# The abstraction of schematic representations from photographs of real-world scenes

HOWARD S. HOCK
*Florida Atlantic University, Boca Raton, Florida 33431*

and

KAREN F. SCHMELZKOPF
*Pennsylvania State University, University Park, Pennsylvania 16802*

Photographs of a downtown street scene were taken from four different locations. At each location the photographs were taken in 12 different directions, 30 deg apart. In the acquisition phase of the experiments, subjects learned to identify the photographs in terms of the location from which they were taken. The testing phase took place immediately after acquisition, after a 1-week delay, after a 6-month delay, and on one occasion after a 1-year delay. Subjects were tested for their ability to identify the camera location for previously seen and new photographs, to discriminate previously seen from new photographs (i.e., recognition), and to place the camera locations on a map of the street scene. The results of six experiments provided converging evidence that subjects abstracted a schematic representation of the spatial layout of the scene from the discrete, partially overlapping photographic samples presented during acquisition.

When we perceive a three-dimensional real-world scene we do not extract visual information from the scene in a complete, continuous fashion. Saccadic eye movements and abrupt changes in head and body position produce discrete, partially overlapping samples of the visual field. Nonetheless, our experience of the scene is integrated and continuous. The traditional explanation for this phenomenon asserts that motor information (either efferent or afferent) specifying the direction of gaze for each of a series of fixations is used to "fit" the samples of visual information extracted during each fixation into a continuous, composite representation of the scene (Helmholtz, 1866/1963; Walls, 1951). This explanation, however, has been effectively refuted by Hochberg's (1978) observation that we can integrate successive, partially overlapping visual samples in perceiving motion pictures despite discontinuous transitions in camera location and direction. The discontinuous transitions are integrated without the perceiver's having access to information specifying either the locations or "directions of gaze" of the camera.

Furthermore, if subjects are given preliminary information concerning the identity of the stimulus that is being sampled (either a verbal cue or a "long shot" of the entire scene), even nonoverlapping visual samples can be integrated into a continuous perceptual representation (Hochberg, 1968, 1978). The latter observation indicates that the integration of discrete samples does not require a process of matching overlapping information in the samples to form a composite, pictorial representation of the stimulus. Instead, it has been argued that the continuous perceptual representation that is formed in viewing motion pictures is based on a three-dimensional schematic layout that is abstracted from the sequence of discrete visual samples presented to the perceiver (Hochberg, 1968, 1978; Neisser, 1967).

The research reported in this paper provides empirical evidence for schema formation in the integration of discrete samples of visual information. Our everyday experience of the visual world and the perception of motion pictures provide convincing evidence for perceptual continuity under conditions of discrete sampling. However, there is much that remains to be learned about the nature of the memory representations that result from the discrete sampling of a scene. Our memory for a previously experienced scene could involve specific representations corresponding to the particular visual samples and/or an integrated, schematic representation that is abstracted from the discrete samples.

The first step in distinguishing between specific and schematic memory representation would be to determine how well novel visual samples of the scene can be identified as part of the scene. The limitation here,

however, is that the novel samples could be identified by virtue of either their similarity to the specifically stored, original samples (i.e., via stimulus generalization) or their consistency with a schematic representation of the scene (i.e., they could be generated from the schematic representation). Distinguishing between specific and schematic memory representation requires further experimental manipulation.

The experiments performed in this study used Posner and Keele's (1970) concept formation methodology to study memory representation for real-world scenes. Posner and Keele were concerned with the abstract representation of ill-defined visual concepts involving dot patterns. The members of each of their concepts were dot patterns generated by relocating the dots from a common, base pattern. Some members of each concept (the originals) were presented in the acquisition phase and then in the subsequent testing phase of the experiments. The bases and other members of each concept (the novels) were presented only during the test phase. The test phase occurred immediately after acquisition for half the subjects, and after a 1-week delay for the other half. Posner and Keele found that the effect of delay on classification was greater for the originals than for the bases and novels; the superiority of the originals during immediate testing was significantly reduced with delayed testing. The relatively small difference in classification accuracy between the originals and novels after a 1-week delay indicated that specific representations were of minimal utility in distinguishing original from novel concept members. Nonetheless, subjects could still accurately classify the patterns according to concept membership. This provided evidence for the emergence of nonspecific, abstract representations that governed subjects' classification performance during delayed testing.

Adapting Posner and Keele's (1970) methodology to scene representation involved taking photographs of a downtown street intersection from four different camera locations. Each camera location functioned as a different concept. The membership of each concept was generated by varying the direction of the camera (in the horizontal plane) for each photograph. Twelve photographs were taken at each location, in 12 directions, 30 deg apart. Each set of 12 photographs thus composed a full panoramic view of the street scene from each camera location. During the acquisition phase of the experiment, subjects learned to group six of the photographs from each location according to the camera location from which they were taken. The subsequent testing procedure took place either a few minutes after acquisition or after a 1-week delay. The testing procedure in Experiment 1 required subjects to identify the camera location of photographs from the acquisition set (originals), as well as photographs that were taken from the same locations as the originals but were not presented during the acquisition phase (novels). The testing procedure in Experiment 2 involved a recognition

task that assessed subjects' ability to distinguish between the original and novel photographs. It was hypothesized that the two experiments would provide converging evidence that the utility of specific representations would disappear over the 1-week retention interval, and the identification of camera location would then depend on a nonspecific, schematic representation of the scene. Evidence that the schematic representation formed by the subjects in Experiment 1 was retained over a 6-month interval, and that the representation generalized to the identification of new camera locations, was obtained in Experiment 3. Experiment 4, which replicated the results of Experiments 1 and 2, eliminated foreground cues as a factor in schema formation. Evidence that schema formation occurred during the acquisition phase of the experimental procedure, not during the 1-week interval between acquisition and delayed testing, was obtained in Experiment 5. Finally, the results of Experiment 6 showed that the schematic representation formed by subjects in Experiments 4 and 5 was retained for intervals of up to 1 year. Experiment 6 also replicated the evidence, first obtained in Experiment 3, that the schematic representation generalized to the identification of new camera locations.

## EXPERIMENT 1

### Method

**Stimuli.** The stimuli used in this experiment were 35-mm color photographs taken with a Nikon EL camera at a traffic intersection in Boca Raton, Florida. The intersection was chosen because it was peripherally surrounded by distinctive buildings (both business and residential) but had sufficient open space to eliminate the possibility of a large, nearby tree or building filling an entire photograph. The photographs were taken early in the morning so that occasional traffic could be avoided and the content of the photographs kept constant. Since the sun was low in the sky, cues involving brightness and shadow direction were inherent in the photographs. The four camera locations were located near each corner of the intersection (they are denoted by the letters A, B, C, D on the map presented in Figure 1). The camera, which was mounted on a tripod, was leveled at a height of 161 cm above the ground. At each location the direction of the camera was changed for each photograph (in the horizontal plane), so that a full panorama of 12 photographs was taken at 30-deg intervals beginning at north (0 deg). Although the camera was equipped with an 84-deg lens, cropping in the production of the photographic prints reduced the visual angle of each photograph to 70 deg. The photographic prints were 12.5 cm in width and 8.8 cm in height. They were mounted on stiff cardboard with a .7-cm white border around each edge. For purposes of illustration, a pair of photographs taken at Location C, but differing in direction by 30 deg is presented in Figure 2. Also presented is a pair of photographs taken at Location D, but differing in direction by 60 deg.

The 12 photographs taken at each of the four camera locations were divided into two interleaved sets of 6 photographs, with consecutive members of each set differing in direction by 60 deg. Set 1 comprised photographs taken in directions 0 (north), 60, 120, 180 (south), 240, and 300 deg at each camera location. Set 2 comprised photographs taken in directions 30, 90 (east), 150, 210, 270 (west) and 330 deg at each camera location. Either Set 1 or Set 2 was presented during the acquisition phase of the experiment. During acquisition, therefore,
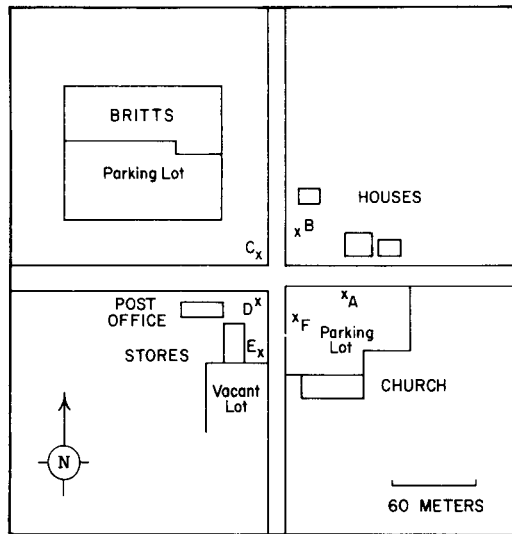
Figure 1. Map of the intersection at which photographs were taken. Photographs from Locations A, B, C, D were presented in all the experiments. Photographs from Locations E and F were presented only in Experiments 3 and 6.

consecutive photographs from the same location differed in direction by 60 deg, and the information in consecutive photographs overlapped on each edge by about 10 deg (14%). Sets 1 and 2 were combined during the testing phase of the experiment. During testing, therefore, consecutive photographs from the same location differed in direction by 30 deg, and the information in consecutive photographs overlapped on each edge by about 40 deg (57%). The alignment of consecutive photographs from the two sets, in relation to the various camera directions, is illustrated in Figure 3.

The net effect of the photography was that virtually all the objects in the street scene (e.g., buildings, trees, parked cars) were "captured" by photographs taken at all four camera locations. That is, the camera locations could not be differentiated on the basis of some objects appearing in photographs taken from one location, but not in photographs taken from the other locations. Rather, camera locations were differentiated by variation in the sizes, shapes, and perspectives of the common set of objects as they appeared in the photographs taken from the different locations.

Procedure. During the acquisition phase of the experiment subjects in one group worked with a deck of 24 photographs comprising the six Set 1 photographs (0, 60, 120, 180, 240, and 300 deg) from each of the four camera locations. The 24-card acquisition deck for the second group of subjects comprised the six Set 2 photographs (30, 90, 150, 210, 270, and 330 deg) from each of the four camera locations. Subjects were told they would



Figure 2. Sample of photographs taken at Location C (top pair) and Location D (bottom pair).
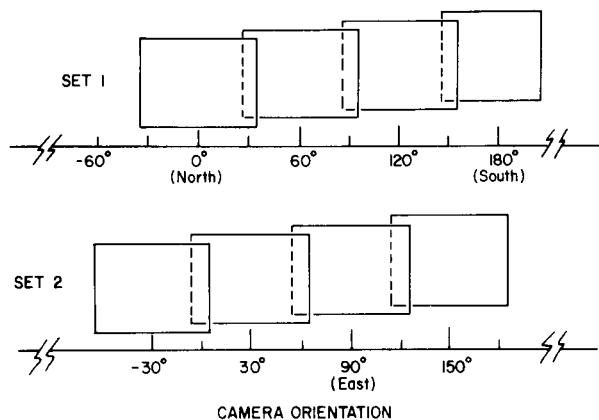
Figure 3. Illustration of the alignment of directions for photographs taken from the same location. The photographs are divided into interleaved sets (1 and 2).

be seeing photographs taken from four different locations within 500 ft of each other and that the photographs from each location formed a panoramic view. The subjects' task took the form of a standard concept formation procedure. They were instructed to sort the photographs into one of four quadrants on a large table, placing the photographs taken from the same camera location into the same quadrant. Printed labels A, B, C, and D, corresponding to each camera location, were placed in the four quadrants (A in the upper left quadrant, B in the upper right, C in the lower left, and D in the lower right). The assignment of camera locations to the quadrants of the tabletop did not directly correspond to the camera locations on the four corners of the intersection.

Each acquisition trial involved sorting the deck of 24 randomly ordered photographs into the four quadrants. Subjects were allowed to change the quadrant in which they placed each photograph anytime during the 10-min interval allowed for each trial. At the end of each trial the experimenter provided corrective feedback by placing the incorrectly sorted photographs into the correct quadrant. The experimenter avoided placing the photographs in consecutive order (e.g., 0, 60, 120 deg, . . .) within each quadrant. Subjects were given 1 min to examine the corrected sort before the photographs were reassembled into a deck and their order rerandomized for the next trial. Following this procedure, subjects were trained to a criterion of two consecutive perfect sorts through the deck. For a trial to be "perfect" all the photographs taken from the same location had to be grouped together and placed in the correct quadrant on the table.

Subjects who successfully reached the acquisition criterion participated in the subsequent testing phase. In the testing phase subjects were instructed to identify the camera locations for the photographs by making the appropriate verbal response (A, B, C, or D). Although the photographs were not sorted into quadrants during the testing phase, the printed verbal labels could be seen in each quadrant (as in acquisition). There was no limit on the time allowed for subjects to respond, but a response was required for each photograph before the next one was presented. The testing stimuli comprised the full set of 48 photographs, with Sets 1 and 2 mixed together in a random sequence that was different for each subject. Set 1 comprised the originals and Set 2 the novels for one group of subjects, and vice versa for the other group of subjects. Within each group, half the subjects were tested a few minutes after acquisition and half were tested after a 1-week delay. Subjects in the delay condition were instructed not to go to the street intersection during the 1-week delay. The data for the one subject who violated this instruction were discarded.

At the conclusion of the experiment, each subject was provided with an 8½ x 11 in. map similar to Figure 1 and asked to place the four camera locations, including labels, on the map.

**Subjects.** Forty-eight students at Florida Atlantic University, 24 males and 24 females, participated in the entire experiment, without pay. Only 32 of the subjects were included in the final data tabulation. Of these, 16 indicated that they were unfamiliar with the street intersection and 16 indicated that they were familiar with it. The experimental design was completely balanced within each group of 16. Sixteen subjects were excluded from the final data tabulation because their degree of familiarity with the intersection was difficult to evaluate. These subjects typically indicated that they had driven through the intersection but had little recollection of the buildings surrounding the intersection.[1] Also excluded from the final data tabulation were six subjects who failed to reach criterion during the acquisition phase.

## Results

The results of the experiment are presented in Table 1. Subjects required an average of 5.5 trials (including criterion trials) to reach the acquisition criterion of two perfect sorts. There was little difference in average number of acquisition trials between the subjects assigned to the immediate testing condition (5.4) and the subjects assigned to the delayed testing condition (5.5). Similarly, there was little difference in average number of sorting trials between the unfamiliar (5.4) and familiar (5.5) subjects.[2]

An average error rate of 14.3% in the verbal identification of camera location was obtained for the original photographs (from the acquisition set) when testing followed immediately after acquisition. This drop-off in performance after subjects were trained to perfect accuracy in the acquisition phase reflected the change from the simultaneous-sorting task used in the acquisition phase to the sequential-identification task used in the testing phase. Subjects' familiarity with the intersection had little influence on the accuracy with which they identified camera location. An analysis of variance on percent identification errors indicated that the difference in identification accuracy between the familiar and unfamiliar subjects was not significant [$F(1,28) = 1.02$, $p > .05$, $MSe = 425.3$]. Although there was a tendency toward more accurate identification for the familiar

Table 1
Experiment 1: Mean Number of Acquisition Trials and Mean Percent Errors in the Identification of Camera Location

| Testing | Acquisition Trials | Identification Errors | | |
| --- | --- | --- | --- | --- |
| | | O | N | N – O |
| Unfamiliar Subjects | | | | |
| Immediate | 5.6 | 18.2 | 33.9 | 15.7 |
| Delayed | 5.3 | 34.9 | 37.0 | 2.1 |
| Familiar Subjects | | | | |
| Immediate | 5.3 | 10.4 | 20.3 | 9.9 |
| Delayed | 5.8 | 34.9 | 38.0 | 3.1 |

*Note—Subjects were either unfamiliar or familiar with the scene depicted in the photographs prior to their participation in the experiment. They were tested immediately following acquisition or after a 1-week delay. O = original; N = novel.*

subjects during immediate testing, the interaction between familiarity and time of test (immediate vs. delay) was not significant [$F(1,28) = 1.13$, $p > .05$, $MSe = 25.5$]. Since the primary concern of the study was with the formation of schematic representations of the scene by subjects who were unfamiliar with it, we wanted to be sure that the statistical results did not depend on an analysis that pooled the data for the two groups. Further statistical analyses were therefore performed separately for the subjects who were unfamiliar with the scene and the subjects who were familiar with it. The results were essentially the same for both groups. That is, camera location was identified more accurately for original photographs (from the acquisition set) than novel photographs when testing occurred immediately after acquisition, but the original-novel difference was relatively small when testing was delayed by 1 week. An analysis of variance on percent identification errors indicated that the interaction between time of test (immediate vs. delay) and prior presentation (original vs. novel) was significant for the unfamiliar subjects [$F(1,14) = 9.86$, $p < .01$, $MSe = 37.0$] as well as for the familiar subjects [$F(1,14) = 5.56$, $p < .05$, $MSe = 14.1$]. Tests of simple effects indicated that, when testing took place immediately after acquisition, the difference in identification accuracy between the original and novel photographs was significant for both the unfamiliar subjects [$F(1,14) = 26.41$, $p < .001$, $MSe = 37.0$] and the familiar subjects [$F(1,14) = 27.78$, $p < .001$, $MSe = 14.1$]. However, when testing occurred after a 1-week delay, the difference between the originals and novels was not significant for either the unfamiliar subjects [$F(1,14) < 1.0$, $MSe = 37.0$] or the familiar subjects [$F(1,14) = 3.70$, $p > .05$, $MSe = 14.1$].[3]

Subjects' placement of camera locations on maps of the intersection were scored on the basis of whether or not the camera locations were assigned, by verbal label, to the correct corners of the intersection. For both the unfamiliar and familiar groups, 13 of 16 subjects correctly placed all four camera locations on the map.[4] The success of most subjects in correctly placing camera locations on the map was not due to a simple translation from the four quadrants of the tabletop used for acquisition to the four corners of the map. This was because the labels assigned to the quadrants did not correspond to the corners of the intersection (e.g., Camera Location C was on the northwest corner of the intersection but was assigned to the lower left quadrant of the tabletop).

### Discussion

For immediate testing, camera location was identified significantly more accurately for the original than for the novel photographs. This indicated that specific memory representations formed for the members of the acquisition set could influence the identification of camera location. Otherwise, an original-novel difference would not have been obtained. When testing occurred

1 week after acquisition, the original-novel difference was not significant, indicating that there was minimal utilization of specific representations with delayed testing. Nonetheless, performance after 1 week (average error rate of 36%) was substantially better than chance (75% error rate). In the absence of an influence of specific representations of previously seen photographs, it could be concluded that nonspecific information was abstracted from the original photographs and generalized to novel photographs that partially overlapped the originals. Subjects' success in placing the camera locations on a map of the intersection provided additional evidence that the abstracted schematic information involved the spatial layout of the scene.

The absence of a familiarity effect in this experiment was not surprising. The identification of camera location for each photograph required relatively precise information concerning distance and perspective. For the subjects in the familiar group, preexperimental familiarity with the scene was limited to experience driving through the intersection. It is conceivable that familiarity would have had a greater impact on performance if it involved visual experience acquired while subjects were walking through the area appearing in the photographs.

### EXPERIMENT 2

In Experiment 1 subjects' use of specific representations in the identification of camera location was indicated, during immediate testing, by more accurate identification for the original photographs (from the acquisition set) than the novel photographs. The present experiment provided a direct test of the availability of specific representations of the photographs presented during acquisition. Instead of identifying camera location for each photograph, subjects were required to indicate whether or not they had seen each photograph before. This recognition test took place immediately after acquisition or following a 1-week delay.

### Method

The photographs, design, and acquisition procedure were the same as in Experiment 1. Either Set 1 or Set 2 was presented during the acquisition phase, and Sets 1 and 2 were combined during the testing phase. If Set 1 was presented during acquisition, the Set 2 photographs served as distractors in the recognition test. If Set 2 was presented during acquisition, the Set 1 photographs served as distractors in the recognition test. The recognition test required subjects to indicate, by a "yes" or "no" response, whether they had seen each photograph before. The subjects assigned to the immediate and delayed testing conditions were balanced in the extent to which they were familiar with the street intersection depicted in the photographs. Sixteen students at Florida Atlantic University participated in the entire experiment, without pay. Another five subjects failed to reach criterion in the acquisition phase and were dropped from the remainder of the experiment.

### Results

The results of the experiment are presented in

| Testing | Acquisition Trials | Recognition Errors | | |
|---|---|---|---|---|
| | | M | FA | Mean |
| Immediate | 5.8 | 21.9 | 26.0 | 24.0 |
| Delayed | 6.1 | 49.5 | 49.5 | 49.5 |

*Note—Subjects were tested immediately following acquisition or after a 1-week delay. M = miss; FA = false alarm.*

Table 2. Subjects required an average of 6.0 trials (including the criterion trials) to reach the acquisition criterion of two perfect sorts. There was little difference in average number of acquisition trials between the subjects assigned to the immediate testing condition (5.8) and the subjects assigned to the delayed testing condition (6.1). There was no indication of response bias in the yes-no recognition test. The recognition data indicated that original and novel photographs were discriminated more accurately when testing occurred immediately after training than when testing took place after a 1-week delay [$F(1,14) = 51.39$, $p < .001$, MSe = 50.7]. Performance in the delayed testing condition was at chance (50%). This outcome was also obtained when a signal detection measure of recognition accuracy was used.

The results of Experiments 1 and 2 thus provide converging evidence that specific representations of particular photographs influenced performance for immediate, but not delayed, testing. This convergence was important because the interaction between time of test (immediate vs. delayed) and prior presentation (original vs. novel photographs) obtained for the identification of camera location in Experiment 1 was of the type that could be removed by a monotonic transformation of the dependent measure: "percent errors" (Loftus, 1978). The results of Experiment 2, however, confirmed that "percent identification errors" in Experiment 1 properly reflected the underlying memory process. For immediate testing, the results of Experiment 1 indicated that specific representations influenced identification accuracy, and the results of Experiment 2 indicated that specific representations were accessible. For delayed testing, the results of Experiment 1 indicated that specific representations had little apparent influence on identification accuracy, and the results of Experiment 2 indicated that specific representations were not accessible. The removal of the interaction of Experiment 1 by a monotonic transformation on the dependent measure would thus result in a pattern of data that would be inconsistent with the interpretation confirmed by the converging evidence of Experiment 2.

## EXPERIMENT 3

This experiment had three purposes. The first was to examine the long-term retention of the schematic representations formed in Experiment 1. To this end,

subjects who were unfamiliar with the street intersection prior to their participation in Experiment 1 were retested 6 months later. None of the retested subjects reported that they had been at the intersection during the intervening months. The second purpose of the experiment was to further examine the generalizability of subjects' schematic representations of the scene. This was done by testing subjects' ability to judge camera locations for photographs taken from camera locations that had not been used in Experiment 1. The final purpose of the experiment was to show that subjects could discriminate between the four camera locations without foreground cues (e.g., differences in the ground surface at different camera locations) that might have contributed to their judgments in Experiment 1. This was accomplished by having subjects judge camera location with foreground information in each photograph masked by an opaque cover.

Since it was unlikely that subjects would remember the verbal labels assigned to the four camera locations after 6 months, the testing procedure was altered in order to eliminate the need for response labels to identify camera location. Instead, subjects were presented pairs of nonoverlapping photographs and were required to judge whether the photographs in each pair were taken from the same or different locations.

## Method

Twenty-four pairs of photographs were formed from the 48 photographs used in the testing phase of Experiment 1. Twelve of the pairs comprised photographs taken from the same camera location (three pairs from each location), and 12 comprised photographs taken from different camera locations. The 12 "different" pairs were composed of pairwise combinations from all four camera locations, with each camera location equally represented. An additional six pairs, three "same" and three "different," comprised photographs taken from new locations. These locations are designated by Points E and F on the map presented in Figure 1. Whether the photographs in each pair were taken from the same or different locations, the information in the two photographs was always nonoverlapping. The minimum difference in angular direction between the pairs of photographs was 90 deg. This eliminated the possibility of subjects' basing their same-different judgments on the relative distances and perspectives of objects common to the two photographs. Finally, in order to eliminate potential information from foreground cues, the bottom 45% of all the photographs used in the experiment was covered by a strip of white, opaque paper.

There were thus a total of 30 same-different pairs of photographs presented, in a different random order, to each subject. Subjects were required to indicate whether the photographs in each pair were taken from the same or different locations. They were cautioned that the photographs in each pair were always taken in different directions, so that they always depicted different objects, regardless of whether they were taken from the same or different locations.

Three groups of six subjects participated in the experiment. The first group comprised six subjects who had been unfamiliar with the street scene depicted in the photographs prior to their participation in Experiment 1, 6 months earlier. The second group comprised six subjects who were unfamiliar with the intersection and had received no previous training with the photographs. The third group comprised six subjects who were

familiar with the intersection but had received no previous training with the photographs.

## Results

Mean percentage errors in the same-different judgment of camera location are presented in Table 3. An examination of the data provided little indication of response bias. There was, at most, a slight tendency toward subjects' making more "different" than "same" responses. The critical finding was that subjects retested 6 months after training performed at a mean error rate of 28.5%, substantially better than the chance rate of 50%. For photographs taken from the old locations (A, B, C, D), the retested subjects' scores ranged from error rates of 12.5% to error rates of 37.8%. In addition to this evidence for long-term retention, the data show that foreground cues were not necessary for the discrimination of camera location. Although subjects reported that they attended to foreground information when they participated in the experiment 6 months previously, they successfully identified camera location in this experiment when the foreground information was covered. Finally, the results for the retest subjects provide evidence for the generalizability of scene schemata. Same-different judgments of camera location were only slightly less accurate for pairs of photographs taken from new locations (E, F) than for pairs of photographs taken from the old locations (A, B, C, D).

The results described above were obtained for subjects who were unfamiliar with the intersection prior to training and had not been to the intersection subsequent to training. For the group of subjects who were also unfamiliar with the intersection, but had received no training with the photographs, same-different judgments of camera location were at chance. The results for these subjects were expected, since the experiment was designed so that there was no information common to any pair of photographs. Performance was also at chance for the group of subjects who were familiar with the intersection but had received no training with the photographs. An analysis of variance on percent errors in the same-different judgment of camera location indicated that neither the difference in performance between old and new locations [$F(1,15) < 1.0$, MSe = 175.7] nor the interaction between location (old vs. new) and groups [$F(2,15) = 1.10$, $p > .05$, MSe = 175.7]

was significant. There was, however, a significant difference among the three groups of subjects [$F(2,15) = 13.15$, $p < .005$, MSe = 228.3]. Newman-Keuls comparisons indicated that the retested subjects performed significantly more accurately than either of the groups that had received no previous training with the photographs ($p < .01$). The difference between the latter two groups was not significant ($p > .05$). Finally, an examination of the data for all three groups of subjects provided no evidence of practice effects. That is, there was no indication that subjects developed a schematic representation of the street scene during the course of the test trials.

## EXPERIMENT 4

The purpose of this experiment was to replicate the results of Experiments 1 and 2, but with three modifications. First, foreground cues that were potentially distinctive to photographs taken from the same camera location were eliminated from both the acquisition and test phases of the experiment. Second, subjects did not have to learn verbal labels corresponding to the four camera locations during the acquisition phase. This was because the testing phase did not require subjects to verbally identify the camera location of each photograph. As in Experiment 3, subjects were presented pairs of nonoverlapping photographs (there were no objects that appeared in both members of a pair) during the testing phase. They were required to judge whether pairs of "original" photographs (from the acquisition phase) as well as pairs of "novel" photographs (not presented during acquisition) were taken from the same location or from different locations. Third, the need to rely on subjects' reports of their familiarity with the intersection was eliminated. In this experiment we could be certain that the subjects, students at Pennsylvania State University, were totally unfamiliar with the intersection appearing in the photographs.

## Method

Procedure. The same photographs of a traffic intersection in Boca Raton, Florida, that were used in the previous experiments were used in this experiment. In order to eliminate foreground cues as a factor in the experiment, the bottom 45% of each photograph was masked by an opaque, white cover during both the acquisition and testing phases of the experiment.

The acquisition phase of this experiment was identical to that of Experiments 1 and 2 in all respects but one. In the previous experiments, printed labels (A, B, C, and D) were used to identify the four quadrants into which subjects sorted the photographs. Subjects learned to group the photographs in the acquisition set according to camera location and also to place each grouping into the quadrant designated by the experimenter. In contrast, no verbal labels were involved in the acquisition phase of this experiment. Subjects again learned to group the photographs according to camera location, but there were no restrictions on the quadrants into which the groupings could be placed.

During the testing phase of the experiment, subjects were presented 24 pairs of photographs formed from the full set of

### Table 3
### Experiment 3: Mean Percent Errors for "Same-Different" Judgment of Camera Location

|  | Camera Location | |
| --- | --- | --- |
|  | Old | New |
| Unfamiliar Subjects—Retested* | 25.7 | 30.6 |
| Unfamiliar Subjects—No Training | 52.1 | 63.9 |
| Familiar Subjects—No Training | 54.2 | 50.0 |

*Retested 6 months after training.

48 photographs. Four kinds of photograph pairs were generated from the orthogonal combination of prior presentation (originals vs. novels) and location (same vs. different). Twelve pairs were formed from the 24 "original" photographs presented during acquisition (either Set 1, which comprised photographs taken in directions 0, 60, 120, 180, 240, and 300 deg at each camera location, or Set 2, which comprised photographs taken in directions 30, 90, 150, 210, 270, and 330 deg at each camera location). Another 12 pairs were formed from the 24 "novel" photographs (Set 2 if Set 1 was presented during acquisition, and vice versa). For each set of 12, six pairs comprised photographs taken from the same camera location and six pairs comprised photographs taken at different locations. Whether the camera locations were the same or different, there were no objects that appeared in both members of a pair. The photographs paired together were taken in directions that differed by at least 120 deg. As indicated previously, this prevented subjects basing their same-different decisions on the relative distances and perspectives of objects common to the paired photographs. For a given subject, all four camera locations could not be equally represented within the six "same" pairs or the six "different" pairs. However, since the pairings of photographs varied from subject to subject, equal representation of the four camera locations within "sames" and "differents" was approximated on a groupwise basis.

The "same" and "different" pairs were presented, one at a time, in random order. Before beginning the testing phase, subjects were cautioned that the photographs in each pair were always taken in different directions, so that they always depicted different objects, regardless of whether they were taken from the same or different locations. For each pair, subjects indicated whether the photographs were taken from the same or different locations (identification test). Before proceeding to the next pair, they also indicated whether or not they had seen the photographs in the preceding portion of the experiment (recognition test). As in Experiments 1 and 2, subjects were tested either immediately or after a 1-week delay.

At the conclusion of the experiment, subjects were provided with a map similar to Figure 1. They were shown samples of photographs taken from each camera location and were asked to indicate the location of the camera on the map.

**Subjects.** Sixteen students, eight male and eight female, from Pennsylvania State University voluntarily participated in the entire experiment, without pay. All were unfamiliar with the street intersection depicted in the photographs. Another three subjects failed to reach criterion in the acquisiton phase and were dropped from the remainder of the experiment.

## Results

The results of the experiment are presented in Table 4. The acquisition phase in this experiment required an average of 7.5 trials (including the criterion trials), 2.0 and 1.5 more than the number of trials required to reach the same acquisition criterion as in Experiments 1 and 2, respectively. Since the acquisition

procedure was otherwise the same, the difference between this experiment and Experiments 1 and 2 could be attributed to the absence of foreground cues in this experiment. This conclusion was consistent with the reports of subjects in Experiment 1 who reported that foreground cues had helped them learn the various camera locations.

Although elimination of the foreground cues resulted in somewhat slower acquisition than in Experiments 1 and 2, the results of the testing phase completely replicated the results of Experiments 1 and 2. As in Experiment 1, there was evidence (from immediate testing) that specific representations of members of the acquisition set could influence the accuracy with which camera locations were identified. After a 1-week delay, however, better than chance identification accuracy was obtained without the involvement of specific representations. That is, same-different judgments of camera location were more accurate for original photographs (from the acquisition phase) than for novel photographs, but only when the testing phase immediately followed acquisition. When testing occurred 1 week after acquisition, the original-novel difference was eliminated. An analysis of variance on percent errors in the same-different judgment of camera location indicated that the interaction between the effects of time of test (immediate vs. delay) and prior presentation (originals vs. novels) was significant $[F(1,14) = 12.50, p < .005, MSe = 38.9]$. Tests of simple effects indicated that, when the testing phase immediately followed acquisition, the originals were identified significantly more accurately than the novels $[F(1,14) = 16.07, p < .005, MSe = 38.9]$, but when the testing phase was delayed for 1 week, the original-novel difference was not significant $[F(1,14) < 1.0, MSe = 38.9]$.[5]

The results for the yes-no recognition test indicated that the original photographs were discriminated from novel photographs significantly more accurately when testing occurred immediately after training than when testing took place after a 1-week delay $[F(1,14) = 73.14, p < .001, MSe = 60.8]$. Performance in the delayed testing condition was again at chance (50%). The results of the recognition test thus replicated the results of Experiment 2, the only difference being that subjects exhibited a strong bias to respond "yes" in the immediate testing condition of this experiment. The data obtained for the same-different judgment of camera

Table 4
Experiment 4: Mean Number of Acquisition Trials, Mean Percent Errors in the "Same-Different" Identification of Camera Location, and Mean Percent Errors in Yes-No Recognition Test

| Testing | Acquisition Trials | Identification Errors | | | Recognition Errors | | |
|---|---|---|---|---|---|---|---|
| | | O | N | N − O | M | FA | Mean |
| Immediate | 7.1 | 11.5 | 24.0 | 12.5 | 4.2 | 22.9 | 13.6 |
| Delayed | 7.9 | 30.2 | 27.1 | −3.1 | 49.0 | 47.9 | 48.5 |

*Note*—Subjects were tested immediately following acquisition, which involved nonsequential feedback, or after a 1-week delay. O = original, N = novel; M = miss, FA = false alarm.

location and the yes-no recognition test thus provided further converging evidence that specific representations of photographs presented during acquisition influenced performance for immediate, but not delayed, testing.

The accuracy with which subjects denoted the locations of the cameras on maps of the intersection was again scored on the basis of whether or not the camera locations were assigned to the correct corner of the intersection. As in Experiment 1, only 3 of 16 subjects failed to accurately place all four camera locations on the map.

## EXPERIMENT 5

In Experiments 1 and 4, evidence that subjects formed a schematic representation of the street scene depicted in the photographs was obtained when testing took place after a 1-week delay. Under conditions of delayed testing, subjects could identify the camera locations for the photographs with better than chance accuracy, and there was no influence of specific representations of previously seen photographs. What remained uncertain was whether the schematic representation for the street scene could be abstracted during the acquisition phase of the experiment or whether the 1-week interval between acquisition and the delayed test was necessary for the schema to be formed. According to the first alternative, subjects could have abstracted a schematic representation of the scene during acquisition, but specific representations of photographs in the acquisition set could still have had a predominant effect on performance during immediate testing. It was possible, on this basis, that modifying the acquisition procedure to promote schema abstraction would result in the formation of a representation that would specify the spatial layout of the scene with sufficient precision for it to predominate over specific representations during immediate testing. This was done by presenting the photographs from each location in consecutive, panoramic sequence (0, 60, 120 deg, etc.) when corrective feedback was provided during the acquisition phase of the experiment. It was anticipated that presenting the photographs in this manner would increase subjects sensitivity to the overlap of information in consecutive photographs and the overall spatial layout of the scene.

## Method

**Procedure.** This experiment differed in design from Experiment 4 in only one respect. During the acquisition phase of Experiment 4, subjects received corrective feedback after each trial. That is, following the subject's incorrect sorts, the experimenter placed the photographs taken from the same location into the same quadrant, but in no particular order. Subjects were given 1 min to examine the corrected sort. The difference in this experiment was that during the corrective feedback, the experimenter placed the photographs taken from the same location into the same quadrant, but now they were presented in panoramic sequence (e.g., for Location A, the 60-deg photograph was placed to the right of the 0-deg photograph, the 120-deg photograph to the right of the 60-deg photograph, and so on).

**Subjects.** Sixteen students, eight male and eight female, from Pennsylvania State University voluntarily participated in the entire experiment, without pay. All were unfamiliar with the street intersection depicted in the photographs. Another five subjects failed to reach criterion in the acquisition phase and were dropped from the remainder of the experiment.

## Results

The results of the experiment are presented in Table 5. The introduction of "sequential feedback" during acquisition reduced the number of trials subjects required to reach criterion from 7.5 in the previous experiment to 4.9 trials in this experiment (including the two criterion trials). Unlike in Experiments 1 and 4, there was no evidence during immediate testing that specific representations of members of the acquisition set influenced subjects' ability to identify camera location. For both immediate and delayed testing, the accuracy of same-different judgments was virtually equivalent for the original and novel photographs. An analysis of variance on percent errors in same-different judgment indicated that the effect of time of test (immediate vs. delay) was significant [$F(1,14) = 6.98$, $p < .02$, MSe = 112.5], but the effect of prior presentation (originals vs. novels) was not significant [$F(1,14) < 1.0$, MSe = 81.12]. The Immediate-Delay by Original-Novel interaction was also not significant [$F(1,14) < 1.0$, MSe = 81.12].

The results for the yes-no recognition test replicated the results of Experiments 2 and 4. That is, recognition performance was significantly more accurate when testing occurred immediately after acquisition than when it took place 1 week after acquisition [$F(1,14) = 207.80$, $p < .001$, MSe = 35.8]. Performance in the delayed

Table 5
Experiment 5: Mean Number of Acquisition Trials, Mean Percent Errors in the "Same-Different" Identification of Camera Location, and Mean Percent Errors in Yes-No Recognition Test

| Testing | Acquisition Trials | Identification Errors | | | Recognition Errors | | |
|---|---|---|---|---|---|---|---|
| | | O | N | N − O | M | FA | Mean |
| Immediate | 5.3 | 17.7 | 17.7 | .0 | 9.4 | 7.3 | 8.4 |
| Delayed | 4.4 | 28.1 | 27.1 | −1.0 | 47.9 | 53.1 | 50.5 |

Note—Subjects were tested immediately following acquisition, which involved sequential feedback, or after a 1-week delay. O = original, N = novel; M = miss, FA = false alarm.

testing condition was again at chance (50%). Although the recognition data indicated that specific representations of photographs in the acquisition set were available during immediate testing, these representations did not influence same-different judgments of camera location (there was no original-novel difference).

Subjects were again competent at denoting the camera locations on maps of the intersection. Twelve of the 16 subjects accurately placed all four camera locations on the correct corners of the intersection.

## Discussion

When testing occurred immediately after acquisition, judgments of camera location were equally accurate for original and novel photographs. Since there was no apparent influence of specific representations of members of the acquisition set, it could be concluded that judgments of camera location were based on a schematic representation of the street scene abstracted during the acquisition phase of the experiment. While the sequential feedback procedure used in this experiment seems to have promoted schema formation, it was not at the expense of the formation of specific representations for members of the acquisition set. The recognition data obtained during immediate testing indicated that subjects could accurately discriminate original from novel photographs.

The results obtained with sequential feedback showed that a schematic representation of the street scene could be formed during acquisition, but we could not rule out this possibility, in the experiments using nonsequential feedback, that the schematic representation emerged during the 1-week interval between acquisition and testing. It seems most likely, however, that both a schematic representation of the street scene and specific representations of particular photographs were formed during acquisition, regardless of whether feedback was sequential or nonsequential. Sequential feedback seems to have resulted in the predominance of the schematic representation when judgments of camera location take place immediately after acquisition. For nonsequential feedback, however, specific representations seem to have predominated when testing occurred immediately after acquisition, with schematic representations influencing performance only after a 1-week delay eliminated the retrieval of specific representations.

Of further interest was the finding that subjects' rate of false alarms (incorrectly saying that a novel photograph had been seen before) was only 7.3% when testing immediately followed acquisition. This low rate of false alarms was of importance because other investigators of spatial representation (Allen, Siegel, & Rosinski, 1978, Experiment 1; Pittinger & Jenkins, 1979) have contended that a high rate of false alarms for novel photographs that "belong" to the same set as the originals constitutes evidence for the formation of an integrated spatial representation. If we accepted their point of view, the

low rate of false alarms obtained in the immediate testing condition of this experiment would have led us to the conclusion that an integrated spatial representation was not formed. Our experiment, however, included both recognition and identification tests. The latter showed, for immediate testing, that camera location could be identified with equal accuracy for original and novel photographs. This provided direct evidence for the formation of an integrated spatial representation. When juxtaposed with the low rate of false alarms obtained in the recognition test, the identification data suggest, contrary to the claims of Allen et al. and Pittinger and Jenkins, that recognition data are not sufficient for the determination of whether integrated spatial representations are formed from partially overlapping photographic samples.

## EXPERIMENT 6

This experiment was performed to replicate the evidence obtained in Experiment 3 indicating that: (1) subjects could retain their schematic representations of the street scene over long intervals, and (2) the schematic representations were generalizable to new photographs from new camera locations.

### Method

The eight subjects participating in this experiment were students at Pennsylvania State University. Four had participated 6 months earlier in Experiment 5, and four had participated 1 year earlier in Experiment 4. They were presented a randomized sequence of 30 pairs of photographs. Twenty-four of the pairs comprised photographs that were presented to the subjects when they first participated in the study. Half of the pairs were composed of photographs taken from the same camera location; half were composed of photographs taken from new camera locations. Six pairs of photographs, three "same" and three "different," were taken from new camera locations (E and F). As in the previous three experiments, the information in the two photographs in each pair, whether "same" or "different," was always nonoverlapping. Foreground cues in the photographs were again shielded from the subjects.

### Results

The results of the experiment are presented in Table 6. The mean error rates for judgments of camera location obtained after delays of 6 months (28.7%) and 1 year (33.9%) were only slightly greater than the error rates that were obtained when testing occurred just 1 week after acquisition (see Tables 4 and 5). Subjects'

Table 6
Experiment 6: Mean Percent Errors for "Same-Different"
Judgment of Camera Location

| Retest After | Old Location | New Location |
|---|---|---|
| 6 Months | 28.1 | 29.2 |
| 1 Year | 34.4 | 33.3 |
| Combined | 31.3 | 31.3 |

performance after these long intervals ranged from error rates of 20.0% to error rates of 40.0% (chance performance was 50%). In addition, there was virtually no difference in performance between previously seen photographs from old locations (A, B, C, D) and new photographs from new locations (E, F). The results of the experiment thus replicated Experiment 3 with respect to both the long-term retention of subjects' schematic representations of the street scene and the generalizability of these representations to new camera locations.

## CONCLUSION

The results of Experiments 1, 4, and 5 indicated that when testing took place 1 week after acquisition, subjects identified camera location with better than chance accuracy. Since this was accomplished with no apparent influence of specific representations of the photographs presented during acquisition, it could be concluded that the identification of camera location was based on nonspecific information abstracted from the discrete photographic samples of the street scene. Chance performance on the yes-no recognition tests administered 1 week after acquisition (Experiments 2, 4, and 5) provided converging evidence that subjects could not utilize specific representations of the acquisition photographs after the 1-week delay.

It is unlikely that the information abstracted from the original photographs involved particular attributes that differentiated between the camera locations. A trivial example of this would have resulted if we had used a different colored filter at each camera location. The likelihood that discriminative features of this kind would be used by subjects was minimized by selecting camera locations such that photographs from all the locations "captured" virtually all the objects (buildings, trees, parked cars, etc.) in the scene. Thus, camera locations could not be differentiated on the basis of some objects appearing in photographs taken from one location, but not in photographs taken from other locations. In addition, it was shown in several experiments that foreground cues that were potentially common to the photographs taken from the same location were not necessary for the accurate identification of camera location.

Instead of features that discriminate between camera locations, it is more likely that subjects' performance was based on the abstraction of schematic representations of the spatial layout of the scene (Hochberg, 1968, 1978; Neisser, 1967). Without a schematic representation, subjects would not have been able to accurately place the camera locations on maps of the scene. Discriminative features or specific representations corresponding to previously seen photographs of the scene could not have been the basis for the scale transformations (from the "large" street scene to the small

map), perspective transformations (from the frontal, picture plane of the camera to the top view required by the map), and symbolic transformations (from real-world objects to the abstract map) required by the map task.

Various investigators, beginning with Bartlett (1932), have contended that schemata can direct the selection and organization of incoming visual information. Thus, nonoverlapping visual samples can be integrated into a continuous perceptual representation if subjects are given prior information that allows them to generate an appropriate schema for the scene (Hochberg, 1968). Of greater relevance to the present research is the observation that schemata can be abstracted in the course of processing a sequence of partially overlapping samples of scenes, and the schema can then direct the further processing of the visual information (Hochberg, 1978). While it is not yet possible to fully specify the contents of scene schemata, Biederman's (1980) research with scene "violations" indicates that rules specifying such interobject relations as support, interposition, probability, object position, and size are inherent in our scene schemata. In their studies of recognition memory for scenes, Mandler and her colleagues (Mandler & Johnson, 1976; Mandler & Parker, 1976; Mandler & Ritchey, 1977) have provided evidence that scene schemata can involve information concerning the spatial composition of scenes (where objects are located) and the spatial relations between the objects. The results of the present research show that scene schemata can involve information specifying the spatial locations of objects that have never appeared in any of the visual stimuli (the camera). Without such information, subjects would not have been able to accurately indicate the locations of the cameras on maps of the scene. For subjects to accurately identify the camera locations of the photographic samples of the scene, the scene schema they form must contain information concerning the projective sizes, shapes, and perspectives of the buildings, trees, and other objects as they appear in the photographs taken from different locations. The scene schemata abstracted by subjects in this study also seem to be generative. They govern judgments of camera locations for photographs taken from new camera directions as well as new camera locations. Finally, the scene schemata are long-lasting. Camera location can be accurately judged 1 year after acquisition.

## REFERENCES

ALLEN, G. L., SIEGEL, A. W., & ROSINSKI, R. R. The role of perceptual context in structuring spatial knowledge. *Journal of Experimental Psychology: Human Learning and Memory,* 1978, *4,* 617-630.

BARTLETT, F. C. *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press, 1932.

BIEDERMAN, I. On the semantics of a glance at a scene. In

M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization.* Hillsdale, N.J: Erlbaum, 1980.

HELMHOLTZ, H. VON. *[A treatise on physiological optics* (Vol. 3)] (J. P. C. Southall, Ed. and trans.) New York: Dover, 1963. (Originally published, 1866.)

HOCHBERG, J. E. In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception.* New York: Holt, Rinehart, & Winston, 1968.

HOCHBERG, J. E. *Perception.* Englewood Cliffs, N.J: Prentice-Hall, 1978.

LOFTUS, G. R. On interpretation of interactions. *Memory & Cognition,* 1978, 6, 312-319.

MANDLER, J. M., & JOHNSON, N. S. Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning and Memory,* 1976, 2, 529-540.

MANDLER, J. M., & PARKER, R. E. Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory,* 1976, 2, 38-48.

MANDLER, J. M., & RITCHEY, G. H. Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory,* 1977, 3, 386-396.

NEISSER, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

PITTINGER, J. B., & JENKINS, J. J. Apprehension of pictorial events: The case of a moving observer in a static environment. *Bulletin of the Psychonomic Society,* 1979, 13, 117-120.

POSNER, M. I., & KEELE, S. W. Retention of abstract ideas. *Journal of Experimental Psychology,* 1970, 83, 304-308.

WALLS, G. L. The problem of visual direction. *American Journal of Optometry and Archives of American Academy of Optometry,* 1951. (Monograph)

## NOTES

1. The data for subjects of uncertain familiarity were completely consistent with the data for the subjects in the familiar and unfamiliar groups.

2. There were no apparent sex differences in acquisition rate, identification of camera location, or accuracy in map placement in this experiment. The same was true in the experiments that follow.

3. Included in the above analysis were the data for six subjects, all from the familiar group, who indicated prior to testing that they could not remember the verbal labels for each camera location. They were told to assign the labels in any way they wished but to do so consistently. Their reversals of the correct labels were confirmed by their postexperimental placement of camera locations on a map of the intersection. The responses for these subjects were therefore scored as correct if they were correct within their own labeling system. An examination of the data for the entire set of subjects indicated no other instances of consistent response-label reversals.

4. The map placements for six of the familiar subjects were scored as correct on the basis of the consistency in their label reversals for both the maps and the primary experimental task of identifying the camera location for each photograph.

5. Unlike Experiment 1, the interaction between time of test (immediate vs. delay) and prior presentation (original vs. novel) in this experiment could not be removed by a monotonic transformation of the dependent measure (Loftus, 1978).