Preattentive Perception of Elementary Three Dimensional Shapes

Jennifer Sun Pietro Perona

California Institute of Technology 116-81
Pasadena, CA 91125
jensun@systems.caltech.edu

CNS memo 29 – August 1993

Abstract

Recently Enns & Rensink (1990, 1991) showed, in a response-time paradigm experiment, that, in some instances, targets differing from their distractors in their perceptual 3D shape can be spotted at reaction times that are only weakly correlated with the number of distractors present. This 3D "pop-out" effect suggests fast, parallel processing of some aspects of shape. Using a 2AFC short duration SOA paradigm with masking, we have confirmed that some shaded elements that can be interpreted roughly as 3D convex shapes that are lit from above, can indeed be discriminated in parallel, in times from 15 to 100 ms. We found that unshaded line drawings of the same shapes, however, require longer display times and are processed serially. While other similar brightness patterns that do not have 3D interpretations are processed serially, the shaded Y junction created by 3 contingent regions of different luminances seems to confer perceptual three dimensionality as well as parallel processing when embedded inside a variety of shapes. Strong pop-out asymmetries suggest that the fast, parallel processing we observed is dependent upon three dimensional perceptual differences. Furthermore, we found that contextual scene information can affect performance. Contextual information that contributes to the interpretation of a consistent 3D scene of normal viewing conditions enhances performance while inconsistent contextual information do not. Our results to date suggest that shaded shapes which are consistent with familiar 3D configurations and lighting conditions can be processed in parallel.

1 Introduction

In the classical view of preattentive vision, only basic features such as color, oriented edges, motion etc. are processed quickly and in parallel, while more complicated stimuli composed of a combination of basic features are supposed to be perceived serially (Beck 1966, 1967, 1982; Olson & Attneave 1971; Julesz 1975, 1984; Treisman & Gelade, 1980). Because the perception of three dimensional shape from shading would require the processing of multiple features, including a variety of oriented edges and luminance levels, it is traditionally believed to require a relatively slow, serial "spotlight of attention" approach. Recently, however, Enns & Rensink (1990, 1991), using a response time paradigm, showed that certain targets differing from their distractors in their perceptual 3D shape "pop-out" with the characteristics of preattentive processing. These surprising results give rise to further questions:

1. What are the relevant features in these perceptual 3D shapes that allow them to be processed in parallel? In particular, is it the shading itself that is important or is it actually the edge boundaries created by the shaded regions?
2. Is the crucial calculation performed on mainly a local corner junction, or is it performed upon the entire shape?
3. Is this process a "hard-wired," local and bottom-up process, or can it be influenced by global and/or contextual information?

Furthermore, we also wondered about the sensitivity of the response time paradigm which Enns & Rensink used. The typical response time they measured ranged from 400 to 700 msec, even for tasks that were considered as easy and processed in parallel. Previous "pop-out" and texture segregation experiments have shown, however, that processing time for parallel tasks generally requires less than 150 msec (Kröse 1987, Gurnsey & Browse 1987). In our series of experiments then, we attempted to confirm some of Enns & Rensink's results using an SOA with masking paradigm instead for a more exact measure of the processing times involved and to provide some further clues toward the understanding of shape perception from shading and boundaries.

2 Methods

We used a two-alternative forced-choice stimulus onset asynchrony (SOA) paradigm with masking. Images were generated on a Silicon Graphics Iris. Stimulus display times ranged from 16 msec to 400 msec depending on the task, and were followed by a blank inter-stimulus interval (ISI) time of 0, 16, or 26 msec and a 200 msec mask. Stimulus screens contained 3, 6, 12, 18 or 24 items of display, with each item spanning approximately 1.5 degrees of visual angle. In screens with 12, 18, and 24 items, spacing between items was approximately 3 degrees, measured from the center of one item to its nearest neighbor, with an additional random jitter of up to 0.3 degrees. For screens of 3 and 6 items, the separation was larger, approximately 7.5 degrees and 4.5 degrees respectively, so as to maintain a comparable maximum eccentricity for all display sizes. One target was present at random among multiple distractors in 50% of the trials. Target-present trials and target-absent trials have the same total number of patterns. Target position was also randomized, but only ones that occured at an eccentricity of less than 6.5 degrees were considered. Subjects were prompted to respond regarding the presence or absence of the target after the mask disappeared. Experiments were presented in blocks of 35 trials, with number of items and duration of display held constant within a block, with the exception of the ones in Experiments 6 & 7 in which the effects of contextual information were investigated. In these, the trials were presented in blocks of 100, with 3 different display sizes intermixed at random. One set of experiments by one subject consisted of about 1800 trials of each pattern tested. Subjects usually trained for 2 days before performance stabilized enough for consideration. Data was collected from a total of nine subjects, all but one naive.

3 Experiments

3.1 Experiment 1 (Shaded Cubes)

In this first experiment, we used a stimulus from Enns & Rensink's 1990 paper. The pattern consists of a shaded Y-junction embedded in a hexagon that is typically interpreted as a cube (See Fig. 1). The distractors can be interpreted as cubes sitting on a surface with lighting from above, and the target can be seen alternatively as a cube with its bottom side exposed and lit from below, or as a concave corner lit from above. This condition with the distractors seen in top-down view and top-lit condition will be referred to as the "normal" condition. In the "reverse" situation, the target and distractor patterns are switched. This is illustrated in Figure 2. Figure 3 shows a typical mask used in these experiments.

The results from this set of experiments show a strong asymmetry in performance between the normal condition and its reverse. Graphs 1 & 2 show typical psychometric curves from one subject for these two conditions. The performance in percentage correct is plotted against display time in msec. In the normal condition task (Graph 1), the curves for the three different display sizes, 3, 12, and 24, fall close together within the same range, suggesting parallel processing. For the reverse condition, however, the curves separate out from each other, indicating serial processing. We estimated the stimulus display duration necessary for 75% accuracy performance, averaged across 6 subjects for the normal condition and 4 subjects for the reverse condition, and plotted this against the number of display items (See Graph 3). Across subjects, for the normal view, performance is consistently fast and in parallel. Duration necessary for processing is virtually independent of the number of distractors. In contrast, for the reverse view, increasing the number of distractors effected a significant increase in the 75% accuracy time. For this task, some subjects do not achieve consistent 75% accuracy even at the longest display times for large display sizes. In these cases, we considered the minimum SOA for detection to be the longest display interval, i.e. the true SOA is at least this long. This asymmetry in performance seems to be correlated with a perceptual asymmetry. All subjects reported that the distractor cubes in the normal

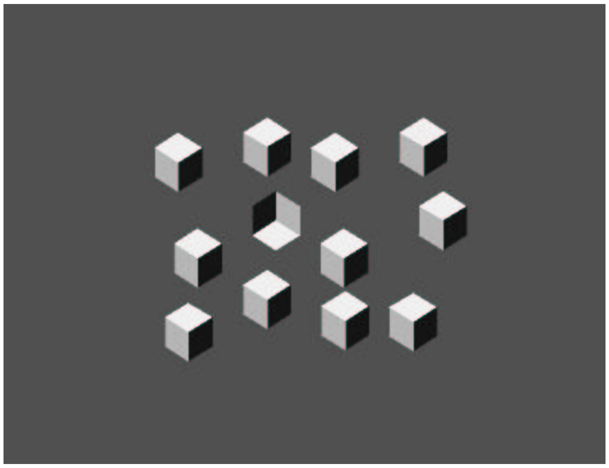


Figure 1: 12-item screen with "normal" condition distractors

Figure 2: "Reverse" condition distractors

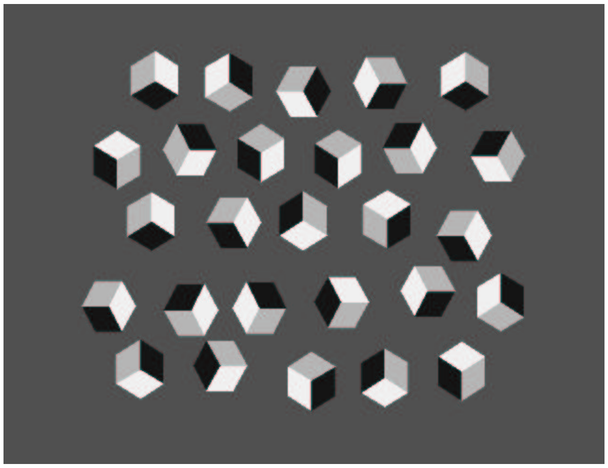
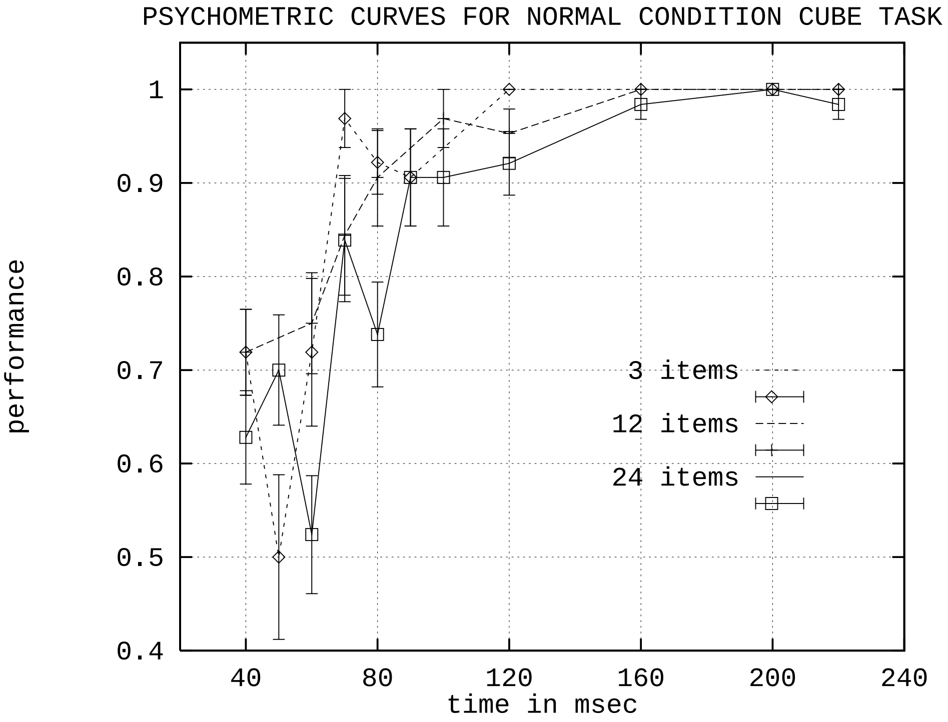
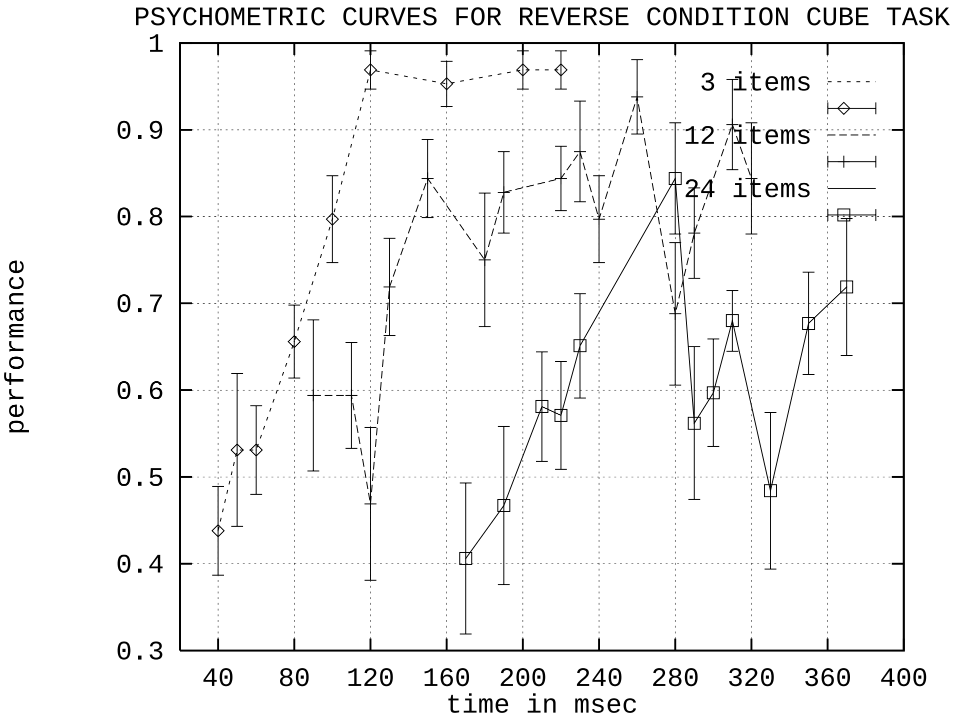


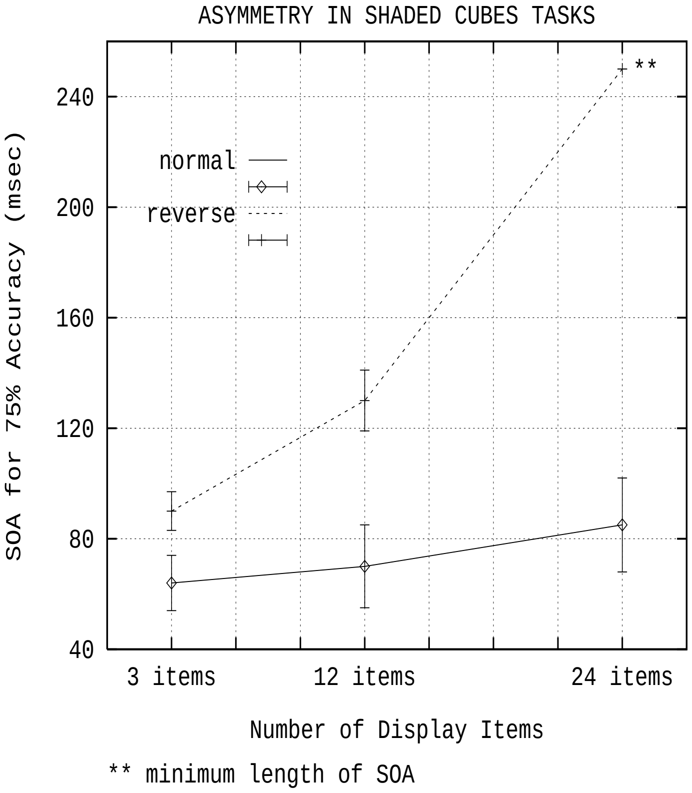
Figure 3: A sample mask used in cubes tasks



Graph 1



Graph 2



Graph 3

condition experiment appeared more strongly three dimensional than those in the reverse condition. The target cube, in the normal condition, looked “strange” or “flat” or, to 2 of the 6 subjects, even “concave” compared to the distractors. The difference between the target and the distractors in the reverse condition was reported to be perceptually much weaker.

3.2 Experiment 2 (Line Cubes)

One of our main goals is to study possible “3D features” that are recognized by the preattentive system. Upon finding that shaded cubes may be processed preattentively, we wanted to see whether the crucial component for pop-out and, arguably, preattentive shape perception is the oriented edges, the shading, or a combination of the two. In their 1991 paper, Enns & Rensink showed that in a response time paradigm, “pop-out” tasks involving line cubes (See Fig. 4) require search times that increased from approximately 500 msec to 700 msec as the display size was increased from 1 to 6 to 12 items. Compared to other line patterns that do not have 3D interpretations and require search times that increased from 500 to more than 1250 msec, the line cubes appeared to be processed in parallel. Combining these results with the results of their 1990 paper, which showed that similar experiments involving shaded cubes, like the ones in Figure 1, give search times of around 400 msec, one might conclude that the line drawing, of a cube in this case, is the essential feature, and that the only purpose of shading is to form luminance borders for edge detection. However, when we conducted our experiments using line elements, our results indicate the contrary (See Graph 4).

The averaged SOAs for 3 subjects each on the line Y-junctions experiment and the line cubes experiment are presented in Graph 4 in comparison to the same normal condition, shaded cubes curve depicted in Graph 3 of Experiment 1. Line Y-junctions are the most difficult. Adding a hexagonal outline gives the percept of a line cube and improves performance. However, we found that line cubes nonetheless require a significantly longer SOA than shaded cubes for all display sizes, with a difference that increases with the number of items of display.

3.3 Experiment 3 (2D Controls)

Since Experiment 2 suggested that shading is a crucial component of preattentive processing, we directed our attention henceforth to shaded patterns. In this experiment we used the same experimental paradigm with various patterns that are shaded in the same black, grey, and white tones as the cubes, but do not have typical 3D interpretations. If these "flat" patterns can also be processed in parallel, then presumably the shaded cubes may also have been processed in parallel using 2D cues only, and the "3D-ness" of the shaded cubes may have nothing whatsoever to do with parallel processing. The following graphs show the plots for three such stimuli which we coined, respectively, the 3-layer torte, the T-junction in rectangle, and the X-junction in diamond, in comparison again with the plot for the normal condition cubes. The 3-layer torte and the X-junction in diamond patterns were originally used by Enns & Rensink (1990). Data was collected from three different subjects for the 3-layer torte experiment and four subjects each for the T-junction in rectangle and the X-junction in diamond experiment. One can see that all the other stimuli gave plots that have more positive slopes. While the slope for the shaded cubes is a little less than 1 msec/item, the slopes for the 3-layer tortes, T-junction in rectangles, and the X-junction in diamonds are 2.5 msec/item, 2.7 msec/item, and 4 msec/item respectively. For three items of display, performance is about equal between the 2D patterns and the cubes, or even easier in the case of the tortes. However, as item number is increased, all the other shapes are more affected than the cubes are. Enns & Rensink (1990) also showed that the 3-layer torte and the T junction patterns are processed serially.

Subjects were always asked how they interpreted the patterns they saw. While the shaded cubes were recognized as such without exception, none of the other patterns shown in this experiment prompted 3D interpretations, except for the 3-layer torte, which one subject voluntarily labeled as stairs. Interestingly enough, this subject also performed better than anyone else on this task, with the 75% accurate SOA's for 3, 12 and 24 items all at around 75 ± 8 msec. Since his perception was different from everybody else's, his data was not incorporated into the plot shown.

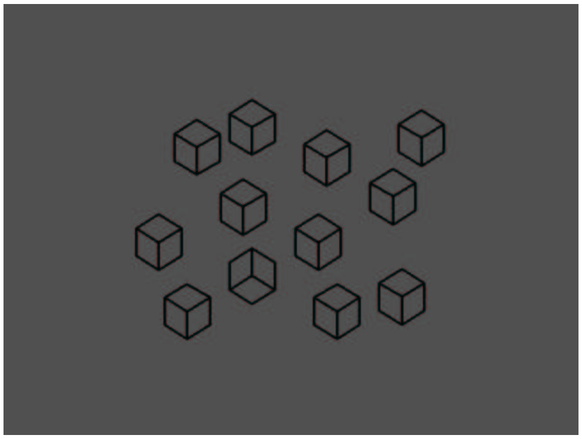


Figure 4: Line Cubes

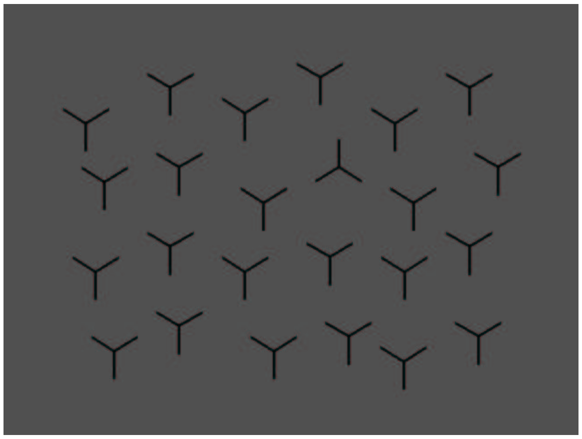
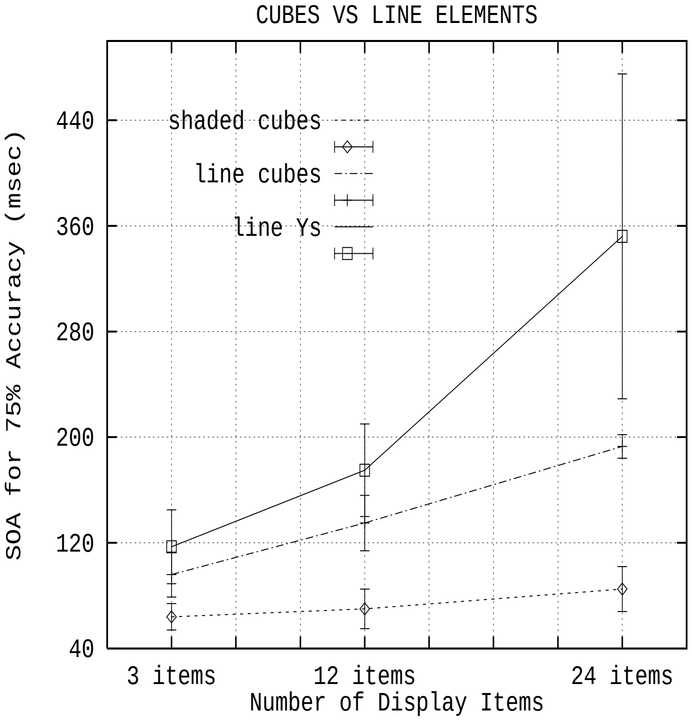
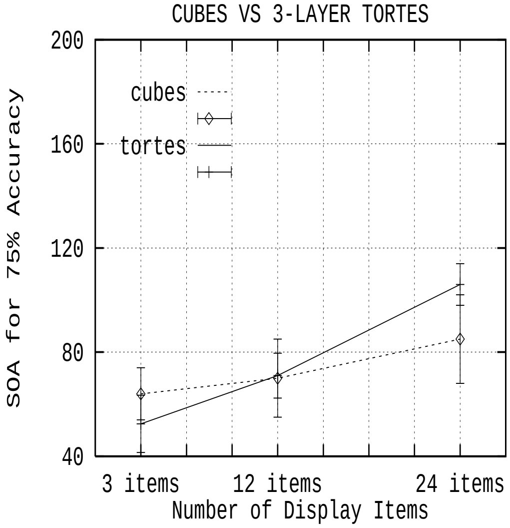


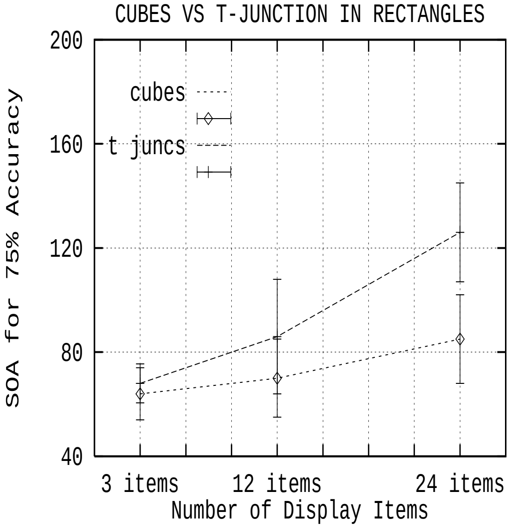
Figure 5: Line Y-Junctions



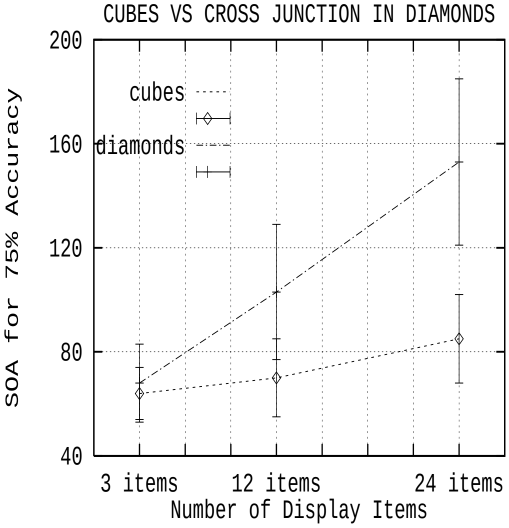
Graph 4



Graph 5



Graph 6



Graph 7

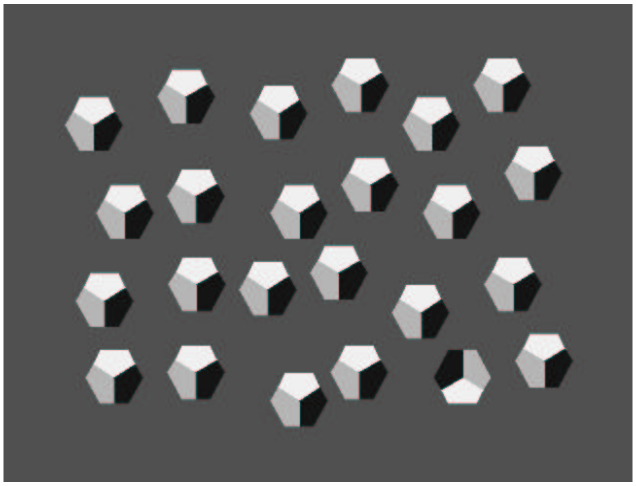
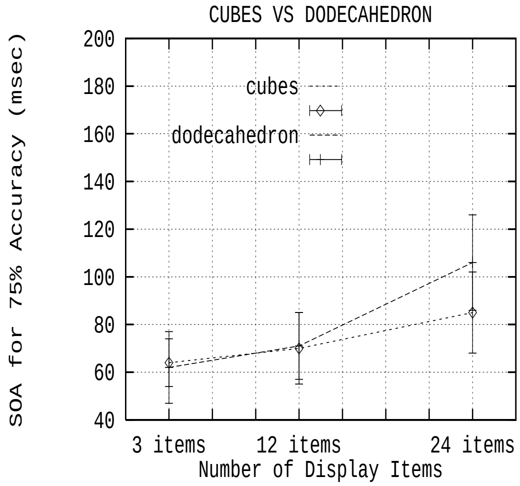


Figure 6: Dodecahedron?



Graph 8

3.4 Experiment 4 (Shaded Y-Junctions)

Since our results suggest that shading, and in particular, the shading pattern for the cube, which can be described as the shaded Y-junction, may be involved in fast, parallel processing, we ask next, what would happen if we embedded this shaded Y-junction in other outlines?

When the Y-junction is embedded in a hexagonal outline that is a 30-degree rotation of the outline of the cube (See Fig. 6), the slope of the line (Graph 8) is slightly more positive (< 2 msec/item) than that for the cubes (< 1 msec/item), but less positive than that for any of the flat, 2D patterns shown in Experiment 3. Testimonial evidence supports this mixed result. Two of the three subjects said that they interpreted the shape as a solid with sides made up of pentagons, a dodecahedron, while one said that it did not look three dimensional at all. However, they all reported that this shape looks "less 3D" then the shaded cubes.

Another pattern we tried is the Y-junction embedded in a diamond (See Fig. 7). All 6 subjects reported seeing the distractors as three dimensional, describing them variously as towers, cut-off pyramids, etc. The target was described either as a hole, or as looking very flat and 2D compared to the distractors. However, all subjects said that this task was hard because the 3D shape was rather unusual. The results, shown in Graph 9, reflect both these observations. The plot is shifted up, an indication of overall difficulty, but the line, with a slope of 2 msec/item, roughly parallels the cube plot and does not diverge much at large display sizes. The target-distractor reversed condition of the pyramids was shown to 2 subjects. This reverse condition proved to be quite difficult for both subjects. Performance was below 75% accuracy for display durations as long as 350 msec for 24 items of display.

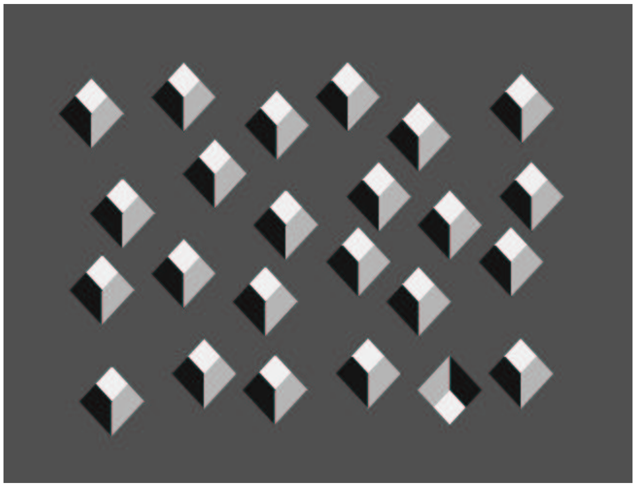
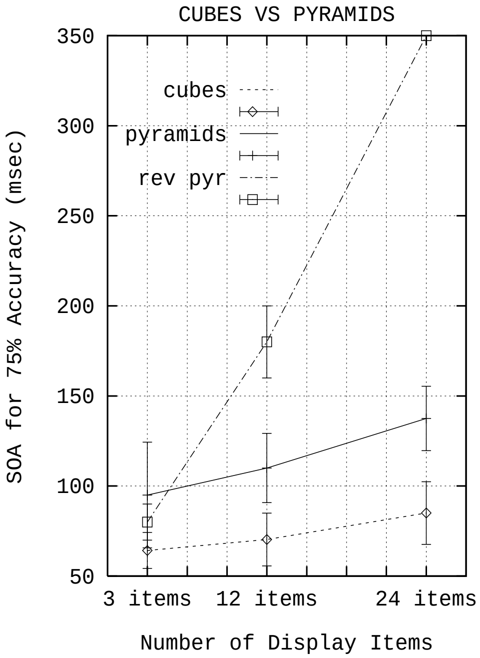


Figure 7: Pyramids?



Graph 9

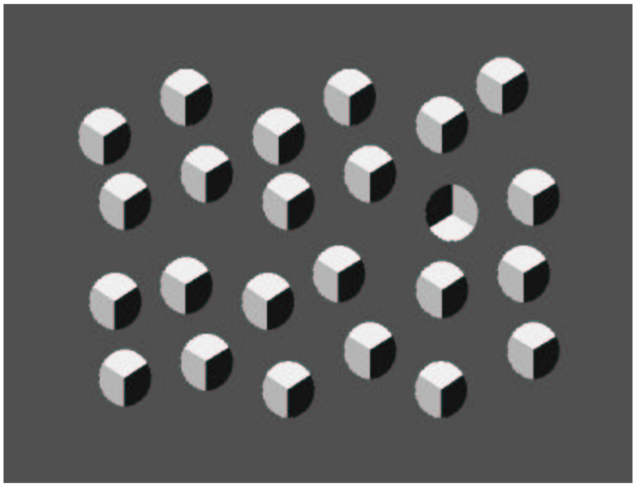
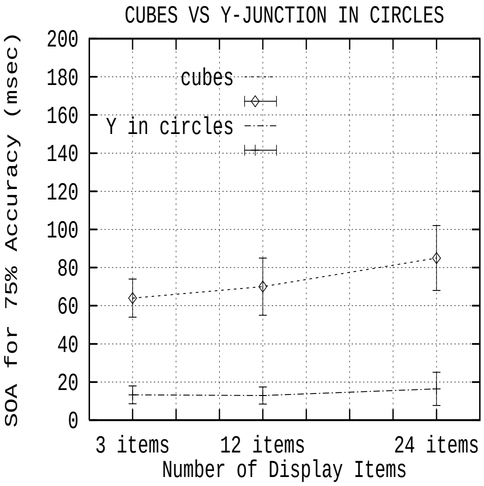
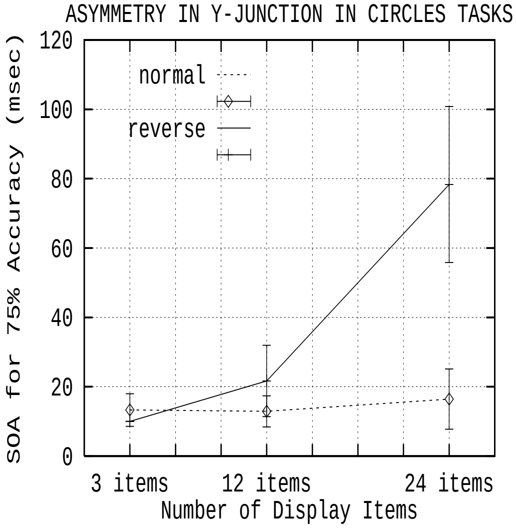


Figure 8: Spot-lit Corners?



Graph 10



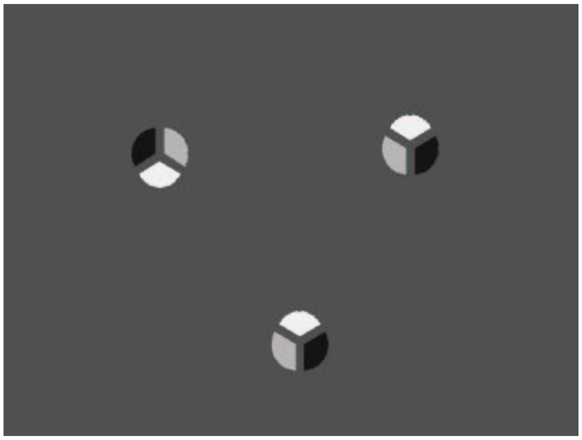
Graph 11

We also embedded the shaded Y-junction in a circular outline (Fig. 8). This was the easiest task of all. For 2 of the 4 subjects who participated in this experiment, the performance is above 85% correct for display durations of 16 msec, which is the shortest duration available on our Iris. For our calculations, we considered those SOA's which are below 16 msec as 10 msec. Graph 10 shows this plot against the cubes plot. The slope for the Y-junction in circles plot is virtually zero. Subjects reported that the patterns looked like corners of cubes or rooms that are illuminated by a circular spotlight.

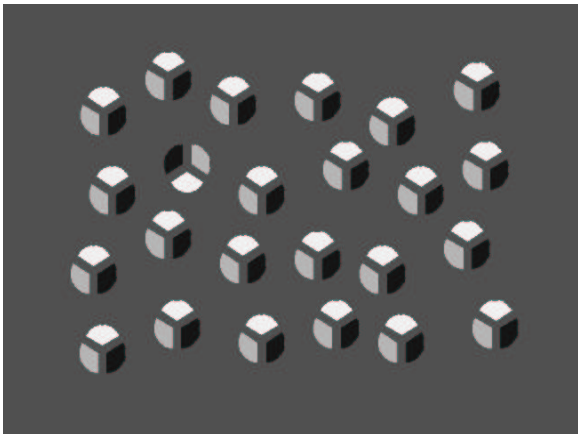
3.5 Experiment 5 (Y-Junction in Circles)

In Experiment 5, we studied in more detail the Y-junction in a circle pattern. To investigate whether this fast, parallel processing is correlated with a 3D percept, we used the target-distractor reversal experiment to see if there is an asymmetry, as in the case for the cubes. Graph 11 shows the averaged results from 3 subjects. The slope here is approximately 3 msec/item. The reverse case is still very fast, but it is no longer parallel.

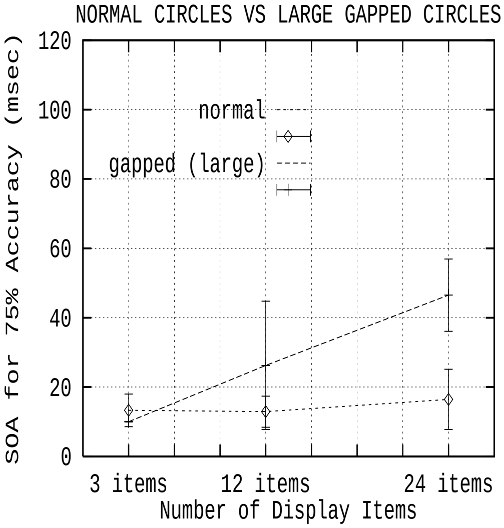
The results so far seem to indicate that the shaded Y-junction is a salient feature in 3D processing. To further explore this idea, we separated the shaded regions from each other with a gap of about 0.2 degrees. Two gapped patterns were used, one has shaded regions of the same area as the no-gap circles, and the other has a total area that is the same as the no-gap circles. Data was collected from 4 subjects for the large gapped circles and 2 subjects for the small gapped circles. Results show that, again, while the performance is still fast, it is no longer parallel (Graphs 12 & 13).



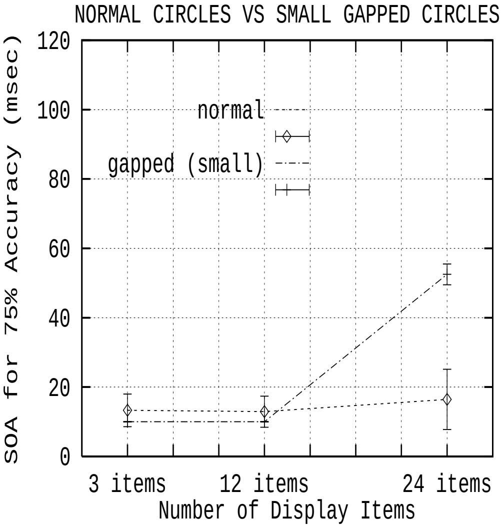
Large gapped circles



Small gapped circles



Graph 12



Graph 13

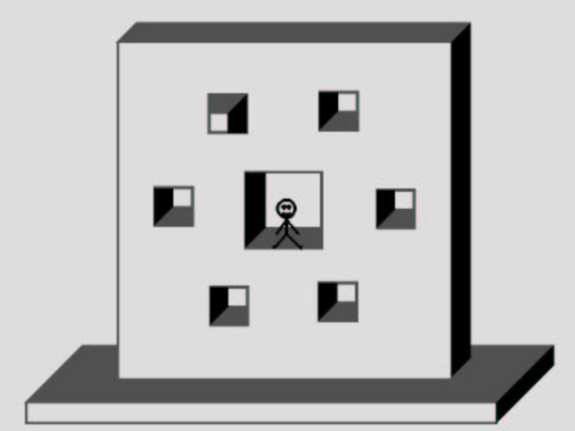


Figure 9: Consistent distractor holes

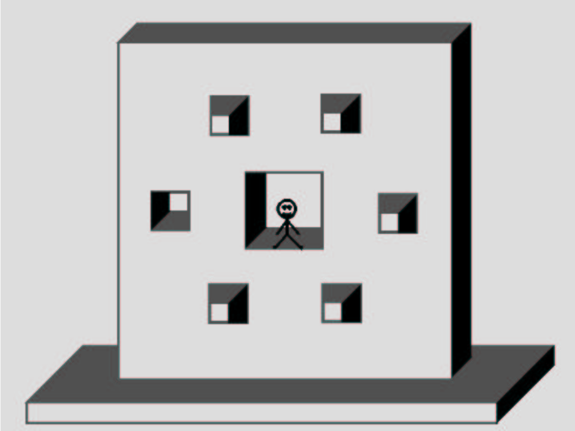


Figure 10: Inconsistent distractor holes

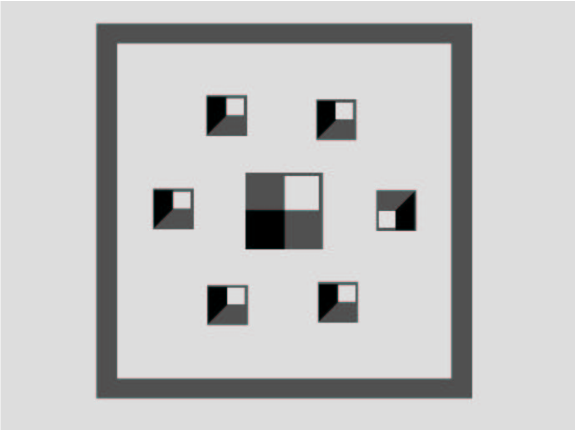
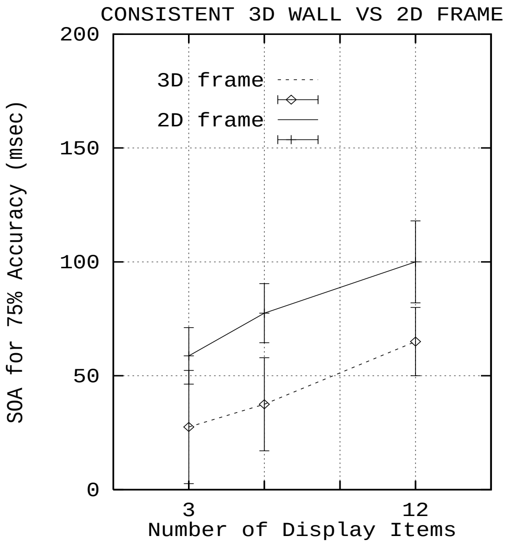
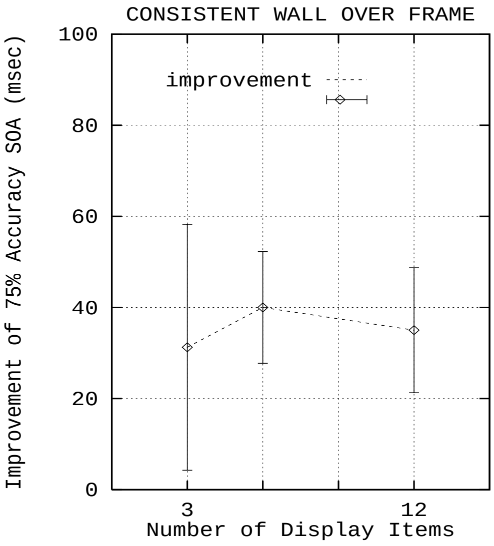


Figure 11: Frame with no 3D interpretation



Graph 14

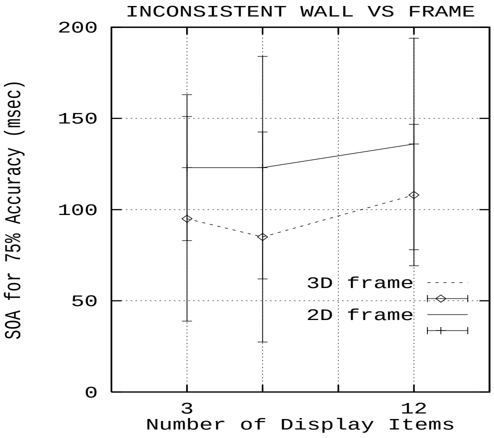


Graph 15

3.6 Experiment 6 (Holes in Wall)

In our last experiments, we investigated whether performance can be affected by background and contextual information that are mediated possibly by top-down processes or global calculations. For a rotated Y-junction in a square, which may be interpreted as a hole, we asked the question: how does a context that has either a consistent or an inconsistent 3D interpretation with respect to its embedded patterns affect performance? Figures 9 & 10 show displays which have distractor holes that are respectively consistent and inconsistent with the 3D wall frame. For control, the same shapes were displayed with an analogous surrounding frame that has no 3D interpretation (See Fig. 11).

Data was collected from 4 subjects for the consistent 3D frame vs. 2D frame exper-



Graph 16

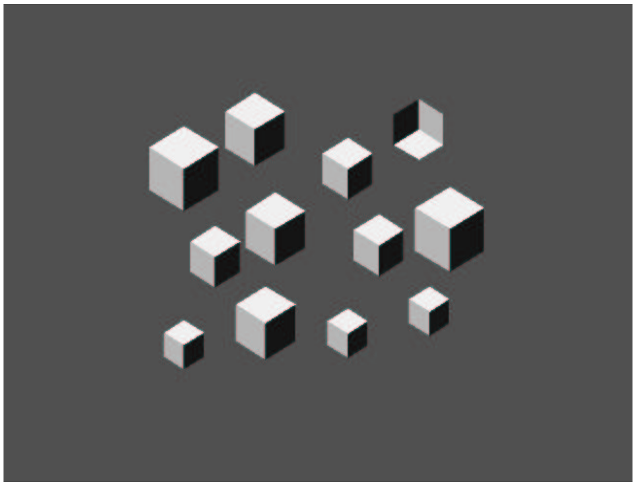
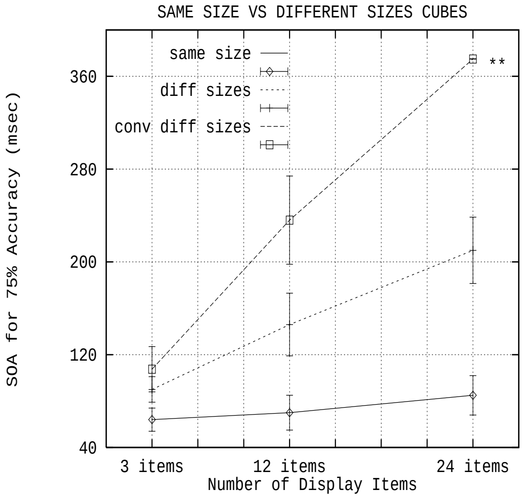


Figure 12: Cubes of different sizes



** minimum length of SOA

Graph 17

iment, and 3 subjects for the inconsistent 3D frame vs. 2D frame experiment. Compared with the controls, the 3D frame that was consistent with the distractor holes facilitated performance significantly, more significantly for 6 and 12 item displays than for 3 (Graphs 14 & 15), while the 3D frame that was inconsistent with the distractor holes did not lead to statistically significant improvements (Graph 16). For the inconsistent frame case, some subjects saw the distractors as protruding cones, which would be consistent with the shading of the frame, instead of inconsistent holes. Other subjects saw the distractors as inconsistent holes only. We suspect that performance may have been facilitated for those who formed the consistent percept, but not for those who formed only the inconsistent percept. Perhaps this dichotomy in perception can explain the large error bars seen in Graph 16.

3.7 Experiment 7

In another inducement experiment we found that perspective can influence performance significantly. While cubes of different sizes displayed in no apparent order are processed serially, as the data from 3 subjects indicate (See Graph 17), cubes of different sizes that are ordered in a gradient, thus mimicking the percept of cubes sitting on the ground, receding off into the distance (See Fig. 13), are processed in parallel. In addition, there is a consistent improvement for all items for all 3 subjects in comparison to a control task consisting of same size cubes (See Graph 18). Subjects reported that the perspective sizing enhanced the 3D percept and made the task easier. In contrast, for what we call the ceiling perspective (Fig. 14), which is a rather unusual viewing condition, we found no consistent improvement .

Not only does the floor perspective view enhance performance when the distractor cubes are in normal, top-lit view, it also enhances performance for the target-distractor reversed case (Fig. 15). Two subjects were used in both of these experiments. Without

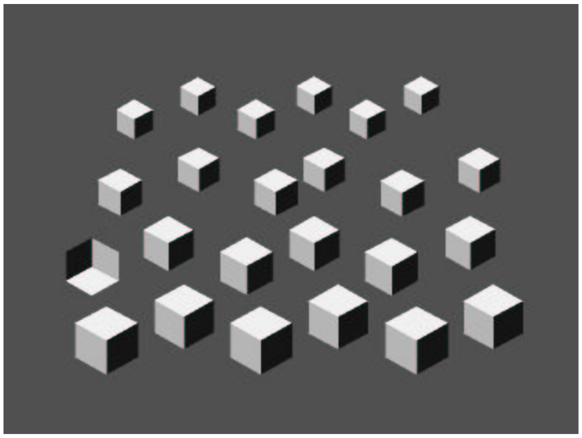


Figure 13: Floor perspective

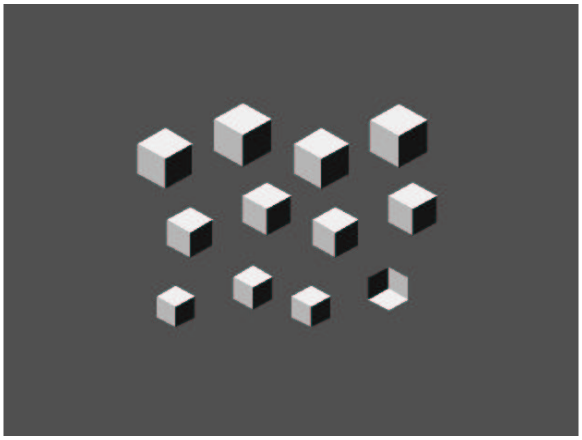
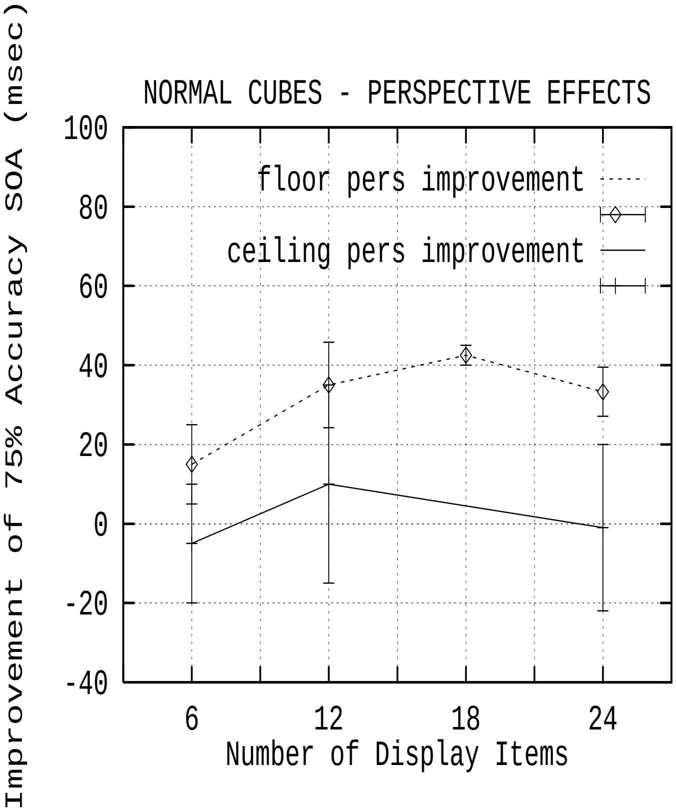


Figure 14: Ceiling perspective



Graph 18

perspective scaling, performance is slow and serial for this task. In floor perspective, there is a consistent improvement across subjects, with larger improvements for larger display sizes (Graph 19). Performance, in a sense, becomes more parallel. Interestingly, as shown in Graph 20, the ceiling perspective in this case (Fig. 16) also improves performance.

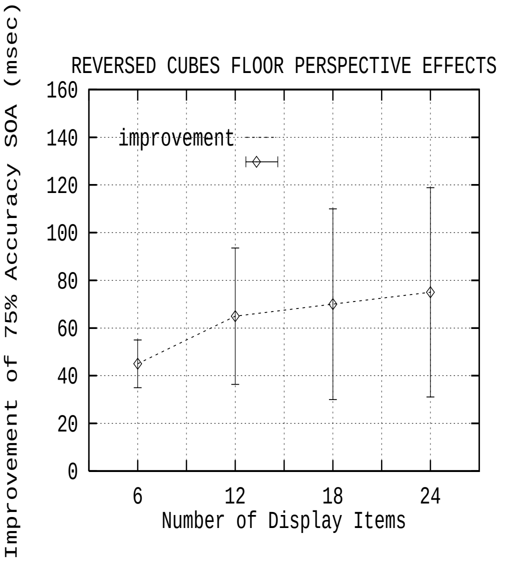
Figures 17 & 18 illustrate our attempt to enhance the perspective effect with backgrounds that suggest a room context. Two subjects participated on experiments which involved a floor perspective as well as a room context. As shown in Graph 21, which compares the performance of between the floor-perspective-only task with the floor-perspective-with room-context task, the room background did not cause any further improvement.

4 Discussion

The major findings of these experiments can be organized as follows:

- Our 2AFC short duration SOA experiments confirmed Enns & Rensink's proposal that three dimensional shape from shading can be processed in parallel and showed that processing times are around 50 to 80 msec, perhaps even as fast as 10 to 15 msec. These fast processing times are comparable to the ones previously reported in other "pop-out" and texture segregation experiments (Kröse 1987, Gurnsey & Browse 1987).
- We believe that this fast and parallel processing is dependent upon 3D information because:
    1. Shaded stimuli that are interpretable as three dimensional are processed in parallel while similar control stimuli that do not have 3D interpretation are not.
    2. Distractor-target reversal experiments that are equivalent in two-dimensional space, differing only in their 3D interpretations, show asymmetry in performance. This asymmetry is seen with the cubes, Y-junction in circles, as well as the pyramids.
    3. 3D contextual information can enhance performance, even in tasks that are already fast and parallel. For tasks that are processed serially, the addition of 3D information seems to also enhance performance. One example is the addition of a hexagonal outline to the line Y-junction in Experiment 2. Instead of increasing the difficulty of the task by complicating the Y-junction, the hexagonal outline improves performance, perhaps by conferring some three dimensionality.
    4. Subjects' reports of 3D perception coincide with performance that indicates fast, parallel processing.
- Unlike the results reported by Enns & Rensink (1991), our results indicate that unshaded line stimuli are processed serially. Other experimental results also support the idea that shading is a crucial component for 3D pop-out. Shaded bubbles, which contain no internal line edges, were shown to be processed in parallel (Braun 1990). We suspect that the response time paradigm, where search times are typically around 500 msec and increases of 100 msec or so as display size increases are considered negligible,

Figure 15: Reverse cubes - floor perspective



Graph 19

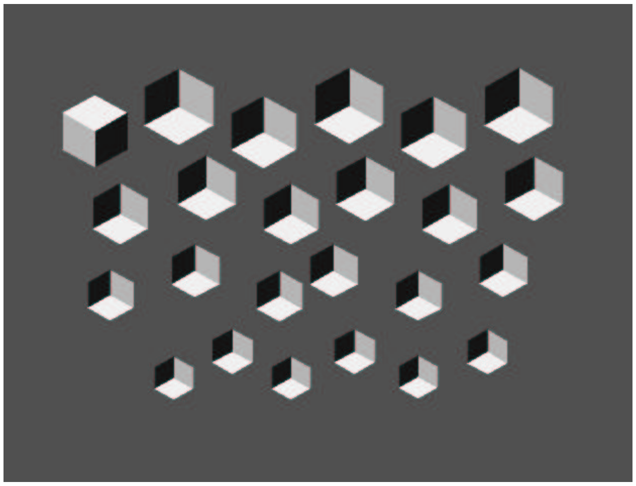
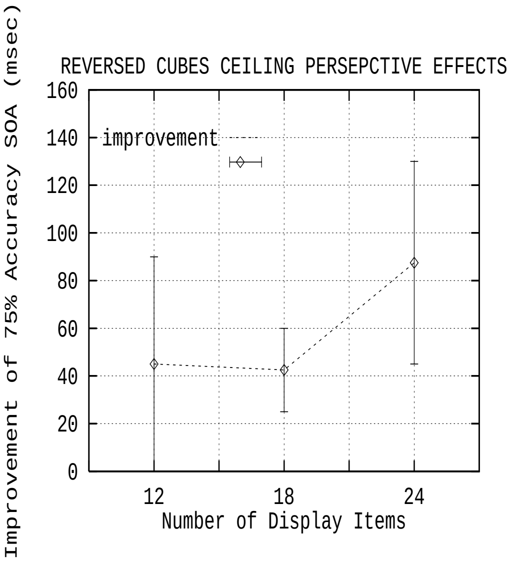


Figure 16: Ceiling perspective



Graph 20

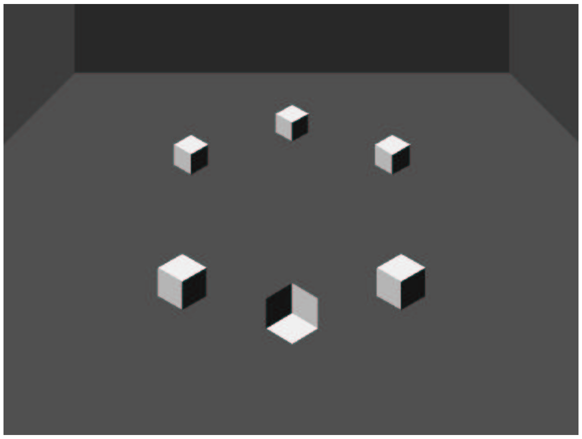
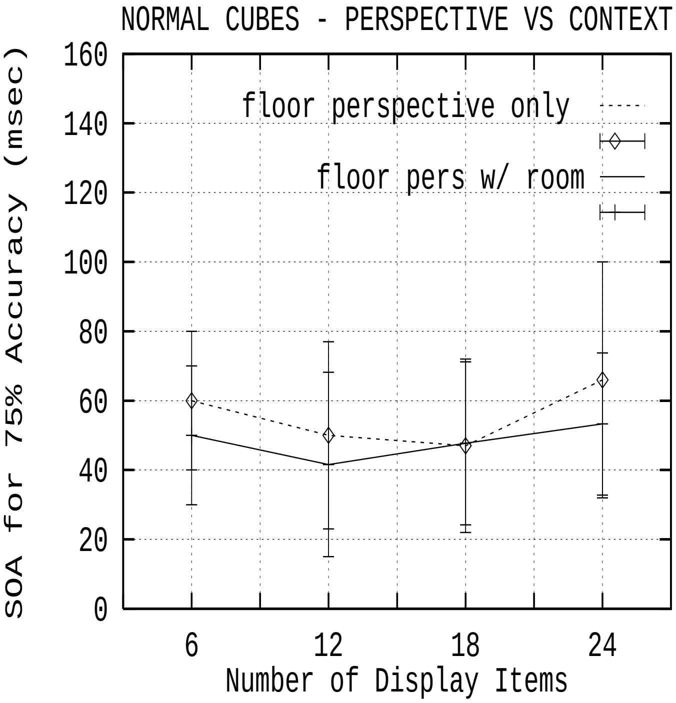


Figure 17: Floor perspective with room

Figure 18: Ceiling perspective with room



Graph 21

may not be sensitive enough to reveal the differences between, for example, the shaded cubes task and the line cubes task.

- Local junctions that define the local shape are processed more quickly than the entire 3D shape. This is shown by the Y-junction in circles experiment in Experiment 4. Circles embedded with a Y-junction, even though they cannot be perceived as whole 3D solids, are processed more quickly than shaded cubes are. It is possible that the shaded arrow-junctions formed by the outer edges of the cube actually "confuse the issue" and hinder the fast processing of the central Y-junction.

- Experiments 6 & 7 show that context information on stimulus configuration can indeed affect 3D perception and processing. If context cues are indeed used, one would expect only consistent cues, as opposed to inconsistent ones, to enhance performance. In Experiment 6, context cues are conferred by a 3D wall frame, and, in fact, performance is (a) significantly improved by the 3D frame that is consistent with the distractor items, and (b) not significantly affected by a 3D frame that is inconsistent with the distractors. Similarly, in Experiment 7, while ceiling perspective, which is inconsistent with the percept of top-lit, "normal" view cubes, did not enhance performance for this task, it did improve the performance for the reverse view, bottom-lit cubes. Perhaps this is because the ceiling perspective provides a more consistent context for the bottom-lit cubes and therefore enhances the percept of convexity which, we believe, may be a key feature for fast, parallel processing.

- We suggest that contextual information affects 3D processing by enhancing a consistent interpretation of a 3D scene via local, bottom-up means instead of top-down, global mechanisms. One piece of supporting evidence is in Experiment 7, where we found that adding a background room context does not affect performance. These early vision mechanisms may be related to those of texture discrimination. In Experiment 6, the 3D frame may be improving performance by adding to the textural density of the scene by providing more similarly oriented, shaded Y-junctions (See Fig. 9). In the task in Experiment 7 where the cubes are of different sizes placed at random, not according to any perspective gradient (see Fig. 12), the uniformity of the scene's textural pattern is disrupted. Perhaps it is this lesser degree of textural uniformity that causes this task to be processed serially, as opposed to the task involving same size cubes, which is processed in parallel. The evidence, however, is preliminary at best, and we hope to better clarify these points in future experiments.

- Our experimental findings are consistent with the intuition that shading patterns which promote a top-lit, convex percept (such as the shaded Y-junction) are processed fast and in parallel, while others (e.g. the upside-down Y-junction) which do not promote such an interpretation, are not. For example, in order for the upside-down Y-rjunction to be seen as convex, it has to be seen as showing a bottom side that is lit from below. If top-down lighting is assumed, it can only be interpreted as concave. A background of stimuli that are easily interpretable as convex would then be processed in parallel, and the disparate target of "unconventional" interpretation would be spotted quickly. With target and distractor stimuli reversed, the task, with a background of difficult to interpret patterns, would require serial processing. This could explain the asymmetry effect we found. Other shading patterns that satisfy the same requirements may presumably be processed in parallel also. One such pattern is the shaded bubble (Braun 1990).

5 Summary

1. Shaded stimuli compatible with a "convex" percept are processed fast and in parallel. Subjects' reports of 3D perception coincide with performance that indicates fast, parallel processing.
2. Line stimuli and shaded stimuli incompatible with a "convex" percept are processed more slowly and serially.
3. Three dimensional context information affects performance.

References

[1] Beck, J. (1966). Effects of orientation and of shape similarity on perceptual grouping. *Perception & Psychophysics, 1,* 300-302.

[2] Beck, J. (1967). Perceptual grouping produced by line figures. *Perception & Psychophysics, 2,* 491-495.

[3] Beck, J. (1982). Textural segmentation. In J. Beck (Ed.), *Organization and Representation in Perception,* pp. 285-317, Hillsdale, NJ: Lawrence Erlbaum.

[4] Braun, J. (1990). Focal attention and shape-from-shading. *Perception, 19,* A112.

[5] Enns, J. T. & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science, 247,* 71-723.

[6] Enns, J. T. & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review, 98(3),* 335-351.

[7] Gurnsey, R. & Browse R. A. (1987). Micropattern properties and presentation conditions influencing visual texture discrimination. *Perception and Psychophysics, 40(3),* 239-252.

[8] Julesz, B. (1975). Binocular depth perception without familiarity cues. *Science, 145,* 356-362.

[9] Julesz, B. (1984). Toward an axiomatic theory of preattentive vision. In G. M. Edelman, W. E. Gall & M. Cowan (Eds.), *Dynamic Aspects of Neocortical Function,* pp. 585-611. New York: Neurosciences Research Foundation.

[10] Kroöse, B. J. A. (1987). Local structure analyzers as determinants of preattentive pattern discrimination. *Biological Cybernetics, 55,* 289-298.

[11] Olson, R. & Attneave, F. (1970). What variables produce similarity grouping? *American Journal of Psychology, 83,* 1-21.

[12] Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97-136.

6 Acknowledgements

We would like to thank Drs. Bela Julesz, Jochen Braun, Jitendra Malik, John Allman, and Petra Stoerig for their comments and suggestions, and volunteer, unpaid subject LH for his contribution. We would also like to thank the NSF for supporting this research and the National Eye Institute for the ARVO/NEI Travel Fellowship Grant to ARVO 1993 where these results were first presented.