COGNITIVE PSYCHOLOGY 24, 295–340 (1992)

Multielement Visual Tracking: Attention and Perceptual Organization

STEVEN YANTIS

Johns Hopkins University

Two types of theories have been advanced to account for how attention is allocated in performing goal-directed visual tasks. According to location-based theories, visual attention is allocated to spatial locations in the image; according to object-based theories, attention is allocated to perceptual objects. Evidence for the latter view comes from experiments demonstrating the importance of perceptual grouping in selective-attention tasks. This article provides further evidence concerning the importance of perceptual organization in attending to objects. In seven experiments, observers tracked multiple randomly moving visual elements under a variety of conditions. Ten elements moved continuously about the display for several seconds; one to five of them were designated as targets before movement initiation. At the end of movement, one element was highlighted, and subjects indicated whether or not it was a target. The ease with which the elements in the target set could be perceptually grouped was systematically manipulated. In Experiments 1–3, factors that influenced the initial formation of a perceptual group were manipulated; this affected performance, but only early in practice. In Experiments 4–7, factors that influenced the maintenance of a perceptual group during motion were manipulated; this affected performance throughout practice. The results suggest that observers spontaneously grouped the target elements and directed attention toward this coherent but nonrigid virtual object. This supports object-based theories of attention and demonstrates that perceptual grouping, which is usually conceived of as a purely stimulus-driven process, can also be governed by goal-directed mechanisms.

Some of these experiments were first reported at the 30th annual meeting of the Psychonomic Society in Atlanta, 1989.

This research was supported by National Institute of Mental Health Grant R01-MH43924, Biomedical Research Support Grant RR07041, and NASA Joint Research Interchange NCA2-413 from the NASA-Ames Research Center to Steven Yantis.

Maz Fallah, Liz Graham, Alyson Holoubek, Nadine Qashu, Ira Silverman, and Vibha Thakral provided excellent technical assistance. I am grateful to Bill Bacon, Alfonso Caramazza, Chris Cherniak, Howard Egeth, Doug Johnson, Jim Johnston, Don King, Toby Mordkoff, Stephen Palmer, Hal Pashler, Zenon Pylyshyn, Brenda Rapp, David Rosenbaum, Saul Sternberg, and Anne Treisman, who provided valuable suggestions for the research or comments on an earlier version of the manuscript.

Correspondence should be addressed to Steven Yantis, Department of Psychology, Johns Hopkins University, Baltimore, MD 21218 (electronic mail may be sent to yantis@jhuvms).

Although the perception of objects in a complex scene is phenomenally immediate and effortless, detailed analysis of recognition times or detection accuracy suggest that object recognition mechanisms are more complex than they seem. This possibility is reinforced by the extreme difficulty of developing computer vision systems that are as rapid, accurate, and flexible as the human visual system is under natural conditions. Among other things, an attentional mechanism is required to guide object recognition and to optimize search (Tsotsos, 1988). In essence, attention serves as an intelligent "front end" to the visual object-recognition system; it delivers goal-relevant aspects of the image to high-level visual mechanisms, thereby making object recognition tractable.

OBJECT-BASED AND SPACE-BASED THEORIES OF ATTENTION

A major goal of research in visual object recognition, then, is to characterize the properties of the visual attention system. This work has led to conceptualizations of visual attention as something like a spotlight (e.g., Posner, 1980), a zoom lens (e.g., Eriksen & St. James, 1986), or a spatial gradient (e.g., Downing & Pinker, 1985; LaBerge & Brown, 1989). These studies have shown that when a response-relevant event occurs away from the focus of attention, the recognition or detection of the event deteriorates as the distance between the focus of attention and the event increases. A common premise in these conceptualizations is that attention is directed to a spatially defined region of the image. Duncan (1984) has referred to these as *space-based* theories of attention (further evidence for which has been reported by Hoffman & Nelson, 1981; Podgorny & Shepard, 1983; Posner, Snyder, & Davidson, 1980; Yantis & Johnston, 1990).

A fundamental postulate of space-based theories of attention states that there is at any given moment only a single convex spatial locus of focused attention, and that everything inside this locus is attended, while everything outside is unattended.[1] From this it follows that (a) there cannot be unattended elements within the locus of attention, and (b) there cannot be attended elements outside the locus. For example, if there are three elements in the visual field arranged horizontally, it is typically not possible to focus attention on the left and right elements without also attending to the center element (for applications of this assumption, see Cohen & Ivry, 1989, and Treisman & Schmidt, 1982).[2]

Hoffman and Nelson (1981) provided evidence for a spatial locus of attention with experiments in which subjects were to determine which of two letters was present in a four-letter display. Each display also contained a small box with one side missing, and subjects were required to determine the orientation of the box figure in addition to making the letter discrimination. Hoffman and Nelson found that the letter discrimination was more accurate when the box figure was adjacent to the letter than when it was not, providing evidence that attention was drawn to the location of the box, enhancing the discrimination of the spatially adjacent letter. Downing and Pinker (1985) found that luminance increment detection latency was fastest for targets that appeared in an expected spatial location, and increased monotonically as the target appeared in more and more distant spatial locations. A similar result involving the distribution of attention to a word or to a single letter was reported by LaBerge (1983).

Space-based theories may be contrasted with what Duncan (1984) called *object-based* theories. Object-based theories assume that attention is directed to one or more objects in the image, regardless of their relative spatial locations. Under most circumstances, of course, all the attributes of an object occur in the same convex spatial region. It is only with clever experimental manipulations that these two accounts can be distinguished. The experiments of Duncan (1984) provide an important instance of this. Subjects were required to report two attributes of one object or one attribute of each of two different objects; in both cases, the two objects were spatially superimposed over fixation. Duncan found that judgments were more accurate when they concerned attributes of a single object than when they concerned attributes of different objects. Neisser and Becklen (1975) and Rock and Gutman (1981) performed conceptually similar experiments.

Prinzmetal (1981) made a related point with a different paradigm. He noted that when a task requires the conjunction of two or more features in order to identify a target element, the presence of conjunction errors (i.e., incorrectly stating that two features that were actually present in two different elements occurred in a single element) signifies a failure of selective attention (Treisman & Schmidt, 1982). Prinzmetal then con-

[1] There are exceptions and qualifications that apply to this claim. For example, it has been proposed that attention can be directed to a concentric ring around fixation (Juola, Bouwhuis, Cooper, & Warner, 1991). Furthermore, attention could in principle be space-based and split into more than one region. I characterize spatial theories as assuming a unitary, spatially convex focus of attention only to distinguish them from technically spatial theories that involve topologically unitary but arbitrary shapes (e.g., a "squiggly ring" that focused attention on every other element of an array of elements arranged in a circle).

[2] In this paper, the term *element* refers to a feature or feature cluster in the visual scene without regard to how it is interpreted or organized perceptually. Elements are indexed by their spatial location in the image. The term *object* is reserved for an element or configuration of elements that has been perceptually interpreted as a coherent object and that is treated as such by the perceptual system. Objects are not necessarily indexed by their spatial locations.

ducted experiments that provided opportunities for conjunction errors and found that feature migrations were more likely to occur between elements that were in the same perceptual group (i.e., one of two collinear arrays of elements) than between elements in different perceptual groups, even though the relative spatial locations of the critical features were the same in both conditions. Thus in Prinzmetal's experiments, it was not the location in space that determined feature migration, but how the elements in a scene were perceptually organized into coherent configurations. Several other experiments have similarly revealed the importance of perceptual grouping processes and object coherence on attention (e.g., Banks & Prinzmetal, 1976; Driver & Baylis, 1989; Fox, 1978; Kahneman & Henik, 1981; Kahneman, Treisman, & Gibbs, 1992; Kanwisher, 1991; Kramer & Jacobson, 1991; McLeod, Driver, Dienes, & Crisp, 1991; Moraglia, 1989; Prinzmetal & Keysar, 1989; Tipper, Brehaut, & Driver, 1990; Tipper, Driver, & Weaver, 1991; Treisman, 1982; Treisman, Kahneman, & Burkell, 1983).

These studies demonstrate that visual selection is object-based under at least some circumstances. Of course, space-based theories and object-based theories are not mutually exclusive. At some level of representation, selection may be primarily space-based, and at another level of representation selection may be primarily object-based. Both types of theories are necessary for a full account of visual selection. However, theories of object-based selection—unlike space-based theories—must explicitly incorporate a mechanism for organizing elements in an image into coherent perceptual objects. One of the goals of this article is to show how the visual system's ability to organize visual elements into perceptual objects is a fundamental aspect of visual selection.

ATTENTION AND PERCEPTUAL ORGANIZATION

The role of perceptual organization in vision and audition was first emphasized by the Gestalt psychologists (e.g., Koffka, 1935/1963; Köhler, 1929/1947; Wertheimer, 1912). The theories advanced early in this century to account for the perceptual phenomena described by the Gestaltists were unsatisfactory for a variety of reasons, and with the growth of behaviorism between the two world wars, interest in perceptual organization waned (Hochberg, 1974, 1979). In the 1950s various efforts to reexamine the issues raised by the Gestaltists appeared, including the information-theoretic approaches of Attneave (1954) and Garner (1962), Johansson's (1950) studies of spatial configurations in displays of points in motion, and studies of object perception by Hochberg and his colleagues (e.g., Hochberg & McAlister, 1953). More recently there has been a resurgence of interest in perceptual organization in vision (e.g., Marr, 1982; Palmer, 1983; Pomerantz, 1981; Pomerantz & Kubovy, 1986) and specifically in selective attention (e.g., Driver & Baylis, 1989; Duncan & Humphreys, 1989; Kahneman & Henik, 1981; Kramer & Jacobson, 1991; Kubovy, 1981; Palmer, 1977; Pomerantz & Pristach, 1989). The question was framed nicely by Kahneman and Henik (1981, p. 183): "If attention selects a stimulus, what is the stimulus that it selects?"

Their answer (and the answer of object-based theories of attention) is that attention selects preattentively defined perceptual objects. Perceptual objects are formed and visual scenes are segmented and interpreted by low-level, stimulus-driven mechanisms of perceptual organization. To the extent that a perceptual object is attended, all of its attributes are also attended. The organizational principles most often invoked include proximity, similarity, common motion, and any of a number of geometric factors such as collinearity, parallelism, and symmetry.

Theories of perceptual organization inform object-based theories of attentional selection, because organizational mechanisms specify the perceptual objects that form the representational basis for selection. Conversely, because (according to object-based theories) attention necessarily selects coherent perceptual objects, grouping may be thought of as a natural byproduct of the process of selection. There is therefore a symbiotic relationship between theories of perceptual organization and object-based theories of attention.

The distinction between stimulus-driven or bottom-up processes on the one hand and goal-directed or top-down processes on the other plays a particularly important role in this symbiotic relationship. This distinction has long been a crucial part of theories of visual selective attention. Goal-directed selection includes directing attention to objects evidencing attributes that are relevant to a current perceptual goal (e.g., Bundesen, 1990); stimulus-driven attentional selection includes attentional capture by certain adaptively significant perceptual attributes independently of current goals or beliefs (e.g., abrupt visual onset; Yantis & Jonides, 1984).

In contrast, perceptual organization has almost exclusively been considered a purely stimulus-driven process (see General Discussion for examples). If object-based theories of attentional selection (which implicitly or explicitly endorse the central role of perceptual organization) are sound, then one might expect to observe a goal-directed or top-down component to perceptual organization in addition to the conventional bottom-up or stimulus-driven one.

The experiments reported below provide new evidence that attention can be directed to perceptual objects that are formed by grouping visual elements, thus supporting object-based theories of attention and verifying the vital role of perceptual organization in visual selection. In addition, the experiments demonstrate that perceptual organization need not be entirely stimulus-driven, but can be imposed on a display by virtue of the perceptual demands of a task.

VISUAL TRACKING AND ATTENTION

The point of departure for these studies is a task first described by Pylyshyn and Storm (1988). Subjects were required to track visually a small number of simple target elements moving quasi-randomly about a display screen among a similar number of identical nontarget elements for several seconds at a time. At various moments during movement, one of the elements would flash, and the subject was to indicate whether the flashed element (the *probe*) was a member of the target set or not. This is a highly demanding attentional task, requiring that subjects continuously keep track of the precise spatial coordinates of as many as five independently moving elements in noise. Subjects tracked the moving elements in these displays with better than 90% accuracy.

In attempting to account for this result, Pylyshyn and Storm (1988) tested several versions of the hypothesis that a single spatially localized focus of attention moved rapidly from one target element to the next during the 5–10 s tracking interval (e.g., Shulman, Remington, & McLean, 1979; Tsal, 1983). This hypothesis was derived from space-based theories which state that visual attention can be directed only to a single spatially convex region. The moving spotlight hypothesis was not supported. This finding was consistent with the predictions of Pylyshyn's (1989) theory of visual indexing (cf. Ullman, 1984b), according to which each of the target elements is tracked by a hypothetical visual index called a *FINST* (for INSTantiation FINger). According to the theory, the target elements were tracked independently and in parallel as they moved by virtue of their being indexed by one of a limited number of FINSTs. When the probe appeared, the subjects determined whether it was indexed or not, and responded accordingly.

Pylyshyn and Storm (1988) concluded that visual tracking is carried out in parallel because it is accomplished preattentively. According to that explanation, the fact that multiple noncontiguous elements could be tracked does not test space-based vs object-based theories of attention because attention is simply not required for tracking; FINSTs are bound to the target elements at the start and remain bound to the targets without attention.

However, another possible interpretation is that an element is tracked reliably only if it *is* attended. This would require that one abandon the assumption made by space-based theories that the locus of attention necessarily corresponds to a small convex spatial region (e.g., a "spotlight"), because Pylyshyn and Storm (1988) tested and rejected this interpretation of their results. Instead, this alternative account holds that attention is directed to the target elements by grouping them into a single higher order virtual object, to which attention is then directed. This is exactly what an object-based theory of attention would predict. However, this interpretation is viable only to the extent that performance can be shown to vary systematically with the degree to which perceptual grouping is possible.

According to the grouping hypothesis, observers initially construct a perceptual representation of a virtual polygon when the target elements are designated, and they continuously update this internal model by comparing it with the display throughout movement (for related ideas, see Cavanagh, 1990; Dawson, 1991; Kahneman et al., 1992; Kosslyn, Flynn, Amsterdam, & Wang, 1990; Lowe, 1987; Marr, 1982, pp. 202–205; and Ullman, 1989). The vertices of the virtual polygon are defined by the instantaneous positions of the elements being tracked. As the elements move, the size, shape, orientation, and position of the virtual polygon change. In general, these changes yield a polygon that is nonrigid, undulating, and periodically collapsing as the target elements move about the screen.[3] The existence of an internal model of the virtual polygon provides a way of determining directly whether the probe is or is not among the target elements.

This account involves two distinct processes. The first process is the initial formation of a perceptual group, which is likely to be governed largely by stimulus-controlled Gestalt laws of grouping like similarity, proximity, common fate, and Prägnanz. The second process is the maintenance of the virtual object during tracking, which is likely to be governed by subjects' ability to dynamically update an internal representation of the element configuration. Group formation can be characterized as relatively stimulus-driven, automatic, and preattentive; group maintenance can be characterized as goal-directed, effortful, and postattentive.[4]

The experiments reported in this article are designed to assess the

[3] A virtual polygon is said to *collapse* when a vertex of the polygon crosses an opposite edge so that the relative ordering of the points on the perimeter of the polygon changes. Shortly after collapse, a different polygon emerges, with an abruptly different shape than in the frames of the sequence immediately preceding collapse. The preservation of identity is "the core of the intuitive notion of a perceptual object" (Kahneman & Henik, 1981, p. 209); collapse therefore may result in the obliteration of the internal model of the virtual polygon.

[4] I use the terms "stimulus-driven" and "goal-directed" rather than "bottom-up" and "top-down," respectively, to emphasize an important functional distinction that is sometimes unclear. Stimulus-driven processes depend only on properties of the stimulus and their interaction with innate characteristics of the perceptual system; for example, attentional capture by abrupt onset appears to be a purely stimulus-driven process (e.g., Yantis & Jonides, 1984; Jonides & Yantis, 1988). Goal-directed processes depend on stimulus factors and in addition on the observer's current goals and knowledge; focusing attention at a location on the basis of a spatial cue is a goal-directed process (e.g., Posner et al., 1980).

hypothesis that subjects group the target elements into a nonrigid virtual polygon and track it during motion. The principal features of the tracking task, which was a variant of Pylyshyn and Storm's (1988) task, were as follows. Ten stationary pluses appeared on a display screen in random locations (or, in some experiments, in canonical locations). A subset of one to five elements flashed on and off several times; this designated the *target set*. All 10 elements then began to move about the screen in multiple directions more or less independently; movement continued for 4–8 s and then stopped. One of the 10 elements was then *probed* by replacing it with a pair of salient concentric squares. The observer's task was to determine whether the probed element was a member of the target set or not. Accuracy, and not speed, was the primary dependent variable.

The empirical strategy was to manipulate factors that influence the extent to which perceptual grouping processes could (a) generate and (b) maintain an internal representation of the virtual polygon made up by the target elements. The generation process was manipulated in Experiments 1–3 by varying the initial configuration of the target elements (canonical vs random), the presentation mode of the target elements (simultaneous vs successive), and the instructions given to subjects (attempt to group vs no grouping instructions). The maintenance of a perceptual group during tracking was manipulated in Experiments 4–7 by placing dynamic constraints on the configuration of the target elements during movement (rigid in 3-space vs nonrigid; nonrigidly convex vs unconstrained), and by varying the degree to which the velocities of the target and nontarget elements were correlated within and between groups.

Two alternatives to the perceptual grouping account will be assessed. The first alternative, provided by space-based theories, asserts that a narrow spotlight of attention is moved from one target location to the next during movement to update the stored coordinates of the target elements. The second alternative, provided by Pylyshyn's (1989) FINST theory of spatial indexing, asserts that the elements are tracked in parallel, but preattentively and independently. These alternative accounts make different predictions about performance in the tracking task as a function of grouping manipulations. By definition, perceptual organization involves the relationships among elements in a display; thus the grouping hypothesis predicts that these manipulations will significantly influence performance. Both space-based theories of attention and the FINST theory assume that such relationships are not pertinent, so they predict that grouping manipulations will not influence performance.

EXPERIMENT 1

Experiments 1–3 manipulate the process of group formation but not group maintenance. In each case, a factor is manipulated that will influence the extent to which the target elements are grouped initially. In Experiment 1, the starting positions of the 3, 4, or 5 target elements were selected either at random or as the vertices of a canonical polygon (i.e., a regular triangle, diamond, or pentagon); the starting positions of the nontargets were always selected at random. Once the elements began to move, there were no differential constraints on the movement trajectories of the elements in the canonical and the random conditions. The hypothesis that the target elements are perceptually grouped into a virtual polygon predicts that canonical starting positions may assist in the grouping process, perhaps by emphasizing that the configuration of the target elements is important. A model that is not sensitive to the configuration of the elements would predict no difference between the random and canonical conditions.

Method

Subjects. Eighteen Johns Hopkins University undergraduates participated in one 50-min session to fulfill an introductory psychology course requirement. All subjects had normal or corrected-to-normal vision.

Stimuli and apparatus. The animated stimulus sequences were displayed on a Hewlett–Packard HP1345A graphics display module controlled by an IBM AT microcomputer. The display device was placed in one field of a two-channel tachistoscope; the other channel contained a continuously illuminated blank white card. By varying the intensity of the light in the blank field, the contrast of the events on the display could be controlled. The luminance of the blank field was 1.4 cd/m^2, and the luminance at the screen of a single display element (plus sign) was 9.4 cd/m^2.

Subjects positioned their heads in a chinrest which controlled viewing distance (58 cm). From this distance, the display screen subtended visual angles of 8.3° in height and 11.2° in width; this was the size of the region within which movement could occur. The display screen itself was not directly visible to subjects; instead, they saw a uniform grey field (i.e., the blank field of the tachistoscope) subtending 11.1° by 14.8° of visual angle upon which the visual events of the experiment appeared to occur. Subjects responded by pressing one of two buttons mounted on a sloped response box placed on the table in front of them.

Each trial consisted of 250 static frames presented one after another for 30 ms each. This yielded an animation sequence 7.5 s in duration. Each frame consisted of 10 small plus signs each of which subtended 0.22° of visual angle vertically and horizontally, and a central stationary fixation square, which subtended 0.11° vertically and horizontally. Each element was surrounded by an imaginary square *cushion*, 0.5° in height and width, into which no other element could enter. The top panel of Fig. 1 illustrates a sample stimulus display with the cushion around one element drawn in dashed lines for purposes of illustration (the cushions were never visible to subjects). The fixation square also had a cushion.

The initial configuration of the target elements was either random or canonical. A sample random configuration appears in the top panel of Fig. 1. The canonical configurations for target-set sizes of 3, 4, and 5, respectively, were an upward-pointing isosceles triangle, a regular diamond, or a regular upward-pointing pentagon (the pentagon is shown in the bottom panel of Fig. 1, with each target element enclosed in a square for illustrative purposes). In each of these cases, the target elements were positioned 2.3° from fixation. The nontarget elements were always placed in randomly-selected locations. There were no constraints on the initial locations of randomly placed elements except that they could not be placed within or "touching" the cushion of any other element.

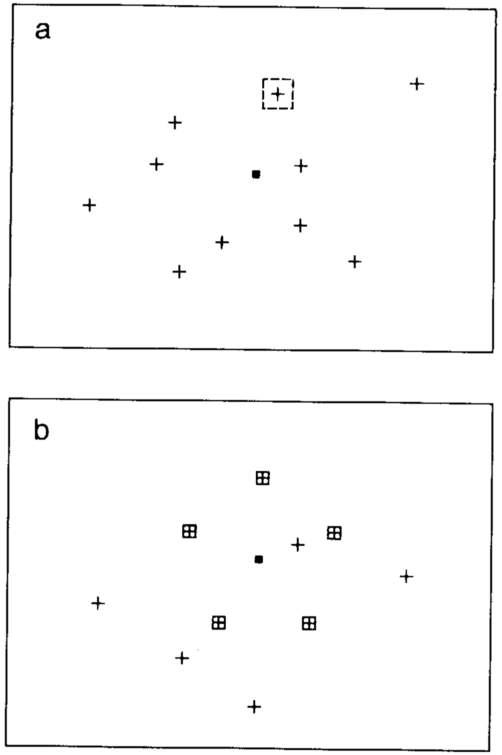

FIG. 1. Two examples of one frame of the animation sequence. In each panel, the central black square represents the fixation point. (a) Random initial configuration. The dashed square surrounding one of the stimulus elements represents that element's cushion; the cushions around each element were never visible to subjects. The large rectangle surrounding all the stimulus elements is approximately the same size as the drawing area of the graphics display used. (b) Canonical initial configuration. The squares surrounding the five targets in this display alternated with the pluses during the target designation phase; the squares and pluses were never visible simultaneously. Scale is approximate.

The position of the elements in each frame of the sequence was computed so as to produce smooth apparent motion throughout each trial. All stimulus sequences were generated in advance of the experimental session and stored on the computer's disk. One sequence was retrieved from disk prior to the onset of each trial, and loaded into display memory. The computer then issued commands to control the rate at which the individual frames were presented. The HP1345A is a vector- rather than a raster-graphics device, providing 1-ms resolution in the timing of frames. Each stimulus sequence was unique.

Procedure. Each trial consisted of three phases: the *target designation phase*, the *movement phase*, and the *probe phase*. In the target designation phase, 10 elements (plus signs) were placed on the screen in canonical or random positions (depending on the experimental condition and whether the element in question was a target or a nontarget), with the constraint that no element could touch the edge of the screen or the cushion of any other element or the cushion of the fixation square. Next, each element was assigned an initial

trajectory direction at random. There were eight possible directions: vertical (up and down), horizontal (left and right), and 45° oblique (up/right, down/right, up/left, or down/left). Each element was also assigned an initial *trajectory duration*, randomly selected from a uniform distribution ranging from 210 to 810 ms in 30-ms increments. Each element was then (invisibly) tagged according to whether it was a target or a nontarget and one of the ten elements (a target on half the trials and a nontarget on the rest) was (invisibly) tagged as the probe for that trial.

To designate the target set, the elements that had been tagged as targets were highlighted by replacing each of them with a square having the same dimensions as the pluses. The target pluses and squares alternated five times for 150 ms per alteration, providing a highly salient specification of the target set. The nontarget elements remained static on the screen during the target-designation sequence.

At the end of the target-designation phase, the movement phase began. Movement was simulated by selecting a new position for each element on the screen in each of the 250 animation frames. There were several constraints that guided the selection of a new position for each element. First, each element was placed 0.14° from its position in the previous frame (and because each frame had a duration of 30 ms, velocity was 4.67°/s). The direction of movement was specified by the element's current trajectory direction. A given element would continue to move in its current direction until one of the following events occurred: (a) the newly selected position placed the element within the cushion of another element or the fixation square; (b) the newly selected position placed the element directly adjacent to or off the edge of the screen; or (c) the element's trajectory duration expired. At the occurrence of any of these events, a new direction and duration were selected at random with the constraint that they could not be the same as the current ones. If the new direction did not resolve a collision conflict, it was discarded and another direction was selected at random from those remaining. Unacceptable trajectory directions were rejected and a new direction selected until the conflict was resolved.[5]

This procedure yielded a sequence of frames in which each element moved in a smooth and continuous linear trajectory for some period of time (210–810 ms or until a "collision"), and then changed direction abruptly and began to move in a new direction. The elements moved independently of one another except in cases of collision (a fairly common occurrence). As stated earlier, the movement phase lasted 7.5 s.

At the end of the movement phase, the probe phase occurred. In this phase, all the elements stopped moving and the probe element (selected earlier) was highlighted by replacing it with a highly salient pair of concentric squares, one the same dimensions as the pluses, the other twice as large.

Subjects were to press the right key if the probe was a target and the left key if it was not. They were to guess if they were not sure. The static probe display (consisting of the fixation square, the probe, and the nine unprobed elements) remained on the screen until the observer made a response or 4 s elapsed, whichever occurred first. If the response was correct, the word "correct" was displayed in the center of the screen for 500 ms; if it was incorrect, the word "error" was displayed; if no response was made, the words "too long" were displayed. After 1 s the next trial began.

Accuracy, and not speed, was emphasized. Subjects reported having no difficulty responding within the 4-s response interval. Subjects were given no special instructions regarding eye position. It is likely that most subjects moved their eyes during the task, although none reported using a special strategy that depended on eye position.

[5] In a few instances, the conflict could not be resolved. In these cases, the sequence was discarded and a new sequence was generated from a different randomly selected starting point. This occurred less than once per 500 trials generated.

Design. Three factors were manipulated in this experiment: the identity of the probe (target or nontarget); the number of target elements to be tracked or target-set size (3, 4, or 5), and the initial configuration of the target elements (canonical or random). Target configuration was alternated between blocks, while probe identity and target-set size were completely crossed and varied within blocks. Subjects completed 6 blocks of 36 trials each, for a total of 216 trials.

Results

Failures to respond within the 4-s response interval were very rare: they occurred only 4 times in 3,888 trials. Trials on which no response occurred were treated as errors.

Subjects responded correctly on about 65% of the trials in this experiment, well above chance (chance responding was 50%). Figure 2 shows accuracy as a function of target-set size, initial configuration, and practice. Within each panel, the results from one-third of the trials is shown: the left, middle, and right panels, respectively, depict the results from blocks 1 and 2, blocks 3 and 4, and blocks 5 and 6. A three-way repeated-measures analysis of variance (ANOVA) was carried out with target-set size (3, 4, and 5), initial configuration (canonical and random), and practice (blocks 1 and 2, blocks 3 and 4, and blocks 5 and 6) as factors. The main effect of target-set size was significant, $F(2,34) = 20.4$, $p < .001$, as was the main effect of practice, $F(2,34) = 3.94$, $p < .05$. The main effect of initial configuration, however, was not significant, $F < 1$. Neither the interaction between target-set size and initial configuration nor that between target-set size and practice was significant, both Fs < 1. However, the interaction between initial configuration and practice was significant, $F(2,34) = 5.15$, $p < .05$.

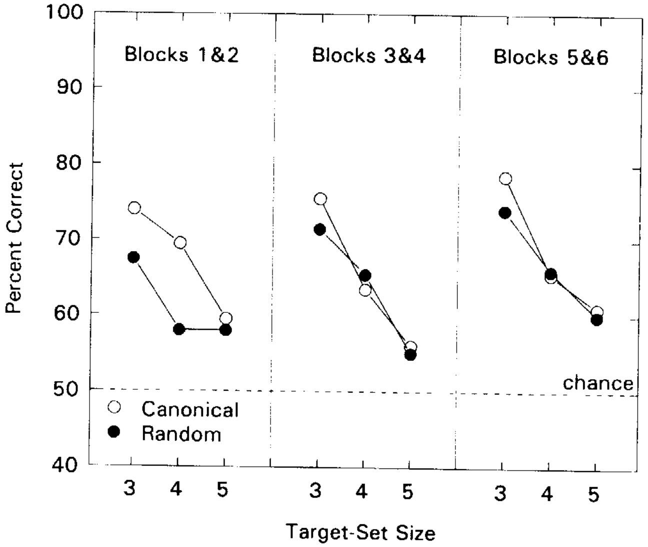

FIG. 2. Results from Experiment 1. In each panel, accuracy is plotted as a function of target-set size for the canonical and random conditions, respectively. The left, middle, and right panels show data from blocks 1 & 2, blocks 3 & 4, and blocks 5 & 6, respectively. Chance responding is 50%.

A contrast analysis of this interaction (Rosenthal & Rosnow, 1985) revealed a significant linear decline in the initial-configuration difference with practice ($F(1,34) = 4.48$, $p < .05$); the quadratic component was not significant ($F < 1$). Early in practice, performance was enhanced by canonical initial configurations, but later the advantage disappeared. This pattern suggests that the grouping strategy emerged at least partly as a result of exposure to the canonical configurations, but once the strategy was discovered, it could be applied in both conditions more or less equally (recall that there was no difference in the trajectories of the canonical and random conditions, and therefore no differential stimulus support for grouping in the two conditions).

Discussion

The results of Experiment 1 are consistent with those of Pylyshyn and Storm (1988) in showing that when observers are required to track as many as 5 of 10 independently moving objects for up to 7.5 s, they are fairly successful. Performance deteriorated as the number of elements to be tracked increased from 3 to 5; tracking three elements was viewed by most subjects as relatively easy, although not effortless, while tracking 5 of 10 elements was universally judged to be difficult if not impossible by some subjects. Many subjects reported adopting a strategy in which only a subset of the target elements was tracked, with the number in the tracked subset growing as they became more practiced at the task.

The analysis of practice effects suggests that although there was no overall advantage of canonical starting position, an advantage was present early in practice. The conclusion is that canonical initial configurations may have suggested a grouping strategy to at least some subjects which served to enhance performance early on, but later in practice most subjects had discovered the grouping strategy and could apply it in both the canonical and the random conditions.

Two alternative accounts of performance in this task must be considered: a sophisticated-guessing strategy and a serial "spotlight of attention" model. Each of these alternatives is analyzed in the General Discussion.

The absolute performance levels observed here are not directly comparable with those obtained by Pylyshyn and Storm (1988) because of differences in the procedures employed. For example, the display screen used by Pylyshyn and Storm had more than four times the area of the one used here, yielding an average element density of less than one-quarter that of the present experiments. The velocities of the elements in Pylyshyn and Storm's experiments changed from moment to moment and from element to element within a trial (ranging from 1.25°/sec to 9.4°/sec); in the present experiments, element velocity was fixed at 4.67°/sec. The most salient difference between the two procedures concerned the probing method. In the Pylyshyn and Storm task, four probes occurred sequentially on each trial; the subject's task was to press a button as quickly as possible immediately after a probe occurred at a position occupied by a target element; thus chance responding was 25%. In the present experiments, one probe occurred after movement stopped, and it was equally likely to occur over a target and a nontarget.

These differences notwithstanding, the results of Experiment 1 corroborate the principle finding of Pylyshyn and Storm (1988): Subjects can successfully track multiple independently moving elements.

EXPERIMENT 2

In Experiment 2, the presentation mode of the target set was again manipulated to influence the ease with which a perceptual group could initially be formed. Targets appeared either simultaneously (as in Experiment 1), or sequentially; the initial spatial configuration of the target elements was always random. The grouping hypothesis predicts that simultaneous target presentation should enhance performance by making it easier to generate a perceptual representation or internal model of the target configurations initially. For example, emergent features of the virtual polygon might only be perceptible if the target elements are simultaneously visible (Pomerantz & Pristach, 1989). Furthermore, sequential presentation could harm grouping in that the individual flashing elements in the sequential condition might capture attention (e.g., Yantis & Jonides, 1984), diverting the subjects' attention from the target configuration as a whole.

Of course, simultaneous presentation might be expected to enhance performance as compared to sequential presentation for another reason as well: sequential presentation necessarily requires that some target elements appear at some interval before trial onset, possibly resulting in diminished memory for some target positions. In order to eliminate the possible memory advantage for simultaneous presentation, simultaneous targets were presented briefly and then followed by a static interval equal to the amount of time elapsing between the presentation of the first *sequential* target and the initiation of movement. In other words, if there were three target elements to be presented, under sequential presentation element 1 would flash for 150 ms, element 2 for 150 ms, element 3 for 150 ms, and then movement would begin; for simultaneous presentations, however, all three target elements would flash for 150 ms, followed by 2 × 150 = 300 ms of the static display.

Method

Subjects. Eighteen undergraduates from the Johns Hopkins University introductory psychology subject pool participated in one 50-min session. All subjects had normal or corrected-to-normal vision. None of the subjects had participated in Experiment 1.

Design and procedure. The design and procedure were the same as in Experiment 1, with the following exceptions. A new between-block manipulation was introduced: in the *simultaneous* condition, the targets were designated by being flashed on and off simultaneously for 150 ms, followed by $150(n - 1)$ ms of the static display, where n is the number of targets in the target set. In the *sequential* condition, each target element was successively flashed alone for 150 ms. Both conditions yielded a target designation phase that lasted $150n$ ms. The initial configuration of the target elements was always random. The two presentation conditions were alternated between blocks; target-set size (3, 4, or 5) was manipulated within blocks.

Subjects completed 6 blocks of 30 trials each.

Results and Discussion

Subjects responded correctly on about 70% of the trials in Experiment 2, well above chance. Figure 3 shows accuracy as a function of target set size for the simultaneous and sequential conditions, respectively. Each panel of the figure depicts the results from one-third of the trials as a function of practice: The left, middle, and right panels show performance for blocks 1 and 2, blocks 3 and 4, and blocks 5 and 6, respectively. An ANOVA revealed a significant main effect of both target-set size and presentation condition ($F(2,34) = 22.4$ and $F(1,17) = 9.2$, respectively, both $p < .001$); the interaction was not reliable ($F < 1$). As before, the main effect of practice was also significant, $F(2,34) = 4.4$, $p < .05$. The interaction between practice and target-set size was not reliable, $F < 1$, but the interaction between practice and presentation condition was significant, $F(2,34) = 3.9$, $p < .05$.

A contrast analysis (Rosenthal & Rosnow, 1985) again revealed a significant linear decrease in the effect of presentation mode with practice ($F(1,34) = 5.16, p < .05$); the quadratic component was not significant ($F < 1$). As in Experiment 1, performance was enhanced by simultaneous initial configurations, but only early in practice. This result leads to the conclusion that the simultaneous initial configuration suggested a grouping strategy early in practice that eventually was discovered and used by all subjects. Because the trajectories of the elements in the two conditions did not differ systematically, however, the extent to which subjects could

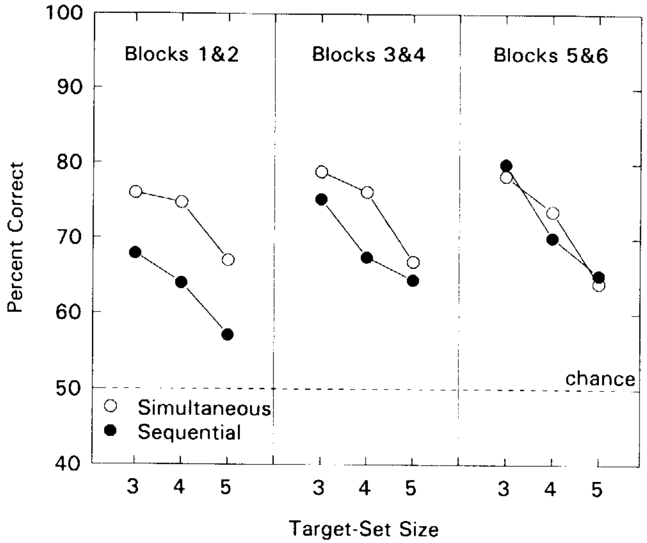

FIG. 3. Results from Experiment 2. In each panel, accuracy is plotted as a function of target-set size for the simultaneous and sequential conditions, respectively. The left, middle, and right panels show data from blocks 1 & 2, blocks 3 & 4, and blocks 5 & 6, respectively. Chance responding is 50%.

employ the strategy to maintain a perceptual group—once it was discovered—was equivalent in the two conditions.

EXPERIMENT 3

The conclusion I have drawn from Experiments 1 and 2 is that subjects discover an efficient grouping strategy after a certain amount of experience with the tracking task, and that this discovery occurred earlier in the canonical condition of Experiment 1 and in the simultaneous condition of Experiment 2 than in the other conditions of those experiments. This conclusion leads to the hypothesis that a more direct manipulation of strategy should yield a similar pattern of results. In Experiment 3, half the subjects were explicitly told about the grouping strategy and were instructed to use it if possible while tracking. The remaining subjects were given the same neutral instructions that subjects received in Experiments 1 and 2. Targets were positioned randomly and designated simultaneously. The hypothesis predicts that subjects given grouping instructions will perform more accurately than subjects given neutral instructions, but only early in practice. As subjects in the neutral-instruction group become experienced with the task, they are expected to discover the grouping strategy spontaneously and use it to enhance their tracking performance, perhaps achieving the level of subjects who receive grouping instructions.

Method

Subjects. Eighteen undergraduates from the Johns Hopkins University subject pool participated in one 50-min session. All subjects had normal or corrected-to-normal vision. None of the subjects had participated in Experiments 1 or 2.

Design and procedure. The design and procedure were the same as in Experiment 1, with the following exceptions. The initial configuration of the target elements was always random, and the targets were always designated simultaneously. All subjects were shown four demonstration trials during the instructions. Half the subjects were given neutral instructions, and half were given grouping instructions. The two sets of instructions were identical up to the point where they were told that they were about to see the four demonstration trials. The neutral-instruction subjects simply looked at the demonstration trials while the experimenter was silent. The grouping-instruction subjects were read the following paragraph during the demonstration trials:

> We have found that people can do better in this task if they think of the target elements as forming the vertices of a changing shape, like a triangle or a rectangle. Imagine that there is an invisible rubber band around the target elements, so that as the targets move, the shape formed by the rubber band changes. If you keep this shape in mind during each trial, it may help you track the targets more accurately.

In all other respects the two set of instructions were identical. Each subject completed 6 blocks of 30 trials each.

Results and Discussion

As in Experiments 1 and 2, subjects responded correctly on about 70% of the trials in this experiment, well above chance. Figure 4 shows percent correct as a function of target set size for the neutral and grouping conditions, respectively. Each panel of the figure depicts performance for one-third of the trials as a function of practice. An ANOVA revealed a significant main effect of target-set size, $F(2,32) = 27.5, p < .001$; neither the main effect of instructions nor the interaction between display size and instructions was significant, both $F < 1$. As in Experiments 1 and 2, there was a significant main effect of practice, $F(2,32) = 21.5, p < .001$. The interaction between practice and target-set size was not significant, $F < 1$. However, the interaction between practice and instructions was significant, $F(2,32) = 4.1, p < .05$. This last interaction can be seen in Fig. 4 as a change in the difference between the grouping and neutral functions in the three panels. Performance was enhanced by grouping instructions only early in practice.

The results of Experiments 1–3 provide converging support for the conclusion that subjects discovered a grouping strategy which helped them perform the task more accurately. In each experiment, the discovery of the grouping strategy occurred more rapidly when the targets appeared in canonical locations, were designated simultaneously, or when

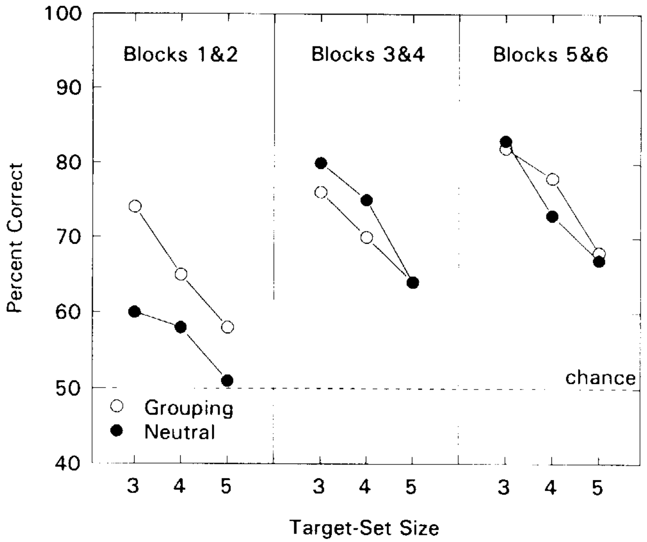

FIG. 4. Results from Experiment 3. In each panel, accuracy is plotted as a function of target-set size for the grouping-instruction and neutral-instruction conditions, respectively. The left, middle, and right panels show data from blocks 1 & 2, blocks 3 & 4, and blocks 5 & 6, respectively. Chance responding is 50%.

grouping instructions were provided. In each case, the advantage dissipated with practice, presumably because subjects in all groups eventually discovered the grouping strategy and used it with the same effectiveness.

Experiments 1–3 involved manipulations that affected the initial formation of a perceptual group, primarily by influencing the salience of a grouping strategy. The effects of these manipulations were temporary and diminished with practice. In the remaining experiments, I manipulate factors concerning the maintenance of a perceptual group during tracking. In each experiment, the trajectory statistics of the elements to be tracked are arranged so as to make dynamic grouping more or less difficult during tracking. Accounts of visual tracking that do not involve grouping predict no effect of these factors.

EXPERIMENT 4

With the displays used in Experiments 1–3, it was quite common for a vertex of the virtual polygon to cross over an opposite edge of the polygon, resulting in object collapse one or more times during a single trial (see Footnote 3). It is at these moments that the virtual polygon loses coherence and one or more of its elements may be lost, because its very identity as an object changes abruptly at these critical moments (Kahneman & Henik, 1981). To the extent that the simplicity of the virtual

polygon contributes to subjects' ability to track effectively, violations of coherence would be expected to result in tracking failures.

In Experiment 4, the configuration of the target set was constrained so as to control whether object collapse was possible or not. In the *unconstrained* condition, the spatial configuration of the target elements was essentially random, as in Experiment 1 (with the exceptions described in the Method section of Experiment 1); this yielded frequent object collapse. In the *constrained* condition, the target elements were required to remain in a nonrigid convex polygon. The convexity constraint yields a virtual object in which the ordinal relations among the vertices of the configuration remain constant throughout movement.[6]

Configurations that are constrained to remain convex will clearly be easier to group than ones that are permitted to collapse; the grouping hypothesis therefore predicts better performance in the constrained condition than in the unconstrained condition.[7] In contrast, any mechanism that tracks the target elements independently predicts no difference between the constrained and unconstrained conditions, because under such a scheme the spatial relations among the target elements are irrelevant.

Method

Subjects. Fifteen subjects participated in Experiment 4. These subjects came from the introductory psychology subject pool, participated in one 50-min session, and had normal or corrected-to-normal vision. None had participated in Experiments 1–3.

Design. The design was the same as in Experiment 1, except (a) the target-set size was 1, 2, 3, 4, or 5, and (b) the primary manipulation now was whether the target configuration was constrained to remain convex or not during movement. This factor was manipulated within blocks. Of course, the convexity constraint only applies to target set sizes 3, 4, and 5. Subjects participated in 5 blocks of 48 trials each, for a total of 240 trials. Within each block, there were 24 unconstrained trials and 24 constrained trials, randomly ordered. Within each of these conditions, there were 3 trials each of target-set size 1 and 2, and 6 trials each of target-set sizes 3, 4, and 5.

Procedure. The procedure was the same as in Experiment 1, with the following exceptions. First, the duration of movement in each trial was reduced from 7.5 to 4.5 s (a total of 150 animation frames were displayed for 30 ms each). This was done so that the increased number of trials could be completed within a session that was not so long as to fatigue subjects. Second, the target sets with three or more elements always began in a canonical

[6] Here, "the configuration" is defined as the convex hull of the target elements. The convex hull is the natural boundary of a set of points. The convex hull of a set of coplanar points is the smallest convex polygon that contains all the points (e.g., Sedgewick, 1983, Chap. 25). For purposes of the present experiments, the elements are positioned so that they always start out forming a convexing polygon, and they are constrained to remain in a convex configuration throughout movement.

[7] In fact, convexity is a stronger constraint than is required for this manipulation: concave but noncollapsing configurations should also yield performance that is superior to collapsing configurations. However, it is possible that the added complexity of concave configurations would limit performance somewhat (see General Discussion).

configuration (triangle, diamond, or pentagon), ensuring that the configuration was always convex to begin with. Third, in the constrained condition, the configuration was monitored throughout movement so that whenever the convexity constraint was about to be violated (this occurred when a given element was about to cross over an imaginary line drawn between the two target elements adjacent to the element in question on the convex hull of the target set), then a new trajectory direction and duration were chosen for that element until the conflict was resolved. Otherwise, the trial events were the same as in the previous experiments.

Results and Discussion

When the target set was constrained to remain convex, performance was reliably better than when it was not. Figure 5 shows the results from Experiment 4 as a function of target-set size and constrained or unconstrained configuration; the left and right panels of the figure display performance from blocks 1–3 and blocks 4–5, respectively. An ANOVA revealed a significant main effect for target-set size, $F(4,56) = 42.4$, $p < .001$, and for configuration, $F(1,14) = 14.8$, $p < .01$; the interaction was not reliable, $F(4,56) = 1.74$, $p > .1$. There was also a main effect of practice, $F(1,14) = 5.1$, $p < .05$, but practice did not interact with the other factors in the experiment, both $F < 1$. The ANOVA was repeated with target-set sizes 3, 4, and 5 only (since the constraint factor was irrelevant for target-set sizes 1 and 2), and the results were qualitatively the same: there was a reliable main effect of target-set size, $F(2,28) = 35.3$, $p < .001$, and of configuration, $F(1,14) = 15.4$, $p < .01$, but the interaction between these factors was not reliable, $F(2,28) = 1.6$, $p > .2$. The main effect of practice for target-set sizes 3, 4, and 5 was also significant, $F(1,14) = 4.8$, $p < .05$, but it did not interact with target-set size or with configuration, $F < 1$.

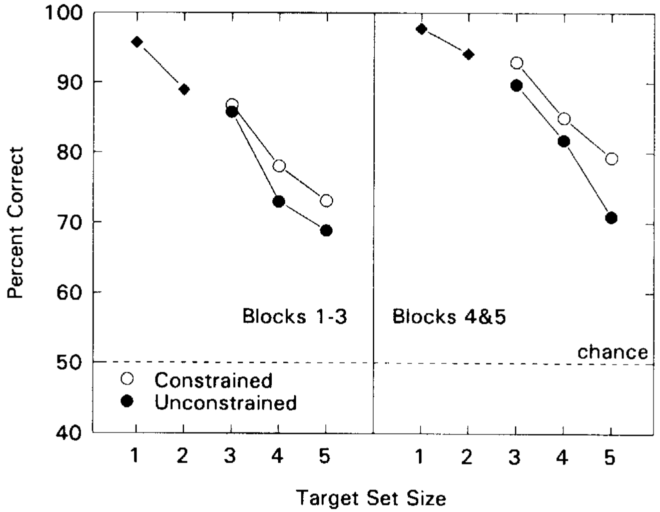

FIG. 5. Results from Experiment 4. In each panel, accuracy is plotted as a function of target-set size for the constrained and unconstrained conditions, respectively. Because convexity is not defined for target sets of one or two elements, target set sizes of 1 and 2 were averaged across conditions and are plotted here with filled diamonds. The left and right panels show data from blocks 1–3 and blocks 4 & 5, respectively. Chance responding is 50%.

That the effect of configuration did not decline with practice contrasts with the corresponding pattern of Experiments 1–3, where the initial advantage for the canonical initial target configuration (Experiment 1), the simultaneous presentation condition (Experiment 2), and the grouping instructions (Experiment 3) disappeared with practice. This is because there was additional stimulus support for grouping in the constrained condition as compared to the unconstrained condition of Experiment 4 throughout the movement phase; this was not the case in Experiments 1–3.

Performance was better overall in Experiment 4 than in Experiments 1–3. This can be attributed at least in part to the reduced duration of the movement phase (4.5 s here vs 7.5 s in Experiments 1–3). The longer the movement phase is, the greater is the probability that the subject will lose one or more elements from the current representation (see General Discussion).

EXPERIMENT 5

In Experiment 4, performance was superior when the configuration of elements was constrained to remain convex than when no constraints were imposed. A side effect of the convexity constraint, however, was that the target trajectories in the constrained condition differed from the nontarget trajectories in various ways. For example, because of the constraints on the position of the target elements, they changed direction more often than nontarget elements did. Furthermore, the target elements were more likely to remain in one quadrant of the display than nontargets were. These two factors might have provided cues to subjects about which elements were targets and which were not, leading to better performance in the constrained condition.

In Experiment 5, this possible artifact was tested by using a variant of the convexity constraint used in Experiment 4. Here, target-set size was fixed at five, and the targets always began in a pentagonal configuration. A randomly selected subset of four targets was constrained to remain in a nonrigid convex configuration during movement (as in Experiment 4). The trajectory of the fifth target was identical to that of a randomly-selected constrained target element from Experiment 4. Thus this *critical target* had trajectory statistics that were identical to those of the constrained elements, but its movements were not constrained with respect to the *current* configuration of targets. The critical target changed direction as often and had a range of motion that was equivalent to the other constrained targets, but it could and frequently did violate the convexity constraint. The critical target was probed on one-fifth of the trials in which a target element was probed.

If attentional tracking depends on the success of perceptual grouping, then performance should be better when the probe is a target that was part of the convex configuration throughout movement than when it was the critical target that violated convexity one or more times during movement. However, if subjects in Experiment 4 made their judgements on the basis of the abovementioned dynamic cues, then performance for the critical target should be the same as for the other targets.

Method

Sixteen new subjects with normal or corrected-to-normal vision were recruited from the unpaid subject pool to participate in one 50-min session. The procedure was the same as in Experiment 4, with the following exceptions. Display size was fixed at 5. Four target elements were randomly selected to remain in a nonrigid convex configuration during movement. The trajectory of the fifth target was specified in advance to correspond to a trajectory selected randomly from among the constrained target trajectories from target-set size 5 trials in Experiment 4. Because this critical trajectory was specified in advance, all other trajectories had to defer to the critical one, yielding more frequent irreconcilable conflicts (leading to restarts) during stimulus generation in this experiment than in Experiments 1–4.

A nontarget was probed on half the trials and a target on the other half. When the probe was a target, it corresponded to the critical target on one-fifth of the trials and to one of the other targets on the remaining four-fifths of the trials. Subjects participated in six blocks of 30 trials each.

Results and Discussion

Mean accuracy ($\pm SE$) when the probe was the critical target was 73.2 ± 3.1%; when it was one of the mutually constrained target elements, accuracy was 82.9 ± 3.3%. The difference of 9.7 ± 2.3% is reliably greater than zero ($t(19) = 4.22$, $p < .001$). This experiment rules out the "dynamic cues" account as the sole explanation for Experiment 4 and shows that an element that violates the convexity constraint is more likely to be lost from the current representation of the target set than one that does not. For example, when the critical target caused the 5-sided virtual object to collapse, the tracking system compensated for this at least part of the time by eliminating the violating vertex from the representation and forming a new virtual object comprising the remaining elements. This result provides further support for the grouping hypothesis.

EXPERIMENT 6

Among the Gestalt laws of grouping, the most prominent in the domain of motion is common fate (Wertheimer, 1912). According to the principle of common fate, elements that move together belong together; the idea is that common motion is unlikely to have been generated by chance, and so there must be some (invisible) common object of which all the visible elements are a part that is causing the visible elements to move in unison.

Common fate, or rigid motion in 3-D space, provided support for perceptual grouping in Experiment 6. Of course, this task would be trivially easy if all the elements were to rotate together. The task employed in Experiment 6 was somewhat more challenging than this. The ten elements were scattered randomly about the display screen. Next the elements were randomly divided into two sets of five elements each; I refer to these as Set I and Set II. Two axes oriented randomly in 3-space were then selected (with the constraint that they were at least 15° apart in orientation). The elements in each set were then constrained to rotate rigidly about the corresponding axis. There were five targets on every trial. On half the trials, constituting the *rigid* condition, all five targets were members of Set I (in other words, the target set and Set I were identical). On the remaining trials, constituting the *nonrigid* condition, three target elements were members of Set I along with two nontarget elements, and the remaining two target elements were members of Set II. Of course, the notion of rigid sets of rotating elements was not articulated to the subjects.

The prediction from the grouping hypothesis is that when the target elements are all members of the same rigid group, performance should be better than when some of the targets are part of one rigid set and some are independently moving as part of another rigid set. Because all elements were always moving in the elliptical paths that define rotation in 3-D space, they could not be discriminated only on the basis of the shape of their trajectories; therefore this experiment contains within it a control for the kinds of dynamic cues that had been possible confounds in Experiment 4.

Method

Subjects. Eighteen undergraduates from the Johns Hopkins University were paid $5 to participate in one 40-min session. None of the subjects had participated in Experiments 1–5, and all had normal or corrected-to-normal vision.

Procedure. There were five targets on each trial. The elements were positioned randomly on the display screen, and divided randomly into Set I and Set II. An axis of rotation was selected randomly for each set with the constraint that they were at least 15° apart and at least 15° from any of the three canonical axes (since rotation around the x or y axes yields simple oscillation and rotation around the z axis yields circular trajectories, both of which are undesirable).

In the rigid condition, the targets were all members of Set I and formed a single rigid group. In the nonrigid condition, three members of Set I were identified as targets, and two members of Set II were identified as targets. The targets in the nonrigid condition therefore did not form a single rigid group, but a complex nonrigid form.

Because of the complexity of the rigidly constrained trajectories used in this experiment, the cushions used in all the other experiments reported in this article were not used here. That is, elements could cross over and/or be momentarily superimposed on one another, leading to possible ambiguities during tracking. Disambiguation was accomplished by the smooth continuation of trajectories in a predictable direction.

As in Experiments 1–5, the 5 target elements were designated at the start of the trial by flashing them on and off several times. Then motion began (rotational velocity was 240°/s) and continued for 4.5 s. Finally, all the elements stopped and one was probed. The probe was a target on half the trials and a nontarget on the remaining trials.

Results and Discussion

Accuracy was 82.2% in the rigid condition and 70.1% in the nonrigid condition, a significant difference, $F(1,17) = 27.9$, $p < .001$. The rigidity constraint yielded much better performance, as predicted by the grouping hypothesis. Models of tracking that do not incorporate a mechanism for grouping cannot easily account for this result.

Practice effects are shown along with the grouping effect in Fig. 6. First, as expected, there was an overall improvement in performance with practice, $F(5,85) = 3.6$, $p < .01$. Second, as in Experiment 4, the superior accuracy in the rigid condition as compared to the nonrigid condition persisted throughout practice: the interaction between condition and block was not significant, $F < 1$.

One might ask why performance was less than perfect in Experiment 6. After all, the Gestalt property of common fate has long been known to yield excellent object recognition and is commonly used to illustrate the visual system's exquisite ability to recover object structure from motion cues alone (e.g., Ullman, 1979). The displays used in Experiment 6, however, were considerably more complex than the ones used in standard structure-from-motion demonstrations. The objects were defined by only five elements, and there were an equal number of noise elements (i.e., points that were not part of the rigid structure being tracked) in the display. It is known that accurate recovery of structure from motion or discrimination of rigid from nonrigid motion depends both on the number of dots used to define the structure and on the amount noise in the display (Braunstein, Hoffman, & Pollick, 1990). Furthermore, the cushions used in the other experiments described in this article to prevent collisions were not incorporated into Experiment 6; this had the consequence that elements could pass through one another during motion, leading to true ambiguities in which elements were targets and which were nontargets.

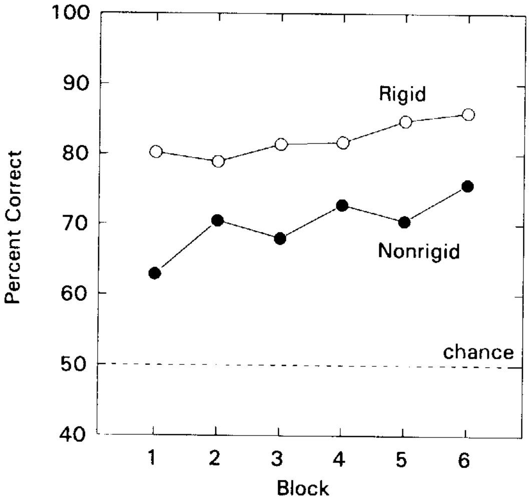

FIG. 6. Results from Experiment 6. Accuracy is plotted as a function of practice for the rigid and nonrigid motion conditions, respectively. Chance responding is 50%.

Informal observations I have made confirm that displays incorporating rigid 2-D target configurations rotating in 3-space with nontargets constrained to move in the picture plane yield virtually perfect tracking performance—the targets "pop out" of the display effortlessly. Unfortunately, this design introduces an unambiguous stimulus cue that discriminates targets from nontargets: elements rotating in 3-space have elliptical trajectories, while elements moving in linear trajectories in the picture plane do not. Because this additional cue could undermine the claim that subjects were actually tracking the targets, the associated design was not used in Experiment 6.

EXPERIMENT 7

In Experiment 6, the trajectories of the target elements were manipulated so as to impose a common fate constraint on their movement that supported perceptual grouping. To further test the extent to which stimulus support can influence the maintenance of a perceptual group, I used a much weaker form of common fate in Experiment 7 by manipulating the relative velocities of the target and nontarget elements. Four different relative velocity conditions were used: all high velocity, all low velocity, target high velocity/nontarget low velocity, and target low velocity/nontarget high velocity. Any velocity difference between targets and nontargets should facilitate grouping; to the extent that grouping is an important factor in the tracking task, a relative velocity difference should enhance tracking performance.

Of course, subjects could in principle use the velocity difference alone to perform the task. This strategy would require determining the relative velocity of the targets and nontargets at the beginning of the movement phase, and then noting the final velocity of the probe element immediately before it stops; this would not necessarily require tracking. In order to prevent this strategy, a pilot experiment was first conducted to estimate the largest relative velocity difference that could not reliably be detected in a nontracking task, and this relative velocity difference was used in the primary experiment.

Fifteen subjects participated in the preliminary experiment. Ten plus signs moved about the screen as in Experiment 1; half the elements had a slow velocity and the other half had a fast velocity. The slow velocity was fixed at 2.96°/s and the fast velocity was 3.70, 4.44, 5.18, or 5.92°/s (yielding velocity differences of 0.74, 1.48, 2.22, and 2.96°/s). Each velocity occurred equally often in each block of trials. There was no target designation phase; instead, the 10 plus signs appeared in random locations for 500 ms, then began to move. At the end of the 7.5-s movement phase, one of the ten elements was probed. Subjects were required to press the right key if the probe was a fast element, and the left key if the probe was a slow element, with each outcome equally likely. Each subject completed 6 blocks of 32 trials.

Mean accuracy ($M \pm SE$) was 52.8 ± 2.1%, 61.8 ± 2.4%, 67.3 ± 2.6%, and 73.9 ± 2.9% for velocity differences of 0.74, 1.48, 2.22, and 2.96°/s, respectively. Performance with a velocity difference of 0.74°/s did not differ reliably from chance, $t(14) = 1.3, p > .1$; in that condition subjects were unable to use the velocity difference to categorize elements as "fast" or "slow." This result is well within the range of velocity-increment threshold measurements obtained by Sekuler (1990) with psychophysically untrained subjects in a related task. These two velocities (2.96 and 3.70°/s) were therefore used as the fast and slow velocities, respectively, in Experiment 7.

Method

Fifteen new subjects participated in Experiment 7. There were four different velocity conditions in this experiment, randomly mixed within each block: (a) both targets and nontargets fast, (b) both slow, (c) targets fast and nontargets slow, and (d) targets slow and nontargets fast. For each of these conditions, the targets were equally often constrained to remain in a nonrigid convex configuration (as in Experiment 4) or not. There were always five target elements in this experiment. Subjects were not informed of the velocity difference until after the experiment, and none reported being aware of a velocity difference even after informed of it.

Results and Discussion

Mean accuracy for the four velocity conditions as a function of whether the configuration was constrained to remain convex or not is shown in Fig. 7. These data were entered into a four-way repeated-measures ANOVA with target configuration (constrained or unconstrained), relative velocity (same or different), absolute velocity (targets fast or slow), and practice (first half of session vs second half) as factors. The main effects of configuration, $F(1,14) = 4.92, p < .05$, and relative velocity, $F(1,14) = 13.06, p < .01$, were reliable, as was the interaction between these factors, $F(1,14) = 4.91, p < .05$. The main effect of absolute velocity was not reliable, $F(1,14) = 2.31, p > .1$. Accuracy improved significantly with practice, $F(1,14) = 5.15, p < .05$, but practice did not interact with any other factor. The only other reliable effect was the interaction between the relative velocity, $F(1,14) = 5.94, p < .05$.

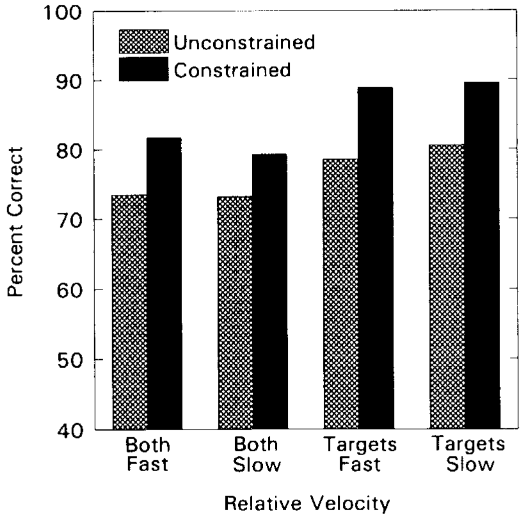

FIG. 7. Results from Experiment 7. Accuracy is plotted for the four velocity conditions (targets and nontargets fast, targets and nontargets slow, targets fast and nontargets slow, targets slow and nontargets fast) for the unconstrained condition (hatched bars) and constrained condition (solid bars), respectively.

The primary result of Experiment 7 was that when a relative velocity gradient was present to support perceptual grouping, performance in the tracking task improved. In replication of Experiment 4, constraining the target elements to remain in a nonrigid convex configuration also improved performance. In fact, the degree to which the convexity constraint improved performance was greater when the velocity gradient was present than when it was not; this is reflected in the significant configuration by relative-velocity interaction. As in Experiments 4 and 6, the improvement attributable to the grouping manipulation did not disappear with practice, providing evidence that the manipulation provided support for the maintenance of the perceptual group during tracking; this contrasts with the findings of Experiments 1–3 in which the grouping manipulation was effective only early in practice, suggesting an effect in the formation but not the maintenance of the group. Overall, the results from prevent this strategy, a pilot experiment was first conducted to estimate the largest relative velocity difference that could not reliably be detected in a nontracking task, and this relative velocity difference was used in the primary experiment.

Fifteen subjects participated in the preliminary experiment. Ten plus signs moved about the screen as in Experiment 1; half the elements had a slow velocity and the other half had a fast velocity. The slow velocity was fixed at 2.96°/s and the fast velocity was 3.70, 4.44, 5.18, or 5.92°/s (yielding velocity differences of 0.74, 1.48, 2.22, and 2.96°/s). Each velocity occurred equally often in each block of trials. There was no target designation phase; instead, the 10 plus signs appeared in random locations for 500 ms, then began to move. At the end of the 7.5-s movement phase, one of the ten elements was probed. Subjects were required to press the right key if the probe was a fast element, and the left key if the probe was a slow element, with each outcome equally likely. Each subject completed 6 blocks of 32 trials.

Mean accuracy ($M \pm SE$) was 52.8 ± 2.1%, 61.8 ± 2.4%, 67.3 ± 2.6%, and 73.9 ± 2.9% for velocity differences of 0.74, 1.48, 2.22, and 2.96°/s, respectively. Performance with a velocity difference of 0.74°/s did not differ reliably from chance, $t(14) = 1.3, p > .1$; in that condition subjects were unable to use the velocity difference to categorize elements as "fast" or "slow." This result is well within the range of velocity-increment threshold measurements obtained by Sekuler (1990) with psychophysically untrained subjects in a related task. These two velocities (2.96 and 3.70°/s) were therefore used as the fast and slow velocities, respectively, in Experiment 7.

Method

Fifteen new subjects participated in Experiment 7. There were four different velocity conditions in this experiment, randomly mixed within each block: (a) both targets and nontargets fast, (b) both slow, (c) targets fast and nontargets slow, and (d) targets slow and nontargets fast. For each of these conditions, the targets were equally often constrained to remain in a nonrigid convex configuration (as in Experiment 4) or not. There were always five target elements in this experiment. Subjects were not informed of the velocity difference until after the experiment, and none reported being aware of a velocity difference even after informed of it.

Results and Discussion

Mean accuracy for the four velocity conditions as a function of whether the configuration was constrained to remain convex or not is shown in Fig. 7. These data were entered into a four-way repeated-measures ANOVA with target configuration (constrained or unconstrained), relative velocity (same or different), absolute velocity (targets fast or slow), and practice (first half of session vs second half) as factors. The main

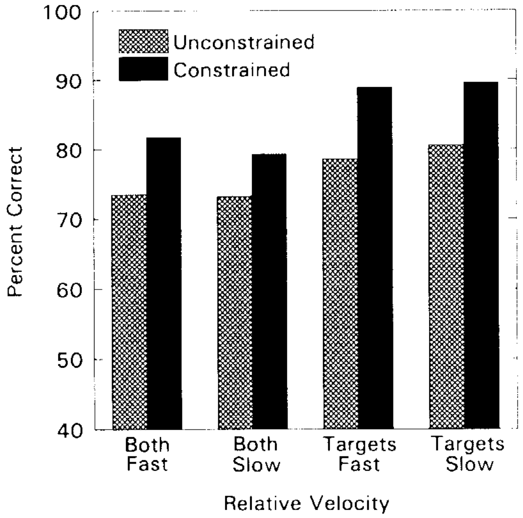

FIG. 7. Results from Experiment 7. Accuracy is plotted for the four velocity conditions (targets and nontargets fast, targets and nontargets slow, targets fast and nontargets slow, targets slow and nontargets fast) for the unconstrained condition (hatched bars) and constrained condition (solid bars), respectively.

effects of configuration, $F(1,14) = 4.92, p < .05$, and relative velocity, $F(1,14) = 13.06, p < .01$, were reliable, as was the interaction between these factors, $F(1,14) = 4.91, p < .05$. The main effect of absolute velocity was not reliable, $F(1,14) = 2.31, p > .1$. Accuracy improved significantly with practice, $F(1,14) = 5.15, p < .05$, but practice did not interact with any other factor. The only other reliable effect was the interaction between the relative velocity, $F(1,14) = 5.94, p < .05$.

The primary result of Experiment 7 was that when a relative velocity gradient was present to support perceptual grouping, performance in the tracking task improved. In replication of Experiment 4, constraining the target elements to remain in a nonrigid convex configuration also improved performance. In fact, the degree to which the convexity constraint improved performance was greater when the velocity gradient was present than when it was not; this is reflected in the significant configuration by relative-velocity interaction. As in Experiments 4 and 6, the improvement attributable to the grouping manipulation did not disappear with practice, providing evidence that the manipulation provided support for the maintenance of the perceptual group during tracking; this contrasts with the findings of Experiments 1–3 in which the grouping manipulation was effective only early in practice, suggesting an effect in the formation but not the maintenance of the group. Overall, the results from

Experiment 7 are consistent with the grouping hypothesis and cannot be accounted for by models that do not include grouping or organizational factors.

GENERAL DISCUSSION

I have reported the results of seven experiments concerning the role of perceptual organization and attention in multielement visual tracking. Each experiment included a manipulation of the extent to which perceptual grouping was possible. In Experiments 1–3, I manipulated factors that influenced the formation of a perceptual group, including the initial configuration of the target elements, whether the targets were designated simultaneously or sequentially, and whether grouping instructions were provided to subjects. In each case a significant improvement in performance occurred, but only in the early stages of practice. In Experiments 4–7, I manipulated common-fate factors that influenced how successfully a perceptual group could be maintained during tracking, including both rigid and nonrigid constraints on the configuration of target elements and modulation in the velocities of the target and nontarget elements. In each the latter experiments, the successful maintenance of the perceptual group during tracking was reflected in improved performance throughout practice.

Together these experiments provide converging evidence that the spatial relations among the targets have a significant impact on the effectiveness of multielement visual tracking. Models in which the configuration of the elements is not relevant cannot capture the regularities in the results I have reported. For example, the FINST theory assumes that indexing is preattentive, and that the observer does not have access either to the content of indexed objects or to the spatial relations among indexed objects (e.g., Pylyshyn, 1989, pp. 67–68). The latter assumption is not supported by the present experiments.

Attention and Perceptual Organization Revisited

Three related conclusions of increasing abstraction are supported by the present results. First, and most concretely, the configuration of the target elements—their spatial relations during motion—has a significant impact on performance in this task. Second, these configural effects support object-based theories of visual attention which hold that the elements to be tracked are preattentively grouped in early vision, and that attention is directed toward this virtual object during tracking. In particular, according to the current account, subjects do not separately track the individual elements or a changing region of space; instead, they attend to and track a single coherent virtual object. Finally, the results provide strong evidence that perceptual organization can be directed by current perceptual goals and knowledge; it need not be exclusively bottom-up or stimulus-driven, as has commonly been assumed. The first of these conclusions is merely a restatement of the results, and needs no further elaboration. The second and third are discussed in more detail below.

Support for object-based theories of attention. Several previous studies have demonstrated that subjects can or spontaneously do segregate stimulus elements in selective-attention tasks using perceptual grouping mechanisms. These experiments have a common theme: they all show that selective attention operates on perceptual objects and need not select on the basis of spatial location alone. For example, Neisser and Becklen (1975) produced a video tape on which two different event sequences were superimposed (a pair of hands playing a clapping game, and a trio of people throwing a ball among themselves). Subjects were required to attend to and note anomalous events in one or the other of the sequences (e.g., one of the ball players temporarily leaving the scene and then returning), and to ignore the other physically overlapping sequence. Subjects could do this easily, and they rarely noticed anomalous events in the unattended sequence.

Similarly, Rock and Gutman (1981) presented to subjects pairs of physically overlapping nonsense shapes, one printed in green ink and the other in red. Each subject rated the shapes of one color according to how pleasing they were. After examining ten such stimuli, subjects were unexpectedly presented with a recognition sheet and asked to indicate which of the shapes they had seen during the rating task. Subjects recognized almost perfectly the attended shapes, and failed to recognize the unattended shapes above chance.

A more recent example of this phenomenon was provided by Kramer and Jacobson (1991), who found that the extent to which a flanking form interfered with responses to a target form depended on whether or not the flanking form was perceived to be part of the same perceptual object as the target form, even when the physical positions of the two forms were the same in each case.

These experiments, together with other studies of perceptual organization in selective-attention tasks, provide support for object-based theories of attention (e.g., Banks & Prinzmetal, 1976; Driver & Baylis, 1989; Duncan, 1984; Fox, 1978; Kahneman & Henik, 1981; Kahneman et al., 1992; Kanwisher, 1991; Moraglia, 1989; Prinzmetal, 1981; Tipper et al., 1990, 1991; Treisman, 1982; Treisman et al., 1983). The present results add to those cited here in showing that selective attention can operate on a perceptual object (the virtual polygon) and need not be directed to other elements that are spatially superimposed on the attended object, as space-based theories assume. This is not to say that selection can never be based on spatial location; there is ample evidence that it can. The claim is that selective attention is not limited to spatially defined representations, but can also operate on object-based representations.

Top-down basis for perceptual grouping. Virtually all discussions of perceptual organization—from those of the Gestalt psychologists (Koffka, 1935/1963; Köhler, 1929/1947; Wertheimer, 1912) to those of recent psychologists (e.g., Banks & Prinzmetal, 1976; Bundesen & Pedersen, 1983; Fox, 1978; Humphreys, Quinlan, & Riddoch, 1989; Kramer & Jacobson, 1991; Olson & Attneave, 1970; Palmer, 1983; Pomerantz & Pristach, 1989; Pomerantz & Schwaitzberg, 1975; Treisman, 1982), psychophysicists (e.g., Casco & Morgan, 1987; Casco, Morgan, & Ward, 1989; Chang & Julesz, 1983; Watt, 1988), and computer-vision theorists (e.g., Ballard, 1984; Chong, Kahn, & Winkler, 1990; Fennema & Thompson, 1979; Lowe, 1985; Mahoney & Ullman, 1988; Marr, 1982; Mohan & Nevantia, 1989; Zucker, 1987)—have considered perceptual organization as an exclusively bottom-up or stimulus-driven phenomenon. According to this approach, grouping depends only on properties of the image, and not on goal-directed factors (e.g., observers' knowledge that certain elements are relevant for a task and others are not). A primary objective has been to characterize the properties of the stimulus that facilitate perceptual grouping; the range of possible answers includes factors like local similarity (e.g., in orientation, brightness, size, or shape), proximity, collinearity, parallelism, symmetry, and common motion—all properties identified by the Gestalt psychologists as important determinants of grouping.

The present experiments have shown instead that grouping can be a top-down, goal-directed process as well. Subjects' tracking performance is based at least in part on an interplay between an internal representation of the current configuration of the elements and the contents of the image, with continuous updating of the representation to match as closely as possible the current stimulus. Previous demonstrations of top-down effects on perceptual organization (e.g., Attneave, 1971; Girgus, Rock, & Egatz, 1977; Peterson & Hochberg, 1983) have relied on observers' ability to intentionally alter the appearance of ambiguous figures like the Necker cube. Peterson and Gibson (1991) in particular have demonstrated the importance of the intentional distribution of attention on the perceptual organization of ambiguous figures. Similar ideas appear in the literature concerning the effects of perceptual set on object recognition (e.g., Pachella, 1975; Steinfeld, 1967). The present results extend these findings to observers' ability to dynamically group elements into a virtual object.

The present experiments provide evidence for a top-down basis for grouping in that there was little or no stimulus support for the segregation and grouping operations; grouping had to be maintained (perhaps effortfully) by continuously aligning an internal representation with the contents of the visual display. The stimulus factors that were manipulated in Experiments 4–7 had their influence by making the alignment process more or less difficult. The key point is that these manipulations *required (goal-directed) tracking to yield grouping effects*. This distinguishes them from those used in experiments like Rock and Gutman (1981) or Neisser and Becklen (1975) in which the two objects or event sequences were always visibly distinguishable in a stimulus-driven fashion.

To illustrate the importance of top-down influences on the organization of the displays I used, I asked a new group of subjects to view displays that either did or did not contain a set of five elements that were constrained to remain convex during motion, in the manner of Experiment 4. These displays differed from the ones used in Experiment 4 in that no target set was defined at the start of motion, and motion began from a random starting point. So the array of ten elements was not grouped into targets and nontargets at the start of the trial in this experiment as it was in Experiment 4. The question was whether subjects could detect a group of convex elements based only on the convexity constraint, and without any top-down help in the form of an initial specification of targets and nontargets. The results were clear: subjects correctly detected the convexity constraint on just 52% of the trials (where 50% is chance guessing); in other words, without a virtual polygon representation to use as a basis for tracking, the stimulus display itself provided no clue as to the identity of the constrained or unconstrained elements. The same can be said for the velocity manipulation used in Experiment 7: the preliminary experiment revealed that the velocity difference alone could not support a bottom-up distinction between the targets and nontargets. (A similar experiment conducted with the displays used in Experiment 6 involving rigid configurations rotating in 3-space revealed in contrast that there was a significant bottom-up component assisting in the grouping process in that experiment.)

An instructive analogy can be drawn between the top-down imposition of a structure on the elements in these experiments and the way in which different cultures have grouped the stars into constellations.[8] The positions of the stars are more or less random, and constellations can be viewed as "arbitrary groupings of stars" (Moore, 1987, p. 104). The groupings are not entirely arbitrary, of course. Certain stellar configurations are "seen" by virtually all cultures (e.g., the Big Dipper; although its name varies from one culture to the next, it is always interpreted as a ladle or pan). These constellations are universal in that they satisfy certain of the classic Gestalt laws of proximity, good continuation, similarity (in brightness), and Prägnanz.

[8] I thank Howard Egeth for suggesting this analogy.

Other configurations are grouped quite differently by different cultures at least in part because of differences in the myths and legends that characterize any culture. Thus the Greek system of constellations, which is the one with which we are most familiar, is completely different from the Chinese, the Egyptian, and the Polynesian systems (Pannekoek, 1961; Schafer, 1978). Brecher (1979) describes an example of the separate contributions of bottom-up and top-down factors in grouping stars into constellations by different cultures. The stars near Sirius—the brightest star in the night sky—were grouped into a bow-and-arrow-shaped constellation by both the ancient Babylonians and the Chinese. For the Babylonians, the arrow was long and Sirius defined its tip; for the Chinese, however, the arrow was short and Sirius was its target. In Western tradition, the stars near Sirius define the constellation Canis Major (the Big Dog). The similar interpretations of the Chinese and Babylonian cultures is thought to reflect a common origin for their astronomical myths. This example illustrates how the observer can impose an interpretation on ambiguous perceptual stimuli, based at least in part on expectations and perceptual set.

Filtering by movement. Recent experiments reported by McLeod et al. (1991) reveal that observers can direct attention to just the moving elements in a display of moving and stationary elements, even if the moving elements do not form a clear group by virtue of common fate (e.g., they move in different directions). The authors concluded that the visual system may have a ''movement filter'' that permits the observer to attend to objects exhibiting a relevant movement attribute (e.g., moving vs stationary or moving up vs moving down). As in the present experiments, McLeod et al. (1991) found the subjects could selectively attend to certain moving elements even when nontarget elements were spatially interspersed with them, providing evidence against exclusively space-based models of selection. The present results differ from those of McLeod et al. in showing that observers can impose a structure on a dynamic stimulus in a top-down fashion without specifying one or more particular movement attributes in advance. In particular, it is unlikely that observers in the present tracking experiments employed a movement filter of the type described by McLeod et al., because no simple or consistent set of motion attributes existed to distinguish between targets and nontargets.

Attending to shapes and regions. A possible prediction of the present account is that nontarget elements inside the virtual polygon formed from the target elements should mistakenly be identified as targets more often than nontargets outside the virtual polygon. This prediction stems from the assumption that attending to a shape necessarily entails attending to the solid region bounded by that shape. However, an object-based account is nonspatial, and so incorporates only the elements and their connections (in this case, the vertices and edges of the virtual polygon), but not necessarily the intervening regions of space. An object-based account would not necessarily predict more false alarms for nontargets inside the final position of the virtual polygon than outside it. In fact, none of the present experiments provided any evidence for such a difference.

Mechanisms of Visual Tracking

In this section, I first describe an object-model alignment mechanism that could carry out the visual tracking task studied here. Using this mechanism as a framework, I then discuss possible sources of performance errors in the task. Then I consider two alternative accounts of performance in this task: a sophisticated-guessing strategy and a moving-spotlight model. Finally, I discuss the relative merits of the alignment mechanism, the FINST theory, and spotlight models of attention.

Object–model alignment. One possible mechanism for the formation of a virtual polygon is that an internal model of the target element configuration is formed at the start of each trial, and the model is continuously compared with the image and updated when necessary. This could be accomplished as follows: First, an internal representation or object model of the initial configuration of target elements (including its shape and location) is generated while the targets flash during target designation; adjacent elements in the object model may or may not be joined with virtual lines to form an explicit virtual polygon. Then, during motion, this representation is updated as the configuration of elements in the image changes. When the probe flashes, the object model is queried to determine whether the probe corresponds to a vertex of the virtual polygon; an appropriate response is then generated. Algorithms for accomplishing model updating include (a) a dynamic variant of the elastic stretching algorithm described by Burr (1983); (b) a variant of Ullman's (1984a) incremental rigidity scheme for recovering structure from motion; (c) alignment-based models of visual recognition (e.g., Lowe, 1987; Ullman, 1989; Yuille, 1991); and (d) token-correspondence approaches to dynamic scene analysis (e.g., Sethi, Salari, & Vemuri, 1988).

This mechanism is related to the visual mechanism that presumably accomplishes mental rotation and other imaging tasks (e.g., Kosslyn et al., 1990; Kosslyn & Pomerantz, 1977; Shepard & Cooper, 1982; Tarr & Pinker, 1989). To see this correspondence, consider Shepard and Metzler's (1971) mental rotation task, in which two 3-D block figures were shown simultaneously and subjects were to determine whether they differed by a rotation only or by a rotation and a reflection of one feature. This task could be accomplished by constructing a mental representation of one of the block figures and rotating it within a visual buffer, continuously comparing it to the other block figure until the two either matched exactly, or matched in all but one feature. The larger the angle of rotation between the two stimuli, the longer the decision took, presumably because larger mental rotations take longer than smaller ones.

The continuous comparison process required in mental rotation, then, is essentially the same as the one required in the present tracking task. The two processes differ in that the internal representation or model of the target-set configuration must not only be compared with the contents of the visual display, it must also be updated so as to correspond as closely as possible with the appropriate elements in the display. A candidate for the updating process is Ullman's (1989) two-stage alignment approach to object recognition in which a viewed object is compared with a set of canonical object models stored in memory. In the first (alignment) stage, all candidate models are transformed (using a limited set of possible transformations) so as to align as closely as possible with the viewed object. In the second (matching) stage, the transformed object model that maximizes a similarity function is selected. A simplified version of this general approach can be adapted to the tracking task described in this article: only a single object model is aligned with the viewed object configuration, so no search is required.

The success of this approach depends on the size of the set of allowable or possible transformations. If there are no limits on the transformations that can occur, then the required computations become implausibly complex. In the present task, the transformations are highly restricted: smooth changes in location, orientation, size, and shape can occur. When the set of transformations is restricted further (e.g., adding the convexity constraint of Experiments 4 and 7), performance improved. The rigidity constraint imposed in Experiment 6 (preventing changes in shape and size) improved performance still further.

The improvement in performance observed under the convexity and rigidity constraints is also consistent with the known effects of complexity in object recognition. Previous accounts of complexity effects in object recognition have stressed the importance of the number of features in the shape (Attneave, 1957), which, in the case of planar polygons, is directly proportional to the number of concavities it contains. A polygon is simple to the extent that it is compact and convex.[9] When a nonrigid configuration exhibits concavities, it may still be coherent (i.e., the ordinal relations among its vertices remain constant), but the existence of concavities results in "arms" and other features. When a configuration violates coherence (e.g., when a vertex crosses over an opposite edge), then the shape of the configuration changes abruptly and either a new object model must be constructed or the violating vertex must be eliminated from the representation.

Sources of error. There are two main sources of error within this framework. First, it is likely that one or more of the vertices of the object will be lost during the course of a trial (e.g., the critical element in Experiment 5), in which case a lower order virtual polygon (i.e., one with fewer vertices than originally specified) may be constructed. When the probe appears on a vertex of the virtual polygon in this situation, a positive response is made, just as when all the elements are successfully tracked. However, if the probe falls on an element not in the polygon, a positive response is made with some probability corresponding to the likelihood that the target element that was lost had been probed, using some version of a sophisticated-guessing strategy (see next section).

A vertex might be lost when the updating process becomes too time-consuming or difficult, which might occur as the complexity of the object increases. If there are several concavities and arms in the object model, it might take longer to update using the alignment methods described in the last section, therefore producing more errors when alignment cannot "keep up" with the continually changing input. Folk and Luce (1987) measured the mental rotation rate for random nonconvex two-dimensional forms (generated using Method 1 of Attneave & Arnoult, 1956), and found slower rotation rates for complex objects than for simple objects. Bethell-Fox and Shepard (1988) reported similar results with objects defined by patterns of filled-in squares in a 3 × 3 matrix, particularly early in practice when the objects were unfamiliar. These studies suggest that perceptual transformations become increasingly difficult as the representation to be transformed becomes more complex. An element might also be lost from the current object model when it violates coherence (i.e., a vertex crosses over an opposite edge). This could disrupt the entire model and cause it to be lost entirely, or it could force the system to drop that vertex so as to maintain coherence among the remaining vertices.

A second source of error in this task is more mundane. The elements in the target set could define a virtual polygon that spanned as much as 11°

[9] Compactness can be quantified as a polygon with a large ratio of area to perimeter, as suggested by Attneave and Arnoult (1956) and by Podgorny and Shepard (1983). In fact, the measure used by Podgorny and Shepard was square-root-area over perimeter; this quantity is maximized by a circle and is shape invariant. Compactness alone does not determine complexity, however, because although a complex figure with many arms will tend to be low in compactness, so will a thin rectangle or ellipse, both of which are demonstrably low in complexity (Attneave & Arnoult, 1956). It is necessary to incorporate other factors, such as the number of features (arms or concavities) in an object, to specify complexity comprehensively.

of visual angle, forcing some elements to be tracked at least 5° from the fovea. The cushion defining how close together elements could drift before bouncing apart was 0.5° in width. These parameters are near the limits of visual acuity (Anstis, 1974). Thus, even assuming perfect tracking, if a nontarget element drifts to within 0.5° of a target 5° from the fovea, then when they subsequently drift apart, the subject might have no idea which of them was the target, in which case he or she would simply have to guess. In Experiment 6, this problem was compounded by the elimination of the cushions altogether, resulting in overlapping targets and producing a real limitation in expected accuracy.

Sophisticated guessing strategies. An alternative to object-model alignment is a guessing strategy that exploits the contingencies inherent in the experimental design. A common strategy used by subjects just beginning the task was to select one of the target elements to track, and follow it with pursuit eye movements. With practice, they were able to include more of the target elements in the set they were tracking. In this section, I analyze a strategy of this sort to determine the extent to which subjects may have used only sophisticated guessing, rather than multielement tracking, to perform this task.

Consider a strategy in which a subset of the elements is selected for tracking. When the number of elements in the subset is one, then performance in the tracking task can be explained without reference to attention, because subjects can simply track the element with their eyes. According to the strategy, if the tracked element is probed, the subject responds positively, and if the tracked element is not probed, then the subject responds positively on some trials and negatively on others. The bottom function in Fig. 8, labeled "1," shows the predicted probability of making a correct response as a function of target-set size if exactly one element is tracked perfectly and probability matching is optimal (the dashed horizontal line represents random guessing performance). The other functions in the figure show predicted probability correct given that subjects perfectly track 2, 3, 4, or 5 targets, respectively, and guess optimally when an untracked elements is probed. In general, performance under this kind of guessing strategy yields performance functions that are monotonically decreasing with target-set size.

Nevertheless, a purely guessing account is unsatisfactory. Consider the observed performance in Experiment 4 (Fig. 5). Except for target-set size 1, percent correct is everywhere significantly greater than predicted by the pure guessing strategy assuming that one element was tracked perfectly. So if the guessing strategy is to work, it has to assume that at least two and perhaps three elements were tracked perfectly and that optimal guessing was employed. However, then we are left with a mystery about

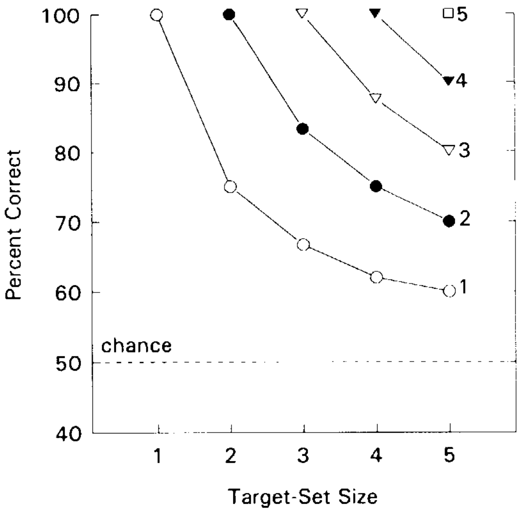

FIG. 8. Hypothetical results based on sophisticated guessing strategy. Curve parameter is number of target elements assumed to be tracked perfectly. Each curve represents predicted probability correct for a given number of elements tracked perfectly, with optimal guessing. For example, the curve labeled "2" shows that when two target elements are tracked perfectly, accuracy is 1.0 when target-set size is 2, and declines to 0.7 when there are 5 targets. See text for further details.

how the tracking was accomplished. Furthermore, the guessing account provides no explanation for the observed grouping effects.

It is very likely that subjects use some form of sophisticated guessing strategy when performing this task; it is equally likely that a guessing strategy is not the only mechanism they employ.

Assessment of the serial attention-switching hypothesis. Most space-based theories of attention allocation incorporate some mechanism for reallocating attention from one spatial region of a display to another. For example, one class of models holds that attention is analogous to a spotlight that moves in an analog fashion across the visual field (e.g., Shulman et al., 1979; Tsal, 1983; but see Eriksen & Murphy, 1987, and Yantis, 1988). As pointed out by Pylyshyn and Storm (1988), a plausible space-based mechanism for visual tracking in the present task is to shift a "spotlight of attention" from one target location to the next during movement.

Pylyshyn and Storm (1988) tested several specific versions of this model and rejected all of them. In this session, I analyze just one possible version of an attention switching model. The model adopted for this test is similar to the first such model tested by Pylyshyn and Storm (1988, p.

186). According to this model, a table of x–y coordinates for the target elements is maintained in memory and is updated as often as possible as the elements move about the display. It is assumed that the starting positions of the elements are entered into the table without error. Once movement starts, the spotlight moves to the last known location of the first element in the table. The coordinates of the element that is currently closest to this location are then entered into the table in place of the previous ones, so long as those coordinates are not already present elsewhere in the table.[10] This procedure continues until the probe appears. The table is then inspected to determine whether one of the sets of coordinates it contains is closer to the coordinates of the probe than to any other element, and the appropriate response is made.

The algorithm will be successful in the tracking task to the extent that the attentional spotlight has a high velocity (defined as the rate at which the coordinates are updated, regardless of how this is done) relative to the velocity of the target elements. This model was simulated using the algorithm for generating stimulus sequences in Experiment 1, and the results of the simulation are shown in Fig. 9. Proportion correct is plotted as a function of the assumed velocity of the attentional spotlight for target set-sizes 2, 3, 4, and 5 (performance is always perfect for one target so long as the velocity of attention is at least as great as the velocity of the elements).

With very slow spotlight velocities (e.g., 10°/s), the probability of a correct response is at chance, or approximately $n/10$, where n is the number of targets and 10 is the total number of elements. To understand this, consider the following example involving target-set size 3. The spotlight starts at the tabled location of Element 1, and then moves to the tabled location of Elements 2 and 3 before returning to the last known location of Element 1. The average path length (i.e., the distance to be transversed by the spotlight of attention in order to update all the elements and return to Element 1) is approximately 15° of visual angle (the average distance between any given pair of target elements across all target-set sizes was about 5°). It takes 1500 ms for attention to move 15° at 10°/s. During this time, Element 1 has moved at least several degrees from its table location (recall that element velocity was 4.6°/s, so its path during the attention movement was 6.9°, although the path need not be uniformly away from the tabled position). By the time the spotlight has moved back to the tabled location of Element 1, the probability is very high that a nontarget has moved closer to that location than Element 1 itself is. Over an entire 7.5-s trial, it is virtually certain that the tabled entries will have been replaced by the coordinates of other elements. On average, over trials, the proportion of coordinates in the table corresponding to targets will equal the proportion of targets in the display (i.e., $n/10$).

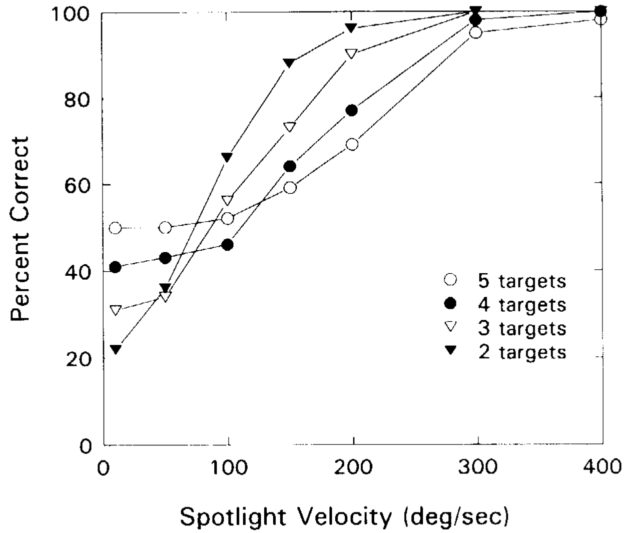

FIG. 9. Accuracy of the simulated serial scanning algorithm as a function of attentional spotlight velocity. Curve parameter is target-set size. Accuracy is equal to the probability over trials that the probe was the closest element to a tabled coordinate pair. See text for details.

As spotlight velocity increases, the probability that the target elements will be sampled before they move too far away from their tabled locations increases monotonically. The fastest functional rate for the spotlight is for all table entries to be updated within the interval of one frame of the sequence (i.e., within 30 ms in this version of the task). The average path length between two target elements was about 5°, so when there are five targets present, a spotlight that travels 25° (five targets) in 30 ms, or 833°/s, should perform perfectly in this task if there is assumed to be no time required to compute coordinates and enter them into the table.[11]

[10] If the coordinates are already present in the table, then the algorithm selects the coordinates of the next-closest element, and so on. This condition is one not employed by Pylyshyn and Storm (1988), and it has the effect that there will be no duplicates in the table of coordinates. This seems reasonable, since the observer knows there are n different elements to be tracked; targets cannot be mistaken for other targets.

[11] The assumption that table updates take no time is highly implausible, especially if they are assumed to involve a certain amount of search to avoid duplication (see Footnote 10). If table updates do take time, then the estimated spotlight velocities would have to be adjusted upward accordingly.

According to the simulation, performance is virtually perfect for all spotlight velocities greater than about 300°/s. The reason performance remains at a relatively high level for velocities of less than 833°/s has to do with the probability that a nontarget will drift closer to a tabled location than a target will within a given amount of time, and this depends on the velocity of the elements and on the properties of the movement trajectories.

For all target-set sizes, performance drops below 50% for velocities less than about 100°/s (see Fig. 9). The reason performance asymptotes below 50% (which is chance in this task) is that the spotlight model as stated above includes no mechanism for guessing; for example, it has no way of knowing when it has "lost" a target item, and therefore it has no basis for probability matching on those trials. A sophisticated guessing strategy like the one described in the previous section would keep the lower tails of these functions at or above 50%.

Figure 9 shows that a model incorporating an attentional spotlight moving at about 150–200°/s that can instantaneously find the nearest element to an attended location and enter its coordinates into the table can account in a qualitative fashion for the results reported here. However, there are several reasons for questioning this account. First, the velocity required to yield reasonable performance in this task is implausibly large.[12] Second, the assumption that searching for the nearest neighbor to a tabled location takes no time is certainly wrong and would add additional required velocity to yield acceptable performance, depending on how long the search is assumed to take. Finally, and most importantly, the model cannot account for the grouping effects observed in the present experiments.

Object–alignment and FINST models compared. The account proposed here, which involves the continuous alignment of an internal model of the configuration of target elements, differs significantly from the one provided by an unadorned reading of the FINST theory (Pylyshyn, 1989; Pylyshyn & Storm, 1988). For example, in the present account, an explicit representation of the configuration of elements is used in tracking; the FINST account holds that the target elements are only indexed, and as it stands the model provides no mechanism by which the spatial configuration of the elements can be used to enhance tracking performance.

[12] P. Cavanagh (personal communication, February 20, 1991) has compared observers' ability to track with attention an arm rotating about fixation to their ability to track the same arm with smooth pursuit eye movements. He found that the largest velocity that could be tracked with attention was about one-fifth the largest velocity that could be tracked with the eyes. Smooth pursuit eye movements are thought to have a maximum velocity of about 100°/s (Hallett, 1986). Although the attention movements required in Cavanagh's experiment are not precisely the same as those that would be required in this task (e.g., they involve tracking a single rotationally moving object), his estimates provide a reasonable upper bound on the velocity of attention movements. Furthermore, Pylyshyn and Storm (1988) conducted an informal meta-analysis of relevant studies, leading them to a rough estimate of 50°/s for the velocity of attention movements.

Nevertheless, the notion of a spatial index token is quite general and a relatively straightforward amendment to the FINST theory might accommodate the present results. For example, one could bind the theory's spatial index tokens to perceptual objects rather than to preattentively defined feature clusters. This would allow indexing to occur (at least optionally) *after* perceptual grouping and possibly after the application of attention.

CONCLUSION

The mechanism proposed in this article for multielement visual tracking is similar in some respects to Kahneman and Treisman's (1984; Kahneman et al., 1992) idea of an "object file." An object file is a temporary representation of a perceptual object that is present in the image. It contains information about the features of an object, as well as information about its location, time of appearance, motion, and any changes in its various attributes over time. The object file is the representational basis for visual selection.

What constitutes an object depends on perceptual grouping processes; in the current experiments, goal-directed grouping segregated target elements from nontarget elements and enabled the formation of an object file containing the changing target configuration. This conception of visual selection emphasizes the fundamental role of perceptual organization in generating object-based representations for attentional operations and other high-level visual tasks. Comprehensive theories of vision and visual attention will therefore require a satisfactory account of perceptual organization to be successful.

REFERENCES

Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research,* **14,** 589–592.

Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review,* **61,** 183–193.

Attneave, F. (1957). Physical determinants of the judged complexity of shapes. *Journal of Experimental Psychology,* **53,** 221–227.

Attneave, F. (1971). Multistability in perception. *Scientific American,* **225,** 62–71.

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin,* **53,** 452–471.

Ballard, D. H. (1984). Parameter nets. *Artificial Intelligence,* **22,** 235–267.

Banks, W., & Prinzmetal, W. (1976). Configurational effects in visual information processing. *Perception & Psychophysics,* **19,** 361–367.

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance,* **14,** 12–23.

Braunstein, M. L., Hoffman, D. D., & Pollick, F. E. (1990). Discriminating rigid from nonrigid motion: Minimum points and views. *Perception & Psychophysics,* **47,** 205–214.

Brecher, K. (1979). Sirius enigmas. In K. Brecher & M. Feirtag (Eds.), *Astronomy of the ancients* (pp. 91–115). Cambridge, MA: MIT Press.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review,* **97,** 523–547.

Bundesen, C., & Pedersen, L. F. (1983). Color segregation and visual search. *Perception & Psychophysics,* **33,** 487–493.

Burr, D. J. (1983). Matching elastic templates. In O. J. Braddick & A. C. Sleigh (Eds.), *Physical and biological processing of images* (pp. 260–270). Berlin: Springer-Verlag.

Casco, C., & Morgan, M. (1987). Detection of moving local density differences in dynamic random patterns by human observers. *Perception,* **16,** 711–717.

Casco, C., Morgan, M. J., & Ward, R. M. (1989). Spatial properties of mechanisms for detection of moving dot targets in dynamic visual noise. *Perception,* **18,** 285–291.

Cavanagh, P. (1990). Pursuing moving objects with attention. *Proceedings of the 12th Annual Meeting of the Cognitive Science Society, Boston* (pp. 1046–1047). Hillsdale, NJ: Erlbaum.

Chang, J. J., & Julesz, B. (1983). Displacement limits, directional anisotropy, and direction versus form discrimination in random dot cinematogram. *Vision Research,* **23,** 639–646.

Chong, C.-Y., Kahn, P., & Winkler, A. (1990). *Perceiver: Target modeling, representation, and perception* (Tech. Rep. ADS-TR3251-01). Mountain View, CA: Advanced Decision Systems.

Cohen, A., & Ivry, R. (1989). Illusory conjunctions inside and outside the focus of attention. *Journal of Experimental Psychology: Human Perception and Performance,* **15,** 650–663.

Dawson, M. R. W. (1991). The how and why of what went where in apparent motion: Modeling solutions to the motion correspondence problem. *Psychological Review,* **98,** 569–603.

Downing, C. J., & Pinker, S. (1985). The spatial structure of visual attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention & performance XI* (pp. 171–187). Hillsdale, NJ: Erlbaum.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General,* **113,** 501–517.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review,* **96,** 433–458.

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.

Erikson, C. W., & Murphy, T. D. (1987). Movement of attentional focus across the visual field: A critical look at the evidence. *Perception & Psychophysics,* **42,** 299–305.

Erikson, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics,* **40,** 225–240.

Fennema, C. L., & Thompson, W. B. (1979). Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing,* **9,** 301–315.

Folk, M. D., & Luce, R. D. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance,* **13,** 395–404.

Fox, J. (1978). Continuity, concealment, and visual attention. In G. Underwood (Ed.), *Strategies of information processing* (pp. 23–66). London: Academic Press.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts.* New York: Wiley.

Girgus, J. J., Rock, I., & Egatz, R. (1977). The effect of knowledge of reversibility on the reversibility of ambiguous figures. *Perception & Psychophysics,* **22,** 550–556.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 10.1–10.112). New York: Wiley.

Hochberg, J. (1974). Organization and the Gestalt tradition. In E. C. Carterette & M. C. Friedman (Eds.), *Handbook of perception* (Vol. 1, pp. 179–210). New York: Academic Press.

Hochberg, J. (1979). Sensation and perception. In E. Hearst (Ed.), *The first century of experimental psychology* (pp. 89–142). Hillsdale, NJ: Erlbaum.

Hochberg, J. E., & McAlister, E. (1953). A quantitative approach to figural "goodness." *Journal of Experimental Psychology,* **46,** 361–364.

Hoffman, J. E., & Nelson, B. (1981). Spatial selectivity in visual search. *Perception & Psychophysics,* **30,** 283–290.

Humphreys, G. W., Quinlan, P. T., & Riddoch, M. J. (1989). Grouping processes in visual search: Effects with single- and combined-feature targets. *Journal of Experimental Psychology: General,* **118,** 258–279.

Johansson, G. (1950). *Configurations in event perception.* Uppsala: Almqvist & Wiksell.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics,* **43,** 346–354.

Juola, J. F., Bouwhuis, D. G., Cooper, E. E., & Warner, C. B. (1991). Control of attention around the fovea. *Journal of Experimental Psychology: Human Perception and Performance,* **17,** 125–141.

Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Erlbaum.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention* (pp. 29–61). New York: Academic Press.

Kahneman, D., Treisman, A., & Gibbs, B. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology,* **24,** 175–219.

Kanwisher, N. G. (1991). Repetition blindness and illusory conjunctions: Errors in binding visual types with visual tokens. *Journal of Experimental Psychology: Human Perception and Performance,* **17,** 404–421.

Koffka, K. (1963). *Principles of Gestalt psychology.* New York: Harcourt, Brace, & World. [Original work published 1935]

Köhler, W. (1947). *Gestalt psychology* (revised edition). New York: Liveright. [Original work published 1929]

Kosslyn, S. M., Flynn, R. A., Amsterdam, J. B., & Wang, G. (1990). Components of high-level vision: A cognitive neuroscience analysis and accounts of neurological syndromes. *Cognition,* **34,** 203–277.

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

Kramer, A. F., & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception & Psychophysics,* **50,** 267–284.

Kubovy, M. (1981). Concurrent pitch segregation and the theory of indispensable attributes. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 55–98). Hillsdale, NJ: Erlbaum.

LaBerge, D. (1983). The spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance,* **9,** 371–379.

LaBerge, D., & Brown, V. (1989). Theory of attentional operations in shape identification. *Psychological Review,* **96,** 101–124.

Lowe, D. G. (1985). *Perceptual organization and visual recognition.* Boston: Kluwer Academic Publishers.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence,* **31,** 355–395.

Mahoney, J. V., & Ullman, S. (1988). Image chunking defining spatial building blocks for scene analysis. In Z. Pylyshyn (Ed.), *Conceptual processes in human vision* (pp. 169–209). Norwood, NJ: Ablex.

Marr, D. (1982). *Vision.* San Francisco: W. H. Freeman & Co.

McLeod, P., Driver, J., Dienes, Z., & Crisp, J. (1991). Filtering by movement in visual search. *Journal of Experimental Psychology: Human Perception and Performance,* **17,** 55–64.

Mohan, R., & Nevantia, R. (1989). Using perceptual organization to extract 3-D structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **11,** 1121–1139.

Moore, P. (Ed.). *The international encyclopedia of astronomy.* New York: Orion Books.

Moraglia, G. (1989). Display organization and the detection of horizontal line segments. *Perception & Psychophysics,* **45,** 265–272.

Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology,* **7,** 480–494.

Olson, R., & Attneave, F. (1970). What variables produce similarity grouping? *American Journal of Psychology,* **83,** 1–21.

Pachella, R. G. (1975). The effect of set on the tachistoscopic recognition of pictures. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 136–156). New York: Academic Press.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology,* **9,** 441–474.

Palmer, S. E. (1983). The psychology of perceptual organization: A transformational approach. In J. Beck, B. Hope, & A. Rosenfeld (Eds.), *Human and machine vision* (pp. 269–339). New York: Academic Press.

Pannekoek, A. (1961). *A history of astronomy.* London: George Allen & Unwin.

Peterson, M. A., & Gibson, B. S. (1991). Directing spatial attention within an object: Altering the functional equivalence of shape descriptions. *Journal of Experimental Psychology: Human Perception and Performance,* **17,** 170–182.

Peterson, M. A., & Hochberg, J. (1983). Opposed-set measurement procedure: A quantitative analysis of the role of local cues and intention in form perception. *Journal of Experimental Psychology: Human Perception and Performance,* **9,** 183–193.

Podgorny, P., & Shepard, R. N. (1983). Distribution of visual attention over space. *Journal of Experimental Psychology: Human Perception and Performance,* **9,** 380–393.

Pomerantz, J. R. (1981). Perceptual organization in information processing. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization* (pp. 141–180). Hillsdale, NJ: Erlbaum.

Pomerantz, J. R., & Kubovy, M. (1986). Theoretical approaches to perceptual organization. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 2, pp. 36.1–36.46). New York: Wiley.

Pomerantz, J. R., & Pristach, E. A. (1989). Emergent features, attention, and perceptual glue in visual form perception. *Journal of Experimental Psychology: Human Perception and Performance,* **15,** 635–649.

Pomerantz, J. R., & Schwaitzberg, S. D. (1975). Grouping by proximity: Selective attention measures. *Perception & Psychophysics,* **18,** 355–361.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology,* **32,** 3–25.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General,* **109,** 160–174.

Prinzmetal, W. (1981). Principles of feature integration in visual perception. *Perception & Psychophysics,* **30,** 330–340.

Prinzmetal, W., & Keysar, B. (1989). Functional theory of illusory conjunctions and neon colors. *Journal of Experimental Psychology: General,* **118,** 165–190.

Pylyshyn, Z. W. (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 multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision,* **3,** 179–197.

Rock, I., & Gutman, D. (1981). The effect of inattention on form perception. *Journal of Experimental Psychology: Human Perception and Performance,* **7,** 275–285.

Rosenthal, R., & Rosnow, R. L. (1985). *Contrast analysis.* Cambridge: Cambridge University Press.

Schafer, E. H. (1978). *Pacing the void: T'ang approaches to the stars.* Berkeley: University of California Press.

Sedgewick, R. (1983). *Algorithms.* Reading, MA: Addison–Wesley.

Sekuler, A. B. (1990). Motion segregation from speed differences: Evidence for nonlinear processing. *Vision Research,* **30,** 785–795.

Sethi, I. K., Salari, V., & Vemuri, S. (1988). Feature point matching in image sequences. *Pattern Recognition Letters,* **7,** 113–121.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science,* **171,** 701–703.

Shulman, G. L., Remington, R., & McLean, J. P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance,* **5,** 522–526.

Steinfeld, G. (1967). Concepts of set and availability and their relation to the recognition of ambiguous pictorial stimuli. *Psychological Review,* **74,** 505–522.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology,* **21,** 233–282.

Tipper, S. P., Brehaut, J. C., & Driver, J. (1990). Selection of moving and static objects for the control of spatially directed action. *Journal of Experimental Psychology: Human Perception and Performance,* **16,** 492–504.

Tipper, S. P., Driver, J., Weaver, B. (1991). Object-centred inhibition of return of visual attention. *Quarterly Journal of Experimental Psychology,* **43A,** 289–298.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance,* **8,** 194–214.

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

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

Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance,* **9,** 523–530.

Tsotsos, J. K. (1988). A 'complexity level' analysis of immediate vision. *International Journal of Computer Vision*, **1**, 303–320.

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

Ullman, S. (1984a). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, **13**, 255–274.

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

Ullman, S. (1989). Aligning pictorial descriptions: An approach to object recognition. *Cognition*, **32**, 193–254.

Watt, R. J. (1988). *Visual processing: Computational, psychophysical, and cognitive research*. London: Lawrence Erlbaum Associates, Ltd.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift fur Psychologie*, **61**, 161–265.

Yantis, S. (1988). On analog movements of visual attention. *Perception & Psychophysics*, **43**, 203–206.

Yantis, S., & Johnston, J. C. (1990). On the locus of visual selection: Evidence from focussed attention tasks. *Journal of Experimental Psychology: Human Perception and Performance*, **16**, 135–149.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 601–621.

Yuille, A. L. (1991). Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, **3**, 59–70.

Zucker, S. W. (1987). The diversity of perceptual grouping. In M. Arbib and A. Hanson (Eds.), *Vision, brain, and cooperative computation*. Cambridge, MA: MIT Press.

(Accepted November 22, 1991)

COGNITIVE PSYCHOLOGY **24**, 341–380 (1992)

Categorization Using Chains of Examples

EVAN HEIT

Stanford University

People can infer unknown features of a stimulus by retrieving memories of similar examples. It is proposed that we can reason from *chains* of examples. For example, stimulus *A* may remind us of *B*, which reminds us of *C*. Information about *C* may then affect reasoning about *A*. A mathematical model for categorization (extended from the context model of Medin & Schaffer, 1978), using multiple-step chains of reasoning and memory for examples, is presented. In five experiments, subjects memorized feature descriptions of fictional people, then made predictions from incomplete descriptions. Various predictions could be made using one-, two-, or three-step chains of reasoning. These experiments varied in terms of stimulus structure, complexity of test questions, and response method (probability estimate or forced choice). The multiple-step context model, with the assumption that people performed one- and two-step chains of inference, successfully accounted for the results of all five experiments.

INTRODUCTION

Inferring Any Feature of a Stimulus

Psychological research on categories has had a limitation. While people can make many inferences about a given stimulus, experiments have typically investigated inferences only about one specific aspect of the stimulus: the designated category label. In most artificial category learning experiments, subjects are only tested by having them provide the label for an unlabeled stimulus.

When we encounter some animal, we might try to infer the category label for its species. But when we encounter a large, growling creature in the dark, we might be more interested in inferring other properties of this creature, such as whether it will attack and how fast it can run, rather than naming it. From a statistical viewpoint, there is no difference between

This research was supported by a National Science Foundation Graduate Fellowship and a dissertation grant from Stanford University. This paper is based on a dissertation submitted to Stanford. Experiments 1 and 2 were presented at the 1990 Meeting of the Psychonomic Society in New Orleans. I am grateful to G. Bower, W. Ahn, J. R. Anderson, J. Greeno, P. Johnson-Laird, E. Markman, D. Medin, M. Pavel, L. Rips, R. Shepard, E. E. Smith, and E. Wisniewski for comments on earlier drafts. Address correspondence to Evan Heit, Department of Psychology, University of Michigan, 330 Packard Road, Ann Arbor, MI 48104, or Evan.Heit@um.cc.umich.edu.

0010-0285/92 $9.00