Perception, 1981, volume 10, pages 573-578

Do background depth gradients facilitate object identification?

Irving Biederman
Department of Psychology, State University of New York at Buffalo, 4230 Ridge Lea Road, Amherst, New York 14226, USA
Received 26 November 1980, in revised form 7 March 1981

Abstract. A number of experiments have demonstrated that an object in a coherent scene can be more accurately identified than when in a jumbled scene or a display of unrelated objects. An experiment was designed to test whether the identification of an object (from a set of six) would be faster and more accurate if the objects were presented against a background gradient that conferred a unifying depth effect on the six objects. No benefit of this background was found. In fact, the accuracy of the identification of objects against a depth background was found to be lower than that of objects which were presented against either no background or a control regular grid background. It is concluded that the perceptual advantage of objects appearing in a well-formed scene derives from semantic relations that are defined between such objects.

1 Introduction

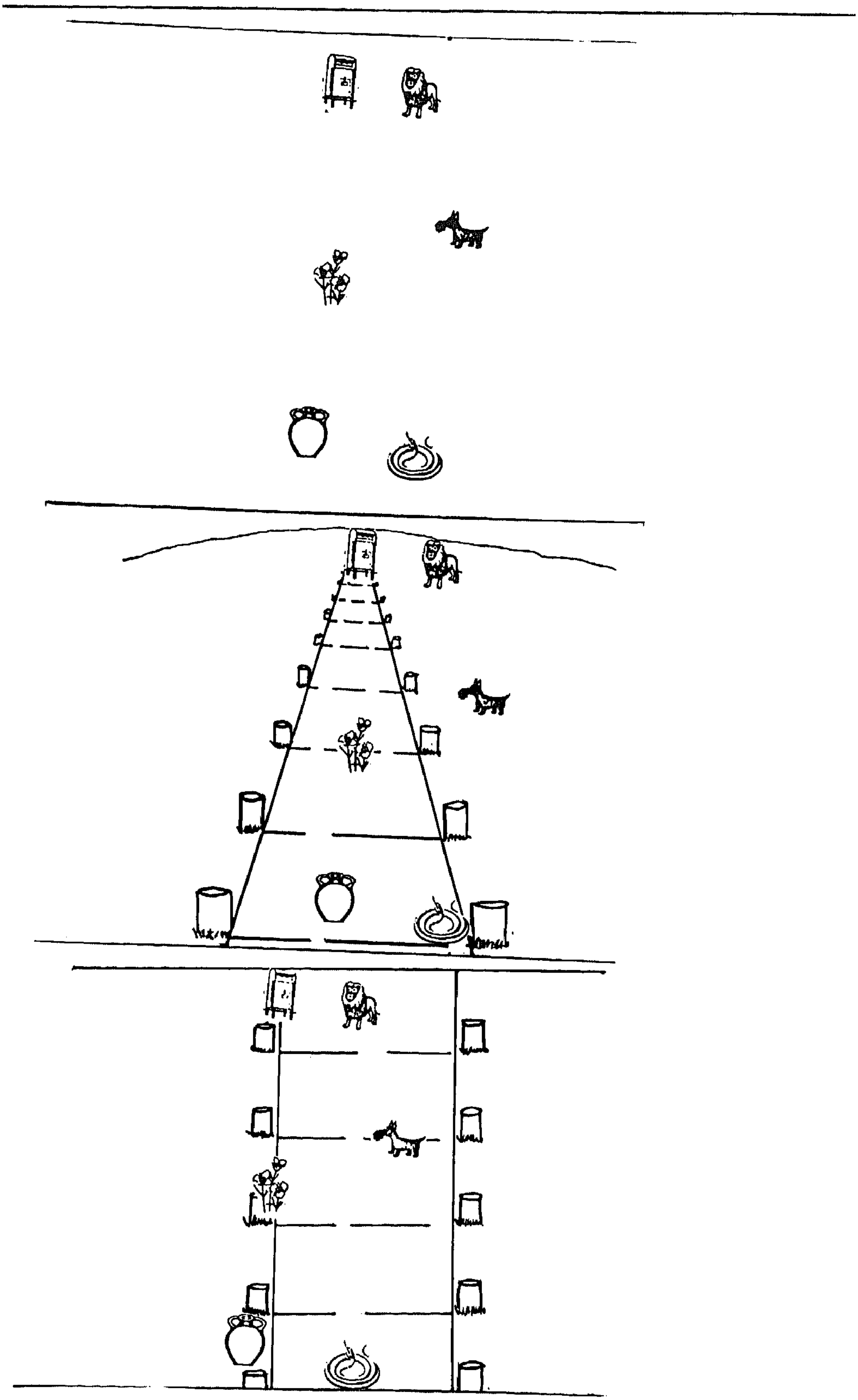
If a picture of a scene is divided into several sections and the sections are rearranged so as to destroy the coherency of an object's context, the perceptibility of that object is reduced (Biederman 1972; Biederman et al 1974). Why does this effect occur?

One possibility is that object identification is adversely affected by the inappropriateness of the *semantic relations* among the objects in the jumbled scene. Thus, for example, if the jumbling places a fire hydrant so that it appears to be in the middle of a street or floating above a truck, then its perceptibility relative to its normal location on a sidewalk will be reduced. Several experiments have confirmed this effect by showing that the perceptibility of objects was reduced when the objects were placed in incongruous positions or appeared to be floating in otherwise intact scenes (Biederman 1981; Biederman et al 1982; Klatsky et al 1980; Teitelbaum and Biederman 1979; Friedman 1979).

The purpose of the present experiment was to evaluate a second possible effect produced by the jumbling. This effect, which is not exclusive of the effect of an object's semantic relations, will be termed the *parsing-facilitation* hypothesis. When pictures were jumbled, not only were the semantic relations among the objects disrupted, but the consistency of the texture and depth gradients was also disrupted. There is some evidence that the visual system, when faced with a number of objects against a background, automatically 'parses' the scene for depth so that the objects conform to a consistent depth perspective as seen from a unique viewing point (Gibson 1966; Farber and Rosinski 1978; Hagen 1976). When a scene is jumbled, this consistency is destroyed, so that a unique viewing point cannot be determined.

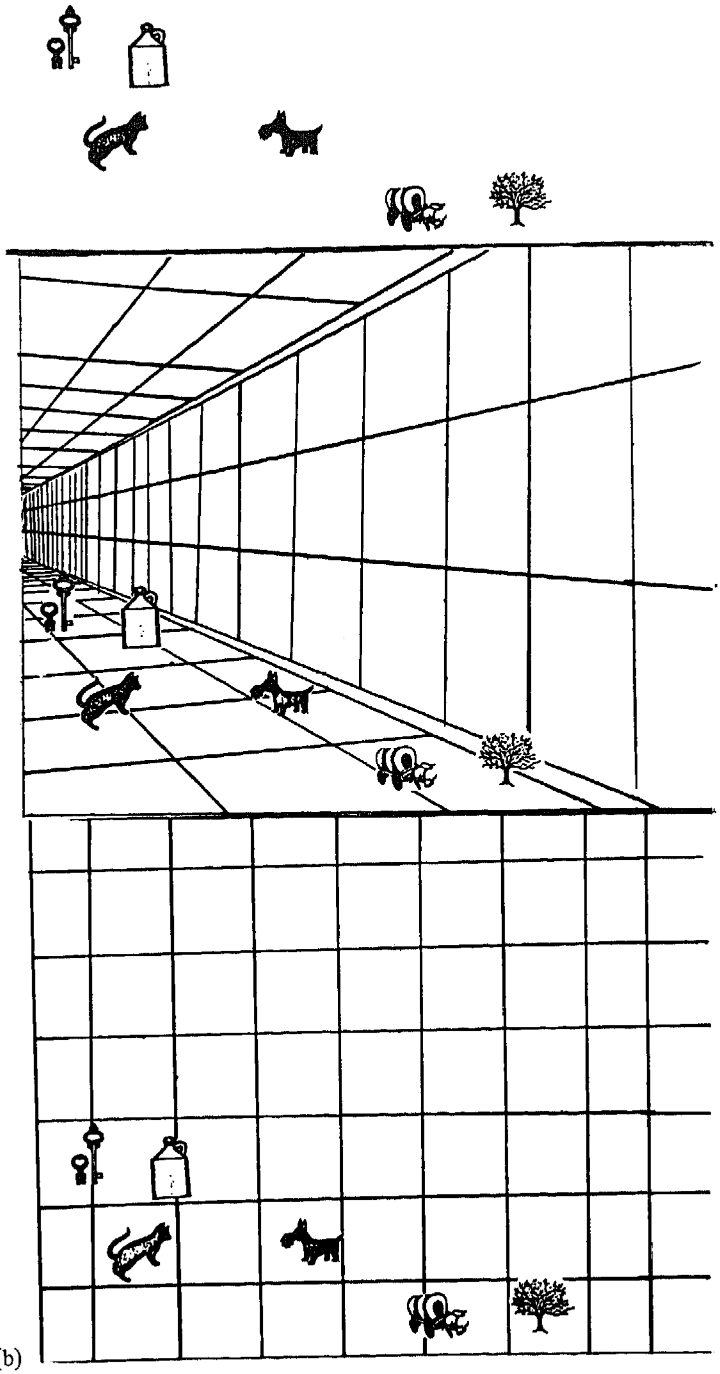
If the capacity required to determine the consistency of depth is also involved in the identification of objects, then the disruption of the depth information would retard identification independently of any disruption of semantic relations.

Although the parsing-facilitation hypothesis has never received direct experimental test, two findings lend it some credence. First, depth information is available sufficiently early to affect same-different judgments and, a fortiori, object identification. When subjects attempted to judge whether two simultaneously presented figures had the same shape, their reaction times (RTs) were slowed if the figures differed in size (Larsen and Bundesen 1978). That depth information can



(a)

affect these judgments has been demonstrated by Blount (1979). He showed that if the figures were superimposed against background depth gradients so that the (objectively) smaller figure appeared to be the same size but only farther away than



(b)

Figure 1. Two sets of six objects shown against blank backgrounds (top panels), a depth background (middle panels), and a control grid background (bottom panels). In set (a) the objects against the depth background ('path') are arranged in a size compatible fashion; in set (b) the objects in the middle panel ('corridor') are arranged so as to violate size relations.

the (objectively) larger figure, the deleterious effect of size disparity on RTs for shape judgments was reduced.

A second set of findings which are compatible with the parsing-facilitation hypothesis comes from those studies demonstrating that memory for a set of objects is facilitated when the objects are arranged into a 'well-formed' scene even if the interactions are semantically implausible (Epstein et al 1960). In memory for images what determines whether there will be a beneficial effect of imagery instructions is not the request to visualize or the plausibility of the relations among the objects, but rather the formation of a spatial interaction among the objects (Bower 1970). As long as the objects are related spatially, there will be a benefit from the imagery instructions (Bower 1970). These results led to theoretical accounts holding that this dramatic benefit of pictorial or imaginal relations on memory depends on the compatibility of such relations with the output of a perceptual parser (Anderson and Bower 1973).

The research literature thus supports a memory effect of depth relations with the possibility of a perceptual effect. The issue, then, is whether the integration of a set of objects into an array with a consistent depth perspective would function as a kind of perceptual mnemonic facilitating the identification of these objects. Specifically, would the presence of background depth gradients (such as those shown in figures 1a and 1b) facilitate object *identification* (not just memory) of a set of unrelated objects?

2 Experiment

2.1 *Subjects*

Thirty-six students, aged 17 to 34 years, participated in the experiment as part of their introductory course requirement.

2.2 *Stimuli*

Drawings of common objects (and creatures) taken from a pool of fifty-four were grouped in twenty-one sets of six objects and mounted on transparent acetate sheets. Each sheet was then Xeroxed with different backgrounds:
(i) A *depth* background which resembled a path, corridor, or tracks and which conferred (by subjective judgment) a depth effect. The path and corridor are shown in the middle panels of figure 1. The depth effect of these gradients was sufficient to produce incongruities in familiar size as shown in the middle panel of figure 1b.
(ii) A control background *grid* which contained about the same number and type of lines as the depth backgrounds but did not induce a depth effect (bottom panels of figure 1).
(iii) A *blank* background (top panels of figure 1).

The three backgrounds with twenty-one different arrangements of six objects yielded sixty-three different scenes. Of the twenty-one depth scenes, eleven induced incongruities in the apparent sizes of the objects, as shown in figure 1b. The sixty-three different scenes were made into slides.

2.3 *Apparatus and procedure*

The slides were projected onto a viewing screen by a two-channel projection tachistoscope which was made by mounting Gerbrand Shutters on two Kodak Carousel projectors. The scenes subtended a viewing angle of 13 deg horizontally and 8·5 deg vertically. All cued objects subtended a visual angle of less than 1 deg.

Subjects were fully instructed as to the nature of the task and were told to respond "as fast and as accurately as possible". The procedure followed a modification of the Averbach and Coriell (1961) poststimulus cueing paradigm. A sequence of events on a given trial went as follows. The subject first read, aloud, the name of the target object from a card taken from a deck of cards of object names. The subject then stepped on a foot-switch to start the trial and 0·5 s later a scene was flashed for 200 ms, immediately followed by a cue, a black dot drawn on an acetate slide, for 500 ms. The cue designated a position in the scene which had been occupied by an object. If the cued object corresponded to the object named on the card, the subject was to press the YES key; otherwise the subject was to press the NO key. RTs were recorded from the onset of the cue to the depression of the key. Two minutes intervened between blocks.

2.4 *Design*
Each subject viewed all sixty-three experimental scenes. The scenes were grouped into three blocks of twenty-one scenes, each composed of equal numbers (seven) of depth, grid, and blank backgrounds. A given array of six objects appeared only once in each block. For a given subject a different object was cued in that array each of the three times it appeared. However, objects, background types, and block orders were all balanced across subjects, so that a given object in a given array was cued equally with depth, grid, and blank background at the same mean block position.

3 Results
Surprisingly, the depth background resulted in the *highest* error rate, 48·7%, compared to 37·8% and 38·8% error rates for the grid and blank backgrounds, respectively ($F_{2,\,270} = 9 \cdot 21$; $p < 0 \cdot 001$). The higher error rates for the depth background did not reflect a tradeoff of speed for accuracy, since the RTs for the depth, grid, and blank backgrounds were 1373, 1355, and 1356 ms, respectively ($F_{2,\,70} < 1 \cdot 00$). Within the depth backgrounds there was no effect of the incongruity in object sizes. RTs and error rates for inappropriately sized and normally sized objects were virtually identical. Error rates did not decline over the three trial blocks but RTs did. From the first to the third block, RTs averaged 1531, 1325, and 1228 ms, respectively ($F_{2,\,70} = 14 \cdot 88$; $p < 0 \cdot 001$).

4 Discussion
The depth backgrounds did not facilitate perception. This effect cannot be explained in terms of interference (eg lateral inhibition or additional display elements) furnished from the depth backgrounds, since there was no difference in either RTs or error rates between the grid and blank backgrounds.

An obvious explanation for the lack of benefit of the depth backgrounds is that perspective information is simply independent of the pattern recognition process for individual objects. Although there might ultimately be an effect of spatial relations on memory, this view would hold that such effects come after objects are identified.

Although this position is compatible with the results from the present experiment, it is inconsistent with the results of experiments, cited in section 1, which showed that the violation of the semantic relations among individual objects in well-formed scenes does retard their identification (eg Biederman 1981).

The remaining possibility is that the spatial relations will have an effect on object identification only to the degree to which they aid in the achievement (or are consequences) of some unitary semantic representation of a scene. That is, if the collection is a scene—say, a kitchen—then the perspective information could aid in the achievement of a schema for the kitchen which, in turn, might facilitate the recognition of some of the objects. With the present depth stimuli, however, no such semantic representation could be achieved. Processing resources directed toward achieving a semantic integration of the objects to the depth background would be wasted. In this sense, the depth backgrounds could have functioned as large, distracting objects.

5 Summary and conclusion

Objects in coherent scenes are more accurately identified than when their contexts are jumbled. This experiment evaluated a depth-parsing contribution to that result by determining whether the presence of a background depth gradient would facilitate the perception of one of six objects presented against that background. No benefit from the presence of depth backgrounds was found. In fact, the accuracy of identification of objects against the depth background was found to be lower than when the objects were presented against either no background at all or a control non-depth grid background. The equivalence of the non-depth grid and blank backgrounds rule out explanations based on lateral inhibition or additional display elements as an account of the absence of a beneficial effect of the depth background. The depth background may simply have served as a large, irrelevant object to which the other objects could not be related. Depth integration will only facilitate the identification of objects if it aids in the achievement of a consistent semantic representation.

Acknowledgements. This research was supported by research grants from the US Army Research Institute for the Behavioral and Social Sciences (Grant MDA903-A-G-0003) and National Institutes of Mental Health (Grant MH33283) to Irving Biederman. The author also held an NIH Postdoctoral fellowship (MH07891). The author thanks Caroline Bowker for preparing the stimuli and running the subjects in a pilot study and Nora E Roberts for running the subjects and coding the data for analysis. Roberta L Klatzky generously provided the object drawings and drew the background gradients for the stimuli.

References

Anderson J R, Bower G H, 1973 *Human Associative Memory* (Washington, DC: Winston)

Averbach E, Coriell A S, 1961 "Short-term memory in vision" *Bell System Technical Journal* **40** 309-328

Biederman I, 1972 "Perceiving real world scenes" *Science* **177** 77-80

Biederman I, 1981 "On the semantics of a glance at a scene" in *Perceptual Organization* Eds M Kubovy,

Biederman I, Mezzanotte R J, Rabinowitz J C, 1982 "Scene Perception: Detecting and judging objects undergoing relational violations" *Cognitive Psychology* **14** in press

Biederman I, Rabinowitz J C, Glass A L, Stacy E W, Jr, 1974 "On the information extracted from a glance at a scene" *Journal of Experimental Psychology* **103** 597-600

Blount J, 1979 "The effect of depth cues and pictorial size on 'sameness' judgments"; paper presented at the meeting of the Psychonomic Society, Phoenix, AZ, November 1979

Bower G H, 1970 "Imagery as a relational organizer in associative learning" *Journal of Verbal Learning and Verbal Behavior* **9** 529-533

Epstein W, Rock I, Zuckerman C B, 1960 "Meaning and familiarity in associative learning" *Psychological Monographs* **74** (4), whole issue

Farber J, Rosinski R R, 1978 "Geometric transformations of pictured space" *Perception* **7** 269-282

Friedman A, 1979 "Framing pictures: The role of knowledge in the automatized encoding and memory for gist" *Journal of Experimental Psychology: General* **108** 316-355

Gibson J J, 1966 *The Senses Considered as Perceptual Systems* (Boston: Houghton Mifflin)

Hagen M A, 1976 "Influence of picture surface and station point on the ability to compensate for oblique view in pictorial perception" *Developmental Psychology* **12** 57-63

Klatsky G J, Teitelbaum R C, Mezzanotte R J, Biederman I, 1980 "Evidence for mandatory processing of contextual information in real-world scenes": paper presented at the meeting of the Eastern Psychological Association, Hartford, CT, April 1980

Larsen A, Bundesen C, 1978 "Size scaling in visual pattern recognition" *Journal of Experimental Psychology: Human Perception and Performance* **4** 1-20

Teitelbaum R C, Biederman I, 1979 "Perceiving real-world scenes: The role of a prior glance" *Proceedings of the Human Factors Society* **23** 456-460