Visual Images Preserve Metric Spatial Information: Evidence from Studies of Image Scanning

Stephen M. Kosslyn
Harvard University

Thomas M. Ball
Johns Hopkins University

Brian J. Reiser
New York University

Four experiments demonstrated that more time is required to scan further distances across visual images, even when the same amount of material falls between the initial focus point and the target. Not only did times systematically increase with distance but subjectively larger images required more time to scan than did subjectively smaller ones. Finally, when subjects were not asked to base all judgments on examination of their images, the distance between an initial focus point and a target did not affect reaction times.

Introspections about visual imagery very often include references to "scanning" across images. Kosslyn (1973) attempted to demonstrate that scanning of images is a functional cognitive process, and his experiment indicated that more time was required to traverse greater distances across mental images. However, in the course of scanning longer distances,[1] people in Kosslyn's experiment also passed over more parts of the imaged object. For example, in scanning from the motor to the porthole of an imaged speedboat, a person passed over the rear deck and part of the cabin; in scanning from the motor to the more distant anchor, one scanned over all of these parts plus the front deck and bow. Given this confounding, then, we have no way of knowing whether Kosslyn's results were a consequence of people actually scanning over a quasi-pictorial, spatial image. One could argue that the image itself was epiphenomenal in this situation and that the apparent effects of distance actually were a consequence of how people accessed some sort of underlying list structure. Parts separated by greater distances on the image might simply be separated by more entries in a list of parts of the object.

The notion that scanning corresponds to processing a list structure, and not the spatial "surface" image (see Kosslyn, 1975, 1976; Kosslyn & Pomerantz, 1977), recently seemed to receive support from Lea (1975). In a typical experiment, people evaluated

[1] We will use terms like *distance* and *size* in referring to mental images, even though images themselves—not being objects—do not have such physical dimensions. Nevertheless, we claim that images represent these dimensions in the same way that they are encoded in the representations underlying the experience of seeing during perception. Thus, we experience images as if we were seeing a large or small object, or one at a relatively near or far distance from us. In addition, the apparent distances between parts of an imaged object are experienced in the same way that one would experience apprehending the distances when seeing the parts of the object. We will use the term *quasi-pictorial* in referring to these sorts of pictorial properties of an image, because an image—not being an object—cannot have the physical properties of an actual picture. For convenience, we will refer to an imaged object that is experienced as being some subjective size as if the image were that size, and we will refer to apparent distances on an imaged object as if they were distances on the image itself.

from memory the relative locations of objects in a circular array. Lea asked his subjects to learn the array via imagery. Following this, they were given the name of one object and asked to name the first, second, or nth item in a given direction. Lea found that the time to respond depended on the number of intervening items between an initial focus point and the target, but not on the actual distance separating a pair of objects in the array. The interpretation of these results is muddied, however, because Lea never insisted that his subjects base all judgments on actual processing of the image itself. That is, subjects were not told to count the items as they appeared in their image but only to count the appropriate number of steps to the target. It is reasonable to suppose that these people encoded the circular array both as a list and as an image. Given that imagery tends to require more time to use in this sort of task than do nonimaginal representations (Kosslyn, 1976), subjects may have actually arrived at most judgments through processing nonimaginal list structures. If so, then it is not surprising that actual distance separating pairs did not affect retrieval times.

The present experiments, then, test the claim that distance affects time to scan images by removing the confounding between distance and the number of intervening items scanned across. If images really do preserve metric spatial information, and images themselves can in fact be scanned, then actual distance between parts of an imaged object should affect scanning time. If the apparent effects of distance observed by Kosslyn (1973) were in fact due to accessing some sort of ordered list, however, then only ordinal relations between parts—not actual interval distances—should affect the time needed to shift one's attention from one part of an image to another.

Experiment 1

This experiment is an attempt to distinguish between the effects of scanning different distances and scanning over different numbers of intervening items. The people who participated in the experiment scanned

visual images of three letters arrayed on a line, "looking" for a named target. Upon mentally focusing on the target, the subject classified it according to whether it was upper- or lowercase. In scanning to the target letter, one had to traverse one of three distances and pass over zero, one, or two intervening letters; letter arrays were constructed such that each distance appeared equally often with each number of intervening items, allowing us to consider each variable independently of the other.

The present claim is that distance per se affects time to scan an image. However, we also expect people to take more time in scanning over more items since each item presumably must be "inspected" as it is scanned over, which requires an increment of time. The present claim does not speak to the issue of which factor affects image scanning more—distance or number of intervening items; we are primarily concerned with demonstrating that effects of distance are not simply an artifact of how many things must be scanned over.

Method

Materials. We constructed two books of stimuli, each containing 36 arrays of letters. Each array consisted of three letters spaced along a 20.32-cm long line. Each array contained two letters of one case, and one of the other; each case (upper and lower) was represented equally often across arrays. Target letters were placed 5.08, 10.16, and 15.24 cm from the point of focus (one of the two ends of the line), and zero, one, or two other letters intervened between the target and point of focus. Intervening items were spaced at equal intervals between the target and focus point. The arrays were constructed such that each distance occurred equally often with each number of intervening items. Each of these nine conditions was represented by 8 arrays, half of which had an uppercase letter as the target and half of which had a lowercase letter as the target. Further, for half of each target type in each condition, the focus point was specified as the left end of the line, and for half it was the right end. We did not use letters whose upper- and lowercases seemed difficult to distinguish (*c, k, o, p, s, u, v, w, x, z*). The remaining 16 letters of the alphabet were used as targets and distractors. Each of these letters appeared at least once as a target in each case, at each distance, and with each number of intervening items, but not with every possible combination of these variables (this would have required

far more trials than we used). The arrays were randomly divided into two sets, which were placed in separate books, and the order of arrays was randomized within each book (with the constraint that no more than three consecutive targets could be of the same case).

We also constructed a tape recording. The tape contained 72 trials of the form "1 . . . cover . . . left . . . A." Each trial was coordinated with an array in the books. The first word named the trial number and was followed 5 sec later by the word *cover* (which was the signal to conceal the array and to construct an image). Two seconds thereafter the word *left* or *right* was heard (indicating point of focus, each word appearing on half of the trials, as noted above). Finally, 3 sec after this, the name of a letter in the corresponding array was heard. Presentation of the letter delivered a pulse to a voice-activated relay that started a reaction time clock (which was stopped by the subject's pressing either of two response buttons). A new number was presented 10 sec after the letter, and the sequence was repeated with a new trial.

Procedure. Written instructions describing the experimental procedure were given to the subject and then were reviewed orally by the experimenter. It was emphasized that we were interested in studying how people process visual mental images, and therefore we wanted the subject always to use an image in performing this task—even if this did not seem the most efficient strategy. These general instructions preceded every experiment reported in this article. Before we are willing to make inferences about imagery from data, we want to be sure that those data were in fact produced via imagery processing.

The subject was told that he or she would soon see simple arrays of letters. We explained that the task was to study an array and then to shut one's eyes and mentally picture the array as it appeared on the page. We would next ask the subject to focus on one end of the image and then to scan to a given letter in the array. As soon as the target letter was clearly in focus, we wanted the subject to "look" at the letter: If it were uppercase, he or she should push one button; if it were lowercase, the other button should be pushed.

Following this, we explained the meanings of the tape-recorded cues that accompanied the arrays. Upon hearing a number, the subjects were to turn to the next page in the book in front of them, which would have that number at the top (pages were numbered consecutively). They should study this array until hearing the word *cover,* at which point they should cover the array with a small piece of cardboard and mentally image the array. While visualizing the array, they then would hear the word *right* or *left,* directing them to "mentally stare" at that end of the line. They should continue to focus at that end until hearing the next word, which would be the name of a letter in the array. At this point the subjects were to scan to the

named letter and classify it according to its case. Eight practice trials (half upper- and half lowercase, in a random order) preceded the actual test trials. The subjects were questioned during these trials to ensure that they were performing the task as instructed. The subjects were asked to perform the task as quickly as possible while keeping errors to an absolute minimum.

This procedure, then, prevented the subjects from initially encoding an array differently depending on the point of focus or the distance of a target or the number of intervening letters. The order of the two books was counterbalanced over subjects, as was the hand (dominant/nondominant) assigned for indicating each case. Each person was interviewed at the conclusion of the 20-min tape recording and was asked to estimate the percentage of time he or she actually followed instructions while performing the task. Further, we asked each subject to attempt to discern the purposes and motivations of the present experiment.

Subjects. Twelve Johns Hopkins University students volunteered to participate as subjects to fulfill a course requirement. Although 2 of these people reported noticing distance effects during the course of the experiment, and 2 people reported observing that it was easier when there were no intervening items, no subject reported noticing both effects, and no subject deduced any part of the hypothesis independently of noticing his or her behavior during the task. Data from 1 additional potential subject were discarded because she estimated complying with the imagery instructions only 60% of the time, and data from another potential subject were discarded because his mean reaction times were more than twice the means of all the other subjects. The 12 remaining subjects reported complying with the instructions at least 75% of the time.

Results

An analysis of variance was performed on the data. Only reaction times from correct responses were used, and errors and wild scores were replaced by the mean of the other scores in that condition for that subject. A wild score was defined as one that exceeded twice the mean of the other scores in that cell for that subject; only one score per cell could be so defined, however. Because we wished to generalize over both subjects and items, we used the Quasi F statistic, F' (Clark, 1973).

As expected, scanning times increased as subjects had to scan further distances to reach the target letter, $F'(2, 30) = 9.89$, $p < .01$. In addition, times also increased when subjects had to scan across more in-
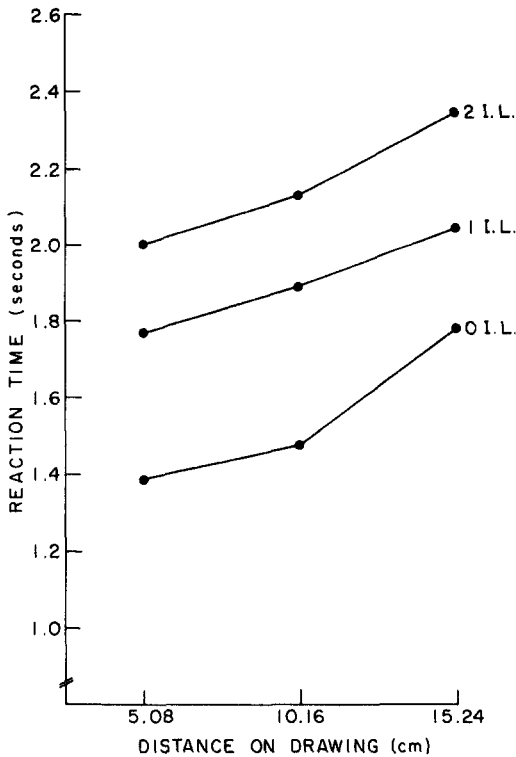
Figure 1. The results of Experiment 1: Classification times when subjects scanned different distances over zero, one, or two intervening letters (I.L.).

tervening letters before reaching the target, $F'(2, 27) = 22.65$, $p < .01$. Interestingly, as is evident in Figure 1, the effects of distance were the same regardless of how many intervening items were scanned over; there was no interaction between the two variables ($F' < 1$). This lack of interaction also indicates, of course, that the effects of intervening items were the same for each of the three distances—which is what one would expect if this effect reflects time necessary to "inspect" each of the intervening letters. Finally, there was no difference in time to categorize letters of different case or to scan left versus right, nor were any other effects or interactions significant.

Errors tended to increase with increasing reaction times. For the 5.08-, 10.16-, and 15.24-cm conditions, errors were .7%, 3.1%, and 1.4%, respectively. Although errors for the 10.16-cm distances were relatively high, they were not significantly higher than the

errors for the 15.24-cm condition ($p > .1$). For the zero, one, and two intervening item conditions, errors were .7%, 2.1%, and 2.4%, respectively. Thus, it does not appear as if speed–accuracy trade-offs affected the data.

Discussion

We found that more time is required to scan further distances across an image. In addition, more time also is required when one scans over more items. Our findings argue against the idea that people were not really scanning a spatial image but rather simply processing a serially ordered list of letters. If so, we should only have found an effect of number of intervening items (if scanning the list were self-terminating). There is no reason to expect such a list to have metric distance from each end to be associated with each letter. Furthermore, we found effects of distance even when the target letter was not separated from the focus point by any intervening letters. Finally, we found that it took the same amount of time to scan right to left as it did to scan in the opposite direction. This last result replicates that of Kosslyn (1973) when his subjects were asked to remember and then to scan visual images (left-to-right scanning was easier, however, when subjects encoded and used verbal descriptions of the pictures instead of images). Thus, image scanning would seem to involve processes or mechanisms different from those highly practiced ones used during reading.

Given the existence of two independent effects of distance and number of intervening letters, one might be tempted to ask which factor is the more important. This is a nonsensical question: By increasing the range of distances, we surely could make distance account for the lion's share of the variance in scanning times—and by decreasing the range of distances, we could diminish the importance of this variable. In addition, we could probably manipulate the importance of number of intervening items by making the distractors more or less difficult to discriminate from the targets. Furthermore, the present claim is not that distance

is more important than other variables, but only that images do preserve metric distance information—and that such information can be used in real-time processing, affecting the operating characteristics of cognitive processes.

One might argue that the effects of distance on scanning time really reflect nothing more than the enthusiastic cooperation of our subjects, who somehow discerned the purpose of the experiment and manipulated their responses accordingly. Although 2 of our subjects did hypothesize distance effects, they claimed to do so by introspecting upon their performance during the task; no subject confessed to consciously manipulating his or her responses. Nevertheless, we would be more comfortable with a task that was more difficult to second-guess and manipulate.

Experiment 2

This experiment involves scanning between the 21 possible pairs of seven locations on an imaged map. Each of these distances was different, and the task seemed sufficiently complex to thwart any attempts to produce intentionally a linear relationship between distance and reaction time. Since the critical question is whether images preserve metric information, it is important that scanning times be a function of some known distance—otherwise, variations in scanning time cannot be taken to necessarily reflect amount of distance traversed. Thus, we wished to ensure that subjects scanned only the shortest distance between two points. In order to do so, we altered the instructions slightly and asked these people to imagine a black speck moving along a direct path across the image. After memorizing the map, these subjects imaged it, focused on a location, and then decided whether a given named object was in fact on the map. If so, the subjects were asked to scan to the named object on the image and to push a button when they "arrived" there; if not, they pushed another button. The time necessary to scan between all possible pairs of locations was measured. As before, we expected times to increase with

distance (although not necessarily linearly, as rates may be variable).

Method

Materials. A map of a fictional island was constructed containing a hut, tree, rock, well, lake, sand, and grass. Each distance between all 21 pairs was at least .5 cm longer than the next shortest distance. The precise location of each object was indicated by a red dot; these locations are indicated by a small *x* in Figure 2.

A tape recording was constructed containing 84 pairs of words. Each location was named 12 times and then followed 4 sec later by another word; on 6 of these trials, the second word did not name a location on the map. The "false" objects were things that could have been sensibly included on the map (e.g., "bench"). On the other 6 trials, the first word was followed by the name of each of the other locations. Thus, every pair of locations occurred twice, once with each member appearing first. The order of pairs was randomized, with the constraint that the same location could not occur twice within three entries, and no more than 4 true or 4 false trials could occur in a row. Presentation of the second word also started a clock. A new trial began 8 sec after the probe word was presented. The test trials were preceded by 8 practice trials naming pairs of cities in the United States for "true" items.

Procedure. The subjects first were asked to learn the locations of the objects on the map by drawing their relative positions. The subjects began by tracing the locations on a blank sheet placed

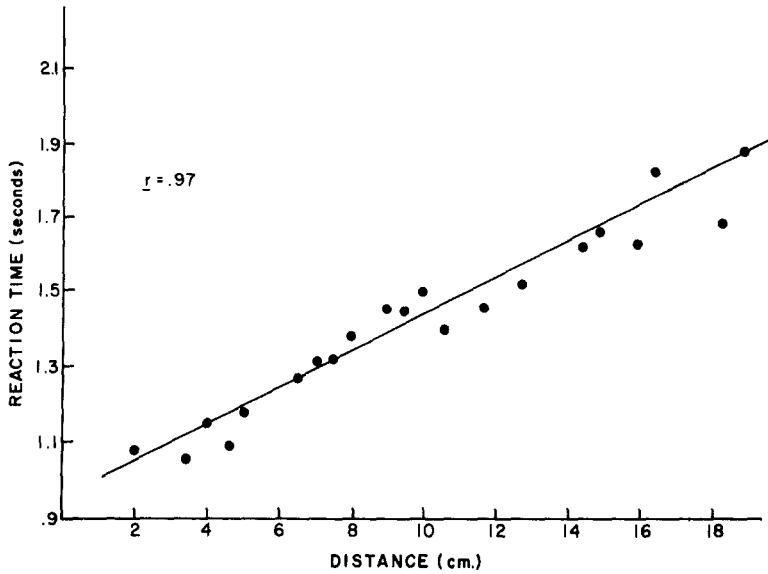*Figure 2.* The fictional map used in Experiment 2.

Figure 3. The results of Experiment 2: Time to scan between all pairs of locations on the imaged map.

over the map, marking the locations of the red dots centered on the objects; this procedure allowed them to see the locations themselves in isolation. Next, they studied the map, closed their eyes and imaged it, and then compared their image to the map until they thought their image was accurate.

The map then was removed, and the subjects drew the locations on a blank sheet of paper. Following this, the subjects were allowed to compare their drawings with the original. This procedure was repeated until all points were within .64 cm of the actual location. Between 2 and 5 drawings were required for subjects to reach this criterion.

Next, subjects were told that they would hear the name of an object on the map. They were to picture mentally the entire map and then to focus on the object named. Subjects were told that 5 sec after focusing on the named object, another word would be presented; if this word named an object depicted on the map, the subjects were to scan to it and depress one button when they arrived at the dot centered on it. The scanning was to be accomplished by imaging a little black speck zipping in the shortest straight line from the first object to the second. The speck was to move as quickly as possible, while still remaining visible. If the second word of a pair did not name an object on the map, the subjects were to depress the second button placed before them. The clock was stopped when either button was pushed, and response times were recorded. As before, we interviewed subjects in the course of the practice trials, making sure that they were following the instructions about imagery use.

Subjects. Eleven new Johns Hopkins University students served as paid volunteers in this experiment. Data from 2 additional people were not analyzed because, when queried afterwards, they reported having followed the imagery instructions less than 75% of the time during the task.

Results

Only times from correct "true" decisions (where a distance was actually scanned) were analyzed. As before, wild scores were eliminated prior to analysis. A wild score was now defined as one twice the size of the mean of the other score for that distance and the scores for the next shortest and longest distances; only one score in any adjacent row of six could be so eliminated. Data were analyzed in two ways, over subjects and over items. We first analyzed each subject's times for the different distances in an analysis of variance. As expected, times consistently increased with increasing distance, $F(20, 200) = 13.69$, $p < .001$. In addition, we averaged over subjects and calculated the mean reaction time for each pair. The best fitting linear function was calculated for these data by the method of least squares; not only did times increase linearly

with increasing distance but the correlation between distance and reaction time was .97. These data are illustrated in Figure 3.

Errors occurred on only 1.3% of the trials and were distributed seemingly at random; more errors did not occur for the shorter distances. Finally, subjects drew maps after the experiment. Not surprisingly, the correlation between the drawn and actual distances between all possible pairs of points was quite high, $r = .96$.

Discussion

Time to scan across visual mental images again increased linearly with the distance to be scanned. This demonstration supports the claim that images are quasi-pictorial entities that can in fact be processed and are not merely epiphenomenal. One of the defining properties of such a representation is that metric distances are embodied in the same way as in a percept of a picture, and the present data suggest that this characteristic is true of visual mental images.

Interestingly, a number of subjects reported that they had to slow down when scanning the shorter distances, because the four objects at the lower left of the map were "cluttered together." The data show no sign of this, however, providing further grounds for taking with a grain of salt subjects' interpretations of their introspections. This experiment seems immune to the potential failings of Experiment 1; somewhat surprisingly, no subjects reported suspecting the hypothesis when it was explained to them afterwards.

Experiment 3

Given the results of the first two experiments, how can we explain Lea's (1975) failure to find increases in reaction times as distances increased? We earlier suggested that this failure was a consequence of his instructions: Subjects were not told to base all judgments on consultation of their images, but only to start off from an imaged location and to "scan" a certain number of objects from there. Although these people initially began with an image, the actual

decisions could have been generated via processing of items in a list. If so, only ordinal —and not interval—relations among items (objects in the array, in Lea's case) should affect time to sort through the list. Effects of actual distance ought to occur only when one scans the spatial image itself, which seems to represent interval information about distance. If we find distance effects even when people do not scan images, we are in trouble: We could not then infer that effects of metric distance implicate scanning of quasi-pictorial images.

A second hypothesis for why Lea failed to obtain effects of distance on time to scan also involves his instructions. Lea did not insist that his subjects always construct the entire array ahead of time; instead, subjects were told simply to image a starting place and then to decide which object was some number of locations away. Perhaps distance only affects time to shift attention between locations in an image when the locations are both "in view" simultaneously. That is, if an entire image is not kept in mind at once, the distance relations between visible and invisible locations may not be represented; these relations could be an "emergent" property of constructing the whole image from its component parts. One might shift to an "invisible" part by generating a sequence of individual images representing intervening locations and not by actually scanning across an image. In this case, interval distance would not be expected to affect time to shift attention between parts.

The following experiment examines the hypotheses described above. In one group, subjects were asked to focus on a given location on an image of the map used in Experiment 2 and then to judge whether a named object was on the map. Unlike the people in Experiment 2, however, these people were not required to consult their images when making their judgments, but simply were asked to reach decisions as quickly as possible. In a second group, subjects also performed the basic task of Experiment 2, but with one major modification: When focusing on the initial location, these people were asked to "zoom in" on it until that

object filled their entire image, causing the remainder of the island to "overflow." These people were told, however, that they must "see" an image of the second named object before responding positively (if in fact it was on the map). The two groups, then, were each instructed to perform in a way that Lea's subjects may have acted spontaneously.

Method

Materials. The same materials used in Experiment 2 were also used here.

Procedure. Subjects in both groups learned to draw the map as did subjects in the previous experiment. The procedure differed from that of Experiment 2 only in the following ways:

1. Rapid Verification Control Group. These subjects were given instructions like those of Experiment 2, except that no mention was made of scanning the image. After focusing on an initially named object, these people were simply to decide as quickly as possible whether the second object of a pair was in fact on the map. As before, subjects were urged to keep errors to a minimum.

2. Image Overflow Group. These subjects were given instructions that differed from those of Experiment 2 in two ways: First, these people were asked to "zoom in" on the initially named object until the rest of the map had "overflowed" (i.e.,

was no longer visible in) their image. Second, they were instructed to be sure to "see" a second named object of a pair before responding positively. These subjects were not told to scan to the second object if it was on the map but only to be sure to "see" it prior to responding; no mention was made of a flying black speck or the like. As before, speed with accuracy was stressed in both groups.

Subjects. Twenty-two new Johns Hopkins University students volunteered as paid subjects in this experiment. Half of these subjects were randomly assigned to one group, half to the other. An additional 3 people were assigned to the Image Overflow Group but were not included, because after the experiment they reported having followed the instructions less than 75% of the time.

Results

Data were analyzed as in Experiment 2. In the Rapid Verification Group, there were significant differences in time to evaluate different pairs, $F(20, 200) = 2.59$, $p < .01$. As is evident in Figure 4, however, times did not increase systematically with distance. In fact, the relationship between distance and verification time was negligible, $r = .09$.

In the Image Overflow Group, in contrast, times did increase systematically with distance. Not only were times to evaluate
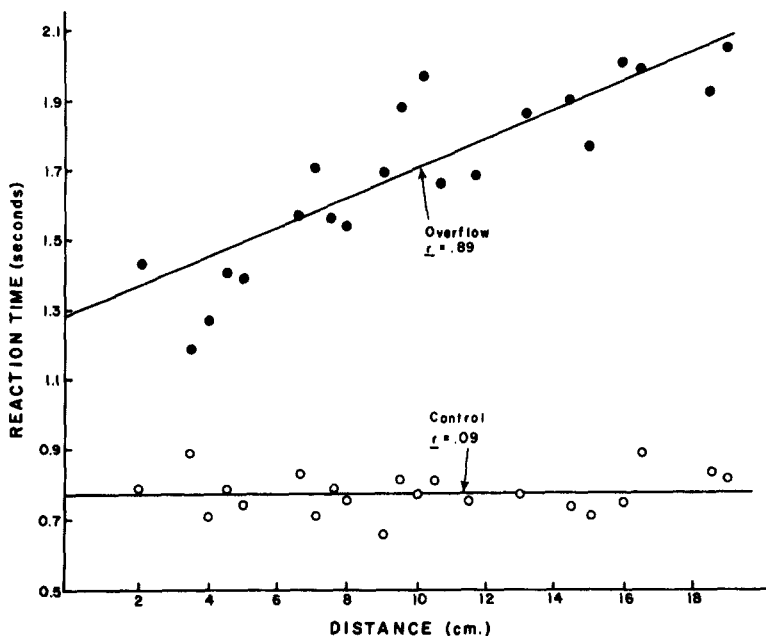

Figure 4. The results of Experiment 3: The effects of distance on response times for the Imagery Overflow and the Rapid Verification Control Groups.

different pairs significantly different from each other, $F(20, 200) = 4.59$, $p < .01$, but there was a respectable correlation between distance on the map and evaluation time, $r = .89$.

We also performed three additional analyses of variance, one comparing the results from each group with the data obtained in Experiment 2 and one comparing the two groups with each other. Not surprisingly, there were less effects of distance in data from the Rapid Verification Control Group than in the Image Overflow Group or in Experiment 2 ($p < .01$ for the interaction of distance and instructions in both cases). In addition, subjects in the Rapid Verification Control Group made decisions more quickly than those in either other condition ($p < .01$ for both comparisons). The comparison between the results of the Image Overflow Group and the findings of Experiment 2 produced a somewhat surprising result, or rather, lack thereof: In this case, the effects of distance were identical for both instructions ($F < 1$). Furthermore, there was no significant difference overall in verification times (the mean for Experiment 2 was 1.428 sec vs. 1.685 sec for the Image Overflow Group), $F(1, 20) = 1.04$, $p > .1$. If "zooming in" increases the subjective size of an image, it should also increase the "distance" between portions of that image; hence, we would have expected that more time should have been required by subjects in the Image Overflow Group.

The error rate in the Rapid Verification Control Group was 3.3%, whereas there were only 1.4% errors in the Image Overflow Group. As before, errors did not tend to increase with shorter distance for the Image Overflow Group, and seemed randomly distributed for the Rapid Verification Control Group. No subjects deduced the purposes or predictions of this experiment.

Discussion

When people were not required to base decisions upon consultation of their images, evaluation times did not increase with the distance between a focus point and a probed object that was in fact on the map. This result allows us to argue against a nonimagery interpretation of the scanning results obtained in the preceding experiments: If the effects of distance obtained previously were due to local activation and scanning through an abstract list structure (e.g., perhaps a graph with "dummy nodes" interposed to mark off increasing distance), then we should have found effects of distance here. Distance per se seems to affect response times only when people actually scan their images. Thus, Lea's (1975) results may simply reflect the fact that his subjects were not told to respond only after seeing the probed object in their image. Clearly, before we draw inferences about image processing from some data, we must be certain that such data were produced when people did in fact use their images. The instructions administered in the present experiments and elsewhere (Kosslyn, 1973, 1975, 1976) seem capable of inducing subjects to use imagery, even if other means of performing a task are available.

Lea's results were probably not a consequence of subjects' not having the entire array in their images prior to processing it, as witnessed by the results of the Image Overflow Group. Although these people only had the focus location in their images, times nevertheless increased with distance to a probed object. We were surprised by these results, which were not expected. This finding seems to indicate that one may construct images such that portions are "waiting in the wings," ready to be processed if necessary. Thus, subjects seemed to have scanned to parts that were not visible initially in their images but were available in a nonactivated portion of the image.

There is one hitch in the above explanation of the data obtained from the Image Overflow Group: If these people "zoomed in" closer to the imaged map than did those in Experiment 2, the subjective distances between parts should have been greater in the Image Overflow condition. If so, then more time should have been required to scan these enlarged images, which was not

the case. One explanation of this disparity rests on a procedural difference between the Image Overflow condition and Experiment 2: Subjects in Experiment 2 were instructed to image a small black speck flying between parts. This task may have required more effort than the simple shift-of-attention instruction used in the present experiment, and thus slowed down scanning. In addition, it is possible that subjects in the two experiments simply scanned at different rates: If people in the Overflow condition scanned relatively quickly, perhaps because distances traversed were on the average relatively large, then we would not necessarily expect any differences in scanning times between the two conditions. The following experiment eliminated the difference in instructions and used a within-subjects design; we hoped that a given person would adopt a constant scanning rate for different materials.

Experiment 4

In this experiment we investigated whether more time is required to scan across subjectively larger images. We worried that if we used stimuli as complex as those included on the map, people might have to "zoom in" (if the image were small) or "pan back" (if it were large) in order to "see" parts clearly. Kosslyn (1975) demonstrated that parts of subjectively smaller im-

Figure 5. The schematic faces used in Experiment 4.

ages are more difficult to identify than parts of larger ones, and this may also be true of parts of "overflowed" images. Not only could difficulty in identifying parts of relatively complex images obfuscate effects of scanning images of different subjective sizes, but people may adjust their scanning rates in accordance with the difficulty in identifying parts. Pilot data lent credence to these fears, encouraging us to use more simple stimuli, where the parts were readily identifiable.

Thus, in this experiment people imaged one of three schematic faces at one of three subjective sizes. The faces had either light or dark eyes, and the eyes were one of three distances from the mouth. These people first focused on the mouth of an imaged face and then shifted their attention to the eyes and decided whether a probe correctly described them. As in Experiment 1, these instructions made no mention of a flying speck or the like. If distances determine scanning times, then subjectively smaller images should be scanned more quickly than larger ones. In addition, the effects of increased distance should become more pronounced with larger images, since when size is multiplied, so are the distances.

Method

Materials. Six schematic faces were constructed. The eyes were 7.62, 10.16, or 12.70 cm above the mouth; for each distance, one face was constructed with light eyes and one was constructed with dark eyes. The faces are illustrated in Figure 5.

Twelve copies of each face were made and used in nine basic conditions, each of which was represented by eight stimuli. These conditions were defined by three subjective sizes—overflow, full size, and half size—and the three distances. Within a condition, half of the faces had light eyes and half had dark eyes. Further, half of the faces with each eye color were paired with the word *dark* and half with the word *light* on an accompanying tape recording, producing an equal distribution of true and false probes. The faces were then randomized and placed in a booklet, with the constraint that no given distance or size could occur twice within 3 trials.

A tape recording also was made. This tape contained stimuli consisting of three parts: First, the number of the trial was given. Second, 5 sec later the word *cover* was presented, followed 1 sec
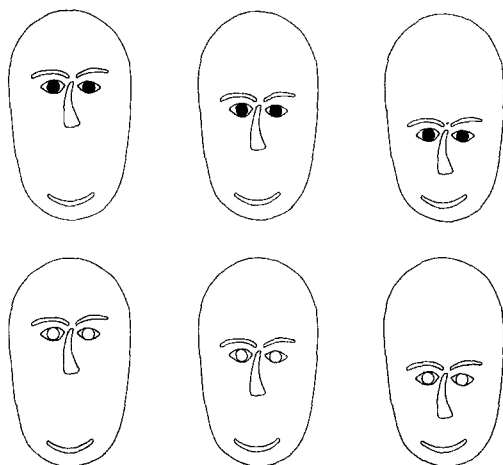
later by one of three cues—overflow, full size, or half size. These stimuli indicated the size at which the subject should construct his or her image. Finally, 5 sec later the word *light* or *dark* was presented, which also started a clock. Ten sec after this, a new number was presented and another trial began. For half of the trials in each size condition, the final word described the eyes of the imaged face, and for half it did not. The 72 test trials were preceded by 8 practice trials.

Procedure. The subjects were told that they were going to see schematic faces one at a time. As soon as a trial number occurred on a tape recording, they should turn to the corresponding page of the book in front of them, exposing a drawing of a face. The subjects were asked to study the drawing well enough to form an accurate visual mental image of it with their eyes closed. After 5 sec, the subjects would hear the word *cover*, at which point they would conceal the face with a small piece of cardboard; shortly thereafter they would hear a size specification, either overflow, full size, or half size. Upon hearing the word *overflow*, the subjects were to image the face so large that only the mouth was visible. Upon hearing *full size*, they were to image it as large as possible while still being able to "see" all of it at once in their image; as soon as this image was constructed, they were to mentally focus on the mouth and wait there until hearing the next stimulus on the tape. Upon hearing *half size*, they were to image the face at half of the length of the full-size version, again focusing on the mouth. Following this, the subjects were told they would hear either the word *light* or *dark*. At this point, they were to "glance" up at the eyes in their image and see if they were appropriately described. If so, they were to push one button; if not, they were to push the other. Hand of response was counterbalanced over subjects; as before, the clock stopped as soon as either button was pushed, and response times were recorded. Subjects were asked to respond as quickly as possible, but always to base decisions on inspection of the image (as in Experiment 1). During the 8 practice trials preceding the test items (half true, half false, including all three size conditions and all three distances), the subjects were asked repeatedly to describe their mental activity, and any misconceptions about the task were corrected.

Subjects. Sixteen new Johns Hopkins University students volunteered to participate for pay; data from an additional subject were discarded because this person reported not following the instructions at least 75% of the time.

Results

Only times from correct decisions were included in an analysis of variance; errors and occasional wild scores (defined as in Experiment 1) were replaced by the mean of the remaining scores in that condition for that subject. As expected, times increased with further separation between the mouth and eyes, $F(2, 30) = 10.81$, $p < .01$. In addition, times increased as subjective size of the image increased, $F(2, 30) = 17.33$, $p < .01$. As is evident in Figure 6, increases in distance did have increasingly larger effects as the subjective size increased; the interaction between size and distance was in fact significant, $F(4, 60) = 3.47$, $p < .025$. Examination of Figure 6 reveals, however, that the effects of distance were not appreciably different in the full-size and half-size conditions. A marginally significant interaction between type of response (true or false) and distance, $F(2, 30) = 2.80$, $.05 < p < .10$, led us to consider separately data from true and false responses. As in the main analysis, distance and size both affected decision times for both types of responses ($p < .01$ in all cases in separate analyses of variance of the true and false responses). However, whereas the effects of distance increased with size for true responses, $F(4, 60) = 5.58$, $p < .01$, they did not increase for false responses ($F < 1$). Furthermore, for true responses there was some difference between the effects of distance in the full-size and half-size conditions. We observed that times increased an average of 109 msec for every additional 2.54 cm separating the eyes and mouth on the face in the half-size condition. On this basis, we predicted that time to scan a face twice as long ought to be 2.050, 2.274, and 2.486 sec, respectively, for the three increasing distances. These predictions were clearly off the mark; a chi-square test comparing these expected results with the observed results was very significant, $\chi^2 = 23.8$, $p < .001$. We then considered the possibility that our subjects adjusted not the length of their images, but the area. If so, then we expected that 1.858, 2.019, and 2.167 sec, respectively, should be required to scan the three distances on a full-sized image; these estimates also failed to fit the data, $\chi^2 = 33.18$, $p < .001$. This failure was much more severe for the middle distance
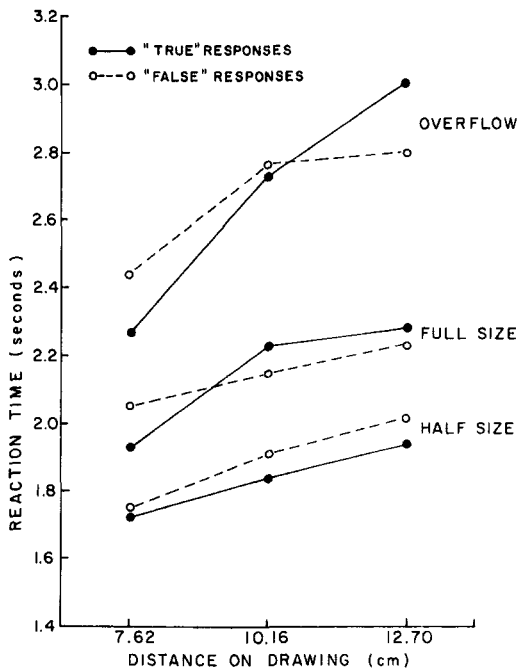
Figure 6. The results of Experiment 4: The time required to classify eyes located three distances from the mouth of faces imaged at three subjective sizes (overflow, full size, and half size).

than the ends (the deviation from the expected for the shortest distance was not significant $[p > .2]$, whereas the deviation for the longest was barely significant at the .05 level). We then speculated that subjects neither halved the lengths nor halved the areas but reduced size by performing some kind of a compromise between the two. Thus, we simply averaged our estimates from the two procedures and discovered that these means did not deviate significantly from the actual observed mean reaction times, $\chi^2 = 5.06$, $p > .05$; again, the best fit here was with the two extreme distances, $\chi^2 = 1.12$, $p > .5$.

Finally, error rates again tended to be positively correlated with reaction times. For true responses, error rates for the 7.62-, 10.16-, and 12.70-cm stimuli were 6.25%, 4.69%, and 6.25% for the overflow condition; 1.56%, 1.56%, and 4.69% for the full-size condition; and 7.81%, 1.56%, and 3.12% for the half-size condition. For false responses, error rates for the 7.62-, 10.16-, and 12.70-cm stimuli were 7.81%, 7.81%, and 10.94% for the overflow con-

dition; 3.12%, 7.81%, and 6.25% for the full-size condition; and 4.69%, 3.12%, and 1.56% for the half-size condition. In two cases, the 7.62-cm items incurred more errors than the 12.70-cm ones (true half size, false half size); t tests evaluating these differences were not significant, but the "true" comparison was marginal, $t(15) = 1.86$, $.05 < p < .1$. Thus, the faces incorporating the shortest distance from the mouth to the eyes may have been evaluated faster than they should have been, because of a lowered response criterion. If so, then the slope of the half-size condition (i.e., effects of increased distance on scanning time) may be steeper than is merited by scanning effects per se. In addition, in one case the errors for corresponding distances were greater for the half size than the overflow condition (true, 7.62 cm); this difference was not significant, $t(15) = 1.00$, $p > .1$, belying a speed–accuracy trade-off here. Finally, no subject deduced the purposes or motivation of this experiment.

Discussion

As expected, people again required more time to scan further distances across their images. This was reflected in three results: First, times increased with further separation between the mouth and eyes of the imaged stimuli; second, more time was generally required to scan across subjectively larger images; and third, there were increasingly large effects of increased distance (on the stimuli) for subjectively larger images. This last result was observed only with "true" responses, however. Although there was some difference in slope (i.e., the effects of increases in distance on the face) between the full-size and half-size conditions, these differences were not as large as would be expected if length were varied. This may have been because (a) people sometimes varied the area of their images and sometimes varied the length, or usually used a compromise of the two measures when determining how to scale the images, and/or (b) people may have performed some other sort of processing when evaluating short distances on subjectively small

images. That is, with the half-size images, the 7.62-cm separation may have seemed so slight that the eyes were visible even as one focused on the mouth. If so, scanning may not have been necessary to evaluate the imaged eyes, and these times thus may have been faster than predicted. This would result in a larger difference between the times necessary to evaluate eyes of faces with short and long distances than we expected—and hence less of a difference in the effects of increased distance in the half- and full-size conditions. The error rates suggested that subjects may in fact have been doing some more rapid, but less cautious, processing for the shortest distance in the half-size condition.

The failure to obtain slope differences for different-sized images on the false trials is not easily explained. There is some evidence, however (see Kosslyn, 1975), that people have more difficulty in using images to arrive at a "false" decision; the present data may simply reflect inconsistent use of imagery on the trials where the probed color was not in fact on the image.

Finally, it is worth noting that the results of this experiment allow us to eliminate one more possible nonimagery interpretation of the scanning effects. That is, one could claim that the closer two objects or parts are, the more likely it is that they will be grouped into the same "chunk" during encoding. Presumably, parts encoded into the same chunk are retrieved in sequence more quickly than parts in different chunks. In this experiment, size of an image was not manipulated until after the drawing was removed, precluding systematic differences in encoding among the three size conditions. Thus, the fact that subjectively larger images generally required more time to scan than did smaller ones seems to run counter to the notion that spatial extent affected scan times only because of a confounding between distance and the probability of being encoded into a single unit.

General Discussion

The present experiments converge in demonstrating that people can scan the distances embodied in images. More time was required to scan further distances, even when the same number of items fell between the focus and target locations. In addition, subjectively larger images required more time to scan than did subjectively smaller ones. Somewhat surprisingly, we found that the effects of distance persisted even when a person "zoomed in" on one part, such that the remainder of the image seemed to overflow. These results suggest that a part of an image may exist "waiting in the wings," ready to be activated into consciousness if needed. Finally, there were no effects of distance on decision times when people did not actually use their images, even though an image had been generated and focused upon. These results taken together indicate that images are pictorial in at least one respect: Like pictures, images seem to embody information about actual interval spatial extents. The present experiments support the claim that portions of images depict corresponding portions of the represented object(s) and that the spatial relations between portions of the imaged object(s) are preserved by the spatial relations between the corresponding portions of the image. These qualities are apparent in our introspections, and the present experiments suggest that people can operate on the representations we experience as quasi-pictorial mental images.

Given our results, how do we account for Lea's (1975) failure to find systematic effects of distance on evaluation times? First, the results of Experiment 3 suggest that Lea's results may simply reflect his failure to ensure that subjects responded only after "seeing" the target in their image. If left to their own devices in making decisions, subjects would probably find a nonimagery strategy to be faster, and such a nonimagery strategy would not result in distance influencing decision times. Second, Lea's task was so difficult (the mean reaction times reach as high as 8 sec) that effects of distance (which are measured in milliseconds, not seconds) may simply have been drowned out by the nonscanning components of this task. Finally, even if imagery were used, Lea's ordered search task, which involved counting successive items, may have induced

subjects to generate a sequence of separate images, each representing an object in the array, rather than to attempt to hold and then scan a complex image (see Kosslyn, 1975, for evidence that more complex images are more difficult to maintain). Weber, Kelley, & Little (1972) report that people can "verbally prompt" sequences of images, and something like this may have occurred in Lea's experiment. If so, then we have no reason to expect distance to affect response times.

In closing, it seems worthwhile to consider briefly two possible conceptualizations of how image scanning might operate. The most obvious notion (two variants of which were suggested by Kosslyn, 1973) is that scanning consists of moving an activated region over a spatial representation, somewhat like moving a spotlight across an unlit billboard. However, the spatial display used in representing images presumably also is used in representing sensory input from the eyes (see Hebb, 1968; Segal & Fusella, 1970) and hence only need represent information from some limited visual arc that corresponds to the scope of the eyes. If so, then one ought to find that one "hits the edge of the billboard" if one scans too far in any given direction. Many people report, however, being able to scan to objects "behind" them in an image and even being able to scan in a seemingly continuous circle across all four walls of an imaged room. There is another way of conceptualizing image scanning that deals with these sorts of observations naturally and easily.

In the Kosslyn & Shwartz (1977) computer simulation of visual mental imagery, it was most elegant to treat scanning as a kind of image transformation, in the same class as mental rotation and size alteration. Here, scanning consists of moving an image across the image display structure, the center of which is posited to be most highly activated (and hence, portions of the image falling under the center are most sharply in focus). In this case, the analogy would be to moving a billboard or sequence of billboards under a fixed spotlight. According to this notion, then, scanning 360° around

one in an image would be accomplished by continuously constructing new material at the edge and shifting it across the image display. If nothing else, this approach may have heuristic value by leading us to look for similarities among scanning and other image transformations.

In conclusion, the present results converge in supporting the claim that the experienced, quasi-pictorial surface image is functional and is not simply an epiphenomenal concomitant of more abstract "deep" processes. Comprehensive models of memory will probably have to include more than the sort of propositional list structures currently in vogue (e.g., Anderson, 1976; Anderson & Bower, 1973).

References

Anderson, J. R. *Language, memory, and thought.* Hillsdale, N.J..: Erlbaum, 1976.

Anderson, J. R., & Bower, G. H. *Human associative memory.* New York: Wiley, 1973.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior* 1973; *12*, 335–359.

Hebb, D. O. Concerning imagery. *Psychological Review*, 1968, *75*, 466–477.

Kosslyn, S. M. Scanning visual images: Some structural implications. *Perception & Psychophysics*, 1973, *14*, 90–94.

Kosslyn, S. M. Information representation in visual images. *Cognitive Psychology*, 1975, *7*, 341–370.

Kosslyn, S. M. Can imagery be distinguished from other forms of internal representation? Evidence from studies of information retrieval time. *Memory & Cognition*, 1976, *4*, 291–297.

Kosslyn, S M., & Pomerantz, J. R. Imagery, propositions, and the form of interval representations. *Cognitive Psychology*, 1977, *9*, 52–76.

Kosslyn, S. M., & Shwartz, S. P. A simulation of visual imagery. *Cognitive Science*, 1977, *1*, 265–295.

Lea, G. Chronometric analysis of the method of loci. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *1*, 95–104.

Segal, S. J., & Fusella, V. Influence of imaged pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology*, 1970, *83*, 458–464.

Weber, R. J., Kelley, J., & Little, S. Is visual imagery sequencing under verbal control? *Journal of Experimental Psychology*, 1972, *96*, 354–362.