

## Visual attention and objects: New tests of two-object cost

JOHAN HULLEMAN and FRANS BOSELIE  
*NICI, University of Nijmegen, Nijmegen, The Netherlands*

Baylis and Driver (1993) proposed that the positions of object parts are coded relative to the position of the object they belong to and that parts of different objects are not directly coded relative to each other. This theory predicts that it is easier to judge a difference in height of parts belonging to a single object (one-object condition) than of parts belonging to two objects (two-object condition). This two-object cost has been reported in several articles (Baylis, 1994; Baylis & Driver, 1993, 1995). However, in all these experiments, the method that was used favored the one-object condition. In the present experiments we obtained, for the first time, evidence for the existence of two-object cost without such a bias.

There are opposing theories concerning the way attention is directed in visual perception. One theory holds that the visual world is parsed into objects (Duncan, 1984) and that visual attention is directed toward these objects one at a time. A second theory states that visual attention acts like a spotlight: no parsing takes place; rather, everything within the circumference of the spotlight is attended (Posner, 1980).

Duncan (1984) provided experimental evidence to support the first theory. He showed that it is easier to report on two different attributes when they belong to a single object than when they belong to different objects. However, Watt (1988) argued that differences in spatial scale were responsible for the observed two-object cost rather than differences in the way attention is allocated to single or multiple objects.

After reviewing the available experimental evidence, Baylis and Driver (1993) argued that apart from the object-based theory of attention, other lines of reasoning advocate a two-object cost. They combined elements of the computational work by Watt (1988) and the traditional distinction between “what” and “where” in vision; they put these into a single theoretical framework and introduced their hypothesis on hierarchical coding of location. This hypothesis states that two types of reference frames are routinely derived by the visual system: a scene-based frame, in which the locations of objects are coded relative to each other, and object-based frames of reference, in which the locations of parts of each single object are coded. The hypothesis predicts that it is easier to judge

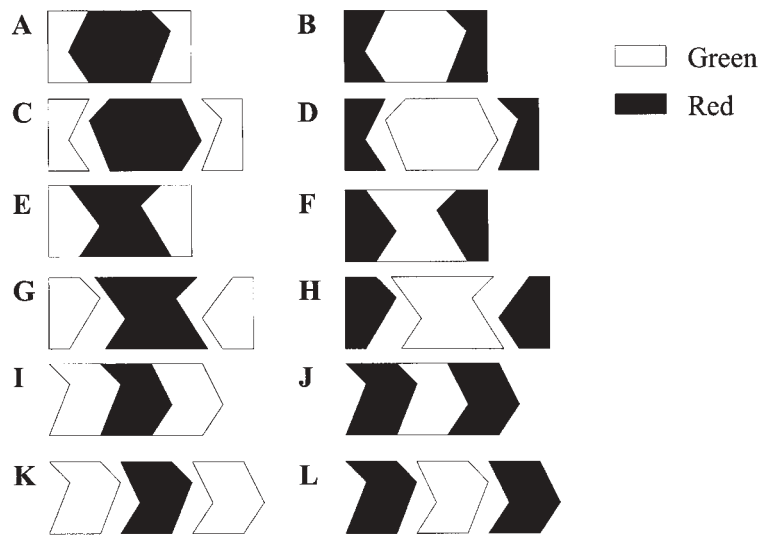
the relative location of parts of a single object than to judge the relative location of parts of two objects. The information needed to make this type of judgment is immediately available in the one-object condition. In the two-object condition, however, a transformation of coordinates has to be performed because the parts are only coded relative to their parent objects and not relative to each other.

Hierarchical coding of location is not simply an elaboration of the object-based theory of attention in the spatial domain. Although hierarchical coding predicts a two-object cost, it does so only when certain conditions are met. First, it has to be possible to perform the task within the object-centered frame of reference. Second, the descriptions of the parts within the object-centered frame of reference must be suited to the task at issue. If the routinely derived part descriptions for the single object are not appropriate for the task, whereas those for the two objects are appropriate, then it is possible that the two-object condition will yield faster reaction times. Baylis and Driver (1995) have reported an experiment in which there was such a two-object advantage.

Together with their hypothesis, Baylis and Driver (1993) introduced a paradigm in which explanations for a two-object cost based on the physical properties of the stimuli (e.g., those proposed by Watt, 1988) would no longer be valid. To test the predicted two-object cost, Baylis and Driver used the two-color displays depicted in Figures 1A and 1B. In one condition, subjects were instructed to attend to red. If they did so, they parsed the display in Figure 1A as a single red object on a green background. Under the same instruction, Figure 1B consequently was seen to contain two red objects on either side of a green background. In the other condition, subjects were asked to attend to green; Figure 1B contains the single object, whereas Figure 1A contains the two objects. The task was to judge which of the two inner vertices was lower on the screen. To ensure that the subjects adhered to their color instruction, the displays shown in

---

We are grateful to Mary Peterson and Bradley Gibson for pointing out the weak spots in the display of Experiment 1. The Netherlands Organization for Scientific Research (NWO) is gratefully acknowledged for funding this project. This research was conducted while J.H. was being supported by a grant from the NWO Foundation for Behavioral Sciences (575-62-062). Correspondence should be addressed to J. Hulleman, Nijmegen Institute for Cognition and Information, PO Box 9104, NL-6500 HE, Nijmegen, The Netherlands (e-mail: hulleman@nici.kun.nl).



**Figure 1.** Examples of displays. A–D: The left and right angles of the central region are both convex; used by Baylis and Driver (1993) and by Gibson (1994). E–H: The left and right angles of the central region are both concave; used by Gibson (1994). I–L: The left angle of the central region is concave, and the right angle is convex; used by Baylis (1994).

Figures 1C and 1D were included in the set of stimuli. The correct answers to these incongruent displays are different for the two color groups, and subjects who do not follow their instruction will have a large error rate.

Baylis and Driver (1993) predicted that subjects would respond fastest on displays when they were attending to a single object and that those displays on which the red group was fastest would be the slowest for the green group, and vice versa. They did indeed find these results. Depending on the color instruction, the same display yielded either the fastest or the slowest reaction times. Baylis and Driver claimed that these results showed that the two-object cost they found in their other experiments reflected differences in the processing of single and multiple objects, rather than differences in physical stimulus characteristics.

Gibson (1994) agreed that there are no physical differences when a single display is perceived to contain either one object or two objects. However, he contended that there were relevant perceptual differences between Baylis and Driver's (1993) one- and two-object conditions. The single object was always convex, and this is known to be relevant to figure-ground organization (Kanizsa, 1979). Therefore, the difference in reaction time between the one-object and two-object conditions could have been caused by the fact that the single object, being the most convex part, was always preferred as a figure. To obtain the percept of two objects would take longer because it requires a perceptual reversal of figure-ground relations. This would lead to the observed slower response in the two-object condition.

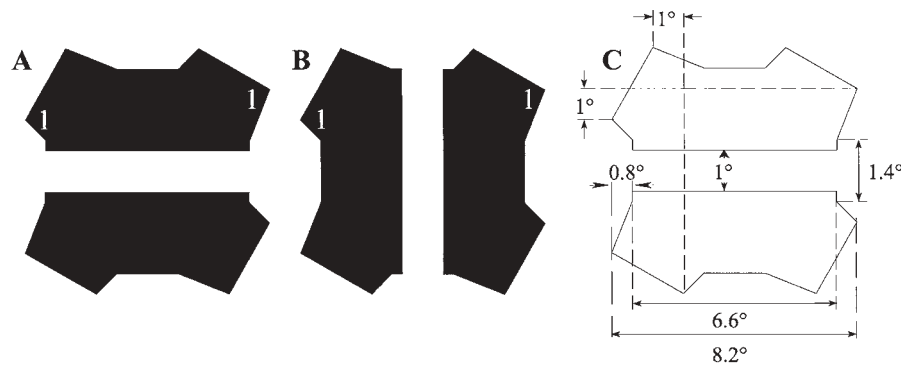
Gibson (1994) performed an experiment in which all of the stimuli shown in Figures 1A–1H were used. He found the same results as those of Baylis and Driver

(1993) when he used their original displays (Figures 1A and 1B), but when he used displays in which the convexity relations were reversed (Figures 1E and 1F), the outcome was also reversed. The two-object condition was faster than the one-object condition. Gibson concluded that it was convexity, rather than the number of objects, that caused the differences in reaction time.

In reply to Gibson (1994), Baylis (1994) reproduced the results of his own experiments, but using displays that did not differ in the degree of convexity in the one- and two-object conditions (see Figures 1I and 1J). He concluded that there was a genuine effect of the difference in number of objects, despite Gibson's results.

Though Baylis (1994), by his new type of displays, did indeed elegantly eliminate the difference in convexity between the central object and the outer objects, still other objections can be raised. First, Gibson (1994) not only showed that convexity favored the one-object condition in the original displays used by Baylis and Driver (1993), but also suggested that the central position of the single object—its "centeredness"—favors the one-object condition, too. The central part of the stimulus has a greater chance of becoming organized as a figure (Rock, 1986), and in the displays of Baylis, the single object is always the central part of the stimulus.

In addition to the two contaminating factors brought out by Gibson (1994), still other factors possibly code-terminated the results of experiments in which Baylis's paradigm and types of displays were used, all favoring the single-object interpretation of a display: (1) The total area of the single object was always only half of the area of the other two objects. (2) The single object was always more surrounded by the other two objects than vice versa. (3) The fixation cross was always presented within



**Figure 2.** (A) Experiment 1, one-object condition. The task-relevant convex vertices at the left and the right sides belong to a single object. (B) Experiment 1, two-object condition. The task-relevant convex vertices at the left and the right sides belong to two objects. (C) The dimensions of the stimulus expressed in terms of visual angle.

the boundaries of the single object. Other things being equal, regions that are more convex, central, small, surrounded, and fixated are much more likely to be seen as figures (Peterson & Gibson, 1994; Rock, 1986).

In fact, it is possible that the type of display used by Baylis (1994; see Figures 1I and 1J) favors the percept of the central region as a figure so strongly that this percept occurs even in the two-object condition. This would yield an underestimation of the two-object cost because the estimate would be based on two single-object judgments. This situation does not have to be detected in the control displays because their appearance is sufficiently different, allowing subjects to adapt their strategy to the display. In sum, despite all experimental effort, the hierarchical coding of location hypothesis has not yet been tested adequately.

Sticking to the kind of display and paradigm used by Baylis and Driver (1993), Gibson (1994), and Baylis (1994) in all their experiments makes it impossible to remove the aforementioned characteristics of the displays, which possibly favor the one-object condition. We therefore looked for another paradigm and another type of display to test the two-object cost hypothesis.

We designed the displays shown in Figures 2A and 2B. The displays are twofold rotationally symmetric. When the orientation of the gap is ignored, the displays are four-fold rotationally symmetric. The subjects' task is to judge the difference in height between the convex vertices at the left and right sides of the uppermost part of the displays. In the one-object condition, the display is oriented as shown in Figure 2A; in the two-object condition, it is oriented as shown in Figure 2B. The one- and two-object conditions differ only in terms of the orientation of the display; the rotational symmetries make the stimuli in both conditions physically the same in all other respects. In both conditions, subjects will parse the display as containing two objects. In the orientation shown in Figure 2A, the task-relevant convex vertices at the left and right sides (indicated by a 1) belong to a single object; in the orientation shown in Figure 2B, they belong to two objects.

## EXPERIMENT 1

### Method

**Participants.** The 12 subjects (10 women and 2 men) were psychology undergraduates at the University of Nijmegen. All had normal or corrected acuity by self-report. They received either course credit or a small payment (\$5) for their participation.

**Apparatus and Materials.** The experiment was conducted on a 486 PC with 15-in. SVGA (800 × 600 pixel) monitor. The shapes (Figures 2A and 2B) were presented in white on a black background. Their maximal extension was 8.2°. The space between the two objects was 1°. The vertical separation of the two task-relevant vertices was always 1°, the highest position a vertex could have was 1.1° below the top of the display, and the lowest position was 3.1° below the top of the display. (See Figure 2C for the dimensions of the display.)

**Design.** The orientation of the display was the only factor. In the one-object condition, the display was oriented horizontally (Figure 2A). A rotation of 90° of the entire display (Figure 2B) yielded the two-object condition. There were equal numbers of one-object and two-object stimuli.

**Procedure.** The subjects sat in a darkened room. Their task was to decide which of the two vertices in the upper half of the display was higher on the screen. When the left vertex appeared higher, subjects had to push the left button of a button box; when the right vertex seemed higher, they had to push the right button. Reaction times were recorded in milliseconds.

The subjects received a written instruction and saw examples of the two types of stimuli on the screen. Stimulus presentation ran as follows: A fixation cross was presented for 500 msec, followed by the display for 150 msec. The subject responded, and 700 msec after the answer, the fixation cross reappeared and a new sequence started. There were six blocks of 80 trials. Half the trials in each block required a response with the left hand, and half with the right.

**Analysis.** The first block of 80 trials was discarded as practice. The first two trials of the other blocks were treated similarly. Thus 390 trials were available for each subject. All of these trials were used in the accuracy analysis. Incorrect trials were not used in the reaction time analysis. Furthermore, in both the one- and the two-object conditions, reaction times that were more than 3 standard deviations away from the mean were trimmed to that value (this concerned 1.7% of the reaction times). The analyses were carried out with SPSS.

### Results

The mean reaction times were 530 and 546 msec for the one- and two-object conditions, respectively. A paired *t* test on the difference in reaction times between the one-object condition and the two-object condition showed

that the mean two-object cost was 16 msec [ $t(11) = 2.83$ ,  $p < .016$ ].

The mean error rates were 7.5% for the one-object condition and 8.8% for the two-object condition, respectively. This makes a speed-accuracy tradeoff very unlikely; a paired  $t$  test on the error rates did not reveal any differences between the one-object and two-object conditions [ $t(11) = 1.64$ , n.s.]. Moreover, there were fewer errors for the one-object condition, which yielded the fastest responses.

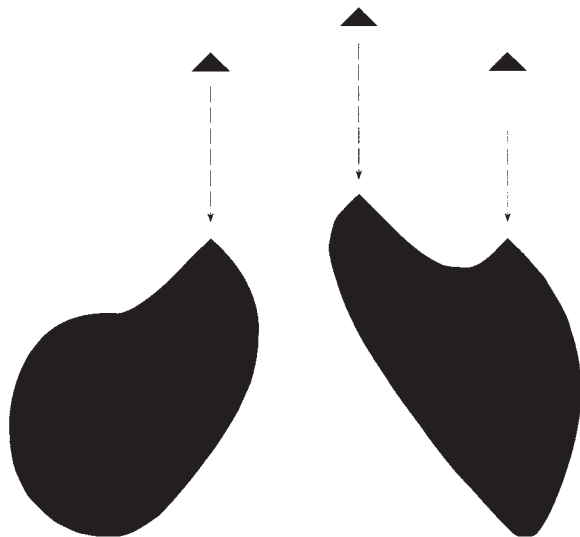
### Discussion

In this experiment, we did find an effect in the predicted direction, but its size (16 msec) is much smaller than that previously reported by Baylis (1994). This could reflect the removal of factors favoring the one-object condition.

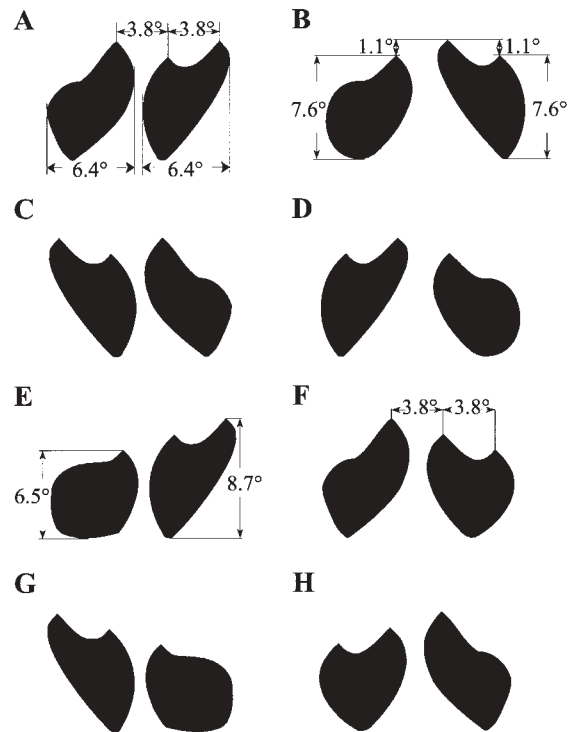
However, as a reviewer pointed out, such a conclusion is not warranted yet, because there are still some critical issues lingering about the stimulus. The orientation of the gap covaries with the manipulation of the number of objects. It is possible that the horizontal gap, by acting as a reference line, favors the one-object condition. This would yield an overestimation of the two-object cost. Another issue is the possibility that subjects saw, in the two-object condition, a single white object partly occluded by a black bar. If so, the 16 msec would actually be an underestimation of the two-object cost.

These possible confounds are inherent to the display type; they cannot be removed. We therefore designed a completely different type of display, shown in Figure 3.

Essentially, the display is formed by a symmetrical configuration of three identical isosceles triangles, with the left and right triangles both being positioned lower than the central one (see Figure 3, upper part). Given such a configuration of triangles, both the spotlight theory and the hierarchical coding-of-location hypothesis predict that the time needed to judge the vertical position of the central triangle relative to either the left or the right triangle would be the same. However, the predictions diverge when one of the two outer triangles—either the left or the right one—is part of a single object together with the cen-



**Figure 3.** Upper part: Manufacturing of the displays for Experiment 2 starts with three identical isosceles triangles, two of which (the left and the right ones) are positioned symmetrically relative to the third (the central one). Lower part: The triangles are then connected by smooth curves. The top vertices of the triangles are the only vertices left. When this display is mirrored vertically around the central triangle, there is a congruence of triangle position.



**Figure 4.** Examples of the displays used in Experiment 2. **A–D:** In the experimental condition, the correct answers are the same for the two on-screen instructions. Notice that C and D are the mirror images of A and B. **E–H:** In the control condition, the correct answers are different for the two on-screen instructions. G and H are the mirror images of E and F. There is a large height difference between the central and the outer vertices of all the displays shown. The displays with small height differences look almost identical, except that the 1.1° is replaced by 0.75°, the 7.6° by 7.95°, and the 6.5° by 7.2°.

tral triangle (double vertex shape), and the remaining triangle is incorporated into another object (single vertex shape; see Figure 3, lower part). The hierarchical coding-of-location theory predicts a two-object cost. Judging the relative height of triangles that do not belong to the same object will take more time than comparing the height of triangles that do.

Examples of the stimuli are shown in Figure 4. The double vertex shape always contains the central triangle. The shape is on the left of the display when it contains the left triangle, and it is on the right when it contains the right triangle. The three top vertices of the triangles are the only vertices left in the display because the triangles are connected by smooth curves. The subjects' task was to judge whether the central vertex is higher or lower on the screen than the outer vertex to be taken into account at that trial. The subject was informed regarding which of the two outer vertices was relevant by an on-screen instruction (L or R) presented before the appearance of the display. An L indicated that the central and the left vertices had to be compared, and an R indicated that the central and the right vertices had to be compared.

The on-screen instruction and the outer triangle incorporated into the double vertex shape together determine the object condition. When the on-screen instruction was L and the left triangle was incorporated into the double vertex shape (double vertex left), the subject performed a single-object judgment. When the on-screen instruction was R, the same display yielded a two-object judgment. When the double vertex shape was on the right, the relation between on-screen instruction and object judgment was reversed.

If the hierarchical coding-of-location hypothesis is correct, one would expect an interaction between double vertex position and on-



screen instruction. When the double vertex shape is on the left, an on-screen instruction L should result in smaller reaction times than an instruction R; when the double vertex shape is on the right, the opposite should hold. The spotlight theory does not predict any interaction.

To vary the double vertex position, we used mirror images. Varying the double vertex position precludes a confound between position and object condition. This is in contradistinction to Baylis and Driver's (1993) displays, in which the two objects were always at the outside of the displays. Because we used all possible combinations of mirror image and on-screen instruction, we did not have to assume that either L and R responses or mirror images were equivalent. We did not expect any effect of mirror image, but such an effect would not be pertinent to the two-object cost hypothesis in any case. We only had to worry about possible biases in a single display that would favor one of the object conditions over the other. There were some biases pointed out in the displays of Experiment 1, but we believe that they are absent from the present displays. Because the three vertices are made from identical isosceles triangles, biases in the spatial scale of the comparisons are impossible. Moreover, the distance between the two vertices that have to be compared is the same for both object conditions, so the retinal eccentricity is identical, too. Finally, the two objects in a display have both the same area and the same width.

## EXPERIMENT 2

### Method

**Participants.** The 11 subjects (7 women and 4 men) were psychology undergraduates at the University of Nijmegen. All had normal or corrected acuity by self-report. They received either course credit or a small payment (\$5) for their participation.

**Apparatus and Materials.** The experiment was conducted on a 486 PC with 15-in. SVGA (800 × 600 pixel) monitor. The shapes were presented in white on a black background. There were two types of experimental displays (Figures 4A and 4B) and two types of control displays (Figures 4E and 4F). The control displays were included to ensure task compliance. There were two height differences to prevent subjects from using the strategy of comparing retinal locations. We used mirror images to vary the position of the double vertex shape. Combining these factors yielded a total of 16 displays. Another 16 displays were constructed following the same rules yielding a total of 32 different displays. Examples of the displays are shown in Figure 4. All shapes have equal area and equal width (6.4°). The horizontal distance between the top vertices of the triangles was always 3.8°. The vertical distance between the top vertices of the triangles was either 0.75° (small height difference) or 1.1° (large height difference). The heights of the shapes were between 6.5° and 8.7°.

The shapes were constructed by fitting curves to the triangles. The connection between curves and triangles was smooth, which means that the only vertices left in the display were the top vertices of the three triangles.

**Design.** There were four within-subjects factors in the experiment, each with two levels: the position of the double vertex shape (left or right), the on-screen instruction (L [left] or R [right]), the height difference between the vertices (0.75° or 1.1°), and the condition (experimental [Figures 4A–4D] or control [Figures 4E–4H]).

**Procedure.** The subjects sat in a normally lit room. Their task was to decide whether the central vertex was higher or lower than the outer vertex (the on-screen instruction, displayed before the presentation of the fixation cross, determined which of the two outer vertices the central vertex had to be compared with). When the central vertex appeared higher, subjects had to push the upper button of a button box, and when the central vertex seemed lower, they had to push the lower button. Reaction times were recorded in milliseconds.

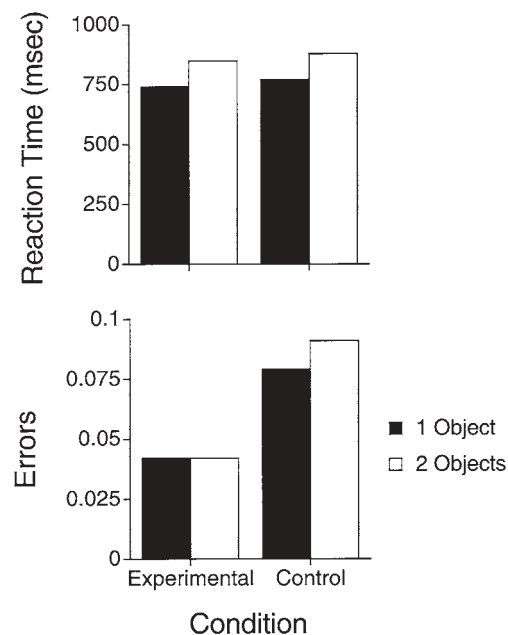
The subjects received a written instruction and a training session consisting of 80 trials that contained stimuli not used in the experimental session itself. Stimulus presentation ran as follows: The on-screen instruction (either an L or an R) was presented for 2,000 msec. A fixation cross was then presented for 500 msec, followed by the display for 150 msec. The subject responded, and 1,000 msec after the answer, a new on-screen instruction appeared and a new sequence started. The central vertex always appeared at the position of the fix-

ation cross. The position of the fixation cross itself was varied within a 0.75° square. There were six blocks of 64 trials. Half the trials required an upper response, and half required a lower response.

**Analysis.** There were 384 trials available for each subject. Of these, 192 were experimental trials, and the rest were control trials. Trials with a reaction time below 300 msec or above 2,000 msec were discarded (2.7%). All of the remaining trials, both correct and incorrect, were used in the accuracy analysis. Incorrect trials were not used in the reaction time analysis.

### Results

Mean reaction times and mean error rates are shown in Figure 5. A four-way (condition × on-screen instruction × height difference × double vertex position) repeated measures analysis of variance (ANOVA) on the reaction time data yielded significant effects for condition [ $F(1,10) = 6.75, p < .03$ ], a significant interaction between condition and on-screen instruction [ $F(1,10) = 10.55, p < .01$ ], and the predicted significant interaction between double vertex position and on-screen instruction [ $F(1,10) = 13.52, p < .005$ ]. Because of the interaction between condition and on-screen instruction, we decided to analyze the conditions separately. In the experimental condition, only the expected interaction between double vertex position and on-screen instruction was significant [ $F(1,10) = 12.61, p < .006$ ]. All the other interactions and main effects were nonsignificant ( $ps > .09$ ). The control condition yielded similar results: a significant interaction between on-screen instruction and double vertex position [ $F(1,10) = 13.39, p < .005$ ]. All the other effects were nonsignificant, although the main effect of on-screen instruction was only barely so [ $F(1,10) = 4.68, p < .056$ ].



**Figure 5. Experiment 2: Mean reaction times (upper panel) and proportion errors (lower panel) for the experimental and control conditions. The one-object condition is shown in black, and the two-object condition is shown in white.**

A four-way repeated measures ANOVA on the accuracy data revealed significant effects of condition [ $F(1,10) = 11.11, p < .008$ ] and an interaction between condition and double vertex position [ $F(1,10) = 6.48, p < .03$ ]. We again analyzed the two conditions separately. A three-way repeated measures ANOVA on the accuracy data of the experimental condition yielded a significant effect of on-screen instruction [ $F(1,10) = 5.67, p < .039$ ]. All the other effects were nonsignificant ( $ps > .3$ ). A three-way repeated measures ANOVA on the accuracy data of the control condition did not yield any significant effects ( $ps > .1$ ).

In the experimental condition, there was both a small nonsignificant difference in reaction time in favor of the R on-screen instruction and significantly more errors in the R on-screen instruction, whereas in the control condition there was a nearly significant difference in reaction time in favor of the L on-screen instruction and no difference in accuracy. It seems that, in the experimental condition, subjects increased their reaction speed during the R on-screen instruction at the cost of accuracy. Because L and R instructions were balanced in the one- and the two-object conditions, this possible speed-accuracy tradeoff had no influence on the testing of the hierarchical coding-of-location hypothesis.

## GENERAL DISCUSSION

Baylis and Driver (1993) rightly stressed that to avoid a bias, stimuli in the one- and two-object conditions must be as similar as possible physically. They designed a paradigm that allowed the one- and two-object stimuli to be physically identical, with the only difference between the two conditions being one of instruction to the subjects. However, the advantage of the physical identity of their stimuli was obtained at the cost of uncontrolled differences of the perceptual tasks their subjects had to perform in the two conditions. Clearly, the perceptual tasks in both conditions must be as similar as possible, too. In fact, it is perceptual identity that one should strive for. Physical identity is only a means to that end, but it is certainly not sufficient. Gibson (1994) has already suggested that the difference in reaction time between the one- and two-object conditions in Baylis and Driver's experiments at least partially reflects the difference in time required to produce a preferred versus a nonpreferred interpretation of the same display.

Obviously, the adequacy of Baylis and Driver's (1993) perceptual set paradigm to test the two-object cost hypothesis critically depends on the degree of perceptual ambiguity of the displays during the very short time (150 msec) they are presented; the one- and two-object interpretations of a display must be equally strong. But, as we have argued, the very nature of Baylis and Driver's displays makes sufficient ambiguity unlikely.

We therefore had to take another route. With our displays, we avoided the problems of alleged perceptual ambiguity and of perceptual set enforced by instruction. In our first experiment, we found an effect of 16 msec. However, the displays used were still vulnerable to criticisms. In our second experiment, we removed the confound of gap orientation and object condition, as well as the possibility of an occlusion interpretation in the two-object condition. When we tested the hierarchical coding-of-location hypothesis in this methodologically adequate way, we found an advantage for the one-object condition. Its size was 110 msec. The discrepancy between this result and the outcome of Experiment 1 is explained by the possibility of an occlusion interpretation in the two-object condition (Figure 2B). In that case, the effect of 16 msec could actually be called an occlusion cost.

The two-object cost of 110 msec is much larger than the one reported by Baylis (1994). Considering the fact that we removed all aspects favoring the one-object condition, we would have expected the opposite. As suggested in the introduction, it seems, therefore, that the result reported by Baylis is an underestimation of the true size of the two-object cost. This would be possible if the two-object condition in Baylis's experiment actually favored a mosaic interpretation. The two-object condition would then actually be a perverted one-object condition, and the difference found by Baylis could be traced entirely to the difficulty of (not) getting the right percept.

Another possibility is that the subjects used different strategies in the control and the experimental conditions in Baylis's (1994) experiment. If they followed their task in the control condition (Figures 1K and 1L), but always used the central object in the experimental condition (Figures 1I and 1J), the two-object cost found would be much too small. The experimental and control displays look too different to dismiss this possibility as improbable. Such a strategy would not be reflected in any difference in error rates between the two conditions. The difference in reaction times in the control condition of Baylis's experiment, in which subjects did have to attend the right color to avoid errors, was larger (about 50 msec, still very much smaller than our result), but the eccentricity of the points that had to be compared was different, too.

In the design of Experiment 2, there was no such difference between the control and the experimental conditions. However, in the control condition, task compliance was required in order to avoid errors. The fact that the control condition yielded almost the same results as the experimental condition strengthens the confidence in the estimate of 110 msec. First, it suggests that the subjects probably complied with their task instructions in the experimental condition, too; otherwise a different size estimate would have been found. Second, because two conditions yielded the same estimate, the true size of the two-object cost is probably indeed in the neighborhood of 110 msec.

The two-object cost follows naturally from the hierarchical coding-of-location hypothesis, without the necessity of an exact characterization of the nature of the frames of reference. Nor is it necessary to give a precise definition of the parts of an object. However, to develop the hierarchical coding-of-location hypothesis further, some serious consideration should be given to these topics. Baylis and Driver (1995) recently gave a definition of parts of an object based on Hoffman and Richards's (1984) minima rule, but questions concerning the frames of reference are still to be answered. For instance, how is the location of an object described relative to a scene-based frame of reference? What is the orientation of the object-centered frames of reference? Are these orientations independent, or is there a general orientation for all object-centered frames? How are the frames of reference derived? Answering questions like these is a real challenge for the adherents of the hierarchical coding-of-location hypothesis.

## REFERENCES

- BAYLIS, G. C. (1994). Visual attention and objects: Two-object cost with equal convexity. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 208-212.
- BAYLIS, G. C., & DRIVER, J. (1993). Visual attention and objects: Evidence for hierarchical coding of location. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 451-470.
- BAYLIS, G. C., & DRIVER, J. (1995). Obligatory edge-assignment in vision: The role of figure- and part-segmentation in symmetry detection. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 1323-1342.
- DUNCAN, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, **113**, 501-517.
- GIBSON, B. S. (1994). Visual attention and objects: One versus two or convex versus concave? *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 203-207.
- HOFFMAN, D. D., & RICHARDS, W. A. (1984). Parts of recognition. *Cognition*, **18**, 65-96.

- KANIZSA, G. (1979). *Organization in vision: Essays on gestalt perception*. New York: Praeger.
- PETERSON, M. A., & GIBSON, B. S. (1994). Object recognition contributions to figure-ground organization: Operations on outlines and subjective contours. *Perception & Psychophysics*, **56**, 551-564.
- POSNER, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, **32**, 3-26.
- ROCK, I. (1986). The description and analysis of object and event perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 33-1 to 33-71). New York: Wiley.
- WATT, R. J. (1988). *Visual processing: Computational, psychophysical, and cognitive research*. Hillsdale, NJ: Erlbaum.

(Manuscript received April 15, 1996;  
revision accepted for publication March 2, 1997.)