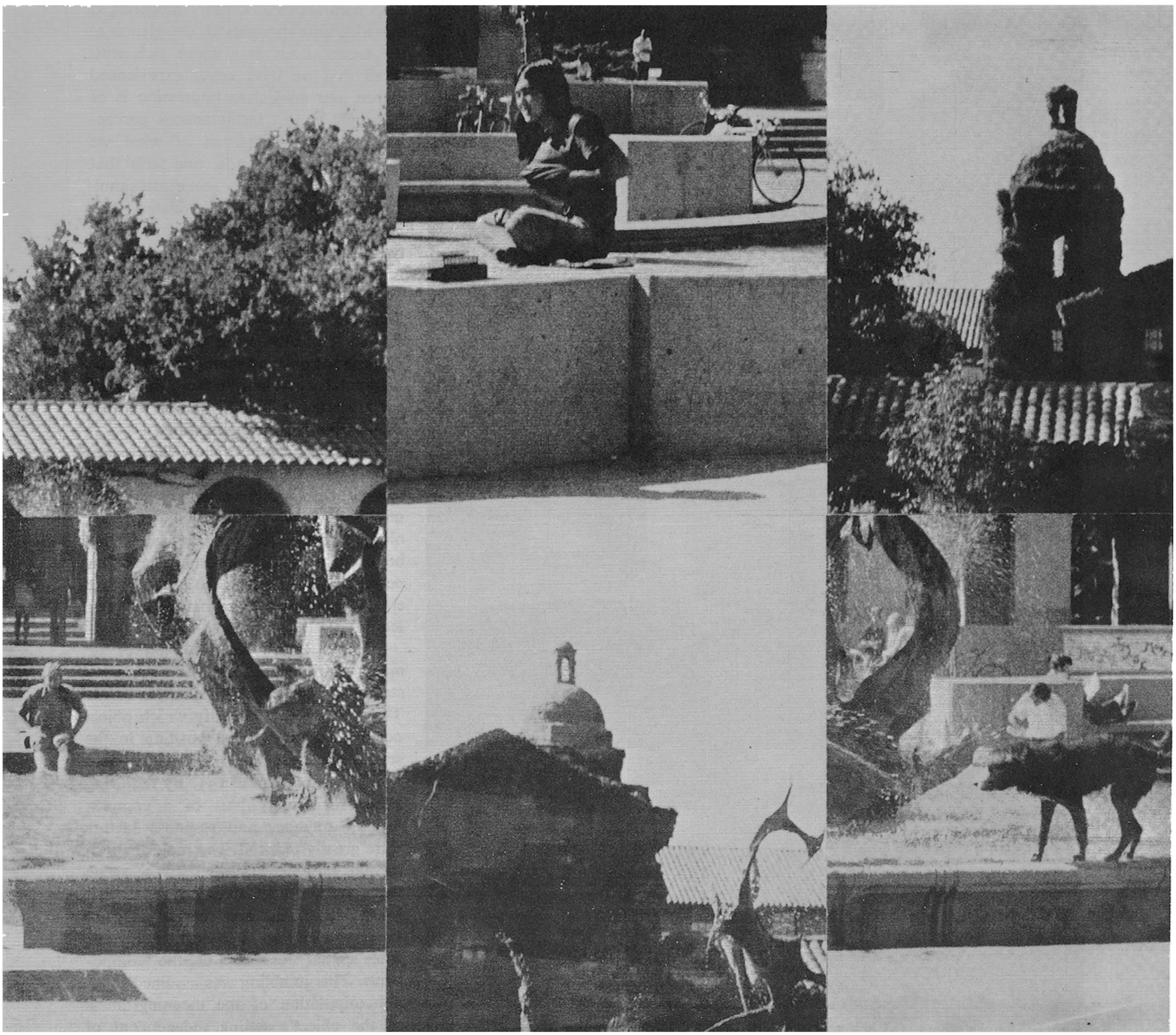
Reprinted from
7 July 1972, Volume 177, pp. 77-80

SCIENCE

Perceiving Real-World Scenes

Irving Biederman



Perceiving Real-World Scenes

Abstract. *When a briefly presented real-world scene was jumbled, the accuracy of identifying a single, cued object was less than that when the scene was coherent. Jumbling remained an effective variable even when the subject knew where to look and what to look for. Thus an object's meaningful context may affect the course of perceptual recognition and not just peripheral scanning or memory.*

In experiments on perceptual recognition, a subject typically sees either a single item surrounded by homogeneous space or an array of unrelated ("random") items. In the real world, such meager perceptual experiences are rare. Outside the laboratory, objects are almost always perceived in some setting or context.

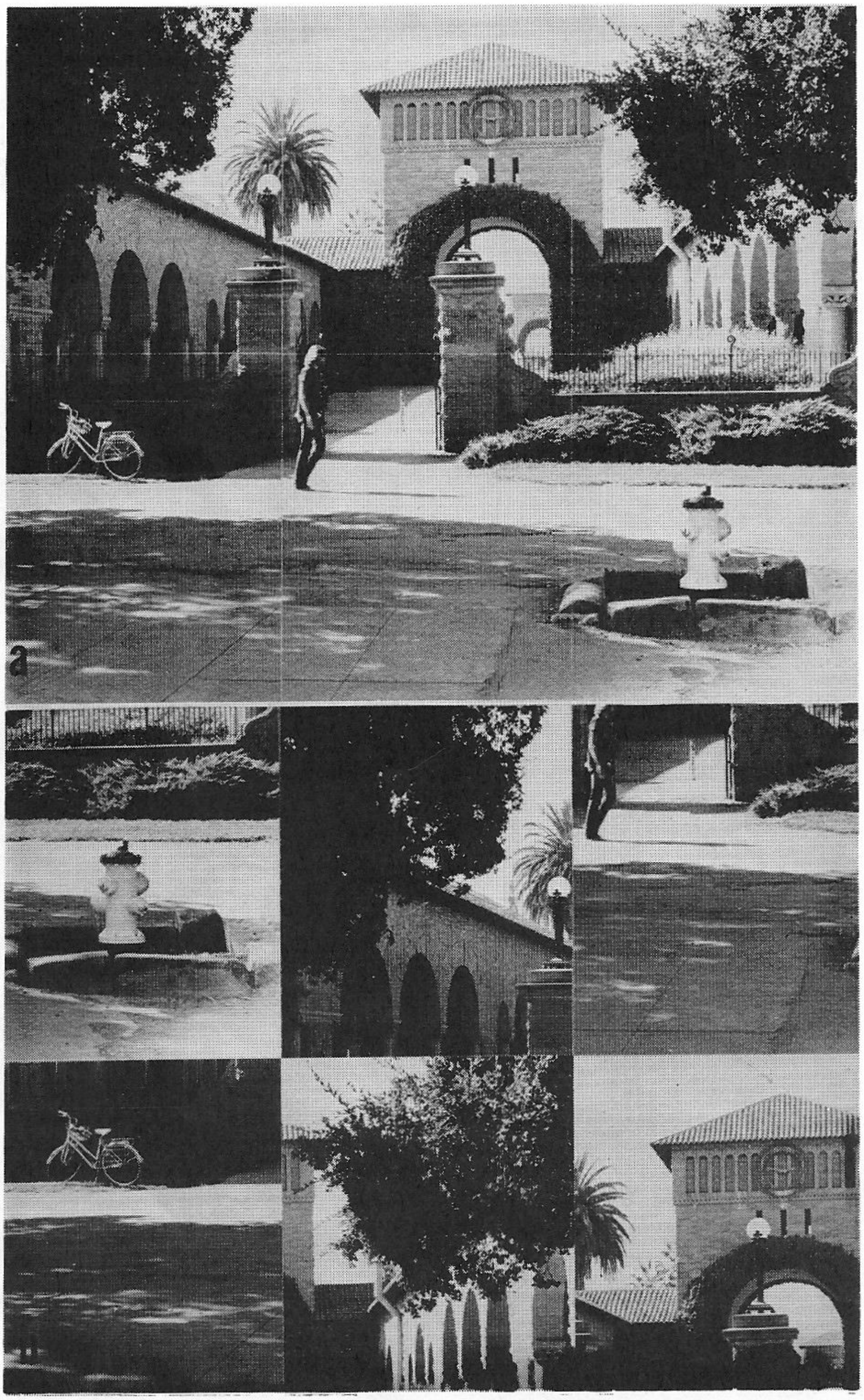


Fig. 1. Sample scenes, (a) coherent and (b) jumbled. Note that the lower left section is the same in both versions. The bicycle would have been the cued object.

Given conventional stimulus displays, it is not surprising that the results of much perceptual research can generally be reconciled with a class of models that hold that the various items of the display are treated as separate entities; that is, they are initially processed independently in a very short-term sensory store (lasting just fractions of a second), and then transferred serially to a longer-term storage system [see (*1*)]. It is in this longer-term storage system that meaningfulness and long-term memory are seen as having their effects.

In contrast to laboratory modeling is the following thought. If we glance at the world, even at a scene rich with detail that we have never experienced before, our subjective impression is of clear and almost instantaneous perception and comprehension of what we are looking at. That is, one feels that the various parts of a scene are simultaneously identified and related. One possible source of this discrepancy between the laboratory and the real world is the presence, in the real world, of a meaningful context. Creatures and things in the real world rarely appear, as they typically do in the laboratory, surrounded only by homogeneous space or unrelated entities. Instead, things occur in some predictable relation to other things, that is, in some setting.

The results I report show that meaningful context does affect perceptual recognition (*2*). A secondary purpose of this study was to advance a methodology whereby real-world scenes could be used as stimuli in experiments on perceptual recognition, so that context effects could be studied more systematically (*3*).

Subjects briefly viewed pictures of many varied scenes: for example, streets, kitchens, desk tops, and so forth. Their task was to identify which object occupied a given cued position in the scene. The technique was derived from Averbach and Coriell (*4*). By requiring a report of only part of a complex display, memory and response factors were greatly reduced. The major experimental variable was whether the scene was coherent or whether it was jumbled —cut into sixths and rearranged (but never rotated) so as to destroy the natural spatial relations of the components. This jumbling was assumed to be a manipulation of the meaningfulness of the object's setting independent of the complexity of the scene.

The scenes were 35-mm black-and-white positive slides. For each scene, two versions—one coherent and one jumbled—were made by photographing a print, 20 by 25 cm, which had been cut into six sections (generally with one horizontal and two vertical cuts) so that at least four well-defined objects were left intact (Fig. 1). The coherent slide was taken after the sectioning, so that the section lines appeared in both versions (*5*). When sections were arranged for the jumbled version, one was left in its original position. This section always contained at least one well-defined object. The position of the section remaining constant was balanced across the different scenes; for example, for one-sixth of the scenes, the top left section was identical in both jumbled and coherent versions.

Subjects viewed slides in a three-channel tachistoscope (*6*). Slides were shown for 300, 500, or 700 msec, and they subtended visual angles of 5° horizontally and 3.5° vertically. An arrow was presented for 300 msec, immediately after the scene in half the trials and immediately before the scene in the other half. The arrow pointed to an area associated with an object. The subject's task was to indicate, by pointing to one of four object pictures, which object had been cued. These object pictures were cut from the original print used in making the scene and were mounted on index cards displayed in a photo album. The cued object was the same in both coherent and jumbled versions of each scene and always came from the section of the scene that remained in its original position.

In addition to the jumbling and cue-order variables, the order in which the subject viewed the scene and the response alternatives was also varied. In the alternatives-before condition, the subject was allowed to peruse the four object pictures before the scene was shown. In the alternatives-after condition, the subject viewed the response alternatives only after he viewed the scene. Thus, in the latter condition a few seconds elapsed between presentation of the scene and presentation of the response alternatives.

Each of 24 subjects viewed four blocks of 32 slides each. Within each block, the slides were equally divided between jumbled and coherent scenes. The four possible combinations of alternative-order and cue-order variables were used, one combination in each block. The first eight slides in each block were considered practice trials and were not included in the data analysis. In the remaining 24 slides, the cued object occurred in each of the six sections an equal number of times (four). All variables were balanced across slides. Each subject viewed only the jumbled or coherent version of a given scene but never both.

Results are shown in Fig. 2. The effects of jumbling [$F(1,22) = 5.17$, $P < .05$], cue order [$F(1,22) = 27.98$, $P < .001$], and alternative order [$F(1,22) = 9.01$, $P < .01$], were all significant (*7*). The effect of jumbling tended to be reduced in the cue-before and the alternative-before conditions, although none of the interactions among jumbling, cue order, and alternative order approached significance. However, the variability of these data was not low enough to warrant acceptance of the null hypothesis. The same must be said for varying presentation times of scenes; the effects were neither consistent nor significant.

The experiment was designed to minimize peripheral-scanning effects (since the cued object was in the same position in both scene versions, and since brief durations and relatively small visual angles were used) and to minimize memory and response effects (by requiring the subject to simply point to one of a small set of well-defined and nameable objects). That jumbling remained an effective variable even when the subject knew where to look (when the cue preceded the scene) and what to look for (when the response alternatives preceded the scene) further limits the roles played by peripheral scanning and memory factors, respectively, in accounting for the jumbling effect. It is most likely that jumbling affected an early, but not peripheral, stage involved in the perceptual recognition of the cued object.

A number of theoretical issues present themselves when one attempts to account for the context effect, that is, the advantage of coherent over jumbled scenes. One issue concerns identification of the functional units involved in the perception of scenes. Is the functional unit an individual object, or does an observer have access to more global units or schema? A second issue is the determination of the locus, in the sequence of processing, where context has its effect. Is it in the initial manner in which objects are physically processed—in the initial segmentation, testing, and weighing of features? Or does the context influence a stage subsequent to

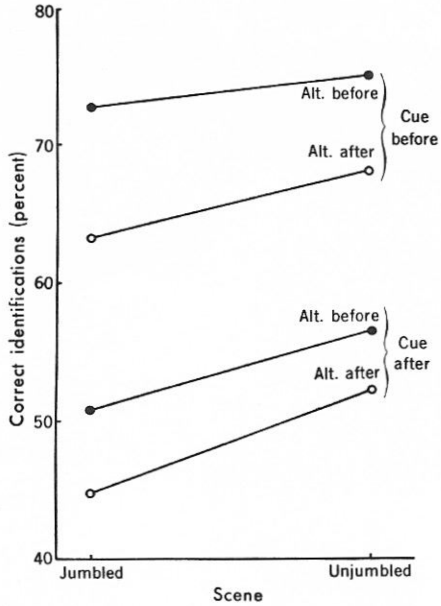


Fig. 2. Correct identifications (mean percentage) as a function of scene version and order of presenting the cue and response alternatives (alt.).

that involved in the physical processing, so that physically ambiguous stimuli are interpreted to be consistent with other aspects of the scene already identified?

This experiment was not analytic for these issues, but Sternberg's additive factors method (*8*), coupled with reaction time measurements in the present task situation, might bring these issues under experimental scrutiny (*9*). For example, if jumbling is affecting a cognitive inferential stage, then an interaction would be expected between the magnitude of the effect of jumbling on reaction times and the magnitude of the effect of the probability of the cued object's being in the scene (*10*). (For example, this probability could be varied by cueing a bowl or a baseball glove on a formally set dining room table.) In a similar manner, interactions between jumbling and (i) the size and contrast of the cued object or (ii) the presence or absence of background and contiguous areas would be expected if jumbling were affecting physical-feature testing or object segmentation, respectively.

IRVING BIEDERMAN*

Department of Psychology, State University of New York at Buffalo, Buffalo 14226

References and Notes

1. D. E. Rumelhart, *J. Math. Psychol.* **7**, 191 (1970).
2. In several studies with rigorous controls for guessing and memory effects, it was shown that an individual letter within a word is more perceptible than that same letter in a

string of random-appearing letters, or even than the letter by itself [D. Aderman and E. E. Smith, *Cogn. Psychol.* 2, 117 (1971); G. M. Reicher, *J. Exper. Psychol.* 81, 274 (1969); D. D. Wheeler, *Cogn. Psychol.* 1, 59 (1970)]. A possible implication of these results, as well as the results I report, is that the functional perceptual unit can be something larger than the individual items (letters or objects).

3. Research with real-world scenes as stimuli has generally been limited to studies of eye movements or memory. The thrust of this report is the effect of an object's setting on its perceptibility in a single glance.
4. E. Averbach and A. S. Coriell, *Bell Syst. Tech. J.* 40, 309 (1961).
5. The end product of four photographic cycles was less-than-optimum clarity. Undoubtedly, performance would have been better if original positive slides had been used and the scenes had been larger.
6. Model GB Auto-Tach, Scientific Prototype Corp.
7. It could be argued that the effect of jumbling is due to the presence of a greater number of object segments, which created visual noise in the jumbled pictures. Since the fragments did not overlap the cued object, it is not clear how this factor could be operative. My observation is that literally adding external noise by scattering segments of other pictures over a photograph—compared to jumbling the photograph—does little to degrade the intelligibility of the scene. That jumbling is primarily affecting the relation of objects is, perhaps, evidenced by the minimum of 10 to 15 seconds of effort needed to determine how the sections of a scene were to be reassembled to the original. A more rigorous test of the noise interpretation would involve stimulus scenes that were made by either cropping out background objects that extended across the section lines or else by drawing scenes in which none of the objects would extend across the section lines.
8. S. Sternberg, *Acta Psychol.* 30, 276 (1969).
9. Reaction times could be measured by providing the subject with a target object before the scene was presented. When the scene and cue are presented, the subject would respond, "Yes," as quickly as possible if the cued object was the target, and, "No," otherwise.
10. The logic of the additive factors method (AFM) holds that if two factors, for instance jumbling and probability, are affecting the duration of separate and independent information-processing stages, then their combined effects on reaction time (RT) should be additive. That is, if jumbling adds 50 msec to the average RT and probability adds 25 msec, then the RT for a low-probability target in a jumbled scene should be, on the average, 75 msec longer than that for a high-probability target in a coherent scene. The AFM may also be applicable to error probabilities, but, in that case, if factors are influencing different information-processing stages, the logarithms of the errors should add.
11. I thank E. W. Stacey, A. L. Glass and S. L. Cook for invaluable assistance and E. E. Smith and G. R. Lockhead for careful readings of the manuscript. Supported by NIMH special fellowship MH 50632-01 and grant 050-7201-A from the Research Foundation of the State University of New York.

On leave during the 1971–1972 academic year at the Department of Psychology, Stanford University, Stanford, California 94305.

December 1971; revised 10 January 1972

COVER

Jumbled scene used in the experimental investigation of the role of coherence in perception. The same object is more difficult to identify in a scene jumbled in this manner than in the original, coherent scene. See page 77. [Irving Biederman, State University of New York at Buffalo; on leave Stanford University, Stanford, California]