COGNITIVE PSYCHOLOGY 24, 175–219 (1992)

The Reviewing of Object Files: Object-Specific Integration of Information

DANIEL KAHNEMAN AND ANNE TREISMAN
University of California, Berkeley

AND

BRIAN J. GIBBS
Stanford University

A series of experiments explored a form of object-specific priming. In all experiments a preview field containing two or more letters is followed by a target letter that is to be named. The displays are designed to produce a perceptual interpretation of the target as a new state of an object that previously contained one of the primes. The link is produced in different experiments by a shared location, by a shared relative position in a moving pattern, or by successive appearance in the same moving frame. An object-specific advantage is consistently observed: naming is facilitated by a preview of the target, if (and in some cases only if) the two appearances are linked to the same object. The amount and the object specificity of the preview benefit are not affected by extending the preview duration to 1 s, or by extending the temporal gap between fields to 590 ms. The results are interpreted in terms of a reviewing process, which is triggered by the appearance of the target and retrieves just one of the previewed items. In the absence of an object link, the reviewing item is selected at random. We develop the concept of an object file as a temporary episodic representation, within which successive states of an object are linked and integrated.

INTRODUCTION

This paper brings together techniques and ideas from two fields that are traditionally separate: (1) the study of object perception and of the con-

This research was supported partly by grants from the Canadian Natural Sciences and Engineering Research Council to Daniel Kahneman and to Anne Treisman, and partly by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant AFOSR 88-0206 to Daniel Kahneman and Grant AFOSR 87-0125 to Anne Treisman. The manuscript is submitted for publication with the understanding that the U.S. Government is authorized to reproduce and distribute reprints for governmental purposes, notwithstanding any copyright notation thereon. We are grateful to Roger Browse, Michael Satterfield, and Ephram Cohen for the development of our software, to Diane Chajczyk, Sharon Sato, and Amy Hayes for their help in programming and running the experiments, to Irvin Rock for helpful comments, to Sherlyn Jimenez, Kathleen Miszuk, and Julia Simovsky for help in preparing the manuscript, and to Marcia Grabowecky for making the figures. Requests for reprints should be addressed to Daniel Kahneman, Department of Psychology, University of California, Berkeley, CA 94720.

0010-0285/92 $7.50

tinuity of object identity through change and motion; and (2) the study of preview or priming effects, where the identification of a stimulus is facilitated if it matches a prime previously seen in the same context. We describe an initial set of studies of an *object-specific preview effect,* so called because the effect of a preview depends on whether the target and the prime are both seen as states of the same perceived object. Two theoretical ideas guided the research: a belief that perceptual objects are essential units of information processing, and the notion that the context within which a stimulus is processed is frequently evoked by the stimulus itself.

Some time ago we proposed an account of object perception as the process of setting up and utilizing temporary "episodic" representations of real world objects, which we call *object files* (Kahneman & Treisman, 1984). These are separate from the representations stored in a long-term recognition network, which are used in identifying and classifying objects. Several lines of evidence motivate this theoretical separation of object files or tokens from the stored types used to label their identity. One is the primacy of objects in determining the allocation of attention. Attention to any one property of an object causes even irrelevant properties of that object to be attended, as in the familiar Stroop effect. Moreover, the division of attention between relevant attributes is facilitated if the attributes belong to the same object (Treisman, Kahneman, & Burkell, 1983). The finding of an object-specific matching effect is a natural extension of these observations: we show that the focusing of attention on a target object not only enhances the salience of all its current properties—it also selectively reactivates the recent history of that object.

The maintenance of the perceived continuity of objects as they move, change, or momentarily disappear requires operations that relate the current state of the object to its prior history. When an object appears in a complex scene, a correspondence process attempts to match it to a particular object seen in the immediately preceding moments. We use the term *reviewing* to refer to the process in which a current target item evokes an item previewed in an earlier visual field. Reviewing facilitates recognition when the current and previous states of the object match, hampers it otherwise.

The next sections develop the theoretical notions of object file and reviewing. We then describe three experimental situations in which object-specific preview effects are observed, and report initial results. The paper concludes with a brief review of related research and theoretical ideas in the recent literature.

Object Files, Movement, and Change

Imagine watching a strange man approaching down the street. As he reaches you and stops to greet you he suddenly becomes recognizable as a familiar friend whom you had not expected to meet in this context. Throughout the episode, there was no doubt that a single individual was present; he preserved his unity (in the sense that he remained the *same* individual), although neither his retinal size, his shape, nor his mental label remained constant. Perception appears to define objects more by spatiotemporal constraints than by their sensory properties or by their labeled identity. The perceptual system is also capable of restoring continuity that has been briefly broken in the stream of sensory inputs. The man who reappears after walking behind a car will normally be treated as the same individual who was seen to disappear, provided that the disappearance was short and that the parameters of motion remain more or less constant.

Discontinuities of sensory input are also produced by movements of the observer—most obviously by movements of the eyes. The issue of object identity and continuity arises at every saccade. When the sensory stimuli change abruptly because of an eye or body movement, the perceptual system faces the task of matching each old object to its immediate history. People's unawareness of their saccades is a testimony to the success with which this task of restoring continuity is performed.

The experienced continuity of a changing object highlights the distinction between two senses of the term "identity." In one sense, the identity of an object is the label conferred on it when it is identified. In that sense, of course, the approaching man does not have the same identity when recognized as a friend as he had when he was assumed to be a stranger. In the other sense of the term, his identity and perceived continuity are precisely what he retains even as his properties and the label or name we give him vary. The ascription of continuity through change is essential to this second notion of identity.

The distinction between the two meanings of identity is related to a contrast between two views of perception (Kahneman & Treisman, 1984). In one view, perception is equated to recognition or identification. A stimulus is said to be perceived when it activates a set of nodes in long-term memory that represent its parts, properties, and categories of membership. (Whether this representation uses distributed or localized codes, and whether it conforms to symbolic or connectionist principles, is largely irrelevant to our discussion). We have called this view the "*display-board model*" of perception; perceptual experience is seen as depending on a succession of states of activation of units in semantic memory—rather like the display board used in some offices to identify the employees currently at work. The display-board model does not provide a natural way of representing the maintained perceptual identity of an object that is successively assigned different labels.

The object-centered approach that will be developed here emphasizes the distinction between identifying and seeing. We adopt the common notion that the visual field is parsed into perceptual objects and a relatively undifferentiated perceptual ground. We then assume that the main end product of perceptual processing of a stationary scene is a set of *object files,* each containing information about a particular object in the scene. Each object file is addressed by its location at a particular time, not by any feature or identifying label. It collects the sensory information that has so far been received about the object at that location. This information can be matched to stored descriptions to identify or classify the object, but it need not be. We can normally see completely novel objects with little difficulty, without knowing what they are. When the sensory situation changes, the information in the files is updated, yielding the perceptual experience of changing or moving objects. A file is kept open so long as its object is in view, and may be discarded shortly thereafter. The system bridges over the discontinuities produced by temporary occlusion, or by saccades, assigning current information to preexisting files whenever possible.

Visual objects are hierarchically organized; a group of dancers can be a visual object, as can an individual dancer, or her right hand. At any instant one of these levels may be dominant in the parsing of the scene. Tentatively, we assume that object files are set up at the preferred level, which is determined by the controlled allocation of attention (LaBerge, 1983; Navon, 1977) or by the automatic effect of bottom-up constraints and grouping factors. We also assume that there is some limit to the number of object files that can be maintained at once, so that the focusing of attention at the lower level causes more of the scene to be pushed into the perceptual ground. Attention to a higher level also has its costs, because the resolution of information within an object file is limited. The file for a complex object will represent its parts and the relations among them, but we surmise that the representation of a part is sketchier than if this part had been allocated an object file of its own, and also that the representation of relations among parts is more detailed than relations among separate objects.

Explicit recognition occurs at the level at which object files are currently set up. To mediate recognition, the sensory description in the object file is compared to stored representations of known objects. If and when a match is found, the identification of the object is entered in the file, together with information predicting other characteristics, its likely behavior, and the responses it should appropriately evoke, both affective and cognitive. The system of episodic object tokens is distinct from the semantic network of nodes and connections that mediates recognition. The identity of a changing object is carried by the assignment of information about its successive states to the same temporary file, rather than by its name or by its properties. Two identical red squares in successive fields may be perceived as distinct objects if the spatial/temporal gap between them cannot be bridged, but the transformation of frog into prince is seen as a change in a single visual object.

Reviewing, Correspondence, and Apparent Motion

Whenever a change in visual input is detected, current information about changing or reappearing objects must be assigned to existing object files; if this fails, a new file must be set up. Three distinct operations are needed to provide perceptual continuity through change. (1) A *correspondence* operation determines, for each object in the terminal display, whether it is "new" or whether it is an object recently perceived, now at a different location (Ullman, 1979); (2) a *reviewing* process retrieves the characteristics of the initial object, now no longer in view; (3) an *impletion* process uses current and reviewed information to produce a percept of change or motion that links the two views (Shepard, 1984).

The phenomenon of apparent motion provides significant information about the functioning of object files. In the basic demonstration of apparent motion a single object is presented at t_0, removed, and eventually replaced at t_1 by a single object in another location. When the spatial and temporal intervals fall within a critical range the perceptual impression is that a single object moves smoothly from the original location to the terminal one. When the exposure of the first stimulus is brief (<130 ms) the object is perceived as moving as soon as it appears. The normal percept of a brief stationary appearance is suppressed in the context of motion.

A critical observation is that in the classroom demonstration of apparent motion the percept of an object moving from one position to another can only be constructed from two successive stationary stimuli *after* the information about the second stimulus is presented. There may be no way of anticipating where the motion will go, or indeed that there will be motion. The object presented in the second display must retrieve (review) a trace of the object in the preceding one. If a close match is found, simple object continuity is perceived. If a physically plausible displacement or transformation could result in a match, the relevant object file may be updated and the transformation may be seen to occur (Shepard, 1984; Warren, 1977). However, if the new stimulus is sufficiently different from all its predecessors, or if the change in location is incompatible with the time interval or with any previous trajectory, a new object file may be opened and the sudden appearance of a new object will be consciously experienced.

When more than one object is present, there is a problem of *correspon-*

dence to be solved before apparent motion can be seen (Ullman, 1979). It is of interest that the similarity of such attributes as shape or color carries relatively little weight in the correspondence process, which is dominated by spatiotemporal contiguity of low-frequency information (Kolers, 1972; but see Green, 1986, 1989). This fits our notion that object files are addressed primarily by spatiotemporal characteristics rather than by properties or labels. It is also important that the set of rules that governs the assignment of histories to current objects is carried out under constraints of coherence: Whenever possible, a one-to-one mapping is preferred, and an object is not necessarily assigned to its nearest neighbor in the previous scene. In the Ternus effect, the perceived direction of movement of several objects can be determined by the location of one newly appearing object, plus the constraint that each object must occupy a location in both displays and be seen in coherent motion between them (Ternus, 1938).

Successive stimuli can be assigned to the same object file even when they are separated by an ISI and a spatial gap that exceed the range of apparent motion. The best example is the amodal completion observed by Michotte in what he called the "tunnel effect." If an object is seen to disappear (with gradual occlusion) and then to reappear some distance away (with gradual disocclusion) subjects report a compelling impression that a single object disappeared into a tunnel or behind a wall, traveled invisibly in the interval, and finally reappeared at the other end. Here again, the perceptual interpretation that bridges the gap between disappearance and reappearance can only be generated after the second event.

The Reviewing Paradigm

The present experiments explore a new paradigm to throw light on the temporal integration of information about objects that move or change. We use evidence of facilitation in naming latency to a repeated letter as a way to investigate how newly appearing objects are matched to possible past appearances within continuing object files.

The general features of the paradigm are the following. A typical experiment consists of two successive displays, respectively labeled the preview field and the target field. The preview field contains two or more different letters. The target field contains a single letter, which is to be named as quickly as possible. The successive displays for several variants of the paradigm are illustrated in Figs. 1–3. In each case, the target is selectively connected to one of the items in the preview field, because they are successively shown in the same place (Fig. 1; Studies 1 and 2), because the target is seen to arrive in apparent motion from the position of one of the preview letters (Fig. 2; Study 3), or because the target is presented in a frame which moved from another location in which it had originally contained one of the preview letters (Fig. 3; Studies 4–7). On

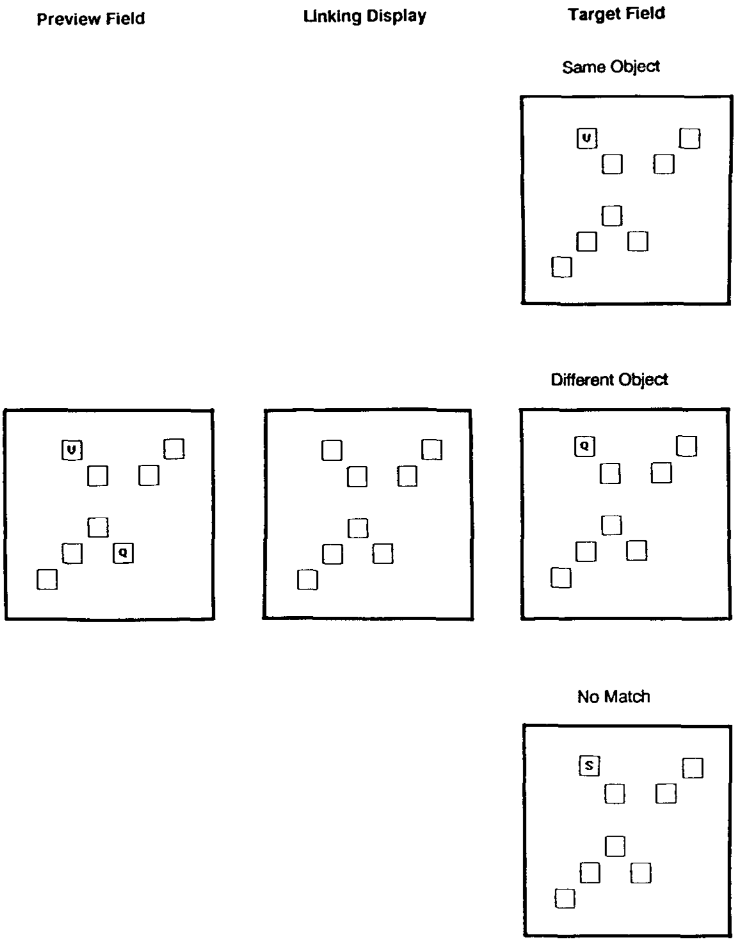

FIG. 1. Examples of displays used in Experiment 1 to test objective-specific reviewing effects. The three headings (Preview Field, Linking Display, and Target Field) show three successive displays, shown at different time intervals. The three vertically aligned displays under Target Field in each case show examples illustrating the three main relations between the previous field and the target field.

some trials, the target letter matches one of the items in the preview field. Three experimental conditions are defined as follows:

Same object (SO)—the target letter matches the preview letter seen as belonging to the same object (or in Experiment 3 as being the same object).

Different object (DO)—the target matches the preview letter seen as belonging to (or being) another object.

No match (NM)—the target matches neither of the preview letters.

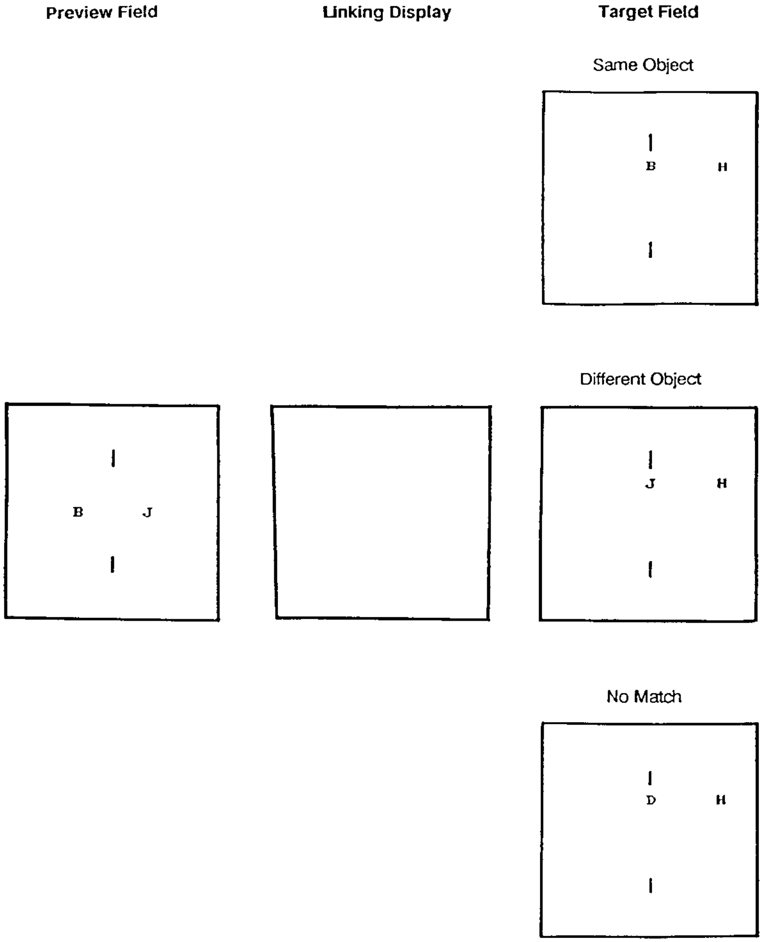

FIG. 2. Examples of displays used in Experiment 3 with apparent motion.

The comparisons of naming latencies in the three conditions yield several useful indices:

The *same-object preview effect* is the difference between naming times in the SO and NM conditions.

The *nonspecific preview effect* is the difference between the DO and NM conditions.

The *object-specific preview advantage* is the difference between the SO and DO conditions.

For most statistical analyses we focus on the last two indices (the first is their sum). The standard result that defines object-specific preview effects is that the latencies are quite similar in conditions DO and NM, and significantly faster in condition SO.

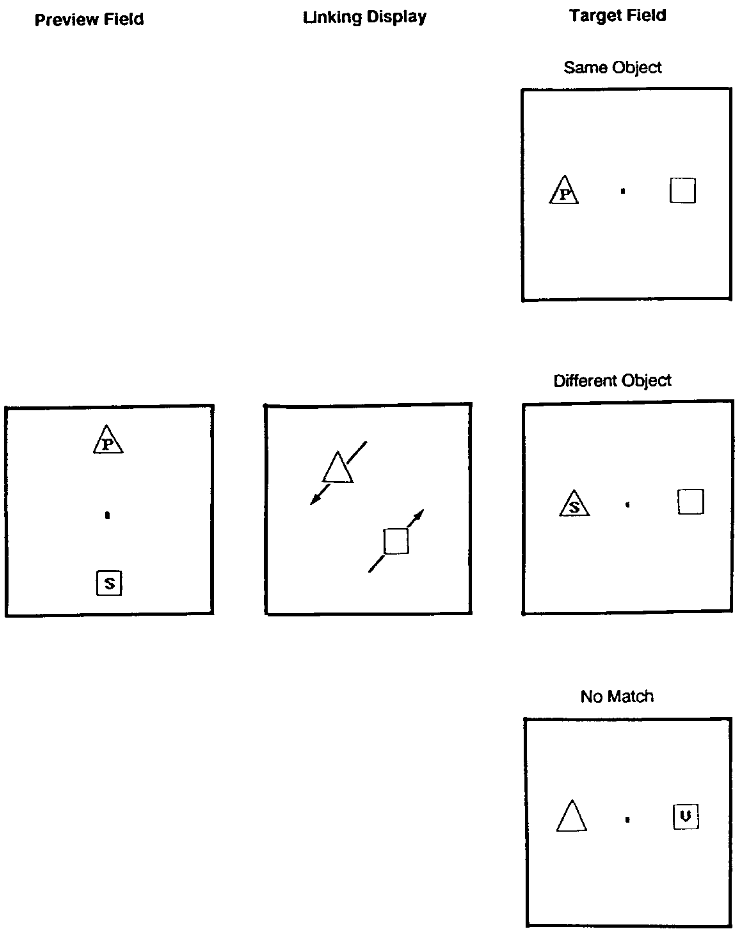

FIG. 3. Examples of displays used in Experiment 4 with real motion in the linking display.

We speak of preview effects rather than priming effects, because the label "priming" can be misleading in several ways. It suggests a beneficial effect of the presentation of a first stimulus on the processing of a subsequent one; but in fact a match between successive stimuli can produce interference as well as facilitation. Furthermore, the facilitation or interference are not necessarily produced by an activation process that is instigated by the "prime" and continues during the ISI between this stimulus and the subsequent target. Matching effects can be produced by the process we call reviewing, a retrieval process triggered by the target, which picks out the trace of a particular past episode (Glucksberg, Kreuz, & Rho, 1986; Kirsner, Dunn, & Standen, 1987; Koriat & Norman, 1988, 1989). The reviewing process described in this paper appears to involve the retrieval by a current stimulus of a plausible prior instantiation, which speeds up or impedes the identification of the current stimulus and the response to it. We test the assumption that, when a match is found within the same single object file, perception of the new stimulus will be faster than when a new object file must be created, or a radical and physically implausible change made to a previously existing file. Matching a previously perceived frog to the prince currently in view may require a few extra milliseconds.

We now turn to a description of a set of experiments using the reviewing paradigm. The experiments were conducted over a period of several years, at the University of British Columbia and at the University of California. Several of the seven studies described below bring together for expository purposes a number of separate experiments, which were most often planned and executed sequentially, and sometimes vary slightly in procedural details.

STUDY 1: REVIEWING WITH STATIONARY DISPLAYS

This study had three goals: the first was to look for object specificity in the letter matching benefit across two displays separated by a temporal interval. (Later experiments add spatial to temporal separation, using moving rather than stationary objects). The second aim was to distinguish the reviewing of object files from node priming, by varying the number of letter tokens independently of the number of letter types. The object-reviewing effect was expected to vary inversely with the number of objects presented (i.e., the number of letter tokens, independent of whether they are instances of the same letter), whereas the amount of node priming might be expected to increase with the number of repetitions of each prime. The third goal was to see whether the duration of the preview display had any effect. Would there be some minimal time required to set up the initial object representations tying the letters to the frames?

We presented eight square frames, and in two, four, or all eight of them flashed a letter either for 250 ms or for 100 ms (see Fig. 1). After an interval of 300 ms, a single letter appeared in one of the previously filled frames. The task was to name this target letter as quickly as possible. In one condition (with a 250-ms initial exposure), all the letters in the preview display were different from each other; in the other two conditions (one with 250 ms and one with 100 ms initial exposure), only two different letters were used, with one, two, or four tokens of each. The main questions were (1) whether the naming latency would be shorter when the target letter matched the letter that had previously appeared in the same frame, compared to when it matched a letter that had appeared in a different frame, (2) whether this object-specific effect depended on the initial exposure duration, and (3) whether the difference in latency to name a matching letter and a different letter in the same frame would depend on the number of letter types or on the number of letter tokens in the first display.

Method

Stimuli. The stimuli were shown on a Mitsubishi G479 monitor controlled by an IBM AT computer and Artist 1 Plus color graphics board. The technical specification of the phosphors, confirmed by direct measurements, indicates decay to 1% of original luminance in less than 30 ms. Subjects were shown a sequence of different displays. In the first display, eight red squares (luminance 9 cd/m^2) were shown on a black background; they were randomly located in a 6 × 6 matrix. Each square subtended 1.1° and the complete matrix subtended 8.9°. The squares remained visible throughout the sequence of events in a trial. After 166 ms, two, four, or eight white letters (luminance 37 cd/m^2) were flashed up inside two, four, or eight of the squares. The letters subtended 0.6° and were shown for 250 or 100 ms, followed by an interval of 300 ms with only the squares present. Then in the final display, a single letter was shown in one of the squares that had earlier contained a letter. In Experiment 1(a) on letter types, each letter in the initial display was different; in Experiments 1(b) and 1(c) on letter tokens, only two letter types were used in the initial display, and each was present as one, two, or four tokens. So a display of eight letter tokens might consist of four L's and four Q's. The preview field was shown for 250 ms in Experiments 1(a) and 1(b), and for 100 ms in Experiment 1(c). Letters were shown in uppercase and were selected from the following set: C,K,L,M,P,Q,S,T,V.

Procedure. Subjects were asked to watch the displays and to name the final letter on each trial as quickly as possible. A voice key detected the response and the computer registered the latency from the onset of the letter. There were three kinds of trials: on same-object (SO) trials, the target letter was the same as the letter that had previously appeared in the same frame; on different-object (DO) trials, the target letter was the same as a letter that had initially appeared in a different frame; on no-match (NM) trials, the final target letter differed from all the initial letters. The target letter never appeared in a previously empty frame. There were nine conditions in each of the three experiments, defined by three display sizes (two, four, and eight letter tokens) and three types of trials (SO, DO, and NM). Each experiment consisted of 12 blocks of 45 trials, giving a total of 60 trials per condition, run in one session lasting about 1 h.

Errors were monitored during the practice trials. They were rare. In the experimental trials the subjects were asked to score their own errors by pressing a key which caused the reaction time to be ignored both for that trial and for the next.

Subjects. Twelve paid subjects (9 women, all students at the University of California) were tested in Experiment 1(a) with two, four, or eight letter types and a 250-ms exposure. Twelve other subjects (8 women) were tested in Experiment 1(b) with two, four, or eight letter tokens, also at a 250-ms exposure, and 12 more (8 women) in Experiment 1(c) with two, four, or eight letter tokens and a 100-ms exposure.

Results and Discussion

The latencies for the various conditions are shown in Table 1, together with two difference measures: the *object-specific advantage* is the difference between naming latency in the same-object (SO) and different-object (DO) conditions; the *nonspecific preview effect* is the difference between the DO condition and the no-match (NM) condition. As is commonly the case in studies of priming, these difference measures do not necessarily reflect a single effect. The object-specific advantage combines a possible benefit of a match between the target and the letter that preceded it, and a possible cost of a mismatch between them. The nonspecific preview measure combines a possible facilitating effect of the prior presentation of the target and a possible reaction of surprise at its appearance in the wrong frame.

In this and in subsequent studies, we carried out three separate ANOVA's: on the mean RT across priming conditions, and on the two preview effects, respectively labeled the object-specific advantage (the difference between SO and DO latencies) and the nonspecific preview benefit (the difference between DO and NM latencies). The two preview effects are not independent, because they share the DO condition, but the separate analyses help the exposition. For convenience, the table also includes the results of a simple t test of each preview effect against zero.

In the ANOVA on mean RT, the main effect of experimental groups (the major columns in Table 1) was not significant [$F(2,33) = 2.23$]. There was a significant effect of display size [$F(2,66) = 84.81$, $p < .001$] but no interaction ($F < 1$).

The object-specific preview benefit was highly significant overall (M = 16 ms; $F(1,33) = 88.37$, $p < .001$). The effect of display size was significant [$F(2,66) = 26.69$, $p < .001$] reflecting a steep decrease of preview benefits as the number of previewed letters increased from two to eight. There was also a significant effect of groups [$F(2,66) = 3.68$, $p < .05$], indicating some reduction of object specificity with the short exposure duration.

TABLE 1
Mean Naming Latencies and Preview Effects in Study 1

Preview letters: Preview duration:	1(a) types 250 ms			1(b) tokens 250 ms			1(c) tokens 100 ms		
Display size:	2	4	8	2	4	8	2	4	8
Same object	474	500	517	473	495	513	520	540	552
Different object	509	517	523	504	517	517	536	552	551
No match	501	519	525	508	521	521	549	560	562
Mean RT	495	512	522	495	511	517	535	551	555
Preview effects									
Object specific	35**	17**	6	31**	22**	4	16	12	−1
Nonspecific	−8*	2	2	4	4	4	13	8	11

** These values are significant at $p < .01$.
* These values are significant at $p < .05$.

The nonspecific preview effect (DO–NM) was slight, but significant overall [$M = 4$ ms; $F(1,33) = 5.37$, $p < .05$]. There was no significant effect of display size and the effect of groups did not quite reach significance [$F(2,33) = 2.90$, $p < .07$].

The central result of these experiments is the finding of a substantial object-specific preview advantage. The response to a target letter was speeded when it appeared in a frame that had contained the same letter in the preview display. With preview displays of two items, the overall advantage of presenting preview letter and target in the same frame rather than in a different one was 27 ms. In contrast, the benefit of presenting the target letter in the "wrong" frame (compared to not showing it at all) was only 3 ms overall. There is some suggestion that the nonspecific benefit from a letter in a different frame was greater at the shorter exposure duration; the average nonspecific preview benefit did reach significance in that group [$t(11) = 2.59$, $p < .05$]. On the other hand, with 250 ms preview, the preview effect is almost entirely object specific. It seems that the specificity may in some conditions take time to become fully established.

A standard account of priming in terms of activation of nodes in semantic memory would predict no difference in priming from the letter in the same and in a different frame. In each case, the nodes for the priming letters would be equally primed, as would the locations in which letters had appeared. Only the specific combination of letter and location was changed in the different-object condition. To explain our results by node priming would require nodes that are specific for every letter in every possible location. Such a model has indeed been proposed by McClelland and Rumelhart (1981) and by Fukushima (1988). However, the experiments we describe later are not amenable to this account, because we observe preview benefits that are object specific without being location specific.

Our form of location-specific priming differs from two other reports which used consistent long-term associations between letters and positions (Banks & Krajicek, 1990; Lambert & Hockey, 1986). Their results demonstrated a tonic effect of expectancy, rather than the phasic integration of information in a temporary representation suggested by the present experiment.

The virtual absence of nonspecific priming in this experiment is a robust observation which recurs in most of our experiments with letters. In view of the ubiquity of repetition priming effects, this is a puzzling result: Why does the advance presentation of the target letter confer no general advantage in naming it, due to preactivation of the relevant nodes in a recognition network? The answer, we suppose, is that the vocabulary of stimuli and responses was quite small; it seems possible that all the items were primed to ceiling by the frequent repetition. In a subsequent experiment that used a large set of words as stimuli, the nonspecific priming was quite strong, although an additional object-specific advantage was still present (Treisman & Kahneman, manuscript in preparation).

The second important finding of our study is that the object-specific advantage was sharply reduced as the number of preview letters increased, and that this display size or load effect was about the same for repeated letter tokens as for different letter types. Although only two different letter types were present in the repeated token displays, the advantage of the same-object over the no-match condition was reduced from 35 ms with two tokens to 8 ms with eight tokens [Experiment 1(b)]. The limit seems to be set by the number of tokens that must be located in their respective frames, not by the number of different letter types to be identified. This result is consistent with the idea that separate object representations must be created for the separate perceptual entities present in the display. When the target letter appears, it is matched only to the earlier letter that formed part of the same object representation. The steep decrease of the object-specific benefit with display size and the reduced specificity with a short preview could reflect a limited rate at which object files can be established. Alternatively, there may be difficulties of retrieval with the more complex displays. The finding of an object-specific preview benefit suggests the existence of a specific, visual memory that preserves the spatial configuration of the initial display. Yet the interval between preview and target (an ISI of 300 ms) should be sufficient to eliminate iconic memory as a factor, according to the usual estimates.

The limited capacity of the memory that is probed in this experiment suggests a relatively long-lived form of postcategorical storage that is nevertheless still visual and object or location specific. However, it is possible that the longer persistence can be explained simply because the criterion for memory is less stringent in our experiment than in a partial report procedure; it is analogous to a measure of "savings" rather than of recall. In the following experiment we attempt to clarify the relation of these results to iconic storage, by comparing an ISI of 700 ms with an ISI of 300 ms.

STUDY 2: EFFECTS OF ISI

In this experiment we used two letter types and two, four, or eight tokens in each display, and tested two ISI's in a within-subject design. The method and stimuli were otherwise identical to those of Experiment 1(b). Eighteen subjects [two of whom had participated in Experiment 1(a)] were given 12 blocks of 45 trials at each ISI, in separate sessions. The order of ISI was counterbalanced. Two subjects were discarded because of exceptionally high error rates (up to 14% in some conditions) and highly variable reaction times. The mean error rate for the remaining 16 subjects was 1.1%.

The results are shown in Table 2. As before, we conducted separate analyses of the mean reaction time and the two preview effects. Reaction time was significantly faster at the longer ISI [$F(1,15) = 28.72$, $p < .001$] and increased with the size of the preview display [$F(2,30) = 95.49$, $p < .001$], but the two variables did not interact [$F(2,30) = 2.20$, NS].

The object-specific advantage (DO–SO) was significant overall [$F(1,15) = 36.62$, $p < .001$]; it decreased significantly with the number of items in the priming field [$F(2,30) = 6.83$, $p < .01$], but neither the effect of ISI nor the interaction of ISI with display size were significant [$F(1,15) = 0.15$ and $F(2,30) = 0.96$, respectively]. The lack of effect of ISI on object specificity would seem to rule out iconic memory as a factor in the reviewing benefits. Whatever visual trace mediates the reviewing effect remains equally effective across an interval within which iconic memory is thought to have completely disappeared.

The nonspecific preview effect (NM–DO) was significant overall [$M = 4.5$ ms, $F(1,15) = 8.69$, $p < .01$], indicating a slight benefit from a preview of the target in a different object. The main effect of ISI was significant [$F(1,15) = 7.49$, $p < .05$]. The linear trend relating increasing nonspecific benefit to increasing display size was not quite significant [$F(1,30) = 3.58$, $p < .06$]. This marginal trend is due mainly to an unusually large increase of naming latency in the NM condition with a long ISI. We surmise that when the subject has had 700 ms to consider a display that consists, say,

TABLE 2
Mean Naming Latencies and Preview Benefits in Experiment 2

Preview duration:	300 ms			700 ms		
Display size:	2	4	8	2	4	8
Same object	527	540	555	506	527	533
Different object	549	554	556	526	534	539
No match	543	556	556	528	546	555
Mean	540	550	556	520	536	542
		Preview effects				
Object specific	22**	14**	1	20**	7	6
Nonspecific	−6	2	0	2	12*	16**

** These values are significant at $p < .01$.
* These values are significant at $p < .05$.

of four K's and four T's, the appearance of a new letter as a target could be somewhat disconcerting.

The main finding of both Studies 1 and 2 is a robust object-specific preview benefit which diminishes sharply with increasing display size. This effect involves a spatial representation of objects, with a sharply limited capacity or limitation on access, but with considerably longer persistence than that attributed to iconic memory. Other researchers have based similar proposals on different lines of empirical evidence. Coltheart (1972), Phillips (1974), Sternberg, Knoll and Turock (1986), and DiLollo and Dixon (1988) all distinguish a schematic short-term visual memory from sensory or iconic storage. Together with ours, their evidence suggests the existence of at least one form of short-term visual storage which, unlike iconic memory, is not tied to spatial position, not subject to masking, and remains available for at least 600 ms. However, the great variety of experimental methods makes it difficult to ascertain whether a single process is involved in all their experiments, and in the priming effects that we observe.

STUDY 3: OBJECT SPECIFICITY IN THE PREVIEW EFFECT WITH OBJECTS IN APPARENT MOTION

Our next concern is to distinguish the existence of object-specific perceptual representations ("object files") from the persistence of information tied to particular visual locations. We dissociate these effects by using objects in real or apparent motion, presenting the target letter in a different location from the matching letter in the initial display. We describe first an experiment in which we used apparent motion to link one of the items in the preview field to the target.

The stimuli in this experiment were two successive displays, each containing two letters, with the second pair displaced diagonally either above or below the first pair (see Fig. 2). No frames were used in this experiment. The preview display was centered between the fixation marks. The target display always included a letter just below the top fixation bar or just above the bottom one, with a second letter either to its right or to its left. The time intervals we used gave viewers a clear impression that a single pair of letters moved coherently from the initial to the final locations. Thus, the two letters in the first display were integrated perceptually with the two in the second display. In Fig. 2, the B in the first frame turns into B in the SO condition, J in the DO condition, and D in the NM condition; the J in the first frame becomes H in all these conditions. Ternus (1938) first described a similar display in which global apparent motion is seen, its direction determined by the relative location of the peripheral item in the second display. This happens whether the shapes are the same or not. Kolers (1972) showed that shape has little influence on apparent motion when the spatial and temporal intervals are optimal.

In our displays too the perceived direction of motion was determined by the location of the peripheral letter in the second display. Nothing in the initial display indicated which of the two previewed letters would be seen to move into the target location and thus to "become" the target letter. We were interested in discovering whether there would be a preview benefit when the target matched one of the initial letters and if so whether it would be object specific, occurring only when the target was perceptually integrated with the matching prime. In this experiment, differences in location could play no role in selective preview effects, since the two preview letters were equidistant from the location of the target. Only the illusory motion linked the target to one previewed letter rather than the other, and the motion was determined only after the appearance of the target. Lingering activation in particular spatial locations could generate only nonspecific benefits.

Method

Stimuli. The apparatus used to control and display the stimuli was the same as in Experiment 1. The letters were red, the fixation marks were white, and the background field was dark. The letters shown on each trial were chosen randomly from the set B,C,D,F,H,J,K,S. Viewed from a distance of 60 cm, each letter subtended 0.9°. Two vertical bars, each subtending 0.4° in length and separated vertically by 3.9°, were used to control fixation. Each trial began with presentation of these bars for 100 ms. Two letters were then shown 1.5° to the left and right of the center of the display, for 100 ms in one session and 1 s in another session. After an ISI of 33 ms (for the 100-ms preview) or 0 ms (for the 1-s preview), two new letters were shown. One of these letters, the target, was presented just below the top fixation bar or just above the bottom one. The other letter in the target field was 3.1° either to the left or right of the target. The target field remained in view until the subject responded.

Procedure. There were eight types of trials in each session, four conditions with the display arriving at the target position from the left side, and the same four conditions with the target display arriving from the right. On SO trials, the target letter was the same as the preview letter that appeared to fuse with it perceptually; on DO trials, the target letter was the same as the preview letter that appeared to move to the more peripheral location; on NM trials the target letter differed from both the initial letters; on asterisks trials, the preview display consisted of two asterisks. The peripheral letter in the final display always differed from the other three. The conditions were randomly mixed in each block of trials. The two sessions with different preview durations were run on separate days in counterbalanced order in a within-subject design. Each session lasted about 1 h.

Subjects were asked to name the letter that appeared between the bar markers as quickly as they could without making errors. Their response times were measured from the triggering of a voice key. Naming latencies were collected in 36 trials in each of the eight conditions. Subjects were given one block of practice trials before each experiment began. As in the previous experiments, they scored their own responses. The correct letter was presented in the center of the screen on each trial, after the experimental displays. Subjects pressed a right-hand key if their response matched the correct letter and a left-hand key if they were incorrect, or if some other sound triggered the voice key before they spoke (a rare event).

Subjects. Thirteen paid subjects (all students at the University of California, 5 women) participated in this experiment.

Results and Discussion

Errors averaged 1.6% across all conditions. The reaction time data are shown in Table 3.

The ANOVA on mean RT showed no effect of either duration of the priming field or the direction of motion, and no interaction [$F(1,12) = 0.32$, and $F(1,12) = 2.95$, respectively]. Separate ANOVA's were run as before on the two preview effects, and a separate ANOVA on the difference between the no-match and the asterisks conditions.

The object-specific benefit was significant overall [M = 15 ms, $F(1,12) = 24.44$, $p < .001$]. There was no main effect of either prime duration or direction of apparent motion [$F(1,12) = 0.37$ and $F(1,12) = 1.35$, respectively], but the interaction of these variables was highly reliable [$F(1,12) = 11.28$, $p < .01$]. Leftward movement gave more object-specific benefit at the short preview exposure and rightward movement gave more at the long. A likely explanation for this interaction is that subjects attended to the letters in the preview field in a left to right order, giving an advantage to the left letter, read first, when the exposure was brief and to the right letter, read last, when the exposure was long.

The nonspecific preview effect was actually a small cost [M = 4 ms; $F(1,12) = 5.89$, $p < .05$] incurred when the target letter was previewed in the wrong location, relative to not previewed at all. The cost may be due to some conflict in the observed direction of motion between the global

TABLE 3
Mean Naming Latencies and Preview Benefits in Experiment 3

Preview duration:	100 ms		1000 ms	
Motion direction:	L	R	L	R
Same object	462	472	467	465
Different object	488	482	473	485
No match	479	479	472	479
Asterisks	467	471	465	468
Mean	474	476	469	474
	Preview effects			
Object specific	26**	10	6	20**
Nonspecific	−9	−3	−1	−6
New letter vs asterisks	12**	8**	7	11*

** These values are significant at $p < .01$.
* These values are significant at $p < .05$.

Ternus effect and a weak tendency to match shapes and reverse the Ternus motion in the DO conditions. There was no significant effect of experimental conditions on the nonspecific preview effect.

The asterisks gave significantly faster latencies than the new letters [$M = 10$ ms, $F(1,12) = 18.70$, $p < .001$] indicating that the object-specific benefit, at least in this display, was due to a cost of assigning different letter identities to the same object, rather than to the benefit of a match. The asterisks were clearly irrelevant to the task and may have been filtered out. Later experiments test the effects of blank previews and of digit previews.

The central finding, as with stationary displays, was the object-specific effect of the preview field. An account in terms of lingering activation in a location can now be ruled out, because the target appeared in a different location from the preview letters, and equally far from the matching letter in SO and in DO trials. The link between target and prime was determined entirely by the location of the irrelevant letter which selected the direction of the illusory motion attributed to the whole display. The processing of the target was selectively affected by the initial letter that was perceptually integrated with it, and largely unaffected by the other letter in the preview field.

A second important conclusion from the present experiment is that the preview effect is determined in a backward process, which is controlled by the target display. The initial displays on SO and DO trials were identical in terms of any forward effect of priming or interference. In both cases two letters were presented, of which one was then repeated as a target. The difference between the conditions arose only after the onset of the target field, because the object correspondence between the two fields was only determined at that time. The preview effect observed in the present study was therefore a "backward" effect (Glucksberg et al., 1986; Kahneman & Miller, 1986; Kiger & Glass, 1983; Koriat, 1981; Seidenberg, Waters, Sanders, & Langer, 1984). A characteristic of the second display (in this case the position of the irrelevant letter) controls the selection of the preview item with which the target will interact. The term "priming" is clearly awkward to describe what is going on here, because it suggests a forward effect from prime to target, rather than a backward selective process.

The results do more than indicate that the target selects the earlier item with which it will interact. They also pin down the basis of selection, and eliminate some plausible alternatives. For example, it would be theoretically possible for the target letter to select the item in the preview field that most closely resembles it; selection by similarity is invoked in many accounts of priming and of episodic retrieval (e.g., Kahneman & Miller, 1986; Kirsner et al., 1987; McClelland & Rumelhart, 1985). In the present experiment, however, selection by similarity would produce equal priming benefits for the SO and DO conditions, since the target-to-be is present in the preview field in both. The absence of any preview benefit in the DO condition shows that similarity played no role at all in this experiment. The position of the letters in the simple configuration to which they belong appears to be the sole basis for selective interaction. It is this spatial property that defines the object file to which successive items are assigned. The rules of correspondence that govern apparent motion in complex displays (Ullman, 1979) also govern the reviewing process that yields object specific preview benefits in the present experiment.

STUDY 4: REVIEWING WITH MOVING FRAMES

In the remaining studies to be described in this paper, we used moving frames to create selective links between items in the preview and target fields. In a typical experiment, the computer presented two outline shapes (frames), then two letters inside the frames (see Fig. 3); the letters then disappeared and the frames moved to new positions, equidistant from the initial positions; finally a single target letter appeared in one of the frames. Reaction time for naming the letter was, as before, the dependent variable. As in the previous experiments, there were three types of display, labeled SO, DO, and NM. A cue (vertical bar markers) appeared during the trial to indicate the position of the target letter.

In Study 4 (which was actually run as a series of experiments with different groups of subjects), we tested a number of variations on the same basic paradigm. We varied the duration of the preview field, the duration of the motion of the empty frames, the contents of the preview display, and the timing of the cue that indicated the position of the target. Most of our studies in this paradigm have used an "early" cue, in which the bars that indicated the position of the target appeared simultaneously with the onset of the motion of the frames. The purpose of the early cue was to direct attention immediately to the target position, to prevent as far as possible selective attention to one of the two preview letters. To check on the possibility that an early cue might direct the subject's attention to one of the moving frames even before the target arrived, we included "late cue" conditions, in which the bars indicating the location of the target appeared simultaneously with the target itself.

By varying the duration of the preview and the speed of motion we explored the possibility that the temporal integration that mediates the reviewing benefit depends on the preview still being only partially processed when the target appears. If the preview is seen as an event that is clearly temporally separate from the target, the two might not be combined. On the other hand, if reviewing represents a retrieval by the later input of its most likely prior instantiation, as we suggested in the Introduction, there is no reason to expect much effect of the preview duration. So long as the preview letters are presented long enough to be identified and linked in separate object files to the frames that contain them, the critical variable should be whether the final letter is seen as a new instantiation of one of the earlier letters. This in turn will depend on whether motion (real or apparent) has linked the appearances of the target and the preview across space and time.

Method

Stimuli. The displays were presented in white on a dark field on a DEC VR17-LC graphics terminal with P40 phosphor, controlled by a DEC PDP 11/34 computer. The display was viewed through a blue filter (Kodak Wratten 47), which cut down visible persistence. Viewing distance was 60 cm.

The following sequence of stimuli was shown on each trial. First two outline shapes appeared, a triangle and a square subtending 2.1° vertically and horizontally, centered 2.5° above and below a fixation point (see Fig. 3). After a delay of 500 ms, capital letters (.6° tall) appeared inside the two frames. The letters were selected from the set B,C,D,F,G,H,K,J,S. They were presented for 20 ms in the "short exposure" conditions, or for 1 s in the "long" conditions. The frames then moved in apparently smooth motion (new images were drawn every 13 ms) to positions centered 4.2° to the left or right of fixation. The triangular frame was always at the top of the display initially, and it moved equally often to the left and to the right; the square frame always started at the bottom of the display and moved to the opposite side. The motion lasted 130 ms in the "fast" conditions and 590 ms in the "slow" conditions. The location of the target was cued by bars above and below it, and the timing of that cue was also a variable in these experiments. The "early" cue was shown with the onset of the preview letters (with the short exposure) or when the frame began to move (with long exposures); the "late" cue appeared at the same time as the target. There were six conditions defined by characteristics of the display sequence, four resulting from the combination of long or short preview duration and early or late cue, and two more with the long preview duration and with slow (590 ms) motion, again with either an early or a late cue. In the conditions combining early cue and short exposure the preview field contained digits rather than letters on a quarter of the trials. Nine subjects in the late-cue, long-exposure condition and all eight subjects in the early-cue, slow-movement condition had empty frames in the preview fields on a quarter of the trials.

Subjects and procedure. The six conditions were run as separate experiments, with some overlap of participants. A total of 71 subjects took part in at least one experiment. Of these, 14 took part in two experiments. Each subject completed one block of 48 trials for practice, followed by five experimental blocks, divided equally among the types of trials included in that experiment.

Results and Discussion

The results of this set of experiments are presented in Table 4. When the statistical analysis called for comparisons across experiments, we treated the various groups as independent, in spite of some overlap of membership. This is generally a conservative procedure.

We analyzed the set of six experiments as a 3×2 design with three display types and two values of the timing of the cue that indicated the

TABLE 4
Mean Naming Latencies and Preview Benefits in Study 4

Motion: Preview duration: Cue	Fast 20 ms		Fast 1 sec		Slow 1 sec	
	Early	Late	Early	Late	Early	Late
N	(14)	(19)	(14)	(21)	(8)	(10)
Same object	484	531	495	543	445	531
Different object	515	544	518	557	486	550
No match	516	556	519	564	484	549
Mean	505	544	511	555	472	543
		Preview effects				
Object specific	31**	13**	23**	14**	41**	19*
Nonspecific	1	12**	1	7*	−2	−1

** These values are significant at $p < .01$.
* These values are significant at $p < .05$.

target position. The analysis of mean RT yielded a highly significant effect of the timing of the cue—as might be expected, the early cue yielded shorter latencies [M = 48 ms, $F(1,80) = 20.31$, $p < .001$]. The three display types did not differ significantly [$F(2,80) = 1.45$], and did not interact with cue timing [$F(2,80) = 0.64$].

The object-specific benefit was highly significant in each group separately, and was, of course, significant overall. The benefit was significantly greater with the early cue than with the late cue [$F(1,80) = 15.70$, $p < .001$]. Again, the display types did not differ significantly [$F(2,80) = 2.23$], and did not interact with cue timing [$F(2,30) = 0.90$]. The nonspecific benefit was barely significant overall [M = 4.5 ms, $F(1,80) = 3.97$, $p < .05$], and marginally larger with the late than with the early cue [$F(1,80) = 3.67$, $p < .06$].

In summary, the results suggest that the object-specific effect is independent of preview duration, slightly increased with slower motion, and larger with an early cue than with a late cue. The nonspecific effect, in contrast, appears to be slightly larger with the late cue.

This pattern of results suggests that on some occasions the presentation of the target causes reviewing of a preview letter that was presented in the other object; the results suggest further that this failure of selectivity is very rare when the early cue is used, and somewhat more common with the late cue—at least with the present display. The early cue was designed to direct the subject's attention immediately to the position in which the target would appear. In the absence of such a cue, attention could be captured on some trials by one or another of the two moving frames, following that frame to its final position, and inducing a bias in favor of the letter it had contained. On one-third of these occasions, the target would actually appear in the frame that the subject did not follow or attend. The occasional captures of attention by the "wrong" frame can only lower performance in the SO condition, because the letter that is reviewed on these trials does not match the target. By the same logic, performance in the DO condition should improve when object selection fails, because the item reviewed will then match the target. In the NM condition, of course, the direction of attention during motion should make little difference because both preview letters differ from the target.

An alternative interpretation of the difference between early and late cues is that presentation of the early cue quickly directs attention to the frame that is moving to the cued location, and thereby diverts it from the other frame and the letter it contains. This account is rather implausible, for several reasons. First, it requires exceptional agility of the mechanism that directs spatial attention. Note that the early cue is a pull cue (Jonides, 1981), of the kind that generally produces an automatic redirection of attention. Note as well that the duration of motion in the fast condition is only 130 ms, which appears too short to contain two movements of attention guided by an inference from the direction of motion. In another experiment with an early cue, we found perfect object specificity even when the duration of the motion was only 55 ms. The preview effect was also highly specific in the apparent motion design of Study 3, which was effectively a late cue situation.

The perfect object specificity of the preview effect with a long-duration preview and an early cue is a significant result. The two letters shown in that field were equal in potential relevance for a full second, regardless of the cueing condition. Although both letters were surely perceived and identified during that time, a preview of the target in the "wrong" object had no effect. This seems to us to be strong evidence against the idea that the preview benefit observed in these experiments is mediated by node activation, because both nodes should have been equally, and probably fully, activated. Note again that the results cannot be explained by location-specific nodes, because the target was never presented in the same location as the prime.

Our aim in varying the speed of motion was to push the boundary conditions still further from any iconic representation by extending the interval between the offset of the previewed letter and the appearance of the target. Would the two still be integrated when the linking motion took nearly 600 ms? The answer is clearly yes; with the early cue the object-specific preview benefit was somewhat larger with the slow than with the fast motion. There is no requirement that sensory traces of the preview letters remain active to ensure a preview benefit. With fast motion, subjects sometimes reported an illusion of seeing the preview letter move with the frame. When the motion took 590 ms, the letter did not remain phenomenologically visible; yet it was still just as likely to be integrated with the target in the process of reviewing.

In this experiment we used different outline shapes as frames, a triangle and a square. This could potentially have been relevant to the object specificity of the preview benefit. However, one condition in Experiment 6(b) replicates essentially the same conditions using identical square frames, and shows the same amount of object-specific benefit. What seems to matter is the perceptual continuity of the frame that links the preview to the target letter, rather than the association between its shape and the letter it contains.

Finally, four of the experiments had additional control conditions, as well as SO, DO, and NM trials: on 25% of trials, two experiments showed digits in the preview field and two experiments showed empty frames. The digits in the short-exposure conditions yielded significantly faster latencies than the no-match letters [$M = 8$ ms, $t(13) = 3.07, p < .01$ with the early cue; $M = 8$ ms, $t(18) = 2.51, p < .05$ with the late cue]. On the other hand, the advantage of the blank fields over the no-match letters did not reach significance (means of 7 ms for long-exposure, late-cue, and 0 ms for the slow-movement, early cue condition]. There seems in these experiments to be a small cost associated with reviewing a different letter rather than an irrelevant symbol or an empty frame, but the main reviewing effect here is due to the advantage of a matching preview letter in the same frame.

STUDY 5: MOVING FRAMES WITH TWO OR FOUR PREVIEW LETTERS

This experiment tests the effect of the number of items in the preview field with moving frames, and extends the comparison of early and late cues to a display that does not encourage attentional following of a particular frame. The experiment was designed to test the generality of the inverse relationship between the size of the object-specific and nonspecific preview effects, which was observed in some conditions of Study 4. We expected that an increase in the number of previewed letters might reduce the selectivity of the priming effect, as it appeared to do in Study 1.

Method

Stimuli. The displays were presented with the same computer and graphics terminal as those of Study 4. As shown in Fig. 4, the four frames that contained the letters were squares, 1.1° a side, which were initially centered at equal intervals on the circumference of an imaginary circle 4.0° in diameter. The position of the set of frames was randomly chosen from three possibilities (with a square at either 12:30, 1:30, or 2:30 on an imaginary clock). The letters shown on a trial were selected from the set of all consonants except G,N,Q,R,V, and W. On 25% of trials the preview letters were replaced by plus signs. The four frames were first shown alone for 500 ms, followed by a 30-ms exposure of either two or four letters (or plus signs). When two letters were shown, they were located at opposite ends of a diameter. The preview letters disappeared and the frames moved along straight lines to new positions, computed by combining an expansion of the circle to 64 mm diameter with a ⅛ turn in a clockwise or counterclockwise direction (see Fig. 4). The display created an impression of simultaneous expansion and rotation of the whole pattern. The frames retained their vertical orientation during the movement, which took 130 ms and consisted of 10 stops. A target letter then appeared in one of the frames. On SO trials the target matched the letter that had previously appeared in that frame; on DO trials it matched a letter previously shown in another frame, on NM trials it was a new letter; on the remaining trials the preview display contained two or four plus signs. Each of these trial types occurred

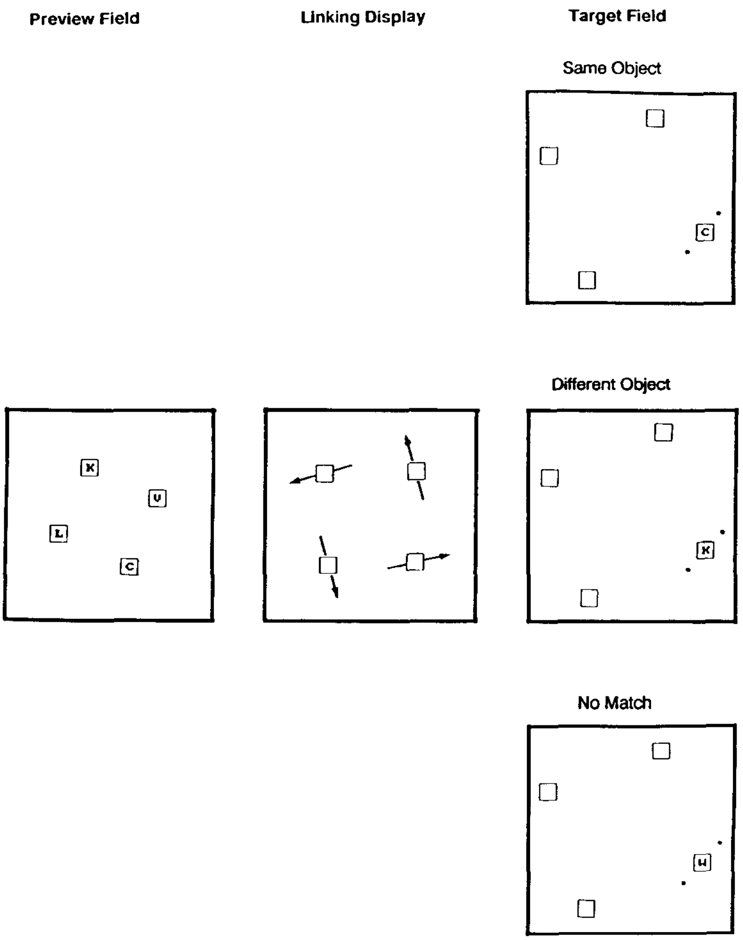

FIG. 4. Examples of displays used in Experiment 5.

equally often in randomized order. The cue consisted of two asterisks presented at opposite corners of the target frame.

Procedure. There were three groups of subjects. One group of 16 subjects was tested with the early cue and two or four preview letters. The other two groups were tested with a late cue; one group of 12 subjects had separate blocks with two or four preview letters, with order counterbalanced; finally, a group of 10 subjects encountered the two types of trials randomly mixed in each block. The aim was to see whether there were any strategic effects which would depend on knowing how many letters to expect.

Results

The results obtained with mixed and with blocked trials were essentially identical, and the data for the 22 subjects tested with the late cue were therefore pooled. The mean naming latencies and preview effects are shown in Table 5. (The final row of the Table is discussed later).

Naming latency was faster with an early cue than with a late cue [M = 50 ms, $F(1,36) = 6.14$, $p < .05$], and faster with two than with four items in the preview display [M = 15 ms, $F(1,36) = 50.17$]. The two variables did not interact ($F < 1$).

The object-specific preview benefit (DO–SO) was significantly greater with two than with four previewed letters [M = 18 ms, $F(1,36) = 10.12$], but there was no significant difference between early and late cue and no interaction between these variables ($F < 1$). The effect of display size is compatible with the findings of Experiment 1(a), with stationary objects.

The nonspecific preview effect (NM–DO) did not quite reach significance [M = 4 ms, $F(1,36) = 3.65$, $p = .07$]. There was no significant

TABLE 5
Mean Naming Latencies and Preview Benefits in Study 5

	Early cue		Late cue	
Number of preview items:	Two	Four	Two	Four
Same object	576	603	623	653
Different object	611	622	663	674
No match	611	626	667	682
Pluses	600	606	642	654
Mean	600	614	649	666
Cost of new letter vs plus	11	20**	25**	28**
	Preview effects			
Object specific	35**	19**	40**	21**
Nonspecific	0	4	4	8*
Total benefit	35	35	48	53

** These values are significant at $p < .01$.
* These values are significant at $p < .05$.

effect of display size [$F(1,36) = 0.46$], or of cue timing [$F(1,36) = 1.02$], and no sign of any interaction between these variables.

There is a notable difference between this study and Study 4 in the effects of cue timing. In the present results, unlike Study 4, the object specificity of preview benefits was hardly reduced by late cueing. We tentatively attribute this finding to the coherent motion of the four frames in the present experiment. The display induced a global percept of an expanding circle, and thereby reduced the tendency to follow one particular frame to its destination before the late cue appeared.

Finally, the comparison of NM trials with plus trials yielded a highly significant difference in favor of the latter [$M = 21$ ms, $F(1,36) = 72.38$, $p < .01$], and a significant interaction with early vs late cue [$F(1,36) = 5.06$, $p < .05$]. The cost of a new letter relative to a plus sign was greater for late cue trials. Thus, in this study there seems to be more interference from nameable letters than from an irrelevant symbol, perhaps because the letters were more perceptually confusable with the target and perhaps because of potential response conflict. However, there was also some facilitation from the identity match in the SO condition relative to the pluses, at least with two preview letters [$t(15) = 3.57$, $p < .01$ for the early cue and $t(21) = 6.62$, $p < .001$ for the late cue].

Discussion: A Model of Reviewing

The discussion of Study 4 introduced the notion of a reviewing process, which usually picks the item presented in the same object as the target, but sometimes selects another item. If we assume that the difference between reviewing a matching or a mismatching item is unaffected by the basis of selection, and that no more than one item is reviewed on any trial, the fraction p of trials on which reviewing is object specific can be estimated from the data for any experimental situation. With two items in the preview field, the object selectivity of reviewing in a particular experimental condition can be estimated by the following ratio, where the labels for experimental conditions denote the corresponding response latencies: $p = (\text{NM–SO})/[(\text{NM–SO}) + (\text{NM–DO})]$.

Allowing for the effect of display size, this random selection model leads to the following expression for p when the preview field contains n items:

$$p = (\text{NM–SO})/[(\text{NM–SO}) + (n-1)(\text{NM–DO})].$$

Applying this measure to the data of Table 4 yields a crude but suggestive indication of the effects of the timing of the location cue and of the number of priming items. The computed values for the proportions of object-specific reviewing trials are .95 and .66, respectively, for two and four items with an early cue, and .90 and .55 for two and four items with a late cue. Object selectivity is nearly perfect with two primes, substantially impaired with four. There seem to be limits to the number of letters that can be bound to specific frames, at least in the limited time available here.

Another measure of interest, labeled *total preview benefit* (TPB), is defined next. The following simple model provides a rationale for this measure. Assume that a single item is reviewed on a fraction r of trials, that no item is reviewed on the other trials, and that reaction time is A ms faster when the target matches the reviewed letter than when it does not. It is easily seen that in the case of a priming field with two items the following equations hold:

$$\text{TPB} = r.A = (\text{SO–NM}) + (\text{DO–NM}),$$

and also

$$(\text{SO–NM}) = p.r.A \text{ and } (\text{DO–NM}) = (1-p).r.A.$$

In the general case of n items in the preview field, these equations become

$$\text{TPB} = r.A = (\text{SO–NM}) + (n-1)(\text{DO–NM})$$
$$(\text{SO–NM}) = p.r.A \text{ and } (\text{DO–NM}) = (1-p).r.A/(n-1).$$

The process model has three parameters, but our design provides no way of separating the true matching advantage A from the fraction r of trials ón which an item is reviewed. We can estimate only the product of these values, which is our measure of total preview benefit, TPB.

We did not report the measures of selectivity and total preview benefits for previous studies, because the added expository burden would not have been rewarded by illuminating findings. First, there were several experiments in which a central assumption of the model appeared to be violated: the model assumes that a match of the target to the reviewed item yields the same benefit of A ms, regardless of whether the match occurs in the SO or DO conditions. This assumption is incompatible with the occasional finding (in Studies 1 and 3) of cases in which performance in the DO condition was reliably worse than in condition NM; we interpreted these results as indicating surprise at the appearance of the target letter in the "wrong" object. Second, the individual estimates of TPB are quite variable, precluding meaningful comparisons across groups.

In the present experiment, the values of TPB were 35 ms for both two and four primes, with an early cue. The corresponding values of TPB with the late cue were 48 and 53 ms. In both cases, the values of TPB appeared to remain constant across different levels of object specificity, induced by differences in the number of previewed letters. The difference in TPB between early and late cue was larger, but not significant in a between-group comparison [M = 16 ms, $t(36)$ = 1.38]. The experiments reported iNEWn Study 6 were conducted specifically to test the constancy of total matching benefits in a within-subject design, which could be sufficiently sensitive to reject it. They explored two other ways of varying the degree of object specificity: in one case the linking motion was replaced by shared color as a possible basis for selective reviewing; in another, the frames moved but not to the final destination of the target.

STUDY 6: A TEST OF THE REVIEWING MODEL

In the model that was just introduced, a target may occasionally [with probability (1-p)] select for reviewing an item that was not presented in the same object. Thus, the process of reviewing is guided by the structure of object relations, but is not necessarily restricted to successive states of the same object. In the next experiment we examine pairs of closely similar displays which differ in the support they provide for object-guided reviewing. In the terms of the minimal formal model introduced above, we set out to create conditions in which the efficiency of selection by object (indexed by p in the model) would vary, in order to test whether total preview benefits (TPB) would remain approximately constant.

Experiment 6(a) used color as a possible basis for selective retrieval of the prime, to replace the linking motion used in Studies 3, 4, and 5. The preview field consisted of two stationary frames in which letters were shown, in two different colors. The target letter sometimes matched the preview letter shown in the same color (SO), sometimes the other letter (DO), and sometimes neither (NM). We compared the preview benefits in this situation to those produced, with a very similar display, where one of the preview letters was linked to the target by movement of its frame.

In Experiment 6(b), the two frames that contained the preview letters always moved, but on unlinked-motion trials the moving frames did not reach the box in which the target was eventually shown (see Fig. 5). This situation provides no obvious basis for selection. Will subjects still retrieve one of the two letters (at random) to match to the target, and if so will they do it less often than with a linking motion or color? Thus the three questions were whether (a) there would still be any reviewing benefit for letters presented in different locations in the absence of linking motion; (b) if so, whether this could be controlled by a shared property (color); (c) finally, whether total preview benefits would be approximately the same in matched conditions.

Stimuli and procedure. Experiment 6(a) was conducted with the color monitors used in Experiments 1–3. The letters in Experiment 6(a) were selected from the set C,K,L,P,Q,R,S,T,V. The sequence of events in the color conditions was the following: two white outline square frames subtending 1.1° a side appeared alone for 500 ms, above and below the center of the screen and separated vertically by 4.6°; characters were shown in the frames for 1 s; 87 ms after the characters disappeared from the frames,

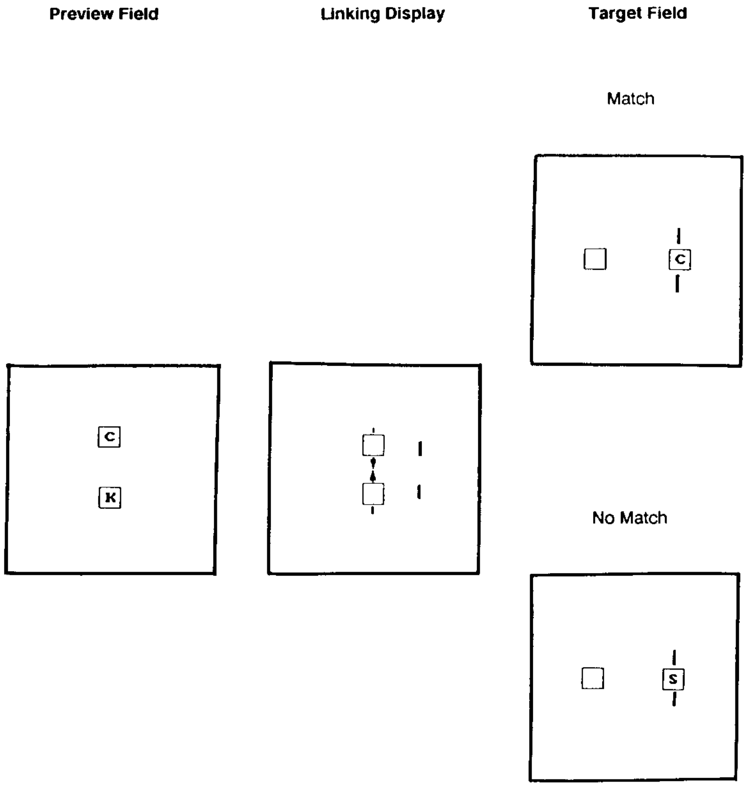

FIG. 5. Examples of displays used in Experiment 6(b).

two new frames appeared to the right and left of the center of the screen and separated by 4.6°; all four frames were present together for 87 ms, to inhibit apparent motion from the first to the second set of frames; the initial frames disappeared, and the new frames were shown alone for 87 ms, at which time a target letter appeared in one of them. No cue to target position was provided.

The sequence in the motion conditions was the same until the disappearance of the preview letters, at which time the frames started to move to their final positions. The motion was carried out in 15 steps and was completed in 250 ms. In the color conditions, one of the preview letters was greenish yellow and the other was purple; the target was shown in one of these colors. In the motion conditions all three letters shown on a trial were in the same color, and the color varied randomly over trials.

Experiment 6(b) was carried out with the same equipment as Experiments 4 and 5. The letters were selected from the set B,C,D,F,H,J,K,S. The frames and letters appeared in white on a dark background. The trial started with presentation of two outline frames, each subtending 2.1° a side, separated vertically (edge to edge) by 2.9°. After 500 ms, two letters appeared in the frames for 40 ms, then fixation bars appeared 3.1° to the left or right of the center of the display, and the empty squares began to move. There were two conditions. The linked motion condition was similar to the motion used in Study 4. In the unlinked motion case, the two frames moved toward the center of the display, until they were 0.8° apart (see Fig. 5). The movement lasted 400 ms. At that time, two frames appeared to the left and right of the center, separated by 4.2°. The frame cued by the fixation bars contained a letter.

Subjects. Two different groups of subjects were tested, 12 (7 women) in the color condition and 9 (5 women) in the nonspecific motion condition.

Results

The mean latencies and preview benefits are shown in Table 6. The standard linking-motion conditions gave the familiar pattern of significant object-specific benefit, and no significant priming in the DO condition.

The color condition of Experiment 6(a) showed significant preview benefit from both the matching and the mismatching colored letters, and no significant difference between the two. [Half the subjects showed more priming from the mismatched color, so the 5-ms difference did not approach significance, $t\ (11) = 1.13$]. The total benefits, 35 ms for motion and 29 ms for color, did not differ significantly ($t = 0.61$, ns).

In the unlinked motion condition of Experiment 6(b) there was no difference between the two objects in the degree to which they resembled or were linked to the target. The only measure of interest in that condition is the sum of preview effects over the two items, which is obtained by doubling the observed priming effect. The result, 29 ms, did not differ

TABLE 6
Mean Naming Latencies and Preview Effects in Studies 6(a) and 6(b)

<table>
<tr><th></th><th>Color</th><th>Motion</th><th>Unlinked motion</th><th>Linked motion</th></tr>
<tr><td>Same object</td><td>527</td><td>544</td><td>536</td><td>476</td></tr>
<tr><td>Different object</td><td>532</td><td>561</td><td>537</td><td>505</td></tr>
<tr><td>No match</td><td>544</td><td>570</td><td>551</td><td>509</td></tr>
<tr><td colspan="5">Preview effects</td></tr>
<tr><td>Specific</td><td>5</td><td>17**</td><td rowspan="2">15**</td><td>29*</td></tr>
<tr><td>Nonspecific</td><td>12*</td><td>9</td><td>4</td></tr>
<tr><td>Total benefit</td><td>29</td><td>35</td><td>29</td><td>37</td></tr>
</table>

** These values are significant at $p < .01$.
* These values are significant at $p < .05$.

significantly from the sum of preview effects for the linked-motion condition [37 ms; $t(8) = 0.70$].

The algebraic model of reviewing was introduced quite tentatively in the discussion of Experiment 5, both because it appeared rather too simple to be true and because it did not give a good account of the effects of preview display size in some of the experiments reported in Studies 1 and 2. The nonspecific preview effect was negative in one condition of Study 1, apparently violating an assumption of the model. The results of Study 2 were quite erratic, because of an unusually steep effect of display size on the control condition NM, which provides the baseline for the estimation of nonspecific benefits. On the other hand, the reviewing model seemed to fit the results of Studies 4 and 5 rather well. The experiments of Study 6 were conducted to obtain a more conclusive test of the model, and their results support it.

The novel assumption of the reviewing model, which is compatible with the notion of a correspondence process, is that the processing of the target letter is affected at most by one item from the preview field. The selection of the preview item that will be reviewed is dominated by a linking motion, but is affected neither by a match of colors nor by a match of the letters across the two fields. We were apparently successful in the attempt to vary the selectivity of reviewing without affecting the total preview benefit. The important conclusion of these experiments is that a shared object is not a necessary condition for the type of preview benefit that we have observed. The object provides a powerful guide to the process of selective retrieval that we have called reviewing but, when no object continuity is present, a new item may retrieve any earlier item, apparently at random.

STUDY 7: RECENCY IN REVIEWING

The evidence presented so far indicates that, when the objects in successive fields are strongly linked, attention to the target letter brings about a reviewing of the immediate history of the corresponding object in the preview field. This reviewing can go back in time for at least 700 ms when no intervening events occur. What happens if intervening objects are presented? Can reviewing access an earlier state of an object that has recently been updated? Our hypothesis suggests this should not be possible. In an attempt to find out, an experiment was conducted in which two sets of letters were successively shown in stationary frames, before these began to move. The same frames were filled in both preview displays, so that the first set of preview letters was overwritten by the second.

Method

Stimuli. The displays were the same as those in Experiment 4 except that (1) the movement consisted of 6 exposures instead of 10 and lasted only 100 ms; (2) two successive pairs of initial letters were presented before the framing shapes moved. The frames were first presented for 300 ms. They remained visible while two pairs of letters were presented for 200 ms each with an ISI of 200 ms between them. The remaining sequence of events was as in Study 4. The cue indicating target position appeared at the same time as the first pair of letters and remained visible until the end of each trial.

Procedure. There were six conditions: (1) SO1: matching letter in first display in the same object as the target; (2) SO2: matching letter in the second display, in the same object as the target; (3) DO1: matching letter in the first display, in a different object from the target; (4) DO2: matching letter in the second display, in a different object from target; (5) NM: no-match with all different letters; (6) different digits (from the set 23457) in both preview displays. These conditions were randomly mixed within blocks. Subjects were given one block of practice and then six experimental blocks of 48 trials each, giving 48 reaction times per subject per condition.

Subjects. A total of 10 students at the University of British Columbia (6 women) participated as paid subjects.

Results and Discussion

The mean naming latencies and preview effects are given in Table 7. The results are clear: of the four items in the preview displays (two each in two consecutive fields), only one yielded a preview benefit. As expected, this was the last letter previewed in the frame that later contained the target [$M = 19$ ms, $t(9) = 3.41$, $p < .01$]. The results for condition SO1, in which the target was previewed earlier in the same object, indicate no benefit whatever ($M = 0$ ms). Since the preview benefit was present and undiminished with stationary objects in Experiment 2 across a blank interval between preview and target of 700 ms, and with moving objects in Experiment 4(b) across an interval of 590 ms, the absence of any benefit from the matching letter in the first field, presented 500 ms

TABLE 7
Mean Naming Latencies and Preview Effects in Experiment 7

Target shown in:	Recent field		Early field
Same object	510		529
Different object	533		532
No match		529	
Digits		525	
	Preview effects		
Object specific	23**		3
Nonspecific	−4		−3

** This value is significant at $p < .01$.

before the target, is unlikely to be due simply to decay over time. The second letter in the same frame appears to wipe out the first, so far as the reviewing process is concerned.

All this makes sense if the purpose of the underlying representation—the object file—is to mediate perception and to control and integrate response tendencies. This function is served economically by maintaining a record of the current state of the environment. The most recent description of an object remains in force unless updated, and it is consulted when the state of the object changes, in order to maintain perceptual continuity. The pattern of results suggests that the information in the object file is overwritten only when new and incompatible information is entered. When the previous information simply disappears, the preview benefit survives apparently intact for at least 600–700 ms, and perhaps much longer.

It would be interesting to see whether the updating rule is strictly object specific rather than location specific. When one object is temporarily occluded by another object, the second should not overwrite the first, if the updating operates on object files rather than physical locations. This could be tested, for example, with displays like those that generate Michotte's tunnel effect (1963), or with displays in which a stationary object is temporarily hidden by an object that appears closer to the observer. Recent experiments suggest that cues to occlusion are available and can be used to modify early visual processing, such as that involved in determining the direction of motion in small apertures (Shimojo, Silverman, & Nakayama, 1988), or the perception of matching shapes (Sekuler & Palmer, 1990).

GENERAL DISCUSSION

This series of experiments has established a robust object-specific benefit: when a new stimulus appears, it will be named faster if it physically matches a previous stimulus seen as part of the same perceptual object. In different experiments we used shared location, relative position in a pattern seen in apparent motion, or a shared frame to link the target selectively to one of the previewed letters. A significant difference between the SO and the DO conditions was found in each case, although the preview displays in these conditions contained the same information and should have induced the same activation in semantic memory.

The most important results of these experiments were the following: (1) with only two objects in the field, the benefit of prior presentation of the target was almost entirely object specific, but selectivity was impaired when the number of items was increased; (2) the limit on performance was determined by the number of different tokens in the scene, not the number of types; (3) the specificity of the preview effect was unchanged when the preview was presented for one full second; (4) the object-specific benefit was not reduced by much slower linking movement, resulting in an ISI of 590 ms; (5) the benefit was wiped out by the subsequent presentation of another letter in the same object; (6) similarity of color did not induce a selective link between the target and a prime; (7) in the absence of links through object identity, the same preview benefits appeared to be distributed equally among the items in the preview display, compatible with an essentially arbitrary process of selection.

Additional experiments to be reported separately produced the following findings: (8) when words from a large vocabulary are used instead of letters, there is a large nonspecific priming effect, as well as a significant object-specific benefit; (9) there is also substantial nonspecific priming, even with letters, when the preview field contains just one letter rather than two or more; (10) differences of case between the preview and the target letter reduced the object-specific preview benefit in some conditions but not in others.

Two further variants of the reviewing design were concerned with the level of information that is integrated in object representations. (11) Feature information as well as letter identity may be collected and integrated in object-specific representations. In one experiment, the priming field consisted of four frames, each containing a single vertical or horizontal line. The frames moved empty to new positions. Three of the frames in the target display contained a line; the fourth contained either a line or a plus sign. Subjects responded to the presence or absence of the plus. They were slowed significantly in responding to the absence of a plus when the two lines shown in each of the frames would have made a plus if superimposed. This result suggests genuine integration of feature information. (12) At the other extreme, we failed to get object-specific benefits in a letter classification task where only the responses were shared between preview and target. This suggests the accumulation and integration of perceptual information rather than the accumulation of response tendencies; however, we do not propose that the possibility of accumulating response tendencies in object files has been ruled out.

We interpret these findings in terms of two theoretical notions: a selective reviewing process, and the object files that are reviewed. Our hypothesis is that the allocation of attention to the target item evokes an automatic process of reviewing, which selects one of the current object files, resulting in facilitation when the target and the retrieved item match, interference when they do not. When the target is assigned by a correspondence process to a particular object in the preview field, this link controls reviewing. In the absence of selective perceptual factors a random item will be reviewed. The information that is retrieved and integrated by the reviewing process may range from elements of shape (the lines that make a plus) to more abstract letter identities. We distinguish the effects of reviewing from effects of node priming. In our account, node priming played no role in the present experiments, presumably because the vocabulary of stimuli and responses was so small that node priming for all items was permanently at ceiling. When we used words we found both an object-specific advantage of about 25 ms for the SO over the DO condition and a large nonspecific preview benefit of about 50 ms in the comparison of DO and NM conditions.

We have stressed three characteristics of the reviewing process that produces the object-specific preview benefits observed in our experiments: it operates backward, it selects only a single item, and it is guided mainly by the features that control the unity and continuity of an object over time, but not by the shape, color, or content of the target. Any priming effect logically involves both a memory trace and a probe, but alternative analyses may focus on one or the other. The present treatment focuses on the function of the target as a probe, rather than on the lingering memory trace produced by the preview. This emphasis is justified by the observation that the SO and DO conditions in several of our experiments are strictly identical until the appearance of the target. Consequently, both the object-specific advantage and the absence of nonspecific priming must be explained by events that occur after that time.

We consider the almost total absence of nonspecific preview benefits in most of the experiments to be a serendipitous outcome of our initial choice of an experimental vocabulary. As already noted, considerable nonspecific priming was obtained in the same design when the stimuli were words drawn from a large vocabulary. However, the recurrent null result obtained in the present studies is a highly instructive result, which seems unlikely to have arisen from a combination of facilitation and interference effects that just happened to cancel out in the DO condition. Instead of the target being primed by several previewed items, and more by the one shown in the same object than by others, we are led to the hypothesis that only one of the previewed items is reviewed, a hypothesis that gained substantial support from the near constancy of total preview benefits observed in Experiments 6(a) and 6(b). Once the hypothesis has been formulated, however, there is no reason to restrict its applications to situations in which nonspecific priming can be eliminated. To account for cases in which nonspecific preview advantages are found, as well as object-specific ones (as in some of our studies with late cues, with no linking motion and with words), we simply give up the assumption that the selection is completely controlled by object relations. Even if reviewing never picked out more than one previewed item, nonspecific benefits would be found whenever the shape or content of the target have a part in controlling the operation of selective retrieval, and also, of course, when cues to object continuity are weak or absent and the selection is made at random. How far this notion of backward processing can be pushed to account for priming effects that are usually attributed to residual or to spreading activation is a matter to be settled by future research. The relation of reviewing to so-called backward priming effects is also a matter to be explored further (Glucksberg et al., 1986; Koriat, 1981).

Related Research: (1) Reviewing and Repetition Priming

Koriat and Norman (1988, 1989) have reported a set of studies that are relevant to our story, and they develop a concept that appears related to the concept of reviewing. They looked at repetition effects across successive trials with an intertrial interval of 500 ms in a classification task involving mental rotation. Their subjects classified letters or digits presented singly or in pairs in different orientations relative to the normal upright. Koriat and Norman found that when a stimulus exactly matched the previous stimulus except for its orientation, classification reaction times were faster than when the stimuli differed. The facilitation decreased with increasing mismatch in orientation, even when the present stimulus was itself upright (although the backward alignment benefit was greater for stimuli that deviated most from the upright orientation). They describe this backward alignment as an automatic "stimulus-induced, perceptual process that responds to the visual congruence between successive stimuli" (p. 491). The match appears to be holistic, since there is no facilitation for two identical digit pairs when the order within pairs is reversed (e.g., 13 and 31). Koriat and Norman suggest that the backward alignment is "designed to detect transformational invariance across successive visual events without necessarily establishing their identities. Like apparent motion, it can rely strictly on the visual correspondence between successive stimuli" (p. 491). We would add the prediction that the benefits would be object specific if the letters or digits were presented in separate frames or locations. In Koriat and Norman's experiments, all the stimuli appeared at fixation.

Kirsner et al. (1987) have offered an account of long-term repetition effects that emphasizes the role of the target as a memory probe. The main interest of these authors is to explain the specificity, both featural and linguistic, of repetition priming in verbal memory tasks. Their theory attempts to account for the evidence that specific conjunctions of surface features and abstract identities are stored and retrieved, whether they are relevant to the task or not. They review evidence that repetition priming in word identification is typically increased when the probe stimuli match the previewed items in type font, in letter case, in phonological form, in modality, and in language; explicit memory for these attributes varies inversely with their importance in mediating priming, suggesting that the priming depends on a failure to differentiate the memory trace from the current probe.

Kirsner et al. (1987) attribute priming to a matching process linking the current stimulus description and the record of an earlier stimulus description. The match speeds up the identification process. They use the terms "description" and "record" to refer respectively to an object file representing a present stimulus and to the memory trace of a past stimulus. The "descriptions" are highly articulated structural descriptions of the object, reached through sensory analysis abstracting information about its physical properties, parsing it into components to give structural elements, and where applicable "redescribing" the object in alternative media such as a phonological code for a visually presented word. "Like object files, they provide a functional location where perceptual information can be stored and organized during interpretation" (p. 149). Kirsner et al. (1987) see word recognition as involving access to "an extremely detailed record of one or more instances" (p. 161). Access is achieved not by a search process but by direct addressing through "the particular combination of codes (i.e., pattern of activity) that constitute the record. When the stimulus description re-creates that combination, the record has been 'discovered' " (p. 161).

There is of course a critical difference between the situations we investigated and those with which Kirsner et al. (1987) were concerned: the continuity of object identity is not a relevant factor in studies of long-term priming, whereas it seems to be the dominant factor that controls reviewing in the present experiments. However, the process of retrieving records that they describe appears to be quite similar to the process of retrieving object files that we have discussed here, except for the change in the basis of selective retrieval.

(2)Trans-Saccadic Integration

The possibility of spatiotopic fusion has been the focus of much important research on the trans-saccadic integration of information (see Irwin, 1991 and Pollatsek & Rayner, 1990 for reviews). The hypothesis that the detailed retinotopic images of successive fixations are brought into register, then combined before being fully processed, was once quite popular, but recent evidence has cast much doubt on it (Irwin, 1991). A series of studies have shown that the information that is integrated across saccades can be moderately abstract. In particular, Pollatsek, Rayner, and Collins (1984) showed that replacing a picture by its mirror image or changing the size of a picture by 10% across fixations did not reduce the benefit of parafoveal preview, and McConkie and Zola (1979) showed that the benefit of parafoveal preview of words was not reduced by arbitrary changes of the case of constituent letters. However, there is also compelling evidence for the maintenance of visual, spatial information across saccades (Irwin, Zacks, & Brown, 1990; Pollatsek et al., 1984). Irwin (1991) opted for an interpretation involving two separate mechanisms. He attributes the more abstract, conceptual facilitation to priming of already existing word and object representations in long-term memory, and the more visual sensory facilitation to short-term visual memory. We offer the alternative suggestion that the object file contains information at several levels. The evidence of object-specific benefits at these levels, which we will present in a separate paper, supports this position.

A recent article by Pollatsek, Rayner, and Henderson (1990) used a design very similar to ours, in conjunction with a study of eye movements. The sequence of events on each trial was as follows: fixation, followed by the presentation of two pictures of easily nameable objects, next to each other, 5° and 10° from fixation; subjects were instructed to move fixation to a mark between the pictures; the onset of a saccade triggered a change in the display, leaving one of the original pictures in view, next to a larger checkerboard pattern. The subjects were instructed to name the remaining picture, and their reaction times were recorded. In the no-switch condition (similar to our SO condition) the picture to be named was in the same spatial location it had occupied in the preview, but its retinal location was changed. In the switch condition (similar to DO) it occupied the spatial location previously filled by the other object. The results indicated a location-specific benefit of 10 ms, and considerable nonspecific priming. The responses in the switch condition were faster by 48 ms than responses on trials where the target matched neither of the two previewed objects.

In terms of the present analysis, which does not assign a special role to location as against object identity, the procedure used by Pollatsek et al. (1990) is somewhat ambiguous—mainly because it is unlikely that the checkerboard was seen as a new state of the object whose position it occupied. If the checkerboard was seen as a new object, it is quite possible that the target object in the switch condition was not perceived as a new token, but as the same object previously shown in the location now occluded by the checkerboard pattern. A similar perceptual ambiguity may affect other experiments in that series, which duplicated the sequence of retinal stimulation of the original studies, with a stationary eye: one field was first shown to the parafovea for 200 ms, followed by the second field at fixation. In this experiment, the SO condition was actually slower than the DO condition, by 19 ms. Once again, there was considerable nonspecific advantage: the SO condition was 31 ms faster than the NM condition. We have not replicated this experiment in detail, but have generated similar displays: The apparent motion they yield appears to us to be ambiguous, with the preview picture often moving into the target picture and the checkerboard appearing as a new object. There is no clear Ternus effect like the global motion one gets with pairs of letters.

The large amount of nonspecific priming observed in these studies could be due in part to this perceptual ambiguity. In addition, node priming was probably more important in that study than in our experiments with letters. Although the vocabulary of pictures was quite small (20 objects were used), the task of naming pictures is harder than letter naming and the reaction times were accordingly much slower than in our experiments. Finally, the correlation between the contents of the preview and target fields was lower than in our studies, which generally used a vocabulary of eight items. In both studies the probability of appearing as the target was ⅓ for each previewed item, but the probability was ⅙ for other items in our study, 1/18 in the study by Pollatsek et al. (1990), a difference that might be reflected in more active processing of the previewed items in their study. The discrepancy between the results of the Pollatsek et al. (1990) study and our own is instructive; follow-up work will be needed to identify the boundary conditions for the object-specific benefit we studied. If our conjectures about the relevant variables are correct, there would be no reason to amend the interpretation offered for our results.

(3) Object Tokens

In the introduction to this paper we contrasted two models of perception: a display-board model in which objects currently in view are represented by the activation of corresponding nodes in long-term semantic memory, and a model in which perception consists of the construction and utilization of the episodic representations that we have called object files. Similar distinctions have been drawn by others, particularly in computational approaches to visual modeling. One early example was the separation between different knowledge sources and the "blackboard" through which they communicate (Reddy, Erman, Farrell, & Neely, 1973). This idea was extended by McClelland (1986) within the connectionist PDP framework to allow more than one word to be identified at a time, using what he called "programmable blackboards." Another example is the idea of incremental representations used by Ullman (1984) in his discussion of visual routines. The incremental representations are temporary structures to which new information is added as it is extracted by visual operations such as boundary tracing, "coloring" or bounded activation, and indexing, at a level of processing that is intermediate between parallel preattentive feature registration and object identification through matching to stored object models.

There are also a few empirical studies that are relevant to the notion of object files. The closest to our paradigm is a set of experiments by Tipper, Brehaut, and Driver (1990) in which they explore the conditions under which inhibition or "negative priming" from a distractor in one display carries over to a target in the next display; in particular they test whether it depends on the two sharing the same location or on their being perceived as the same object. Between the preview and the probe displays, their stimuli appeared to pass behind an occluding surface and to emerge at different locations. The negative priming effect (slower responses to the target when it had previously been seen as a distractor) proved to be object centered rather than tied to either the retinal or the environmental location of the distractor in the preceding display. The authors suggest (p. 503) that "object files can survive occlusion and may be inhibited during selection of another object." Their finding complements ours: whereas we have demonstrated facilitation from a preview of the target, they show interference when a previously unattended (and therefore inhibited) object later becomes relevant. Both demonstrate that attentional effects are tied to objects rather than to locations.

Perhaps the most dramatic demonstration of the ability of the perceptual system to individuate and to track object tokens is described by Pylyshyn and Storm (1989). They presented subjects with either identical stimuli (small white pluses) moving in random directions at randomly varying speeds. Subjects were able to track up to four of these for several seconds with about 85% accuracy in the absence of any individuating properties other than the spatiotemporal continuity of the elements.[1] Pylyshyn (1989) explores the idea that we need the ability to index visual elements—features, locations, parts or objects—at some very early stage of processing before we can make any other information about their spatial relations explicit. He suggests that we have available about four visual indices or FINSTs (short for "fingers of instantiation") which we can attach to features or clusters of features and which can maintain access to these features as they move in space or as we define their relation to other features. These FINSTs are set up and maintained preattentively, and may in fact be used to determine where attention should be directed.

How do FINSTs relate to object files? We might think of them as the initial spatiotemporal label that is entered in the object file and that is used to address it. Our object files contain considerably more information; in fact all the information that defines and describes a particular perceived

[1] Pylyshyn and Storm's subjects may have coded the four moving pluses that they were tracking as the corners of a single global deforming shape. If so, they would, in our object file terms, have been entered into a file representing a single changing object, making their spatial relations explicit, and perhaps representing them as emergent features of the global shape such as its angles, its elongation, and its convexity or concavity. Yantis (1989) reported evidence consistent with this hypothesis.

object. So a FINST might be the initial phase of a simple object file before any features have been attached to it. The story is a little more complex, however, since several FINSTs are needed to define the structure of any object with parts whose spatial relations can vary; so there cannot be a one-to-one mapping between FINSTs and object file addresses. The problems that the two theories are designed to solve are not exactly matched, and, as a consequence, neither are the theoretical constructs. FINSTs are perhaps closer to Marr's concept of place tokens—abstract markers that allow the visual system to treat filled locations independently of the particular features or objects that occupy them. For example, place tokens allow certain spatial relations, such as colinearity, to be made explicit without reference to any other aspect of the elements between which they hold.

The most important perceptual phenomena which led us to postulate the existence of temporary, episodic representations of objects, separate from their descriptions in a long-term recognition network, are the following:

(*1*) *The feature-conjunction problem*. We are unlikely to have a node for the word "fox" written in green uppercase letters and another for the number "162" handwritten in red ink. It is easier to explain with object files than with a display board analogy how arbitrary sets of potentially interchangeable properties can be allocated to the correct objects in the perceptual representation (Treisman & Schmidt, 1982). This has been called the "binding problem," and has been discussed in the context of computational theories of vision that use distributed representations, for which the problem is particularly acute (Feldman & Ballard, 1982; Hinton, McClelland, & Rumelhart, 1986; Strong & Whitehead, 1989). Some way of representing temporary episodic associations of features is needed when several objects are present at once, and it is not immediately obvious how the activation of sets of long term memory nodes could mediate this information. As Hinton et al. (1986) point out, "In a conventional computer it is easy to solve the binding problem. We simply create two records in the computer memory. . . . In parallel networks it is much harder to solve the binding problem." The present results make it clear that, even with coarse coding of conjunctions, as suggested by Hinton et al. (1986), simple location-specific replication of shape or object detectors is insufficient to model human perception of dynamic displays. We suggest that separate records—object files in our terms—may still be necessary.

(*2*) *Constraints on visual attention*. In tasks requiring selective attention, performance is usually efficient when one object must be selected and another ignored as irrelevant. On the other hand, it is difficult to attend selectively to different properties of a single object, like the color and word in the standard Stroop task (Kahneman & Henik, 1981). Display board models would have difficulty explaining why it should be easier to focus on one of two nodes when the properties they code happen to characterize two different objects in the field than when they belong to a single object. Conversely, divided attention is improved when two relevant items are phenomenologically grouped to form a single "perceptual object" (Duncan, 1984; Treisman et al., 1983), as if attention operates on object files as units. Finally, interference from distractors in a letter classification task (Eriksen & Eriksen, 1974) is determined more by grouping through shared motion than by spatial proximity (Driver & Baylis, 1989). Letters that move together may be assigned to a shared object file, making selective attention to one of them more difficult.

(3) *The type-token distinction.* When the scene contains several replicas of the same object, we must form separate identical tokens to represent each instance. If perception depended only on the level of activation of particular nodes for the identities of the objects currently visible, some additional way would be needed to distinguish many small dogs, or many atypical dogs, from one large or typical dog. Interestingly, it appears that the coding of repeated instances may fail with brief presentations or at high rates of sequential presentation, leading to "repetition blindness" (Kanwisher, 1987; Mozer, 1989).

(4) *The perception of moving, changing objects.* Perhaps the most compelling source of evidence that perception is object centered is the observation of continuity and unity across motion and change that we illustrated at the beginning of this paper. Onlookers in the movie can exclaim "It's a bird; it's a plane; it's Superman!" without any change of referent for the pronoun. If the appropriate constraints of spatiotemporal continuity are observed, objects retain their perceptual integrity and unity. Since neither spatial location, sensory properties, nor even the most appropriate label need remain constant, we are forced to attribute any object-specific perceptual phenomena to some form of object-specific representation, addressed by its present location and by its continuous history of travel and change through space over time.

REFERENCES

Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance,* **15**, 448–456.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science,* **6**, 205–254.

Glucksberg, S., Kreuz, R. J., and Rho, S. (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition,* **12**, 323–335.

Green, M. (1986). What determines correspondence strength in apparent motion? *Vision Research,* **26**, 599–607.

Green, M. (1989). Color correspondence in apparent motion. *Perception and Psychophysics,* 45,15–20.

Hinton, G. E., McClelland, J. L., & Rumelhart, D. E. (1986). Distributed representations. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations.* Cambridge, MA: MIT Press/Bradford.

Irwin, D. (1991). Perceiving an integrated visual world. In D. Meyer & S. Kornblum (Eds.), *Attention and performance* (Vol. XIV), in press.

Irwin, D., Zacks, J. L., & Brown, J. S. (1990). Visual memory and the perception of a stable visual environment. *Perception and Psychophysics,* 47, 35–46.

Jonides, J. (1981). Voluntary versus automatic control over the mind's eye movement. In J. Long & A. Baddeley (Eds.), *Attention and performance* (Vol. IX). Hillsdale, NJ: Erlbaum.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization.* Hillsdale, NJ: Erlbaum.

Kahneman, D., & Miller, D. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review,* 93, 136–153.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman and D. A. Davies (Eds.), *Varieties of attention.* New York: Academic Press.

Kanwisher, N. G. (1987). Repetition blindness: Type recognition without token individuation. *Cognition,* 27, 117–143.

Kiger, J. I., and Glass, A. L. (1983). The facilitation of lexical decisions by a prime occurring after the target. *Memory and Cognition,* 11, 356–365.

Kirsner, K., Dunn, J., & Standen, P. (1987). Record-based word recognition. In M. Coltheart (Ed.), *Attention and performance: Vol. XII. The psychology of reading.* London: Erlbaum.

Kolers, P. A. (1972). *Aspects of motion perception.* Elmsford, NY: Pergamon.

Koriat, A. (1981). Semantic facilitation in lexical decision as a function of prime-target association. *Memory and Cognition,* 9, 587–598.

Koriat, A., & Norman, J. (1988). Frames and images: Sequential effects in mental rotation. *Journal of Experimental Psychology: Learning Memory and Cognition,* 14, 93–111.

Koriat, A., & Norman, J. (1989). Establishing global and local correspondence between successive stimuli. *Journal of Experimental Psychology: Learning Memory and Cognition,* 15, 480–494.

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance,* 3, 371–379.

Lambert, A. & Hockey, R. (1986). Selective attention and performance with a multidimensional visual display. *Journal of Experimental Psychology: Human Perception and Performance,* 12, 484–495.

McClelland, J. L. (1986). The programmable blackboard model of reading. In J. L. McClelland, D. E. Rumelhart, & the PDP Research Group (Eds.), *Parallel distributed processing: Vol. 2. Psychological and biological models.* Cambridge, MA: MIT Press.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General,* 114, 159–188.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics,* 25, 221–224.

Michotte, A. (1963). *The perception of causality.* (T. Miles and E. Miles, Trans.). London: Methuen. (Original work published 1946).

Mozer, M. C. (1989). Types and tokens in visual letter perception. *Journal of Experimental Psychology: Human Perception and Performance,* **15**, 287–303.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology,* **9**, 353–363.

Pollatsek, A., & Rayner, K. (1990). What is integrated across fixations? In T. Weymouth & L. T. Maloney (Eds.), *Exploratory vision: The active eye.* New York: Springer-Verlag.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General,* 426–442.

Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance,* **16**, 199–210.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial index model. *Cognition,* **32**, 65–97.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking of multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision,* **3**, 1–19.

Rayner, K. (1978). Foveal and parafoveal cues in reading. In J. Requin (Ed.), *Attention and performance* (Vol. VIII, pp. 149–161). Hillsdale, NJ: Erlbaum.

Rayner, K., McConkie, G. W., & Zola, D. (1980). Integrating information across eye movements. *Cognitive Psychology,* **12**, 206–226.

Rayner, K., & Pollatsek, A. (1983). Is visual information integrated across saccades? *Perception and Psychophysics,* **34**, 39–48.

Reddy, D. R., Erman, L. D., Farrell, R. D., & Neely, R. B. (1973). The Hearsay speech understanding system: An example of the recognition process. *Proceedings of the International Conference on Artificial Intelligence* (pp. 185–194).

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre- and postlexical loci of contextual effects on word recognition. *Memory & Cognition,* **12**, 315–328.

Sekuler, A. B., & Palmer, S. E. (1990). Visual completion of partly occluded objects: A microgenetic analysis. Manuscript in preparation.

Shepard, R. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review,* **91**, 417–447.

Shimojo, S., Silverman, G. H., & Nakayama, K. (1988). Occlusion and solutions to the aperture problem for motion. *Vision Research,* in press.

Sternberg, S., Knoll, R. K., & Turock, D. L. (1986). Direct access by spatial position in visual memory: 2. Visual location probes. *Bell Laboratories Technical Memorandum.*

Ternus, J. (1938). The problem of phenomenal identity. In W. D. Ellis (Ed.), *A source book of Gestalt psychology.* New York: Harcourt, Brace and Co.

Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception and Psychophysics,* **33**, 527–532.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology,* **14**, 107–141.

Ullman, S. (1979). *The interpretation of visual motion.* Cambridge, MA: MIT Press.

Ullman, S. (1984). Visual routines. *Cognition,* **18**, 97–159.

Yantis, S. (1989). *Dynamic multielement attentional tracking.* Talk given at 30th annual meeting of the Psychonomic Society, Atlanta, GA.

(Accepted December 10, 1990)