Protocol

Ernest Greene^{1,*} and Michael J. Hautus²

¹ Department of Psychology, University of Southern California, Los Angeles, CA 90089, USA

² School of Psychology, The University of Auckland, Auckland 1010, New Zealand

* **Correspondence:** Email: egreene@usc.edu; Tel: +1-213-500-0169

Abstract: Many theories have been offered to explain how the visual system registers, encodes, and recognizes the shape of an object. Some of the most influential assume that border lines and edges activate neurons in primary visual cortex, and these neurons encode the orientation, curvature, and linear extent of the shape as elemental cues. The present work challenges that assumption by showing that well-spaced dots can serve as effective shape cues. The experimental tasks drew on an inventory of unknown two-dimensional shapes, each being constructed as dots that marked the outer boundary, like an outline contour. A given shape was randomly picked from the inventory and was displayed only once as a target. The target shape was quickly followed by a low-density comparison shape that was derived from the target (matching) or from a different shape (non-matching). The respondent's task was to provide a matching judgment, *i.e.*, deciding whether the comparison shape was the “same” or “different.” Clear evidence of non-chance decisions was found even when the matching shapes displayed only 5% of the number of dots in the target shapes. Visual encoding

mechanisms allow a shape to be identified when it is displayed at various locations on the retina, or with rotation or changes in size. A number of hierarchical network (connectionist) models have been developed to accomplish this encoding step, and these models appear especially credible because they are inspired by the anatomy and physiology of the visual system. The present work demonstrates above-chance translation, rotation, and size invariance for unknown shapes that were seen only once. This is clearly at odds with connectionist models that require extensive training before a shape can be identified irrespective of location, rotation, or size.

Keywords: shape encoding; translation invariance; rotation invariance; size invariance; model matching

1. Introduction

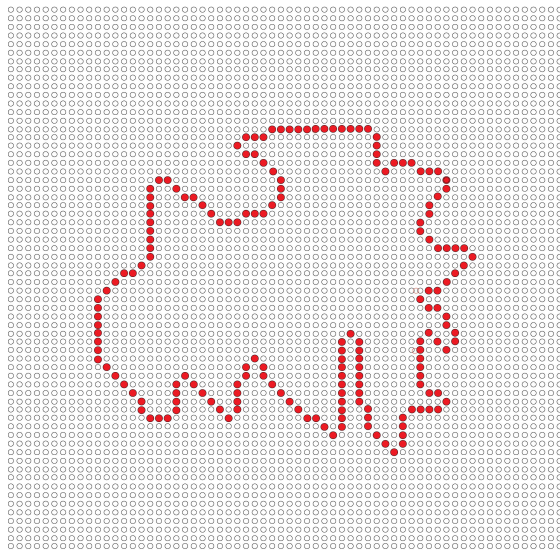
“(C)omplexity of the phenomena to be explained would soon require so complex a network of interconnections and interconnections between interconnections that the original theory would eventually become unrecognizable.” (Wolfgang Kohler [1] 1938).

The Gestalt psychologist Goldmeier E [2] provided insight into cognitive mechanisms for shape processing by presenting alternative stimulus patterns and asking for matching judgments. A great many of these stimuli were dot patterns, and he demonstrated that small changes in dot location, or the absence of just a few dots, could greatly change the matching decisions. These results indicated that the locations of individual dots were registered by the visual system and this information contributed to specifying a given shape.

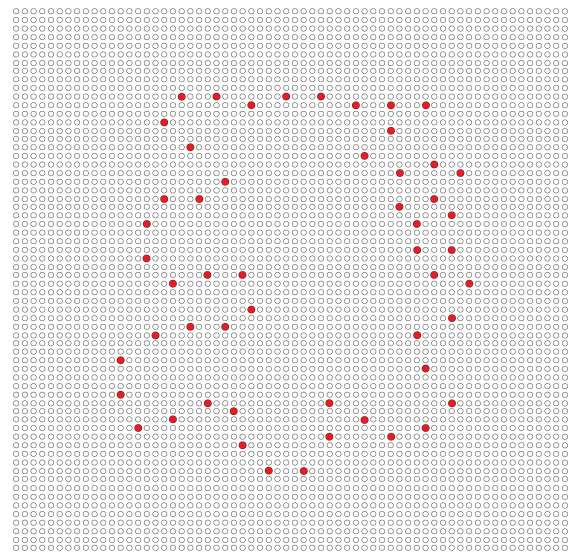
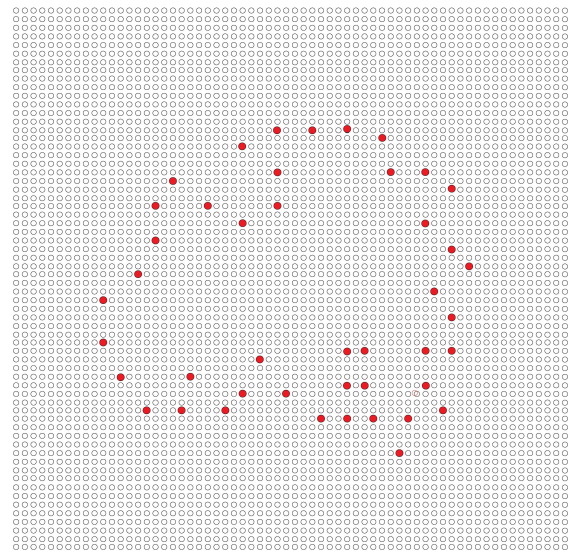
Previous work from this laboratory has also emphasized the role of discrete boundary locations as elemental shape cues. For example, Greene E [3] displayed the outline boundaries of common shapes as a string of dots. He then presented sparse subsets of the dots to determine what portion would be needed for recognition by respondents. Evenly spaced dots provided for recognition at the lowest dot density, with mean recognition-density being 18%. Almost a fifth of the shapes could be identified with densities below 10%.

Greene E & Visani A [4] examined recognition of “fat” letters that were randomly decimated to provide low-density stimuli, each being displayed with a 10-microsecond flash of all the sampled dots. All letters in the inventory could be identified when the stimulus provided an 18% density, and almost half could be named with a density of 10%.

Matching Shape



Target Shape



Non-matching Shape

Figure 1. Examples of shape displays. Each panel illustrates the 64 x 64 array of light-emitting diodes (LEDs) in the display board, with red dots showing the LEDs that were simultaneously flashed to display a given shape. The boundary of each target shape was displayed as a chain of adjacent dots that formed a single continuous loop, as illustrated in the left panel. The target shapes that were displayed to a given respondent were selected at random (without replacement) from the shape inventory, and each was shown only once. The matching shape, shown in the upper right panel, displayed a portion of the target dots, 25% in this example. The dots to be displayed were chosen by an algorithm that provided for even spacing. The lower right panel shows a non-matching shape, this being a

random selection (without replacement) among the shapes that were not used as targets.

Greene E [5] varied the density of “thin” letters, wherein the strokes forming each letter were represented by a single-file chain of adjacent dots. Here also, each letter pattern was shown for 10 microseconds. Recognition of the letters – large or small – was above chance at 3% density, and at 15% density the probability of recognition was in the 70% range.

It is interesting that such minimal boundary information can elicit recognition of shapes. Perhaps this is possible only because the shapes are well known. Letters are thoroughly over-learned, and recognition of known objects was possible at the lowest density for those that were especially common. It would be useful to examine the ability to encode and compare shapes that have not been stored in long-term memory, which can be described as “unknown” shapes.

The present experiments drew from an inventory of unknown shapes, each being rendered as a continuous string of dots forming an unbroken loop. An example is provided in Figure 1. A given shape was displayed as a “target shape” with an ultra-brief, simultaneous flash of all the dots. This was quickly followed by display of a “comparison shape” that was either a low-density version of the target shape or a low-density version of a different shape. These are designated as “matching” and “non-matching” shapes, respectively. Comparison shapes were varied in density of boundary dots, as illustrated in Figure 2. The respondent’s task was to say whether or not the comparison shape appeared similar to the target shape.

The three experiments also examined whether the encoding and comparison of shapes, each seen briefly only once, would allow for translation, size, and rotation invariance.

Experiment 1 displayed the target shape in one quadrant of the display board and the comparison shape in the opposite quadrant. Experiment 2 displayed the target shape in the center of the board and the comparison shape provided an enlarged version of the same pattern. Experiment 3 displayed the target shape and comparison shapes in opposite quadrants, but also with various amounts of rotation of the comparison shape.

All data were processed with signal detection theory methods that are designed to account and adjust for response bias. Overall results showed: a) substantial ability of the visual system to quickly encode and compare shapes without benefit of long-term memory; b) correct above-chance matching decisions even when the boundaries of comparison shapes displayed very few dots; c) above-chance performance wherein the experiments varied location, size, and rotation of comparison shapes in relation to target shapes. These results serve as a challenge to models that use hierarchical connectivity to achieve shape encoding.

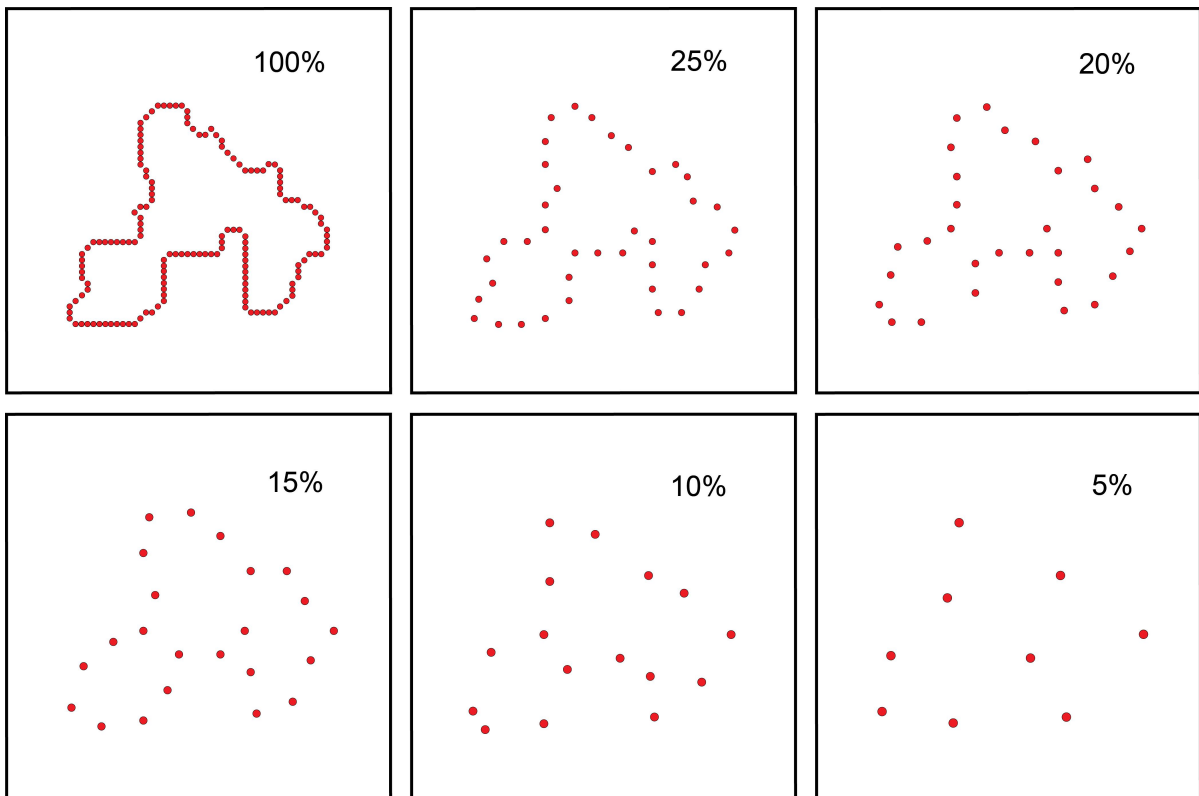


Figure 2. Illustration of dot-density treatment levels. The target shape (upper left panel) was always shown at 100% density. Matching and non-matching comparison shapes were displayed with five levels of density, from 5% to 25%. In each experiment the dots were flashed against a dark background and with low room illumination, so they were very conspicuous. Dots in the illustrations have been enlarged to partially compensate for the difference in perceptibility. Any perceived differences in area as a function of dot density are illusory.

2. Materials and Method

2.1. Research authorization and respondent participation

The research protocols were approved by the USC Institutional Review Board. Twenty-four respondents contributed data, eight different respondents for each of the experiments. Four were male and 20 were female. Ages ranged from 18 to 26 years, with mean age being 19.6 years. Before testing, each was informed that their participation was completely voluntary and that they could discontinue testing at any time. Each respondent was able to complete the experiment for which he or she was recruited and all data that was gathered once protocols were fixed has been included in statistical analysis.

2.2. Display equipment and implementation of protocols

Shapes were displayed with a 64x64 array of light-emitting diodes (LEDs), which is described as the display board. The stimulus shapes were provided by 10 μ s, simultaneous flashes of all the dots of a given shape. A respondent thus perceived a pattern of dots that surrounded a zone, which allowed perception of a shape. Respondents viewed the board from a distance of 3.5 m, so each dot subtended a visual angle of 4.92 arc'. Dot-to-dot spacing was 9.23 arc', and the total span of the display board at the viewing distance was 9.80 arc°.

Flash and fixation-point intensities were specified in radiometric units – microwatt per solid angle (μ W/sr). Ambient illumination of the test room was 10 lx, which is in the mesopic range. The LEDs have a peak wavelength of 630 nm (red) and the range of wavelengths is quite narrow, so it is likely that only L-opsin cones were stimulated.

Experimental protocols were implemented by Mac G4 Cube instructions, written as Tk/tcl custom applications. The computer instructions were further interpreted in machine language through a Propox MMnet101 microcontroller running at 16 MHz. This provided for control of flash timing with 1 μ s resolution.

2.3. Inventory of shapes

An inventory of 450 unknown shapes was created, each consisting of a sequence of dots that formed an outline boundary. Each dot in the sequence was adjacent to one dot on each side, such that one could transition from one dot to the next around the perimeter of the shape and return to where one started having encountered each dot in the shape only once. Each shape was “unknown” in the sense that there was a deliberate effort to provide arbitrary turns, arcs, and straight sequences, producing a boundary that would not resemble any known object.

The number of dots used to generate shapes ranged from 100 to 269, with the mean being 168 dots. Distance from the centroid to dots ranged from 12.8 to 22.3 dots, mean distance being 16.9 dots. At the viewing distance of 3.5 m, this corresponds to a range of 2.0 to 3.5 arc°, with the mean being 2.6 arc°.

2.4. Description of task

A matching judgment protocol was used in each of the experiments. Each trial of a given experiment required two successive displays, the first showing the target shape and the second showing the comparison shape. Target shapes were randomly chosen from the inventory, and each

respondent saw a given target shape only once. The comparison shape for a given trial was either a low-density version of the target shape, designated as a matching shape, or a random choice from among the shapes that were not chosen for display as targets (designated as a non-matching shape). The non-matching shapes were subject to the same density reductions as were the matching shapes. The respondent's task was to say whether the comparison shape matched the target shape, as described more completely below.

Across experiments, each trial was preceded by display of a fixation point, consisting of four dots that were at the center of the display board. Each dot of the fixation point provided steady emission of light at an intensity of $0.1 \mu\text{W}/\text{sr}$, and emission was quenched 200 ms prior to display of the target shape. Then the dots of the target shape were simultaneously activated for 10 μs , with intensity of the flashes being $1000 \mu\text{W}/\text{sr}$. The comparison shape was displayed 300 ms after the target shape using the same timing and intensity.

For Experiment 1, on each trial the target shape was presented in one quadrant of the display board, positioned so that the dots on one side of a given shape, and on the top or bottom, were displayed by LEDs on the outer edge of the quadrant (Figure 3). Which corner to use was chosen at random on each trial. The comparison shape was displayed in the quadrant that was on the opposite diagonal. For example, if the target shape was shown in the upper right corner of the board, the comparison shape was shown in the lower left corner.

Experiment 1 manipulated the density of comparison shapes across five levels, specifically 5%, 10%, 15%, 20% and 25% density. For a given trial the density the dots to be flashed, and the starting point for selecting which dots to include, were chosen at random by a method that also maximized the net spacing among dots.

For a given respondent, three hundred shapes from the inventory were chosen at random to be target shapes, with the remaining 150 shapes being used as the non-matching shapes. Thirty matching and 30 non-matching shapes were further assigned to each of the five density levels, so the assessment of density treatment effects for a given respondent were based on 30 matching trials and 30 non-matching trials -- a total of 300 trials.

Experiment 2 was designed to evaluate size invariance (Figure 4). Each target shape and comparison shape was centered on the board. Specifically, the addresses of boundary dots of a given comparison shape were used to calculate the location of the centroid, this being the mean of the X address values and the mean of the Y address values. The centroid of a given target shape was then positioned at the center of the display board, which therefore determined the locations of boundary dots.

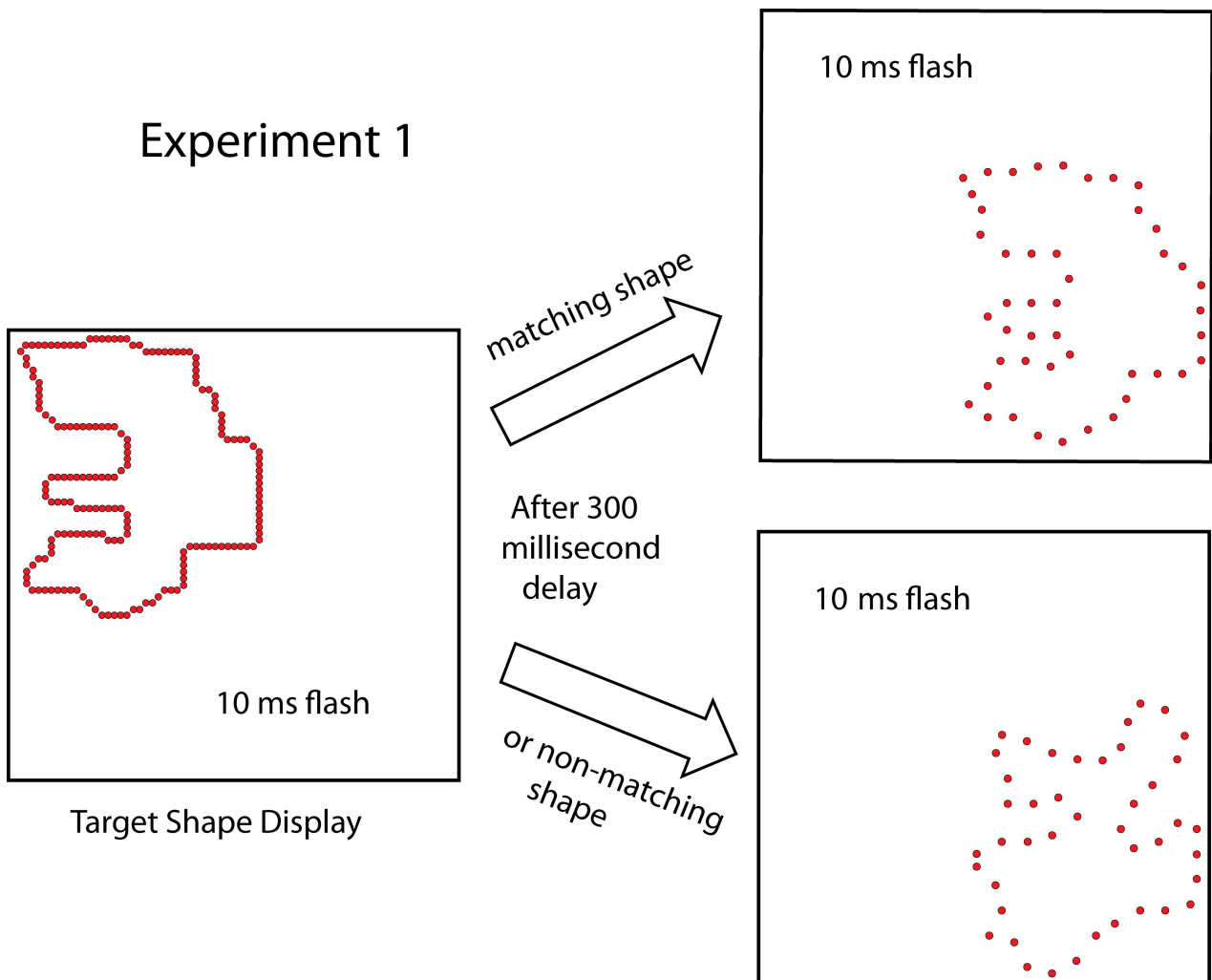


Figure 3. Evaluating translation invariance. In Experiment 1 the target shapes were displayed with a simultaneous 10 μ s flash from all the dots forming the boundary, requiring also that the shape be positioned in a corner of the display board, which corner being determined at random. Three hundred milliseconds later the low-density comparison shape was flashed for 10 μ s, with half the trials providing a matching shape and the other half providing a non-matching shape. A given comparison shape was displayed in the opposite corner from where the target shape had been shown, assuring that above-chance responding would reflect some degree of translation invariance in the processing of shape cues.

The matching shapes were enlarged versions of the target shapes, but at much lower densities. They were generated using the Experiment 1 treatment levels, these being 5%, 10%, 15%, 20% and 25% samples from the target shapes. Here also, the density and the starting point for dot selection was randomly chosen on each trial. Then for each sampled dot, the length of the vector from centroid

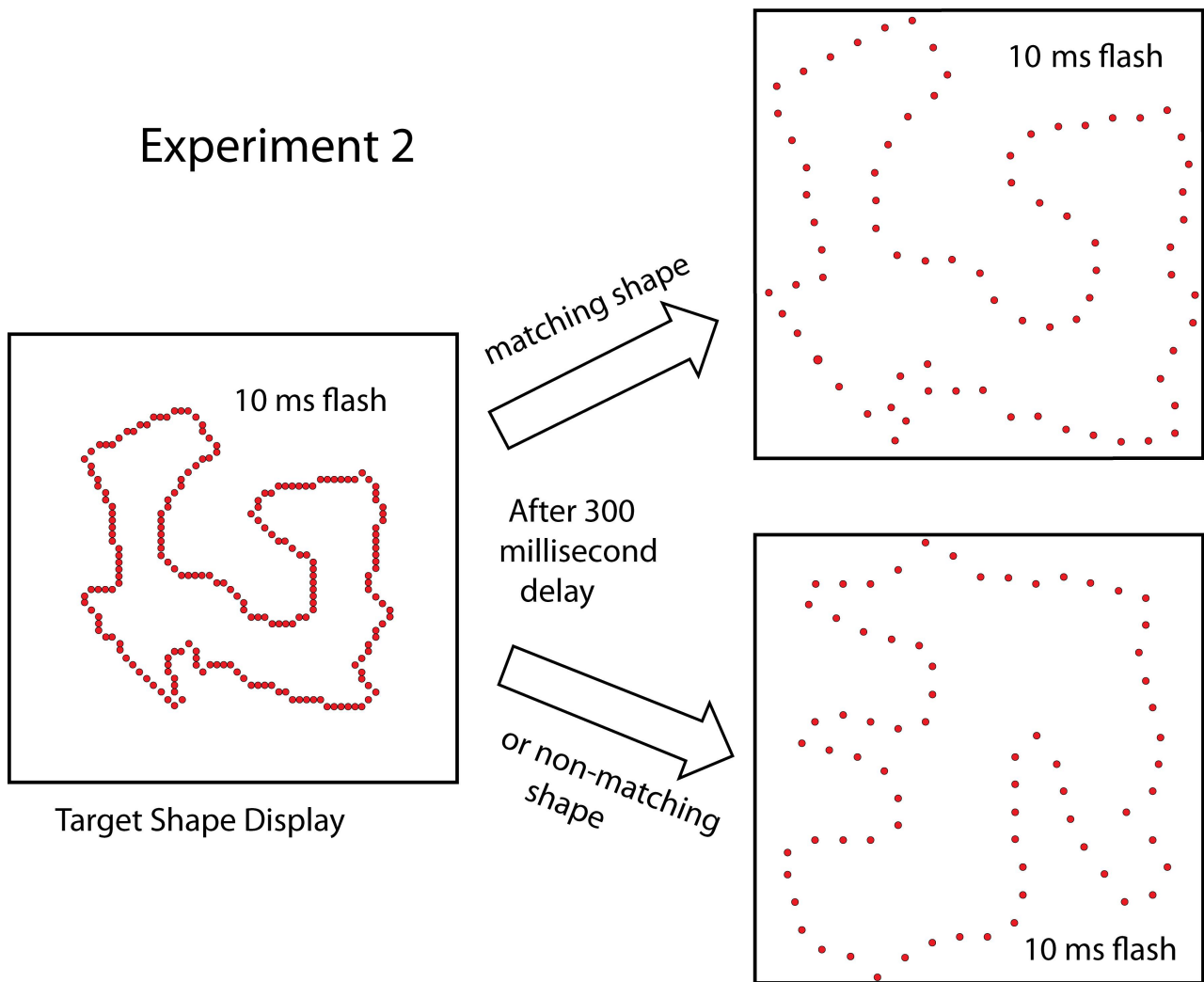


Figure 4. Evaluating size invariance. In Experiment 2 the target was centered on the display board. A given target shape was sampled to produce a low-density version, then the vector (distance and angle) from the centroid to each dot was multiplied by 1.4 to yield a resized version of the target shape, this being designated as the matching shape (upper right panel). The same steps were applied to non-target shapes to generate enlarged non-matching shapes. Above-chance responding on this task would reflect some degree of size invariance in the processing of shape cues.

to that dot was calculated using the Pythagorean theorem, and then this value was multiplied by 1.4. This yielded a longer vector, and the array dot that was closest to the tip of this vector was included in the boundary of the matching shape. This provided a matching shape that could be said to be 40% larger than the target shape, though this would not likely specify the difference in area. Non-matching shapes were created using the same method. As in Experiment 1, thirty matching and non-matching

shapes were presented to a given respondent at each of the five density levels, so (30 matching + 30 non-matching) x 5 densities = 300 total trials.

Experiment 3 was designed to evaluate rotation invariance (Figure 5). As in Experiment 1, target shapes and comparison shapes were displayed in diagonally opposite quadrants of the display board. All comparison shapes were displayed with a 15% density and the experimental variable was degree of rotation of the matching shape. On each trial, this required selecting dots in the target shape to derive a 15% subset, calculating the centroid to dot vector for each of the dots, then rotating the vectors and displaying the array dot that was closest to the tip of each vector. Rotations were applied only to matching shapes. Given that the orientation of the non-matching shape was irrelevant, there was no reason to apply rotation. Rotation levels varied from 0° to 180° in 30° increments.

Rotations were clockwise and counterclockwise in equal number, this being a covert variable that was not retained in data analysis. Twenty shapes were presented to a given respondent at each of the seven rotation levels, so (20 matching + 20 non-matching) x 7 rotations = 280 total trials.

2.5. Respondent judgments

Respondents were asked to maintain a steady gaze on the fixation point as the stimuli were being displayed. To initiate a given trial, the experimenter clicked an on-screen button that quenched the fixation point and then sequentially flashed the target shape and the comparison shape. The respondent voiced that the pair members were the “same” or that they were “different.” This was recorded by the experimenter using on-screen buttons on the computer monitor, after which the next trial was delivered. The experimenter had no information about which shape was delivered on a given trial, or whether the judgment of the respondent was or was not correct.

It should be emphasized that using the term “same” was for convenience. Each matching shape was a degraded version of the target, so clearly it was not identical. The goal was to see which manipulation, *i.e.*, density, location, size, rotation, would change the boundary structure to the point where it might not be judged as having the same overall shape as the target. This was explained to each respondent, and the experimental results affirm that the concept of “sameness” did serve to distinguish matching from non-matching shapes as well as provide consistent changes in the probability of a “same” judgment as a function of density.

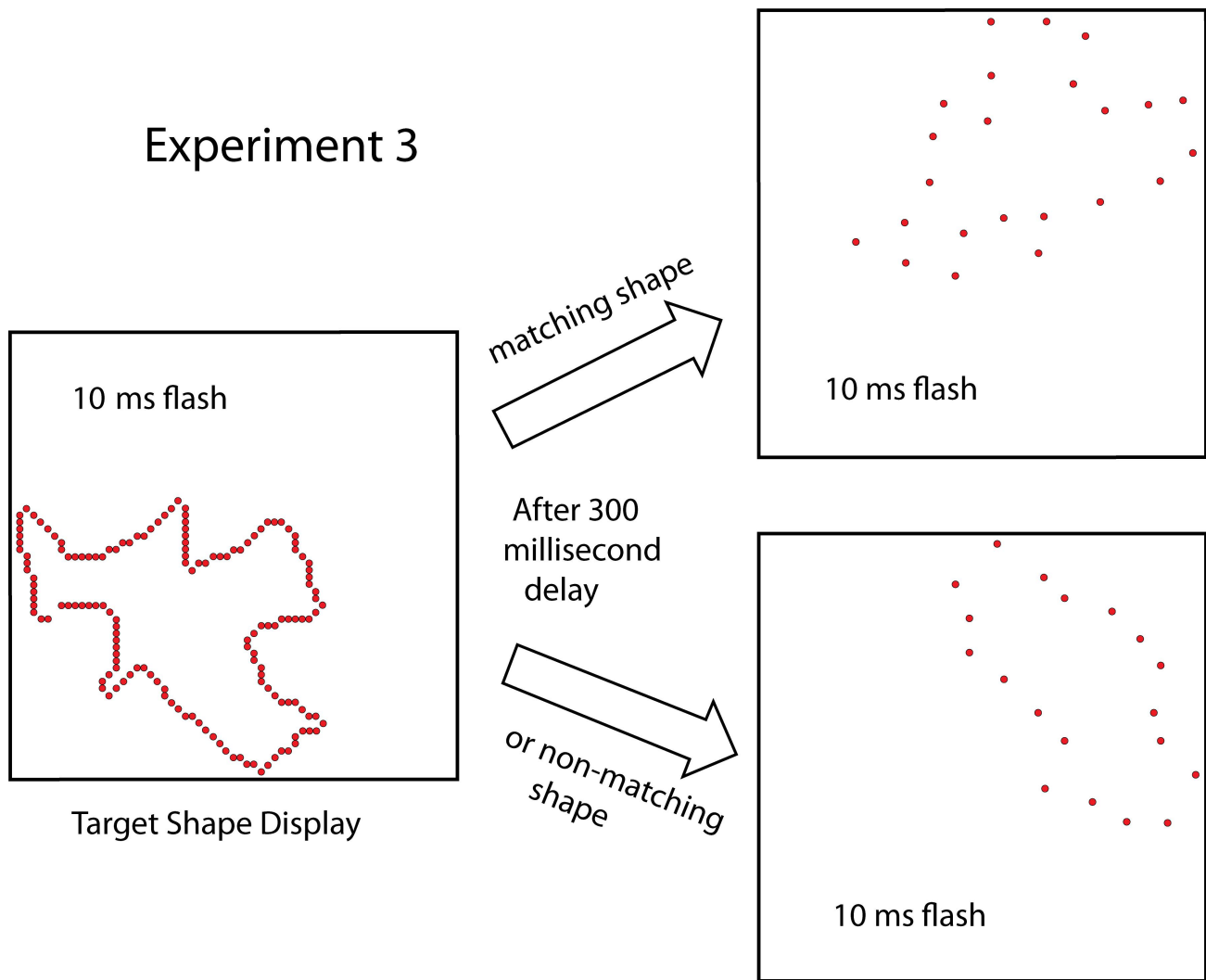


Figure 5. Evaluating rotation invariance. In Experiment 3 a given target shape was again displayed in a randomly selected corner of the display board and the comparison shape was displayed in the opposite corner. Display conditions were the same as in Experiment 1, except that all comparison shapes were shown at 15% density and the matching shape was submitted to clockwise or counterclockwise rotation for up to 180°. The matching shape shown here has been rotated by 60°. Above chance responding on this task would reflect some degree of rotation invariance in the processing of shape cues.

2.6. Statistical analysis

Response bias, this being the tendency for a respondent to prefer a response of “same” or “different”, can have a major impact on commonly-used measures of performance such as proportion

correct. Signal detection theory provides a framework to provide measures that are corrected for response bias. The most commonly reported bias-corrected measure is d' . What is frequently overlooked is that the structure of the task determines how d' should be calculated. Most researchers adopt the bias-correction formula $d' = z(H) - z(F)$ that is appropriate for a Green DM & Swets JA [6] detection-theory task, also called the yes-no task. This calculation is not appropriate for the matching-judgement task, which is procedurally equivalent to a yes-no reminder task. As detailed by Hautus MJ and associates [7], this kind of task allows for two decision strategies, these being the decision rules that the respondent applies to the evidence on a given trial. For the current research the difference decision strategy applies, which is described in the context of this task by Macmillan NA & Creelman CD [8]. We based our calculations on this strategy, for which the bias-correcting formula is $d' = \sqrt{2}(z(H) - z(F))$. In this formula, H is the proportion of “same” judgments to matching shapes (alternatively, the proportion of correct judgments to matching shapes), F is the proportion of “same” judgments to non-matching shapes (alternatively, one minus the proportion of correct judgments to non-matching shapes), and $z(\bullet)$ is the inverse-normal transform [7]. Values of F and H were adjusted for values of 0 or 1 (which would otherwise lead to $d' = \pm\infty$) prior to calculation of d' . We adopted the log-linear correction for this purpose [9]. Bias correction requires the combination of response information from both matching and non-matching shapes. Measures based on one or the other types of shapes are contaminated by response bias.

In addition to the bias-free measure of performance, d' , we calculated the variance associated with d' . The variance was determined directly from the binomial variation of the sampled proportions, namely the false-alarm and hit rates. A tabulation method can be employed, as demonstrated by Miller J [10] for the yes-no task. Similar calculations were performed for the present task, using SDT Assistant software [11]. The resulting variance associated with a particular estimate of d' was then employed to generate confidence intervals for that estimate, *e.g.*, $d' \pm 1.96\sqrt{\sigma^2}$, where σ^2 is the estimated variance.

While an estimate of d' and its variance are relatively easy to determine for the data from an individual respondent, there is also a correct method for summarizing performance across respondents. Data from different respondents should not be directly combined prior to bias correction, as this includes statistical bias into the resulting overall estimate of performance [8, 12]. Rather, d' for individual respondents was determined, as described above, and the resulting estimates of d' averaged to provide the group estimate. Average d' is statistically unbiased. Given that each of the N respondents has a d' estimate with an associated variance, the variance of their average d' is the sum of their individual variances divided by N^2 . This group variance estimate was used to develop

confidence intervals for group performance for a given condition.

It may be more intuitive to express performance as a proportion. Standard measures, such as proportion correct, $p(c)$, are influenced by response bias. However, $p(c)$ can be adjusted to remove this bias. It is no surprise that this adjustment requires the determination of d' in the first instance. The adjusted measure is called $p(c)_{\max}$ because it is also the maximum value of $p(c)$ obtainable for a particular level of sensitivity, d' . For the current task,

$$p(c)_{\max} = \Phi\left(d'/(2\sqrt{2})\right), \quad (1)$$

where $\Phi(\bullet)$ is the cumulative distribution function of the normal distribution. We transformed the estimates of d' and their confidence intervals to estimates of $p(c)_{\max}$ and their confidence intervals. The latter are typically asymmetrical. We primarily report our results in terms of $p(c)_{\max}$ values and 95% confidence intervals. The range of values for $p(c)_{\max}$ is from 0.5 for chance performance, to 1 for perfect performance. Note that the log-linear correction applied prior to the calculation of d' naturally carries through to the consequent value of $p(c)_{\max}$.

3. Results

3.1. Experiment 1

For Experiment 1 respondents received the added information that each comparison shape would be presented at a different location than the target shape. Based on these instructions as well as initial display trials, they had to assess whether or not the comparison shape matched the target shape irrespective of the spatial displacement. Each respondent provided judgments at each level of density yielding false-alarm (F) and hit (H) rates, each being based on 30 trials. Each (F , H) pair was converted to an estimate of d' and its associated variance using the tabulation method described above. Each d' value combines information from judgments about matching and non-matching shapes to provide a bias-corrected measure of performance. We do not present these individual-respondent results directly. However, all 40 d' estimates were positive and 34 of the 95% confidence intervals had positive lower-bounds (thereby excluding a d' of zero from the interval). Thus 85% of the matching judgments by individual respondents produced performance levels significantly above chance. Even at the 5% density level, half of the respondents demonstrated performance significantly above chance.

Individual d' estimates were then averaged across respondents for each level of density and the

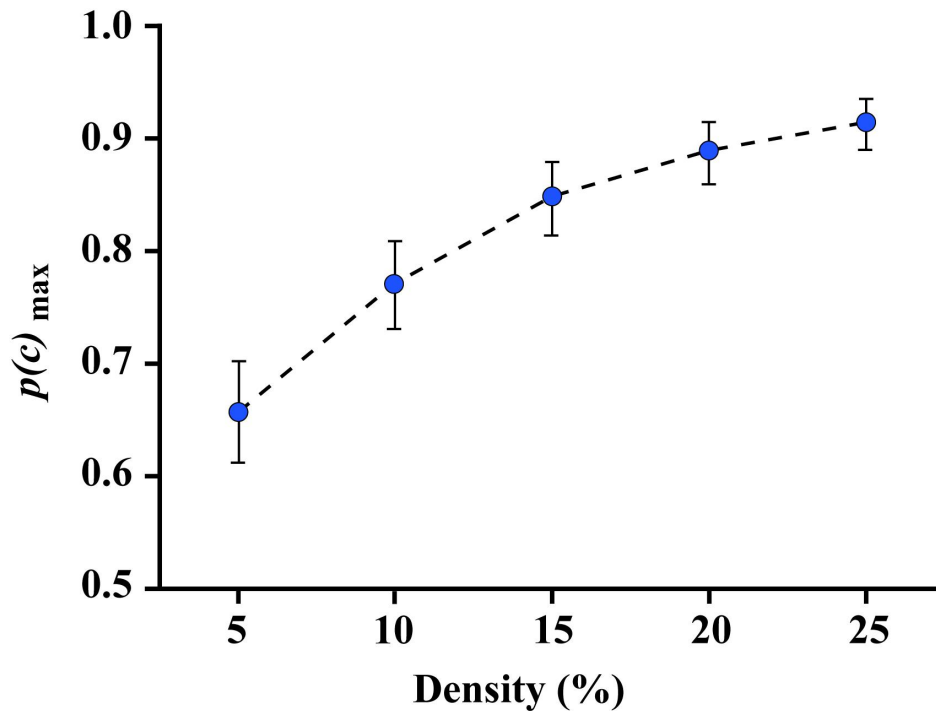


Figure 6. Judgments of spatially translated shapes remain above chance as dot density is reduced. Group means for Experiment 1 are plotted at each level of dot density, expressed in terms of the bias-free index $p(c)_{\max}$ and including 95% confidence intervals. The performance at every level of density was well above chance, affirming that unknown shapes can be quickly encoded and matched even when minimal boundary information is provided. Further, the results indicate that the comparison mechanism provides for translation invariance.

corresponding variance calculated. The variance was used to generate 95% confidence intervals around the estimate of d' . These d' estimates and their confidence intervals were then transformed to their equivalent proportional representation, $p(c)_{\max}$, using Equation 1. It is again worth noting that on the $p(c)_{\max}$ scale, a score of 0.50 represents chance decisions, and a score of 1 indicates judgments that were always correct, with the full range having been corrected for response bias.

The mean $p(c)_{\max}$ values across respondents, along with 95% confidence intervals are plotted in Figure 6. Performance is well above chance even at the lowest dot density. Each of the task conditions differed from chance performance at a significance level of $p < 0.001$.

For the higher density conditions, the curve begins to asymptote, perhaps at a level below $p(c)_{\max} = 1$. If true, this may relate to ceiling effects caused by the relatively small number of trials (60) undertaken by each respondent at each level of density.

Experiment 1 demonstrates that unknown shapes can be encoded and compared within moments,

making it possible for render valid matching decisions within a few seconds. Further, it demonstrates that a matching shape that consists of a very sparse set of boundary dots can provide above-chance performance even when only 5% of the boundary is displayed. This clearly has implications for the nature of encoding and comparison mechanisms, though the exact nature of those operations is not yet known. The experiment also shows that the visual system can compare the shapes even though they were displayed at different locations on the retina. Each target shape was seen only once, so the translation invariance does not depend on comparison to memory stores.

3.2. Experiment 2

In Experiment 2 each dot of a comparison shape was repositioned to be at the tip of an elongated vector, thus providing an enlarged version as the comparison shape. The basic characteristics of the data collected were the same as Experiment 1 and an identical analysis was undertaken. For individual respondent data, all 40 estimates of d' had positive lower-bounds for their 95% confidence intervals. Each of the task conditions differed from chance performance at a significance level of $p < 0.001$. Group means, in terms of $p(c)_{\max}$, are plotted in Figure 7.

Shape-matching performance in Experiment 2 was even higher than in Experiment 1 at every level of dot density, significantly so at the three lowest densities where the 95% confidence intervals did not overlap.

One might infer that the ability to provide useful shape cues given this spatial manipulation of dot location suggests that a shape summary is based on the location of dots in relation to the centroid of the dot pattern. In other words, one might hypothesize that dot patterns are summarized by converging the angle and distance of each dot relative to the centroid of the pattern, and this summary allows for normalization of values or other means to compensate for distances. How else can one explain size invariance, and in particular, a size-adjusting mechanism that allows very sparse boundary markers to be compared to a target shape?

3.3. Experiment 3

In Experiment 3 all the dots of comparison shapes were at 15% density, and the matching options were repositioned by rotation around the centroid of the shape. The data structure was slightly different from the first two experiments. Each respondent completed only 40 trials at each angle of rotation, yielding F and H values based on 20 trials. There were seven angles of rotation, for a total of 280 trials – slightly below the 300 trials that were provided in previous experiments.

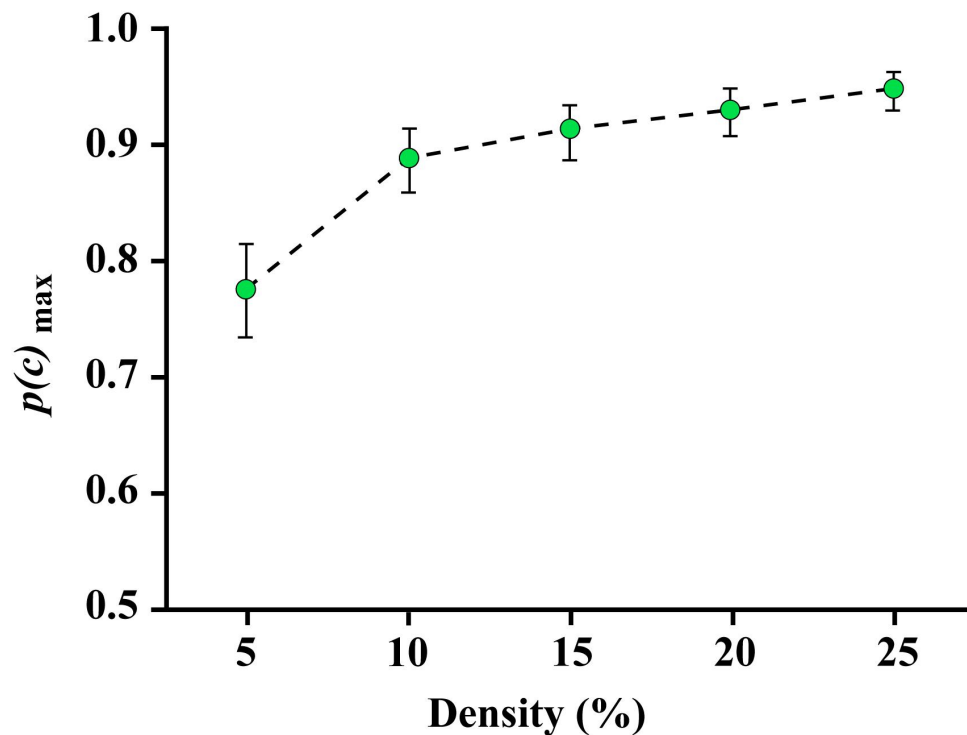


Figure 7. Judgments of resized shapes remain above chance as dot density is reduced. Matching judgments for the group have been plotted for each level of dot density in Experiment 2, expressed in terms of the bias-free index $p(c)_{\max}$ and including 95% confidence intervals. For this experiment the matching shape was enlarged relative to the target shape, yet $p(c)_{\max}$ values were well above chance. This affirms that the shape-comparison mechanism provides for size invariance even when the shape information is not stored in long-term memory.

Performance was generally lower in this experiment. Nonetheless, of the 56 estimates of d' , all but two were above zero. Further, 42 (75%) of the d' estimates had positive lower-bounds for their 95% confidence intervals. Group means, expressed in terms of $p(c)_{\max}$, are plotted in Figure 8.

The zero-degree condition of Experiment 3 provided the matching shape at the same orientation as the target shape, so finding that this condition provided the highest $p(c)_{\max}$ score confirmed our expectations. The score was lower than was found for the 15% density condition of Experiment 1, *i.e.*, 0.75 vs 0.83. Nonetheless, there was overlap of the confidence interval for the two experiments, so these results might be viewed as a successful replication of effects.

The plot in Figure 8 shows above-chance matching judgments at each of the rotation angles. The profile of $p(c)_{\max}$ changes as a function of angle, and additional work will be required to explain those

differences. Nonetheless, the basic finding of above-chance decisions at every angle calls for an explanation for how the repositioned dots could be seen as matching the target shape. The rotation of dot position relative to the centroid did not eliminate the ability to judge the marked locations as being derived from the target shape. This argues for generation of a shape summary that reflects the locations of dots relative to the centroid of the shape, as well as a mechanism for modifying angular relations within the summary.

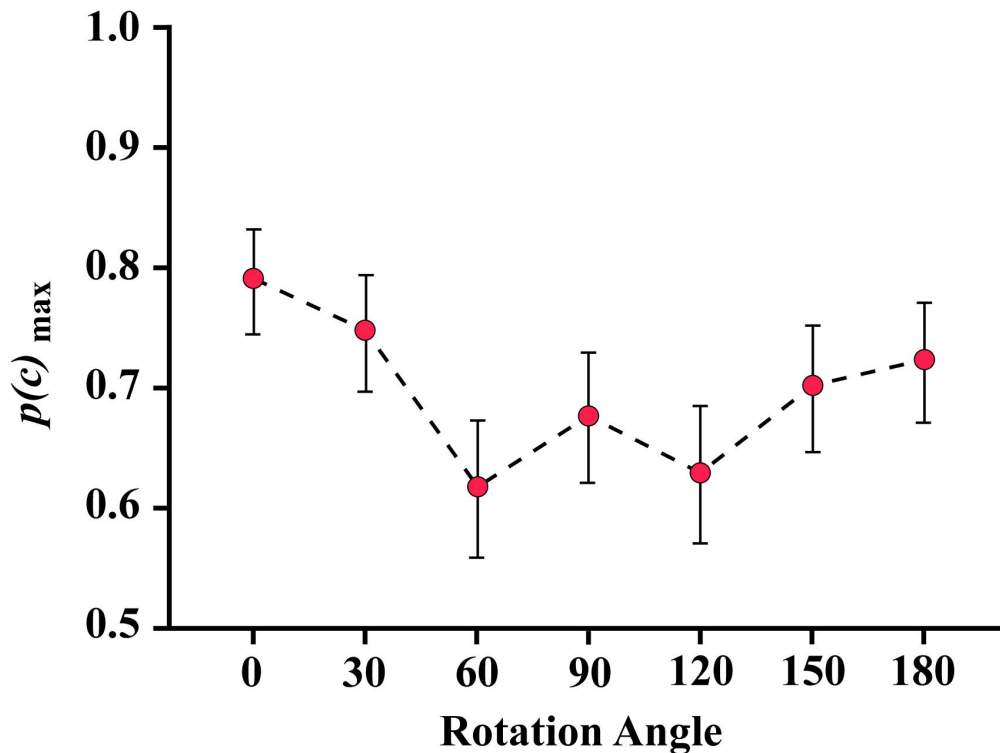


Figure 8. Judgments of comparison shapes remain above chance across all rotations. Estimates of performance for the group have been plotted at each angle of rotation in Experiment 3, expressed in terms of the bias-free index $p(c)_{max}$ and including 95% confidence intervals. All mean performance was well above chance, affirming that the shape-comparison mechanism is rotation invariant even when none of the shape information is being held in long-term memory.

4. Discussion

We have tried to be circumspect in describing the task without using the word “recognition.” At this point, however, it would be good to affirm that the perceptual and working memory mechanisms that mediate those judgments can also be described as providing for shape recognition. The term “recognition” is most often used in discussing access of information being held in long-term memory.

Here the target shape is being retained in working memory for only a second or two -- the time needed to register the comparison shape and decide if it is a match to the target shape. However, the basic encoding operations are likely the same as for naming a known shape, so at this point we see no special benefit in formulating new terminology to avoid describing the process as recognition.

The present results provide a challenge to widely held theories for how shapes are encoded and compared to achieve recognition. Our concern is with neuronal substrates, and the mechanisms by which neurons mediate this process are most clearly articulated by those who have advanced formal computational models. The focus of this discussion, therefore, will be on evaluating the assumptions and requirements of formal models, citing only a few that have received the most attention and endorsement by psychologists, neuroscientists, and machine-vision engineers.

A cornerstone principle that is common to these models is that edges and line segments serve as elemental shape cues, and one of the earliest operations of a shape-encoding system is to register those elemental cues. Seminal evidence and concepts have been provided by Hubel DH & Wiesel TN [13, 14], Selfridge OG [15], Sutherland NS [16], Binford TO [17], Barlow HB [18], Milner PM [19], Palmer SE [20], and Marr D [21].

A common view is that early and essential shape encoding is accomplished by orientation-selective (simple) cells of primary visual cortex. Many have adopted an alternative shape primitive, this being some form of spatial frequency grating, such as a Gabor patch [22-28]. Either way, a major goal is to encode the orientation of contours, so for simplicity we will refer to the filters at this level of processing as orientation-selective cells, whether observed, implemented as computational steps, or as a math formalism.

A substantial amount of modeling has built on the early computational concepts, much of it focused on how to improve filtering principles to deal with complexities in the visual scene. Edges are often indistinct, shadows can provide false boundaries, and the system needs to know which lines and edges belong to which object. We can identify 2D and 3D shapes in the presence of motion, or when seen at various sizes, or with rotation. There are substantial challenges to be overcome in modeling and/or providing algorithms that can effectively accomplish these goals.

A number of the recent advances have provided better ways to reach effective decisions. For example, Edelman S [29] evaluated similarity to provide for better categorization of novel 3D objects. Cooke and associates [30] followed up with multidimensional scaling techniques that allowed comparison of what cues were more important for visual and tactile perception. Hayworth KJ [31] focused on filter characteristics that would accomplish binding the features of an object. Rodriguez-Sanchez and associates [32] proposed multi-cue filter attributes that would register the curvature of a contour. Hopfield JJ [33, 34] and Maass W [35] proposed stimulus encoding methods

that use time-to-fire rather than rate of firing as a way to speed up the encoding process and explain short-latency responses of the nervous system. McClelland JL [36] provides a model that implements Bayesian logic to better explain human efficiency in letter and word recognition. Guan and associates [37-39] have developed algorithms that make better use of depth, color, and other shape cues, which are especially efficient for doing scene analysis on mobile devices.

Some of the modeling effort has been directed to explain how the visual system can make use of global information cues such as fragmented contours. Ullman S [40] was the first to provide a formal model for interpolating among contour fragments. A series of studies followed improved on the basic methods [41-45] and the Pizlo group [46] has provided useful background information about this approach, in addition to formulating a novel means for segregating boundary fragments from background.

Many of the formal models are combinatorial, wherein recognition of a shape is based or has a substantial requirement for enumerating shape features. It has been noted that combinatorial models are unworkable because the astronomical number of shapes that we can identify irrespective of retinal location, rotation, and differential size. This has been called a “combinatorial explosion” [48]. We agree with this criticism, but want to focus on other problems for which our data are more relevant.

Our main concern is with biological plausibility of the models. We object to the proposal that the activity of orientation-selective cells provides elemental shape cues, or more specifically, that orientation, curvature, and linear extent (length) of the stimulus contours are essential to recognition. The judgments made in the present experiments were based on shape information that was provided by discrete dots that were spaced with a density as low as 5%. There is no evidence that activation of orientation-selective neurons by a sequence of widely spaced dots will deliver information about orientation, curvature, or linear extent. Models that use biologically realistic orientation-selective cells to register contours would not be activated by a sparse complement of dots.

The Shapley laboratory [47] used drifting sinusoidal gratings to activate orientation-selective neurons in primary visual cortex of Macaque monkeys, mapping specifically for receptive field length. We don't have comparable information for humans, but it is generally thought that at this stage of visual processing the neuroanatomy is quite similar. They found that the excitatory field for 30 of the 31 observed neurons was less than 2.5 arc° in length. Greene E [3] examined recognition of a wide range of known shapes, *e.g.*, animals, plants, vehicles, furniture, tools. These were displayed with dots, including a condition that provided substantial spacing of adjacent dot locations. He reported substantial levels of recognition even when the separation of all dots was greater than 2.5 arc°. This would provide no more than one dot as the stimulus to an orientation-selective cell, which means that it could not generate any orientation, curvature, or length information.

A follow-up study [49] found that dot-strings designed to activate orientation-selective cells were no more effective at eliciting recognition than an equivalent number of dots displayed at random locations. This reinforces the argument that activation of orientation-selective cells is not an essential first step in the shape-encoding process.

The present results demonstrate that the visual system can derive shape information from a sparse complement of boundary markers, these being discrete dots that were briefly flashed. That information allows the respondent to determine whether the implied shape was derived from the contours of an unknown shape that was flashed only moments before. The individual dots did not provide any information about orientation and it seems unlikely that the visual system “regenerated” the missing contours.

A simple thought experiment, discussed more thoroughly elsewhere [50], should help illustrate the point. Three dots that are equidistant from one another will be perceived as having a triangular shape. They will provide a triangular pattern even if the dots are spaced more than 2.5 arc° apart, and the perception is consistent whether one is looking directly at the middle or if the middle of the dots is a bit eccentric to the fixation point. Note that when the middle is somewhat eccentric to fixation, two dots will stimulate the hemiretina on one side of the fovea and the third dot will stimulate the other hemiretina. Under these conditions the crossed and uncrossed projections from retina to primary visual cortex carry the stimulation to different hemispheres. One hemisphere will register two dots and the other hemisphere will register the third dot.

The only information being provided to the hemisphere that registers the single dot is its location within the visual field. The dot does not provide any cues about orientation, length, or curvature. Even for the hemisphere receiving two dots, those cues would not be elicited if the dots are separated by more than the length of the largest receptive field of orientation-selective cells. The three-dot pattern is not providing any stimulation that would allow the orientation, length, or curvature of all dot pairs to be registered, yet the respondent will see the pattern as forming a triangle.

One might quarrel that the three-dot pattern is not really a shape, and the demonstration is not relevant to recognition of continuously bounded 2D or 3D forms. But then one must reject the many studies of fragmented figure completion, or illusory figures, as being irrelevant to the discussion of shape perception. If one displays the boundaries of the triangle with a string of dots and then progressively reduces the density, a respondent will identify each version as a triangle until the depletion leaves only two dots. We see no meaningful distinction that can be made between a sparse pattern of dots that implies the shape and continuous contours that also elicit the perception of a triangle.

The point of the thought experiment is that registering the locations at which boundary markers

lie is a necessary and sufficient basis for perceiving shape. Which brings us to another fundamental issue that is universally ignored by computational modelers. All formal models that deal with non-local shape information, including translation, rotation, and size invariance assume that the location of a stimulus within the image can be specified using address values. The most common coordinate system is Cartesian, wherein a given location in the image plane can be designated with an x value for row and y value for column. A polar coordinate system, where the address values specify angle and distance, is also popular. Mathematics is a well-developed discipline that has formulated a vast array of methods for manipulating address information, and the use of mathematical formalisms as computational tools has proven to be amazingly effective for image processing, and more specifically for effecting shape and object recognition. The assumption of models has been that the stimulated locations could be specified within a coordinate system, and these values could then be manipulated using any and all mathematical tools at one's disposal. But are the assumptions biologically plausible?

The problem is that we have no basis for thinking that a given neuron within a population "knows" its location and can inform others of that location by delivering address values. It is not sufficient to invoke evidence that genetics and developmental mechanisms have provided precise positioning of specific cell types. The fact that the anatomy is precisely constructed doesn't mean that the cell can generate a signal that informs others of its location. Upon being stimulated by two dots that lie at some distance from one another, how can neurons within the population deliver information about their absolute and relative locations, as required by the formal models?

Some engineers who are working on neuromorphic models in machine vision have solved this problem with "smart retina" circuitry. For example, the Delbruck lab describes a camera array that will respond to a local light transition by sending the address of the activated pixel to a data bus [51]. But neurophysiologists have not provided any evidence that a stimulated neuron will generate a signal that says where it lies within the population. We have minimal basis for thinking that a discrete local stimulus, *e.g.*, a flash dot, can deliver address values.

Does the ability of the visual system to successfully execute accurate saccadic eye movement provide concepts for how the visual system could bypass the need for addresses? For express saccades one can reasonably assume specific connectivity between the sensor array and the motor array to accomplish the eye movement [52, 53]. But with voluntary saccades the respondent can derive vectors that cannot be readily explained by neuronal connections. For example, a brief stimulus can be flashed at some distance from fixation and the respondent can be instructed to look to a location that is twice that distance. The respondent can successfully comply. Or the instruction can be to look to a location that is at the same distance but in the opposite direction. Again, the

respondent can do so. Or the instruction can be asked to look to a location that is at right angles to the direction at which the flash occurred, or at right angles but using twice the distance. Any number of behavioral tests can demonstrate generation of voluntary saccades as though the addresses of stimulated locations were being provided.

A two-pulse saccade protocol [54, 55], can provide even more striking examples of voluntary control that cannot be readily explained without assuming address values. Two dots can be flashed sequentially, one to the right of fixation and then one above fixation, both being delivered and then vanishing before the eye has an opportunity to move. The respondent is instructed to saccade to each flash, first to the right and then to the one above the original fixation. The second saccade, from the location of the first flash to the second, is a diagonal vector. The visual system can successfully execute that vector as though it had been calculated from addresses within a Cartesian coordinate system, that calculation being known as the Pythagorean Theorem. Simple math, but how did the neurons accomplish this task without the benefit of addresses? We repeat the criticism that formal computational models assume addresses without having any evidence that neuron populations can generate and signal address values.

Returning to shape encoding concepts, it appears that most modelers think that the ability to do precise mapping of connections from one population to another solves the problem. The information passes from one population of neurons to another through anatomical connections, and selectivity of response for a given shape is developed by forming new connections or altering the strength of connections within the network [56]. The connectionist models essentially bypass the address issue by asking the system to connect in ways that register shape attributes irrespective of their spatial locations.

The Neocognitron model by Fukushima K [57] has been extremely influential, as well as the VisNet model developed by Rolls ET [58, 59]. The ability to provide location, rotation, and size invariance has been an important consideration for plausibility of the modeling. Fukushima K [57] developed the basic mechanism by which this would be accomplished. He achieved invariance by using multiple layers of neurons, having widespread interconnectivity of one layer to the next, tailoring the influence of connections to provide selectivity of response, and then delivering numerous training trials with the stimulus being varied across the dimension that should be ignored, *e.g.*, spatial location. Refinements to this approach have been proposed by Rodriguez-Sanchez AJ & Tsotsos JK [32], Riesenhuber M & Poggio T [60], Pasupathy A & Connor CE [61], Suzuki *et al.* [62], and Pinto and associates [63].

It is no small consideration that connectionist models require multiple (usually many) training trials. Combining information from various image locations depends on changing how the neurons

are connected or producing differential strength of the connections across several cell layers. This is brought about by training trials, with each trial altering the strength or positioning of countless connections.

It should be noted that even recognition of a shape at a fixed location using a connectionist model requires training trials. It is assumed that the proper linkage for identifying a given shape is not preprogrammed. One must provide an example of each shape that is to be discriminated, and connectivity among the neurons must be progressively adjusted to discriminate among the alternative shapes. So even the simplest neural network is designed to have widespread connections that are to be modified by training so that a particular combination of elemental line segments will be recognized as a specific shape. The encoding process itself, *i.e.*, the operations that would distinguish one shape from another, are based on creating changes in connectivity among the neurons, which provides for storage. Unless the shape information is stored, it is not encoded, *i.e.*, summarized. The connectionist models require numerous training trials before shapes can be discriminated.

The present work demonstrates that a matching shape can be judged as similar to an unknown target shape that was displayed only once affirms that training trials are not needed by the visual system. The displays were ultra-brief and responses were given within a second or two. This indicates that encoding is done quickly, perhaps within tens of milliseconds. Masking studies show that a mask becomes progressively less effective at blocking shape perception, with the impairment being minimal in roughly one hundred milliseconds [64].

It is also very significant that the visual system can accomplish shape encoding and recognition even when the unknown shape to be compared is placed at a different location than the target, or is enlarged, or is rotated. It should be obvious that the experimental results are not an anomaly produced by use of low-density comparison shapes. If the matching and non-matching shapes had been displayed at 100% density, the decisions would have been far more accurate. Indeed, one reason for adjusting density was to reduce the amount of shape information being delivered, which avoids ceiling effects and provides balanced variance estimates that support hypothesis testing.

A final objection to the formal models is the extreme emphasis on cortical anatomy and physiology for explaining basic shape perception. A large majority of the published articles are focused on the neuronal populations that are present in primate brains. This allows the model to be structured with numerous cell layers, which makes it easier to demonstrate effective recognition along with translation, size, and rotation invariance. It would be difficult for a connectionist model to manifest this encoding flexibility if the model was restricted to just two or three cell layers. Yet shape and pattern discrimination has been demonstrated in animals that have much smaller brains.

Fish can identify shapes and patterns. Damselfish can discriminate between predators and

non-predators from the shape of the mouth and the distance between the eyes [65]. Siebeck and associates [66] found that discrimination of facial patterns was based on shape, not color. Juvenile coral fish do not recognize predators whereas adults do, which indicates that the discrimination must be learned [67]. Damselfish can learn to discriminate various kinds of 3D stimuli [68]. Newport and associates [69] trained human faces as positive and negative stimulus cues and then tested the fish for discrimination of these faces in relation to forty-four novel faces. All fish reached peak levels of discrimination in the 77% to 89% range. They could successfully discriminate even when faces were equated for color and brightness, and with auxiliary cues being excluded by using an oval mask that showed only the central features of the human face. The processing of shape cues is likely being done in the retina and/or in the optic tectum.

Aside from formal experimental findings, it is inconceivable that countless generations of fish could have successfully lived and reproduced over millions of years without being able to encode and identify the diverse shapes that they encountered. Effective interaction with their environment would require that they quickly register the shapes of predators, prey, and conspecifics. Survival would be best assured if they could register and store the shapes and positioning of territorial landmarks. These shape discriminations would be done without benefit of the large neuronal populations of the primate brain.

We acknowledge that the visual skills of primates (including humans) are superior to fish, and the need to identify objects in the face of clutter and other sources of image degradation makes far greater demands for complex neuronal mechanisms. But here it appears that the connectionist models require multiple layers and extensive training because the encoding principles are so inherently unwieldy that simpler models are unable to provide effective discrimination even for basic shapes and patterns.

At present, we have no evidence that the neurons in an array can deliver information about their location by using address values. However, it is not inconceivable that the functional equivalent of distance and angle information can be derived through anatomical and physiological mechanisms. There is exciting work being done using large-array electrodes that can monitor interactions taking place within a population [70, 71]. It is possible that spreading waves of activity within the retina or optic tectum could register relative distances among boundary markers, or the distance from those markers to a centroid. This could provide the information needed to generate a shape summary, as has been previously suggested [3, 50, 72, 73].

5. Conclusion

The present experiments demonstrate that the visual system can quickly encode and compare shapes that have not been previously stored in memory. Here the shape information was provided by

discrete dots that marked boundary locations. Bias-correcting data analysis found that judgments were above-chance even when the shapes being compared displayed a small fraction of the total number of boundary markers. Further, the decisions were well above chance even when the low-density comparison shape was displayed at a different location from the initial target shape, or when it was enlarged, or when it was rotated in relation to the target. These findings are a challenge to formal models for how the nervous system processes shape information.

Acknowledgments

David Nyberg designed the circuitry of the display system. Jack Morrison of Digital Insight fabricated the system, wrote the firmware and software required for its operation, calibrated emission intensity and duration, and contributed in numerous other ways. Respondents were tested by Hannah Nordberg, Alyssa Brill, Denisse Carrasco, and Onyinye Onwuzulike. This research was supported by the Neuropsychology Foundation and the Quest for Truth Foundation.

Conflict of Interest

All authors declare that there are no conflicts of interest in this paper.

References

1. Kohler W (1938) *Gestaltprobleme und Anfänge einer Gestalttheorie*. Jahresbericht über der Gesellschaft Physiologie. Translated by: Ellis WD, *A Source Book of Gestalt Psychology*, London: Routledge & Kegan Paul: 58.
2. Goldmeier E (1936/1972) Similarity in visually perceived forms. *Psychol Issues* 8: 1-135.
3. Greene E (2007) Recognition of objects displayed with incomplete sets of discrete boundary dots. *Percept Mot Skills* 104: 1043-1059.
4. Greene E, Visani A (2015) Recognition of letters displayed as briefly flashed dot patterns. *Atten Percept Psychophys* 77: 1955-1969.
5. Greene E (2016) Information persistence evaluated with low-density dot patterns. *Acta Psychol* 170: 215-225.
6. Green DM, Swets JA (1966) *Signal Detection Theory and Psychophysics*. New York; Wiley.
7. Hautus MJ, van Hout D, Lee HS (2009) Variants of A Not-A and 2AFC tests: Signal Detection Theory models. *Food Qual Prefer* 20: 222-229.
8. Macmillan NA, Creelman CD (2005) *Detection Theory: A User's Guide*, New Jersey: Lawrence Erlbaum.
9. Hautus MJ (1995) Corrections for extreme proportions and their biasing effects on estimated values of d' . *Behav Res Methods Instrum Comput* 27: 46-51.
10. Miller J (1996) The sampling distribution of d' . *Percept Psychophys* 58: 65-72.

11. Hautus J (2012) SDT Assistant (version 1.0) [Software]. Available from <http://hautus.org>.
12. Hautus MJ (1997) Calculating estimates of sensitivity from group data: pooled versus averaged estimators. *Behav Res Methods Instrum Comput* 29: 556-562.
13. Hubel DH, Wiesel TN (1959) Receptive fields of single neurons in the cat's striate cortex. *J Physiol* 148: 574-591.
14. Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195: 215-243.
15. Selfridge OG (1959) *Pandemonium: A Paradigm for Learning in the Mechanization of Thought Process*, London; HM Stationary Office.
16. Sutherland NS (1968) Outlines of a theory of visual pattern recognition in animals and man. *Proc R Soc Lond B Biol Sci* 171: 297-317.
17. Binford TO (1971) Visual perception by computer. *Proc IEEE Conf Syst Control*, Miami, FL.
18. Barlow HB (1972) Single units and sensation: a neuron doctrine for perceptual psychology. *Perception* 1: 371-394.
19. Milner PM (1974) A model for visual shape recognition. *Psychol Rev* 81: 521-535.
20. Palmer SE (1975) Visual perception and world knowledge: notes on a model of sensory-cognitive interaction. In: Norman DA, Rumelhart DE eds, *Explorations in Cognitions*, San Francisco: WH Freeman & Co.: 279-307.
21. Marr D (1982) *Vision: A Computational Investigation into the Human Representation and Processing of Information*, New York: Freeman: 51-79.
22. DeValois RL, Devalois KK (1991) Vernier acuity with stationary moving Gabors. *Vision Res* 31: 1619-1626.
23. Panda R, Chatterji BN (1996) Gabor function: an efficient tool for digital image processing. *IETE Tech Rev* 13: 225-231.
24. Kohonen T, Oja E (1998) Visual feature analysis by the self-organizing maps. *Neural Comput Appl* 7: 273-286.
25. Pettet MW, McKee SP, Grzywacz NM (1998) Constraints on long range interactions mediating contour detection. *Vision Res* 38: 865-879.
26. Pennefather PM, Chandna A, Kovacs I, et al. (1999) Contour detection threshold: repeatability and learning with 'contour cards.' *Spat Vis* 12: 257-266.
27. Taylor G, Hipp D, Moser A, et al. (2014) The development of contour processing: evidence from physiology and psychophysics. *Front Psychol* 5: e719
28. Dong X, Chantlier MJ (2016) Perceptually motivated image features using contours. *IEEE Trans Image Process* 25: 5050-5062.
29. Edelman S (1999) *Representation and Recognition in Vision*, MIT Press.
30. Cooke T, Jakel F, Wallraven C, et al. (2007) Multimodal similarity and categorization of novel, three-dimensional objects. *Neuropsychologia* 45: 484-495.
31. Hayworth KJ (2012) Dynamically partitionable autoassociative networks as a solution to the neural binding problem. *Front Comput Neurosci* 6: e73.
32. Rodriguez-Sanchez AJ, Tsotsos JK (2012) The roles of endstopped and curvature tuned computations in a hierarchical representation of 2D shape. *PLoS ONE* 7: e42058.
33. Hopfield JJ (1995) Pattern recognition computation using action potential timing for stimulus representation. *Nature* 376: 33-36.
34. Hopfield JJ (1996) Transforming neural computations and representing time. *Proc Nat Acad Sci*

- 83: 15440-15444.
35. Maass W (1997) Fast sigmoidal networks via spiking neurons. *Neural Comput* 9: 279-304.
 36. McClelland JL (2013) Integrating probabilistic models of perception and interactive neural networks: a historical and tutorial review. *Front Psychol* 4: e503.
 37. Guan T, Wang Y, Duan L, et al. (2015) On-device mobile landmark recognition using binarized descriptor and multifeature fusion. *ACM Trans Intell Syst Technol* 7: e12.
 38. Wei B, Guan T, Duan L, et al. (2015) Wide area localization and tracking on camera phones for mobile segmented reality systems. *Multimed Syst* 21: 381-399.
 39. Pan H, Guan T, Luo Y, et al. (2016) Dense 3D reconstruction combining depth and RGB information. *Neurocomputing* 175: 644-651.
 40. Ullman S (1976) Filling-in the gaps: the shape of subjective contours and a model for their generation. *Biol Cybern* 25: 1-6.
 41. Sha'ashua A, Ullman S (1988) Structural saliency: the detection of globally salient structures using a locally connected network. In: *Proc 2nd Intern Conf Comput Vision*, Clearwater FL: 321-327.
 42. Kellman PJ, Shipley TF (1991) A theory of visual interpolation in object perception. *Cogn Psychol* 23: 141-221.
 43. Shipley TF, Kellman PJ (1992) Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Percept Psychophys* 52: 97-106.
 44. Cormen TH, Leheron CE, Rivest RL, et al. (2001) Single-source shortest paths and all pairs shortest paths. In: *Introduction to Algorithms*, MIT Press & McGraw-Hill: 580-642.
 45. Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: evidence for a local "association field." *Vision Res* 33: 173-193.
 46. Kwon TK, Agrawal K, Li Y, et al. (2016) Spatially-global integration of closed, fragmented contours by finding the shortest-path in a log-polar representation. *Vision Res* 126: 143-163.
 47. Sceniak MP, Hawken MJ, Shapley R (2001) Visual spatial characterization of Macaque V1 neurons. *J Neurophysiol* 85: 1873-1887.
 48. Bowers JS (2009) On the biological plausibility of grandmother cells: implications for neural network theories in psychology and neuroscience. *Psychol Rev* 116: 220-251.
 49. Greene E (2008) Additional evidence that contour attributes are not essential cues for object recognition. *Behav Brain Funct* 4: e26.
 50. Greene E (2016) How do we know whether three dots form an equilateral triangle? *JSM Brain Sci* 1: 1002.
 51. Lichtsteiner P, Posch C, Delbruck T (2008) A 128x128 120 dB 15 μ s latency asynchronous temporal contrast vision sensor. *IEEE J Solid-State Circuits* 43: 566-576.
 52. Robinson DA (1964) The mechanisms of human saccadic eye movement. *J Physiol* 174: 245-264.
 53. Zimmermann E, Lappe M (2016) Visual space constructed by saccadic motor maps. *Front Human Neurosci* 10: e225.
 54. McSorlely E, McCloy R, Williams L (2016) The concurrent programming of saccades. *PLoS ONE* 11: e0168724.
 55. Bhutani N, Sengupta S, Basu D, et al. (2017) Parallel activation of prospective motor plans during visually-guided sequential saccades. *Neurosci* 45: 631-642.
 56. Feldman JA, Ballard DH (1982) Connectionist models and their properties. *Cogn Sci* 6: 205-254.

57. Fukushima K (1980) Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol Cybern* 36: 193-202.
58. Rolls ET (1992) Neurophysiological mechanisms underlying face processing within and beyond the temporal cortical visual areas. *Phil Trans R Soc* 335: 11-21.
59. Wallis G, Rolls ET (1997) Invariant face and object recognition in the visual system. *Prog Neurobiol* 51: 167-194.
60. Riesenhuber M, Poggio T (2000) Models of object recognition. *Nature Neurosci* (S3): 1199-1204.
61. Pasupathy A, Connor CE (2001) Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86: 2505-2519.
62. Suzuki N, Hashimoto N, Kashimori Y, et al. (2004) A neural model of predictive recognition in form pathway of visual cortex. *BioSystems* 76: 33-42.
63. Pinto N, Cox DD, DeCarlo JJ (2008) Why is real-world visual object recognition hard? *PLoS Comput Biol* 4: e27.
64. Hancock PJB, Walton L, Mitchell G, et al. (2008) Segregation by onset asynchrony. *J Vision* 8: 1-21.
65. Karplus I, Goren M, Algorn D (1982) A preliminary experimental analysis of predator face recognition by *Chromis caenuleus* (Pisces, Pomacentridae). *Z Tierpsychol* 58: 53-65.
66. Siebeck UE, Parker AN, Sprenger D, et al. (2010) A species of reef fish that uses ultraviolet patterns for covert face recognition. *Curr Biol* 20: 407-410.
67. Karplus I, Katzenstein R, Goren M (2006) Predator recognition and social facilitation of predator avoidance in coral reef fish *Dascyllus marginatus* juveniles. *Mar Ecol Prog Ser* 319: 215-223.
68. Siebeck UE, Litherland L, Wallis GM (2009) Shape learning and discrimination in reef fish. *J Exp Biol* 212: 2113-2119.
69. Newport C, Wallis G, Reshitnyk Y, et al. (2016) Discrimination of human faces by archerfish (*Toxotes catareus*). *Sci Rep* 6: e27523.
70. Greschner M, Field GD, Li PH, et al. (2014) A polyaxonal amacrine cell population in the primate retina. *J Neurosci* 34: 3597-3606.
71. Greschner M, Heitman AK, Field GD, et al. (2016) Identification of a retinal circuit for recurrent suppression using indirect electrical imaging. *Curr Biol* 26: 1935-1942.
72. Greene E (2007) Retinal encoding of ultrabrief shape recognition cues. *PLoS One* 2: e871.
73. Greene E (2016) Retinal encoding of shape boundaries. *JSM Anat Physiol* 1: 1002.



AIMS Press

© 2017 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)