COGNITIVE SCIENCE

Cognitive Science 25 (2001) 3–36

http://www.elsevier.com/locate/cogsci

Comparative visual search: a difference that makes a difference

Marc Pomplun[a,*], Lorenz Sichelschmidt[b], Karin Wagner[c], Thomas Clermont[d], Gert Rickheit[b], Helge Ritter[d]

[a] *University of Toronto, Department of Psychology, 100 St. George Street, Toronto, Ontario, Canada M5S 3G3*
[b] *Faculty of Linguistics, University of Bielefeld, P.O. Box 10 01 31, 33501 Bielefeld, Germany*
[c] *Faculty of Psychology, University of Bielefeld, P.O. Box 10 01 31, 33501 Bielefeld, Germany*
[d] *Faculty of Technology, University of Bielefeld, P.O. Box 10 01 31, 33501 Bielefeld, Germany*

Abstract

In this article we present a new experimental paradigm: comparative visual search. Each half of a display contains simple geometrical objects of three different colors and forms. The two display-halves are identical except for one object mismatched in either color or form. The subject's task is to find this mismatch. We illustrate the potential of this paradigm for investigating the underlying complex processes of perception and cognition by means of an eye-tracking study. Three possible search strategies are outlined, discussed, and reexamined on the basis of experimental results. Each strategy is characterized by the way it partitions the field of objects into "chunks." These strategies are: (i) Stimulus-wise scanning with minimization of total scan path length (a "traveling salesman" strategy), (ii) scanning of the objects in fixed-size areas (a "searchlight" strategy), and (iii) scanning of object sets based on variably sized clusters defined by object density and heterogeneity (a "clustering" strategy). To elucidate the processes underlying comparative visual search, we introduce besides object density a new entropy-based measure for object heterogeneity. The effects of local density and entropy on several basic and derived eye-movement variables clearly rule out the traveling salesman strategy, but are most compatible with the clustering strategy.

* Corresponding author. Tel. +1-416-978-3990; fax: +1-416-926-0921.
E-mail address: marc@psych.utoronto.ca (M. Pomplun).

PII: S0364-0213(00)00037-9

1. Introduction

1.1. Visual search

To date, attempts at investigating the cognitive processes involved in visual search have led to an immensely large number of empirical studies (cf. Brogan, Gale & Carr, 1993). The standard task for subjects in visual search experiments is to state whether or not a display contains a designated item that differs from others with respect to a single dimension such as color or shape (*feature search*) or several dimensions such as color and shape (*conjunctive search*). With unlimited exposure time, the data that usually accrue from this sort of task are error rates and, more importantly, reaction times—the times subjects typically take to respond "yes" when a target is present or "no" when it is absent.

Generally, searches of visual displays are held to embrace an ensemble of perceptive processes which expose parallel as well as serial characteristics and which involve dynamic shifts of attention. The key observation is that with feature search, reaction time is usually independent of the size of the item set while with conjunctive search, reaction time increases linearly with the number of distractor items in the display.

This pattern of findings has been taken as suggesting two modes of visual processing which correspond to consecutive stages: an initial, preattentive stage during which many items are analyzed in parallel with respect to only few specific features, and a subsequent stage during which the set of items is analyzed sequentially with respect to feature combinations. Two relevant theories are the "feature inhibition theory" (Treisman & Sato, 1990), which maintains that during early processing likely distractor items are eliminated from the search set by means of inhibition based on the analysis of single features, and the "guided search model" (Wolfe, Cave & Franzel, 1989), which maintains that during early processing a likely target region or likely target items are activated a priori on the basis of feature similarity.

Yet another theory emphasizes the role of perceptual grouping in visual search. Following Pashler (1987), who suggests that conjunctive search proceeds on a cluster-by-cluster basis, processing within clusters of items being exhaustive and parallel but processing between clusters being sequential, this approach (Duncan & Humphreys, 1989, 1992) maintains that visual search is guided by the global similarities that hold within the item set.

More recently, several studies have gone beyond reaction time measurement. They have analyzed eye movements in visual search, with results that corroborate the guided search model. They found a strong bias towards fixating on items that share certain features (especially color) with the target, both in humans (e.g., Williams & Reingold, submitted; Scialfa & Joffe, 1998) and monkeys (Motter & Belky, 1998). Also, some current models of visual search are based on oculomotor data rather than on reaction times (e.g., Rao & Ballard, 1995). These and other studies demonstrate that eye-movement measurement is eminently suited for investigating visual perception. Making eye movements while viewing a scene is absolutely natural, and modern equipment which allows subjects to move their head relatively freely does in no way perturb normal viewing (Kowler, 1990).

Eye-movement measurement has major advantages over reaction time measurement. First, eye-tracking systems yield information not only about the duration but about the time course of search. Second, temporal data are supplemented by spatial ones, thus enabling researchers to reconstruct subjects' gaze trajectories in detail. Third, the abundance of spatiotemporal data allows for more exact empirical validation and more precise modeling of vision than does reaction time measurement. After all, eye movements may be an even more direct overt manifestation of cognitive processes in vision than are latencies. Low-level factors such as item size or item density as well as high-level factors such as item gestalt or item function influence both the length of saccades and the duration of fixations (see Rayner, 1998, for a review). According to findings in reading research (Just & Carpenter, 1987), the total fixation time spent on any single item or item cluster can generally be considered a valid measure of the processing time for that particular object.

Despite the promising results of eye-movement recording in visual search, we believe that adherence to the standard visual search paradigm could lead to unnecessary and inappropriate limitations of the research endeavor. Standard visual search tasks usually require only a small number of eye movements; large parts of the display can be processed within a single fixation. Therefore, gaze trajectories yield only coarse information about a subject's strategy to find the target. It is virtually impossible, for instance, to test the hypothesized distinction between two successive stages in visual search (see above) on the basis of eye movements alone.

Moreover, not all of the important components of visual perception are involved in a standard visual search task. Perhaps the most important element that is neglected is *memory*. In visual search, people have to keep in memory some representation of a particular object in order to be able to identify the target item among non-target items. The standard visual search paradigm thus mirrors a situation in which somebody is looking for a well-known object. The memory structures that are relevant in such situations must include information about the designated target object and its features plus information about the actual scene. Unfortunately, the standard visual search paradigm is not particularly well suited to deepen our understanding of the role of memory in visual perception since the designated target is hardly ever systematically varied within experiments.

In everyday life there are other—not necessarily less common—situations that require purposeful search. A common example of the kind of visual search addressed here is the "original and fake" kind of picture-picture comparisons occasionally found in magazines. Such matching tasks resemble many real-world situations that arise from a discrepancy between an actual state and a target state. Construction, for instance, requires that the current situation continually be compared to (and adjusted according to) the goal. While the standard visual search task requires the viewer to hold in mind a representation of a single target item, a matching task requires the viewer to keep two structured sets of items in memory at once. Consequently, the representations that are functionally relevant in matching tasks are rather complex since both the actual and the target scenes have to be memorized for comparison. Below, we shall show that a different variant of visual search can make the use of these representations transparent.

1.2. A new experimental paradigm: comparative visual search

In light of the above considerations, we would like to introduce a new experimental paradigm which may yield informative data about visual search and hopefully give insight into the underlying cognitive processes.

The paradigm involves tracking the eye movements of subjects performing a matching task. The task is to detect the only difference between two halves (hemifields) of a display. Each half of the display contains a number of simple geometrical objects such as colored squares, circles, or triangles. The right display half is a translated copy of the left one; it is identical to the left one with respect to object number, object position, object form, and object color—except for one object which differs from the corresponding object in the other half either in form or in color. The subjects' task is to search the display, comparing the halves in order to detect the mismatch. On identifying the target object, they are to press a mouse button.

Comparative visual search follows the tradition of *picture matching* (Humphrey & Lupker, 1993). In picture matching experiments, subjects are typically shown pairs of pictures (presented either simultaneously or sequentially) and asked to indicate whether or not they feature the same object. The "classic" findings from picture matching studies are, among others, that "yes" answers are faster with identical views of an object than with different views (Kelter, Grötzbach, Freiheit, Höhle, Wutzig & Diesch, 1984; Klatzky & Stoy, 1974), that response times are a sinusoid function of the difference in terms of depicted angle, thus corroborating the notion of mental rotation (Rock, Wheeler & Tudor, 1989; Shepard & Cooper, 1982), and that for some objects "canonical views" such as frontal or profile view are particularly easy to process (Cooper, Biederman & Hummel, 1992; Palmer, Rosch & Chase, 1981). Recently, a large number of studies have used the sequential variant of picture matching to investigate the phenomenon of *change blindness* (e.g., Rensink, O'Regan & Clark, 1996; Simons & Levin, 1997). Subjects have been found to miss even large changes between successively presented, real-world images, demonstrating a surprisingly limited visual memory.

These findings raise two questions which also pertain to comparative visual search. First, there is the question of *representational format*. How must objects be represented in memory to enable people to recognize them in pictures? The solutions that have been proposed range from the notion of a single object-centered format (Marr, 1982) to a view-specific type of representation (Tarr & Pinker, 1989). Still other approaches advocate multiple representations that comprise view-specific as well as view-invariant plus semantic-conceptual formats (Ellis, Allport, Humphreys & Collis, 1989). It makes sense to postulate multiple representational formats, all of which can be transformed into each other and all of which can be functional according to the task and the objects in question. For the recognition of familiar objects, for instance, a view-invariant abstract representation may suffice; for a picture matching task, a view-specific format may be more appropriate (cf. Lawson & Humphreys, 1996).

Second, there is the question of the *cognitive processes* involved in picture matching. Which processes must people perform to recognize pictures as different views of the same object? Basically, picture matching encompasses the comparison of object sets that are more

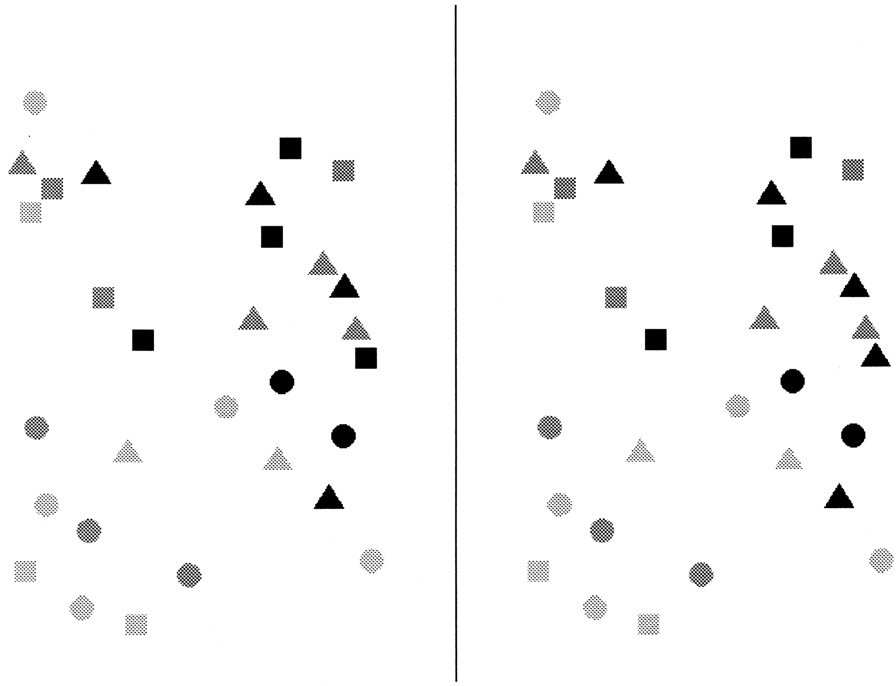

Fig. 1. Example of a randomly generated stimulus display.

or less similar. According to recent research, comparison involves an alignment of global structures; a process which yields overall commonalities, differences related to the commonalities, and differences not related to the commonalities. In particular commonalities and the differences related to them appear to be crucial to similarity judgments (Markman & Gentner, 1996). Comparison at the global level facilitates the determination of both matching and mismatching features of corresponding objects. Goldstone (1994), for instance, has observed that subjects who saw scenes that were composed of two pairs of butterflies were more sensitive to matching features such as wing shade when there were certain global alignments such as the arrangement of the butterflies than when there were not.

In the context of the comparative visual search paradigm, the objects we use are abstract with two low-level features, namely color and form. Fig. 1 shows an example stimulus. Since the spatial arrangement of objects is important in comparative visual search, and since the stimuli have no semantic content, it is unlikely that abstract, view-invariant representations play a role in comparative visual search. With respect to comparison, the prerequisites for an efficient comparative search are met because the spatial arrangement of the objects in both hemifields of the display are identical. However, since the number of objects in each hemifield is relatively large and visual working memory is limited (e.g., Simons & Levin, 1997), comparative search must proceed in several steps. Accordingly, saccades within one hemifield should be distinguishable from longer saccades between hemifields. Also, both response time and oculomotor parameters should, to some extent, depend on the global and local structure of the display, in particular on the attributes that hold for the fixation point in question.

Let us consider three basic strategies that subjects might apply. These global strategies are extreme in that they are not likely to be found in their pure form; however, they may serve to illustrate the range of possibilities and provide three useful points of reference for a discussion.

2. Candidate strategies for comparative visual search

2.1. The traveling salesman strategy

A first strategy could be termed the "traveling salesman" strategy. Its name is derived from the traveling salesman problem (TSP), which is a well-known paradigm in computer science: Starting from his home, a salesman has to visit a number of certain cities before he can return. Of course he wants to save time and energy, so he tries to find the shortest roundtrip that is possible. If subjects use an analogous strategy for their scan paths, they would be expected to prefer object-to-object paths of minimal overall length.

In comparative search, however, the situation is different from the standard TSP: First, subjects do not need to return to the starting point after having scanned all of the objects in the display. Second, the task of comparing corresponding objects influences subjects' strategy. Due to the limited capacity of working memory it is not possible to first memorize all information given in one of the hemifields and then compare it to the other hemifield. Instead, subjects have to switch between the hemifields during the search process and thus deviate from the "optimal" scan path in terms of the TSP. Hence, there is a trade-off between memory usage and scan path minimization. Let us assume—for the benefit of simplicity—that subjects memorize only one object at a time. What could we expect their scan paths to look like?

The first object is memorized in hemifield A, and then compared to its counterpart in hemifield B. In order to minimize the length of their scan paths, subjects are likely to memorize the next object in hemifield B rather than to switch back to hemifield A and memorize the next object there. This means that after the memorization of the first object, subjects are likely to process successively two objects in the same hemifield: The object stored before is verified and a new one is stored.

While the search process is likely to start at the top of the display, the problem for the viewers is to find the optimal scan path—the one that touches every object pair with the shortest possible saccades. On the assumption that two fixations fall to each object pair, the average number of fixations as well as fixation duration should be constant, and there should always be *two* successive fixations within the same hemifield.

2.2. The searchlight strategy

Another strategy could be to scan the display in a "searchlight manner": Given a foveal area which is fixed in size, each fixation would cover a certain subregion of the display.

While the size of the area covered with each fixation would be constant, the number of objects covered would depend on the density of the objects in that particular subregion. Thus, topology now determines which particular objects are processed during any fixation. If the size of the area covered with each fixation were decreased, the searchlight strategy would eventually turn into the traveling salesman strategy; on the other hand, if the size of the area were considered variable, the searchlight strategy would become indiscriminable from clustering, a further candidate strategy discussed below. The problem here is a spatial one—to scan the display with a minimal number of fixations so that every object up to the target object is visited at least once (and, at best, only once), while keeping saccades as short as possible. Similar to the traveling salesman strategy, two successive fixations per "visit" to each hemifield would be expected. With the searchlight strategy, the spatial distribution of objects in the display should be an important determinant of fixation duration and saccade length; one might expect longer fixations on subareas with a high object density, which is also where saccades should be preferentially aimed at.

2.3. The clustering strategy

A third strategy could be to proceed cluster by cluster: With each fixation, subjects might process a certain number of objects. These clusters would be assumed to be the maximal subset of objects next to the fixation point which can be processed below a specific "effort threshold." It is plausible to assume such a threshold, because working memory and visual attention limit the subjects' capacity of processing. How should the effort of processing and its threshold be defined? We can consider at least three extreme cases:

(1) Definition by number: A cluster consists of the k objects closest to the fixation point.
(2) Definition by distance: A cluster consists of all objects within a circle of radius r around the fixation point.
(3) Definition by object attributes: A cluster consists of those objects that have the same color and form as the one next to the fixation point.

While definition (1) accounts for the limited capacity of working memory, it ignores the objects' features, i.e. objects of the same color or form could be processed with the same effort as objects of mixed colors and forms, which is rather implausible. Furthermore, the distances between the objects are not considered; for example, two objects with a distance of 15 degrees of visual angle between them could be processed as efficiently as two objects with a distance of 2 degrees between them. Definition (2) is more adequate with respect to object eccentricity, however, it does not consider any parameters of working memory. This definition would turn the clustering strategy into the searchlight strategy. Finally, definition (3) takes into account an influence of object features on processing effort. It is inadequate, however, in its representation of memory capacity: Any number of objects could be memorized at the same time.

Since none of the above criteria is sufficient by itself to define a plausible threshold for the effort of processing, it seems reasonable to *combine* the criteria: The effort should increase with the number of objects, with their eccentricity, and with the entropy of their attributes. If the local object features enable the use of a large cluster, one or more within-hemifield saccades may be employed during its memorization and comparison.

With such a clustering strategy, one would expect the average number of fixations to be smaller and fixation duration to be longer than with the traveling salesman strategy; to some extent, fixation behavior should depend on the information content, or entropy, of the objects in a cluster. Saccades should be relatively short since the majority of saccades can occur within hemifields. Again, in order to avoid unnecessary saccades, subjects are supposed to perform successively verification and memorization of clusters within the same hemifield.

2.4. Strategy issues

All three candidate strategies can be characterized in terms of a minimization of a cost function. While the traveling salesman strategy can be viewed as an attempt to optimize search by keeping the overall length of the scan path to a minimum, the searchlight strategy can be taken as an attempt to optimize search by minimizing the total number of fixations in an exhaustive scan of relevant subregions while keeping constant the area covered with each fixation. In contrast, the clustering strategy can be viewed as an attempt to optimally exploit the capacity of working memory, achieved by grouping the objects in such a way that memorizing clusters can proceed with as little effort as possible. In face of the limited processing capacity, the objective behind the clustering strategy is to subdivide the objects into as few clusters as necessary, which comprise as many objects as possible that are maximally alike. So, the economical principles featured in the three strategies are different.

A clear difference lies in the segmentation of the set of objects to be compared. First of all, the strategies differ in the amount of information processed with each fixation. The number of objects covered per fixation is one for the traveling salesman strategy. In contrast, both the searchlight strategy and the clustering strategy maintain that with each fixation a variable number of objects is processed. According to the searchlight strategy, that number is determined by the object density at the fixation point since the size of the area covered is thought to be fixed. According to the clustering strategy, however, the number is determined within the limits of fixed-capacity working memory by some capacity oriented variables such as the variance, or entropy, of objects in terms of location, color, and form.

In addition, there are qualitative differences between the strategies. In order to keep the length of the scan path to a minimum, it would suffice to roughly analyze the spatial layout of the objects to determine their location. Thus, information about the spatial location of the objects is necessary and sufficient for calculating an optimal scan path in accordance with the traveling salesman strategy. Employing the searchlight strategy, however, is more exacting. The searchlight strategy presupposes that subregions have been coarsely analyzed as to the density of objects, in order to fixate on subregions with many objects by preference and to avoid fixating on empty ones. Thus, the information necessary and sufficient for the searchlight strategy is spatial location plus local density. An even more comprehensive

Table 1
Hypotheses about expected values and factors for the candidate strategies (by dependent variables)

Variable	"Traveling Salesman"	"Searchlight"	"Clustering"
Number of fixations	$=n$ (objects per hemifield)	$< n$	$< n$
Fixation duration	constant	f(density)	f(density, entropy)
Saccade length	f(density)	f(density)	f(density, entropy)
Successive fixations	$=2$	$=2$	>2; f(density, entropy)

analysis is necessary to license the clustering strategy. A cluster can be defined as a group of objects that is similar with respect to dimensions such as location, color, or form. It follows that various feature dimensions of the objects must be analyzed in conjunction in order to achieve perceptual grouping. The clustering strategy must take into account not only where the objects are located and how far they are apart (density) but also whether or not neighboring objects are alike in color or form (entropy). This raises the question of to what extent we also have to consider the costs for performing the cost minimization itself.

From the above considerations it follows that the traveling salesman strategy could be characterized by the occurrence of approximately as many fixations as there are objects in each hemifield: On average, a mismatch can be detected after scanning 50% of the object pairs (disregarding any extra fixations due to detection failures or ascertaining). In contrast, both the searchlight strategy and the clustering strategy should take fewer fixations since more than one object can be processed with each fixation. A summary of these predictions is given in the first row of Table 1.

Predictions can also be made for fixation duration. In the case of the traveling salesman strategy, the duration of fixations should be constant. In the case of the searchlight strategy, fixation duration should depend on spatial location, that is, on object density. In the case of the clustering strategy, object density should also have an effect, in addition to variance in object features like color or form (entropy). Since it only makes sense to talk of variance when a fixation covers more than one object, we expect an interaction of object density and entropy. The second row of Table 1 provides a summary of these expectations.

Saccade length should be a function of the spatial layout of the display, that is, of the local density of the objects, for any of the three candidate strategies. For the clustering strategy, saccade length could be affected by interactions. As a cue to clustering, similarity in location (object density) may be diminished by variance in object features (entropy). These expectations are summarized in the third row of Table 1.

Finally, the number of fixations that occur in succession before changing to the other hemifield should be higher with a clustering strategy than with any of the other search strategies. This is because, if the local parameters (low entropy and density) allow the subjects to process large clusters, the information gathered during several consecutive fixations might be accumulated in working memory before proceeding to comparison. In addition, the factors that determine fixation duration should also determine the number of successive fixations within the same hemifield with a clustering strategy. A summary is given in the last row of Table 1.

3. Method

3.1. Subjects

The subjects ($N = 16$) were students of various fields at the University of Bielefeld. They were paid for their participation. All subjects had normal or corrected-to-normal vision; none had pupil anomalies, and none were color blind.

3.2. Materials

The stimulus displays were presented on a computer screen with a resolution of 640×480 pixels. The displays showed patterns of simple geometrical objects on a black background. These objects appeared in three different forms (triangles, squares, and circles) and three different colors (fully saturated blue, green, and yellow). The objects were about 0.6° of visual angle in diameter. Color brightness was adjusted so that the objects were approximately equally luminous as indicated by pre-tests. The object locations were randomly generated with a constraint prohibiting object contiguity or overlap. The random distribution was chosen such that it slightly tended to create regions of similar objects in order to make it possible to study eye movements within homogeneous vs. heterogeneous object distributions (cf. Appendix A).

Each stimulus display was 24° wide and 16° high and vertically bisected into two hemifields with 30 objects each. The objects in each half were equally balanced for color and form. The hemifields were translationally identical in the color, form and spatial distribution of the 30 objects—with one exception: One object differed from its "twin" in the other hemifield, either in color or in form.

3.3. Apparatus

Eye movements during comparative visual search were measured with the OMNITRACK 1 system (Stampe, 1993). OMNITRACK 1 is a non-invasive imaging eye tracking system consisting of a 486DX2-66 type computer equipped with image processing hardware and software plus a headset. Subjects were seated about 60 cm away from a 17″ color monitor. They wore a headset equipped with two miniature infra-red video cameras which yielded real-time information about pupil and head movements, so that any head movements during viewing did not impair the accuracy of eye-movement measurement.

From the camera data, actual fixation points on the screen were calculated at a frame rate of 60 Hz. Only fixations that lasted for at least five frames (i.e. 83 ms) were registered. The absolute precision of the system lay within 0.7 to 1.0 degrees of visual angle, but with the help of a neural network interface, the error was reduced to 0.5 degrees, corresponding to about 0.6 cm or 12 pixels on the screen (Pomplun, Velichkovsky & Ritter, 1994).

Prior to experimentation, a calibration procedure was performed by making the subject fixate on specified points on the screen. Also, the system had to be recalibrated from time to time to compensate for the sliding of the head set due to subjects' movements.

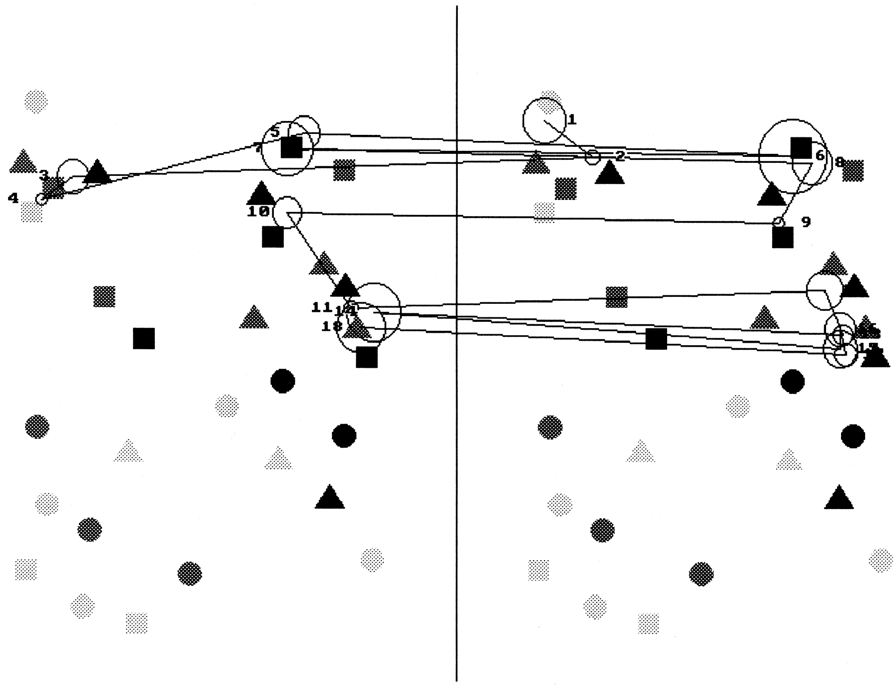

Fig. 2. Example display of Fig. 1 with the plotted visual scan path chosen by one of the subjects. Fixations are numbered; circle size signifies fixation duration.

3.4. Procedure

Subjects were tested individually. Their task was to find the only difference between the two halves of each display. They were to press a mouse button placed in front of them as soon as they had detected the mismatch.

Each subject viewed 50 stimuli. After every tenth stimulus, the eye tracker was recalibrated. The stimuli were newly generated for every subject so that none of the patterns occurred twice. Subjects knew that the critical mismatch would be either in color or in form; they did not know, however, when to expect which kind of mismatch. 25 of the 50 trials had the difference in color and 25 had the difference in form. If the experimenter noticed that subjects did not find one or more targets, indicated by a discrepancy between gaze and target position during their manual reaction, additional trials were appended to replace the incorrect ones. This happened in fewer than 2% of the trials.

4. Results

Technically speaking, the design of the study was a factorial with repeated measures on all independent variables. Independent variables were the type of mismatch (color vs. form) and three local stimulus parameters for each fixation point registered: Object density $\varrho(\mathbf{p})$, color entropy $S_c(\mathbf{p})$, and form entropy $S_f(\mathbf{p})$ at the point **p** in question. An important feature of these local parameters is that they are completely uncorrelated. That is, the object density in a particular area does not influence the probability of finding high or low color entropy or of finding high or low form entropy, at that same location. Details concerning the definition and calculation of independent variables can be found in Appendix B.

Furthermore, motivated by the eye-movement patterns observed in pre-tests, a distinction between two successive phases in comparative search was introduced: During a first phase, termed *search and comparison*, people appear to perform a quick scan in order to efficiently locate the mismatch. On encountering a "suspicious" region, a second phase, *detection and verification*, is initiated. The transition from the first to the second phase is operationally defined to occur when the following two conditions are met: (a) the subject's gaze position gets closer than 50 pixels—corresponding to approximately two degrees of visual angle—to any of the target objects and (b) a manual reaction is registered within the next two seconds.

This definition was derived from the qualitative observation that subjects' eye-movement patterns change in a distinctive way after the detection of a possible target. It seems to take subjects about one to three seconds to verify their suspicion and to press the mouse button. There might be more accurate ways to define the transition between the phases, possibly by splitting the search process into an even more detailed series of stages. Lacking such improved measures, we have to rely on the current operational definition to show two things. First, that search is done in phases at all, and second, that these phases can provide a basis for a useful analysis.

The current study analyzed the two phases individually, in order to separate search and comparison processes from the cognitive processes involved in ascertaining a mismatch.

Dependent variables fall into two categories. Basic dependent variables were measures which are commonly obtained in eye-movement based visual search studies: reaction time (RT), number of fixations (NF), fixation duration (FD), and saccade length (SL). Derived dependent variables were tailored to the specific purposes of this study. These were measures that might prove essential for an understanding of comparative search. They were: number of successive fixations within the same hemifield (FW), probability of missing the target (PM), and area coverage per fixation (AC). For details concerning the calculation of dependent variables see Appendix C.

The data recorded during the search and comparison phase were subjected to analyses of variance. Repeated measures analyses were performed using conservative Huynh and Feldt adjustment of degrees of freedom, and multiple planned comparisons were done by simple effects analyses plus conservative Bonferroni-adjusted means tests if so required. Level cut-points for density and entropy were set at the lower and upper third marks (low vs. medium vs. high). The density range was split at the values 1.0 and 2.0, whereas the entropy scale had cutpoints at 0.6 and 0.85. The global variables RT and NF could only be related to the factor "type of mismatch"; thus they were each entered into a one-way analysis of variance. Each of the variables FD, SL, FW, and AC was entered into a four-way analysis of variance (type of mismatch, object density, color entropy, and form entropy). The results showed that the factor "type of mismatch" had no significant effect on any of these variables, which is not surprising because subjects did not know the type of difference before detecting

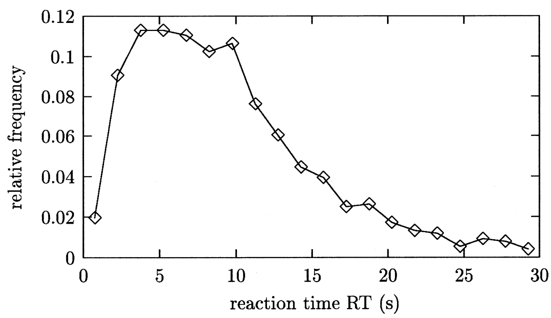

Fig. 3. Histogram of reaction times.

it. Therefore, data were collapsed over "type of mismatch", reducing the proportion of missing data to a value of about 1%.

The effect of phases of processing on NF, FD, SL, and FW was tested by performing a series of one-way analyses of variance, because the amount of data from the detection and verification phase was not sufficient to include any other factors. Additionally, the data from the detection and verification phase were entered into an analysis of variance with the factor "type of mismatch." The analysis of PM needed a special design which will be explained below.

4.1. Basic dependent variables

4.1.1. Reaction time

On average, subjects needed 10 950 ms to process a display. A histogram of reaction times in the experiment is shown in Fig. 3; the most remarkable feature is a plateau of short reaction times between three and ten seconds. Differences in color were detected faster (9 903 ms) than differences in form (11 997 ms) ($F(1;\ 15) = 8.66$; $p = 0.010$). Since RT is a global measure, it is not possible to test for the effects of local parameters. Also, the distinction between the two phases does not make sense in an RT analysis, because the latter phase is defined on the basis of RT.

4.1.2. Number of fixations

Averaged over all subjects and both types of mismatch, subjects needed 39.6 fixations to finish a trial. Processing a display that had the difference in color generally took fewer fixations (35.7) than processing a display that had the difference in form (43.6) ($F(1;\ 15) = 9.07$; $p = 0.009$). Similar to RT, NF cannot be related to any local stimulus parameters.

Unlike RT, NF was split into fixations during the search and comparison phase (NF_s) and fixations during the detection and verification phase (NF_v). Search and comparison required an average of 35.2 fixations. When the difference was in color, subjects managed with fewer fixations (31.3) than when the difference was in form (39.1) ($F(1;\ 15) = 8.39$; $p = 0.011$). In contrast, detection and verification required an average of 4.5 fixations, regardless of the type of difference ($F(1;\ 15) = 0.07$; $p = 0.788$).

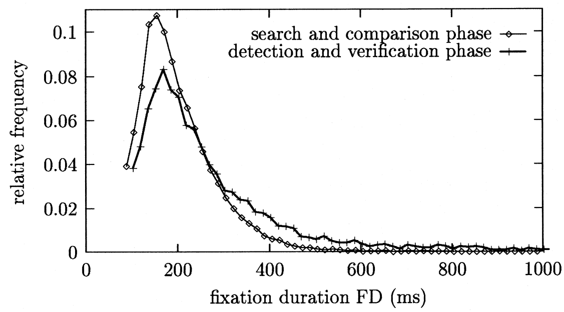

Fig. 4. Histogram of fixation durations.

4.1.3. Fixation duration

In the comparative visual search task, fixations had a mean duration (FD) of 207.2 ms. The average in the search and comparison phase (FD_s) was 197.3 ms; the average in the detection and verification phase (FD_v) was 286.0 ms which is significantly longer ($F(1;\ 15) = 32.45;$ $p < 0.001$). Fig. 4 shows a combined histogram of FD_s and FD_v.

Since long lasting fixations are generally taken to indicate extraordinary cognitive load, we have calculated the proportion of fixations in excess of 500 ms ("long fixations"). This value was taken from literature (Velichkovsky, 1995). The proportion of long fixations turned out to be about 15 times higher in the detection and verification phase (11.4%) than in the search and comparison phase (0.75%) ($F(1;\ 15) = 33.05;$ $p < 0.001$).

FD is a measure that lends itself to be related to local parameters of the display. The analysis showed that FD was in fact a function of object density ($F(2;\ 30) = 8.77;$ $p = 0.001$); no other effects were significant. In Fig. 5, FD is plotted against $\varrho(\mathbf{p})$. The subjects' fixations in high density regions (213.6 ms) were longer than those in medium density regions (197.4 ms) ($F(1;\ 15) = 14.52;$ $p = 0.002$) and those in low density regions (191.9 ms) ($F(1;\ 15) = 9.88;$ $p = 0.007$).

4.1.4. Saccade length

The histogram of saccade lengths (SL), given in Fig. 6, suggests that a distinction be made between two types of saccades: saccades linking fixations within the same hemifield (about

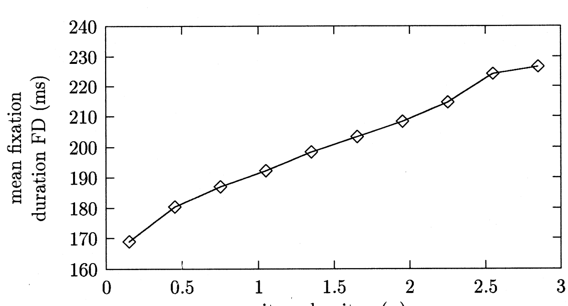

Fig. 5. Mean fixation duration as a function of local object density at the fixation point.

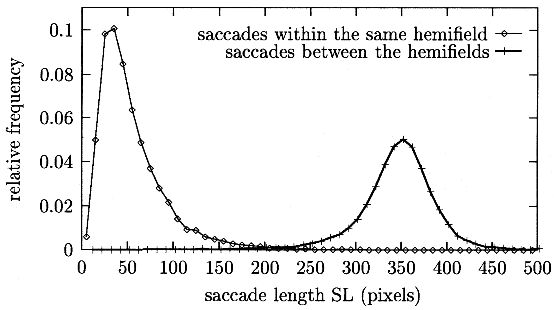

Fig. 6. Histogram of saccade lengths.

10 to 150 pixels long) and saccades passing the boundary between hemifields (about 250 to 450 pixels long). In the analysis below, only the first type of saccades is taken into account in order to reconstruct the way in which subjects have subdivided the set of objects to be compared.

The mean length of saccades across all subjects was 55.7 pixels. On average, saccades during search and comparison (SL_s) spanned 56.8 pixels while saccades during detection and verification (SL_v) were shorter, namely 40.8 pixels in length ($F(1; 15) = 42.12$; $p < 0.001$).

SL can also be related to local display parameters. However, the fact that saccades—unlike fixations—have a linear spatial extension constitutes a problem: Where should the local parameters of a saccade be measured? Striving for detailed information, we performed two planned analyses of variance, based on (a) the starting points of the saccades and (b) the end points of the saccades respectively as a reference.

Analysis (a) revealed two significant effects. For one, SL depended on local object density at the starting point ($F(2; 30) = 35.79$; $p < 0.001$). In medium density regions, saccades were longer (58.9 pixels) than in high density regions (48.8 pixels) ($F(1; 15) = 53.69$; $p < 0.001$); even longer saccades (64.1 pixels) were to be found in low density regions ($F(1; 15) = 10.06$; $p = 0.006$). For another, SL showed a significant effect for form entropy ($F(2; 30) = 6.40$; $p = 0.005$). Saccades were longer when they started in regions of low form entropy (59.3 pixels) and medium form entropy (57.6 pixels) than when they started in regions of high form entropy (54.8 pixels) ($F(1; 15) = 10.87$; $p = 0.005$ and $F(1; 15) = 1.52$; $p = 0.004$ respectively). The effect of color entropy showed the same tendency as form entropy, but did not reach significance ($F(2; 30) = 2.99$; $p = 0.065$).

Analysis (b) revealed an influence of local object density on SL ($F(2; 30) = 72.31$; $p < 0.001$), as saccades ending in medium density regions were longer (54.9 pixels) than those ending in high density regions (46.4 pixels) ($F(1; 15) = 58.59$; $p < 0.001$). Saccades ending in low density regions were even longer (72.3 pixels) than those ending in medium density regions ($F(1; 15) = 52.91$; $p < 0.001$). However, analysis (b) did not reveal any significant effects of color or form entropy. Fig. 7 illustrates SL as a function of local object density, whereas Fig. 8 shows SL as a function of local color and form entropy.

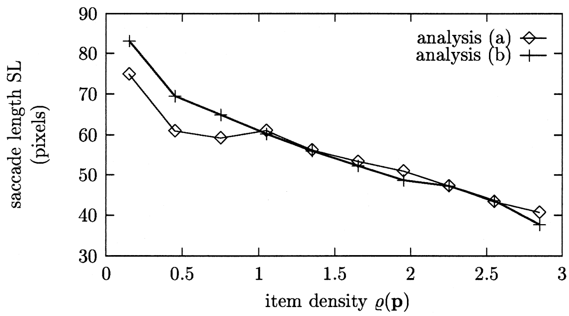

Fig. 7. Mean saccade length as a function of local object density at its starting point (analysis a) and at its target point (analysis b).

4.1.5. An intermediate summary

In view of the complex results obtained so far, a brief intermediate summary seems to be useful. First of all, the type of mismatch to be detected made a difference to reaction time and to the number of fixations made in each trial. Generally, *color mismatches* were detected more easily, i.e. faster and with fewer fixations, than *form mismatches*. Also, the number of fixations clearly exceeded the expected values even for the case of a TSP strategy. Even when considering only the search and comparison phase, subjects took more fixations than there were objects in each hemifield. This observation can be interpreted as meaning that people occasionally failed to detect a mismatch at first sight, and so parts of the display had to be scanned twice.

As for local parameters, both fixation duration and saccade length were affected by object density. The observation that fixations took longer in high density regions is not compatible with a traveling salesman strategy. It is, however, in line with the predictions derived from the searchlight strategy: Provided that the focal area is relatively constant in size, fixations in high density regions cover a relatively large number of objects, and so representing these should take longer. It is in line with the clustering strategy as well: In high density regions more objects can be memorized or compared per fixation: Accordingly, such fixations should take longer.

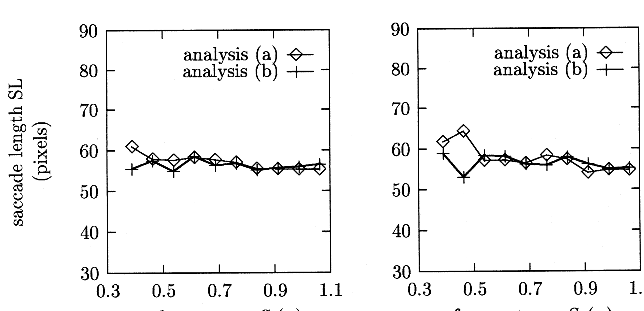

Fig. 8. Mean saccade length as a function of color entropy (left) or form entropy (right) at its starting point (analysis a) and at its target point (analysis b).

The fact that saccades were shorter in high density regions concords with the predictions of all three strategies since the distance to the next suitable fixation point—be it the next object on the "salesman's" path, a neighboring area of fixed size, or a neighboring cluster—should be rather short. The observed effect of form entropy and the tendency of color entropy at the starting point of a saccade to influence the saccade's length indicates that the number of objects processed at a time also depends on their informational content. Conceivably, processing the objects in uniform areas is undemanding, so subjects tend to focus on areas of higher entropy within each hemifield of the display.

All in all, there is no support for the "strong" version of the traveling salesman strategy, but there is evidence in favor of the searchlight strategy and, in particular, the clustering strategy. At any rate, local object density as well as entropy are to be taken into account as determinants of fixation behavior in comparative visual search.

4.2. Derived dependent variables

4.2.1. Successive fixations within hemifields

Eye-movement measurement does not only allow one to distinguish search and comparison from detection and verification, but it also enables researchers to reconstruct the scan paths that lead to the detection of mismatches in comparative searches. By separating saccades that occur within the same hemifield from those that occur between hemifields, it is possible to itemize the individual steps in the course of comparison as well as to identify the objects currently attended. Consecutive saccades within the same hemifield link those fixations which pertain to the objects to be compared in that particular step. Thus, the number of successive fixations within the same hemifield (FW) is of particular relevance to a detailed investigation of the strategies which people follow in comparative search tasks.

On average, subjects made 2.45 fixations before shifting over to the other hemifield. An analysis of variance showed that subjects switched between hemifields more rapidly after encountering a mismatch ($F(1;\ 15) = 93.29;\ p < 0.001$)—the number of successive fixations within one hemifield during search and comparison (FW_s) was 2.57, while the corresponding value for the detection and verification phase (FW_v) was 1.91.

FW proved to depend on local object density ($F(2;\ 30) = 13.99;\ p < 0.001$): When objects were widely dispersed, so that the first in a series of fixations would cover only a few of them, people took more fixations (2.58) before shifting to the other hemifield than when the first fixation landed in a region of medium density (2.45) ($F(1;\ 15) = 12.80;\ p = 0.003$). Even fewer fixations (2.30) were found after touching a high density region ($F(1;\ 15) = 6.34;\ p = 0.024$). The influence of object density on FW is plotted in Fig. 9.

Also, FW was affected by form entropy ($F(2;\ 30) = 3.46;\ p = 0.045$): The more the objects in the vicinity of the fixation point varied in form, the more fixations people took before shifting to the other hemifield (FW at low $S_f(\mathbf{p}) = 2.39$; FW at high $S_f(\mathbf{p}) = 2.51$) ($F(1;\ 15) = 13.62;\ p = 0.002$). In contrast, color entropy had no significant effect on FW. No interactions between the factors were found. Fig. 10 illustrates FW as a function of the local entropy values.

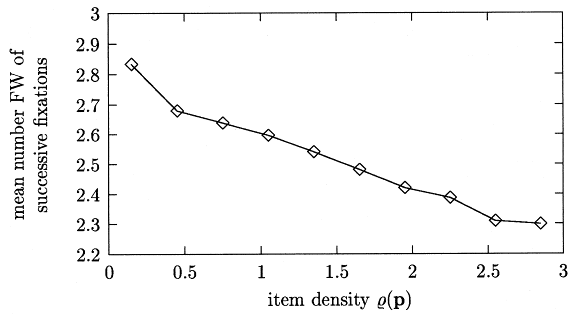

Fig. 9. Mean number of successive fixations within the same hemifield as a function of object density at the first fixation point.

4.2.2. Probability of missing the target

As mentioned in connection with RT and NF, subjects did not always detect the mismatch when first fixating both target objects or their vicinity. This was the case in 12.9% of all trials, where the vicinity of a target was defined by a maximum distance of 50 pixels (two degrees), as explained in Section 4. In other words, the probability of detecting the mismatch was about 87.1% once each of the target regions was consecutively fixated.

Which factors influence PM? We checked three local factors (object density, color entropy, form entropy) and one global factor (type of mismatch—color or form). Each local factor was divided into two levels ("low" vs. "high"), based on the mean of its values at the centers of the two target objects. The entropy cutpoint was set at 0.85, the density cutpoint at 1.5. Splitting up each factor into three levels—as in the analyses described above—was not feasible because of the low number of available data.

The four-way analysis of variance revealed that subjects failed to detect the mismatch more often when the difference was in form (15.6%) than when the difference was in color (10.2%), but this effect did not reach significance ($F(1;\ 15) = 3.57$; $p = 0.078$). Local object density, however, had a significant effect on PM ($F(1;\ 15) = 9.52$; $p = 0.008$). High object density increased the rate of missing the target (15.3%) compared to low object density (10.5%). Furthermore, the analysis revealed an effect of local form entropy: Targets

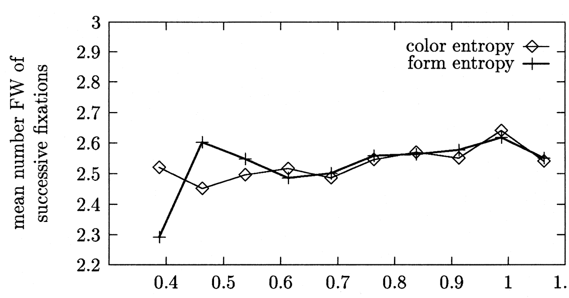

Fig. 10. Mean number of successive fixations within the same hemifield as a function of color or form entropy at the first fixation point.

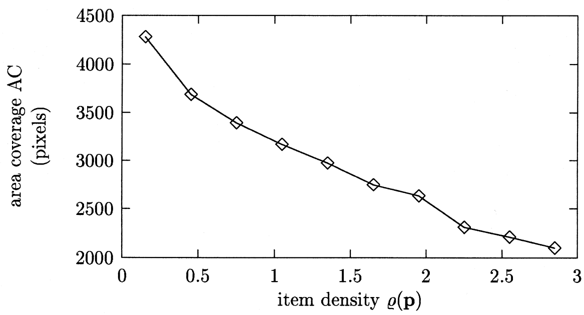

Fig. 11. Mean covered area per fixation as a function of the local object density at the fixation point.

were more frequently missed (14.6%) when the target area had high form entropy than when it had low form entropy (11.2%) ($F(1;\ 15) = 6.91$; $p = 0.019$). No other effects or interactions were found. This means that form "disorder" around the target objects led to a higher probability of missing the mismatch, regardless of whether it was in color or in form.

4.2.3. *Area coverage per fixation*

This derived variable provides a measure of the average area covered by one fixation (AC). It is defined on the basis of eye-movement patterns (see Appendix C for an explanation). AC may serve as a rough estimate of the efficiency subjects exhibit in scanning the display.

In the experiment, AC had a grand mean of 2858 pixels. The analysis of variance yielded main effects for all "local" factors: object density ($F(2;\ 30) = 46.63$; $p < 0.001$), color entropy ($F(2;\ 30) = 7.33$; $p = 0.003$), and form entropy ($F(2;\ 30) = 6.64$; $p = 0.004$). As for object density, AC was large (3516 pixels) in low density areas, but smaller (2751 pixels) in medium density areas ($F(1;\ 15) = 44.99$; $p < 0.001$), and even smaller still in regions of high density (2309 pixels) ($F(1;\ 15) = 14.87$; $p < 0.002$). As to color entropy, AC was significantly larger (3137 pixels) for fixations with low $S_c(\mathbf{p})$ than for fixations with high $S_c(\mathbf{p})$ values (2597 pixels) ($F(1;\ 15) = 12.73$; $p = 0.003$). Moreover, fixations in medium $S_c(\mathbf{p})$ regions turned out to cover more area (2841 pixels) than those in high $S_c(\mathbf{p})$ regions ($F(1;\ 15) = 6.02$; $p = 0.027$). Finally, with regard to form entropy, AC was larger (3072 pixels) when $S_f(\mathbf{p})$ was low but smaller (2633 pixels) when $S_f(\mathbf{p})$ was high ($F(1;\ 15) = 8.87$; $p = 0.009$). In addition, AC reached a higher value with medium form entropy (2870 pixels) than with high form entropy ($F(1;\ 15) = 8.51$; $p = 0.011$). Fig. 11 and 12 illustrate the area covered per fixation decreasing with growing local complexity of the display.

4.2.4. *Summary*

Derived dependent variables shed more light on the strategies that people employ in comparative visual search. First, local parameters once again proved to be significant determinants of fixation behavior: Object density and entropy played a role in determining the number of consecutive fixations within one hemifield, the average area covered per fixation, as well as the probability of missing the target. Second, people's actual gaze

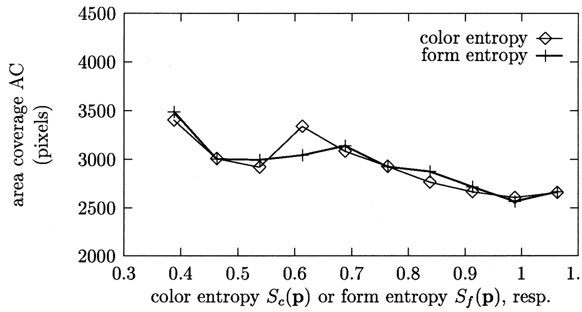

Fig. 12. Mean covered area per fixation as a function of the local entropy values at the fixation point.

trajectories can be reconstructed more precisely if one takes into consideration the conditions under which saccades between hemifields occur. In fact, the number of consecutive fixations within hemifields was generally larger when proceeding from a low density region than when proceeding from a high density region. This observation may suggest again that people prefer fixating complex regions over fixating less complex ones. However, the number of consecutive fixations *increased* with growing local form entropy. This finding can be explained by the fact that the scanning of a high form entropy region requires foveal processing and thus a large number of fixations. Additionally, these observations suggest that color entropy and form entropy influence processing in different ways. Also, the effects of density, color entropy, and form entropy on the average area coverage per fixation demonstrate that, in determining their eye movements in comparative search, people take into account economical principles. They tend to optimize working memory load so as to manage with the least number of fixations in a trial.

Again, the varying probabilities of missing the target can be taken as indicating that color mismatches can be detected more readily than form mismatches. Moreover, high form entropy turns out to increase the probability of missing the target, whereas color entropy has no significant effect. The dependence of PM on the type of mismatch is reflected in a corresponding difference in the additional search time.

5. General discussion

What insights about the process of comparative visual search does the analysis of eye movements yield? First, we have obtained information about the *global* time course of comparative search. Second, the *local* effects of stimulus parameters on eye movements have yielded a variety of data about the cognitive factors that control the search behavior.

As to the global analysis, eye-movement measurement has rendered it possible to split the process of comparative search into two successive phases—a first phase labeled “search and comparison,” and a second one labeled “detection and verification.” The two phases differ significantly with respect to eye-movement characteristics. First of all, the average fixation duration is longer in the second phase. In particular, the proportion of fixations in excess of 500 ms, to be considered as a measure of the subjects’ mental effort (Velichkovsky, 1995), is increased by a factor of 15. This suggests that during the second phase, people concentrate on identifying and processing the mismatching objects rather than quickly scanning the hemifields in search of a difference—as they do during the first phase. Moreover, average within-hemifield saccade lengths are shorter in the second phase, indicating that the subjects' attention is now focused on a small "suspicious" object or region. This conjecture is corroborated by the observation that during phase two there are fewer successive fixations within the same hemifield than there are in phase one. It must be noted, however, that the criterion for determining the transition between the stages was chosen intuitively and might not be optimal (see Section 4).

Another important result is the explanation of a reaction time difference, i.e. the finding that mismatches in color are processed faster than mismatches in form. Since subjects did not know in advance whether the mismatch to be detected was in color or in form this difference appears to be somewhat striking. Yet we have found a plausible answer to this. Subjects appear to "miss" differences in form more often than differences in color. A miss, in this context, means a failure to detect the mismatch when scanning the vicinity of the target and to continue with search and comparison. The eye-movement data are well in line with this consideration: The probability of detecting the target was 89.8% with respect to color mismatches and 84.4% with respect to form mismatches. In fact, we found that on average subjects "wasted" only 1248 ms because of missing color mismatches, but 2855 ms by missing form mismatches. This effect is likely to be the main reason for the difference in reaction time. The remaining discrepancy of about 400 ms may be explained by the assumption of different durations of the verification phase. We cannot fully rule out, however, that this finding may, to some degree, reflect differences within the color or form dimension (i.e., between the specific colors and forms used in the experiment) rather than differences between the color and form dimensions.

As to the local analysis, a detailed investigation into comparative search processes has been performed, focusing on how viewers' eye movements are determined by local display characteristics, namely by *local object density*, *local color entropy*, and *local form entropy*.

The average length of saccades has proven to be inversely proportional to local object density. This is intuitively plausible because people tend to direct their saccades at objects, yet in low density subregions, objects are located relatively far from each other which makes people produce relatively long saccades. The fact that the length of a saccade is partially determined by the degree of form entropy and—marginally significantly—by the color entropy at its starting point, is also in line with the idea that processing economy is a highly important principle in comparative search. With respect to the end point of a saccade, its length depends on local object density but not on any of the entropy values. This result can be explained by retinal eccentricity: The end point of a saccade is determined before its realization; it is usually located in a subject's parafoveal region, where it is possible to coarsely estimate the local density in advance but not the local entropy.

These findings extend the conclusions reached by other researchers who have presented evidence that saccade lengths change in correspondence with changes in the position of the target as well as in the degree of overall visual heterogeneity (e.g. O'Regan, 1989; Jacobs, 1991): Though processing economy is a fundamental aspect in visual search, it is not a central, fixed capacity mechanism which limits performance but rather a flexible one. The course of visual search is controlled by a mechanism that is capable of adapting to local parameters of the display. While this account holds for both the standard search paradigm and for comparative search, there is one crucial difference: Using grid-like displays that did not vary in object density, the studies mentioned above have shown that the parameters controlling saccade length apply to the spatial layout and to overall form entropy. The present experiment, however, using randomly generated displays that did vary in local object density, shows saccade length to be primarily determined by the spatial distribution of the objects, regardless of any other features. This pattern of results suggests that it might be useful to distinguish between parameters related to spatial location and parameters related to object features, which will be discussed below.

Another basic dependent variable, fixation duration, exhibited a strong dependence on local object density. The average duration of fixations increases linearly by about 50% along with local object density values. Fixation duration in no way depends on local color or form entropy.

Again, this observation goes beyond the results of previous research. Previous work has shown a relationship between fixation duration and the spatial layout of the objects to be processed, even in such detail as to find that global density measures, such as average distance, predict fixation duration less exactly than does minimum object distance (cf. Jacobs, 1991; Nodine, Kungel, Toto & Krupinsky, 1992). The present study goes beyond averaged quantities and provides insight into the dynamic nature of this relationship: In the course of search, fixation duration is determined on the fly, depending on the *local* spatial parameters of the display. It is conceivable that the time spent on a fixation varies along with the number of objects covered by it. This would imply, however, that the specific features of the objects are not necessarily computed before proceeding to the next fixation. On the one hand, this finding suggests that subjects, for reasons of efficiency, do not memorize and compare the objects one by one but in chunks (the size of which probably varies). On the other hand, in determining relevant object configurations, subjects appear to merely localize the objects—regardless of their color or their form—by chunks, gathering only the coarse information required for memorization or comparison.

Thus, the basic dependent variables, saccade length and fixation duration, mainly yield information about the low-level processes involved in comparative visual search, in particular, about the perception of object clusters and their comparison by means of working memory. In order to consider higher level processes controlling eye-movement behavior, such as the planning of search paths and strategies in the utilization of memory, we have to turn to the derived dependent variables in the present experiment.

The number of successive fixations within the same hemifield has been shown to decrease when local object density increases. This finding can be attributed to principles of economy in the usage of capacity-limited working memory. People appear to be able to process more objects per fixation in high density subregions than in low density subregions. Accordingly, they need fewer fixations to “fill up” their working memory when searching high density areas. So the factor which is best suited to model the amount of information processed in each fixation is object *density*, a local parameter which merely reflects the spatial location of the objects to be scanned.

In addition, however, eye movements in comparative search are also influenced by object *entropy*, a local parameter which presupposes that the identity of the objects has been established. That is, their features have been identified and the appropriate color and form values have been computed.

The influence of local entropy on search and comparison becomes most transparent when considering the area covered per fixation. Not unlike the notion of "grain size"—operationally defined via the minimum distance between neighboring objects (cf. Jacobs, 1991), this measure gives an indication of the human visual span or "focus size" during the solution of the search task. The area covered per fixation is inversely related to local object density, color and form entropy. While the strong dependence of area coverage on object density can plausibly be explained along the same lines as saccade length, the influence of both color and form entropy signifies that the search strategy depends on establishing the identity of the objects. The more complex a particular region of the display is with respect to the color and form of its objects, the more fixations are required to gather a sufficient amount of information to detect a mismatch. Therefore, we arrive at the conclusion that density and entropy exert different and largely independent effects on the processing characteristics in visual search.

Interestingly, the factors density and entropy are closely linked with a major hypothesis about the organization of the visual system into one subsystem dealing primarily with location (the "where?" subsystem) and one dealing primarily with object identity (the "what?" subsystem) (cf. Trevarthen, 1968; Mishkin, Ungerleider & Macko, 1983; Velichkovsky, 1982; Bridgeman, van der Heijden & Velichkovsky, 1994). Density is the factor most closely associated with the functioning of the "where?" subsystem: The computation of object density involves the evaluation of spatial locations only, with no need to discriminate among object types. The discrimination of object types, on the other hand, is crucial for the computation of entropy; therefore, this factor is most closely associated with the operation of the "what?" subsystem (note that we use a normalized entropy measure that is fully independent from object density, cf. Appendix B). The different effects of the factors density and entropy can be interpreted as indirect evidence for the usefulness of the distinction between "where?" and "what?" and they provide additional insight into the characteristics of the two processing streams.

The current findings provide at least a rough outline of the cognitive processes involved in comparative search. The study indicates that comparative visual search proceeds in a "perpendicular fashion," much as is pictured in the sample scan path provided in Fig. 2 (see page 13). The actual number of fixations conceivably depends on working memory capacity and on the number of objects processed per fixation. That number, again, is determined by the local stimulus parameters object density, color entropy, and form entropy, and some of their interactions. After gathering as much information as is appropriate in terms of memory load, subjects tend to direct their gaze with one long horizontal saccade to the corresponding cluster in the other hemifield. The object features in that cluster can then be compared to the stored representation, working memory can be cleared, and the first cluster can be tagged as analyzed. Processing may now proceed to the next cluster, which is determined by strategic considerations (in order to avoid returns) as well as tactic ones (on the basis of local object density), and so forth. If, however, search and comparison yields a "suspicion," the viewer will switch to a detection and verification phase, which is characterized by fewer but longer fixations.

The picture of comparative visual search that emerges from the experiment does not provide any support at all for a traveling salesman strategy which proceeds object by object: In contrast to the corresponding predictions, saccade length and fixation duration do depend on local stimulus parameters. To some extent, the results are compatible with the clustering and the searchlight strategy. That is, with each fixation, a limited number of objects are processed before proceeding to the next cluster. Saccade length and fixation duration patterns correspond to the predictions made on the basis of the searchlight strategy. These basic measures of oculomotor behavior are affected by the local parameters relating to the spatial configuration of the objects to be scanned. The area coverage per fixation, however, exhibits a determination pattern that closely resembles the predictions made on the basis of the clustering strategy. The effects of local object density, color entropy, and form entropy suggest that issues of economy play a part in determining the search path. Within the limit set by the capacity of working memory, the number of objects that can be processed with each fixation depends on object similarity in terms of location, color, and form. The main effect of local object density signifies that the more objects a cluster comprises, the closer they are to each other. The main effects of color and form entropy signify that the more objects a cluster comprises, the more likely they are to be homogeneous in terms of color and form. The large number of fixations per trial (35.2), however, seems to contradict our expectations with regard to the clustering strategy. As mentioned above, one reason may lie in the fact that people sometimes miss the target which makes them scan the display again. Another explanation might be subjects' tendency to minimize the use of working memory which makes them accept the trade-off of an increase in saccades. Ballard, Hayhoe & Pelz (1995), for instance, obtained a similar result with a hand-eye construction task. For both a construction and a comparative search task, working memory usage seems to be substantially more "expensive" than eye movements, resulting in more fixations per trial than predicted. All in all, the empirical data favor the clustering strategy over its competitors.

Not in line with our initial hypotheses, no interaction of entropy with density has been found in any basic or derived variable. This means that the effect of entropy is not—as we assumed—restricted to regions of high object density. Conceivably, the maximum distances between neighboring objects were not sufficient to reduce entropy effects in regions of low density in our experiments. Therefore, it seems that the size of the attentional focus can be adapted to specific situations: The focus of attention can, for example, cover a large area in regions of low object density. Due to this flexibility, the entropy of object features plays an important role in determining the search strategy, even if the objects are located far apart from each other.

Altogether, the results of the present experiment suggest that oculomotor behavior in comparative visual search is determined on the basis of cognitive economy. The clustering strategy employed reflects the existence of a "where?" subsystem to provide a basis for chunking, and of a "what?" subsystem to provide a basis for comparison. One question that remains to be answered, then, is how cluster size can be modeled best. Further studies will elaborate on this question.

We think that these results provide good evidence for considering comparative visual search as a promising new paradigm for the future study of cognitive processes. By providing rich information about visual behavior during an important class of cognitive tasks, it opens up new possibilities to formulate and test new cost functions. This allows an extension of the analysis of cognitive processes in terms of principles of cognitive economy—a means of analysis that has proven fruitful in the past. We hope that this may open a new road towards a better understanding of cognitive processes—if not the avenue, certainly a way to go.

Acknowledgment

The authors gratefully acknowledge the comments of two anonymous reviewers and the share of Elena Carbone and Boris M. Velichkovsky in the preparation of the paper in hand. Their thoughtful criticism and helpful commentary has contributed considerably to its improvement. We would like to thank Alex Gunz and Martin Herold for improving the readability of the present paper. The research reported here was funded by the German Science Foundation (Collaborative Research Center 360 "Situated Artificial Communicators," project B4).

Appendix A. Algorithmic generation of stimuli

Wishing to avoid the influence of any a-priori structural information, we might be tempted to use simple random distributions for the locations and features of the objects. Such a simple choice, however, has several problems, necessitating specific refinements:

- Objects may overlap and thus constitute new, "hybrid" forms which would likely induce distinct recognition processes and to violate the simple parametric structure of the stimuli. Hence, objects were set at least 20 pixels apart. Aside from this constraint, objects were randomly placed within a rectangle of 260×400 pixels in each hemifield.
- It may happen that in some stimuli specific colors or forms are strongly overrepresented. Unbalanced sets of objects can lead to deviating search strategies, biasing the experimental results. Hence, we constrained each of the three colors and the three forms to be equally represented, i.e. there must be 10 triangles, 10 blue objects, etc. No specific combination of color and form was allowed to appear more frequently than others.
- A homogeneous random function is unlikely to yield large regions with the same color or form. The occurrence of such regions, however, is an important precondition for investigating the effects of object uniformity vs. object variety on eye movements. Thus, we used a random function that tended to create regions with the same color or form more likely than a homogeneous random distribution would do.

Obviously, the first two of these criteria can easily be achieved, whereas the third one requires mathematical definitions of color and form *segregation*, plus an algorithm being capable of generating color and form distributions according to these two segregation parameters. In order to make the generation of stimuli transparent and replicable for other scientists, a detailed mathematical description is given in the following.

As a starting point, we give a formalized description of the stimulus patterns in one hemifield: A pattern is a set of N objects

$$\mathbf{o}^{(n)} = \begin{pmatrix} o_x^{(n)} \\ o_y^{(n)} \\ o_f^{(n)} \\ o_c^{(n)} \end{pmatrix}, \ n = 1, \ \ldots, N, \tag{1}$$

where $(o_x^{(n)}, o_y^{(n)})$ is the pixel position of the object's center on the screen, $o_f^{(n)}$ is the object's form (1 = square, 2 = triangle, 3 = circle), and $o_c^{(n)}$ is the object's color (1 = blue, 2 = green, 3 = yellow). In the current experiments, N was set to 30.

Now the variable *form segregation* (α_f) is introduced. It can be defined as the quotient of the mean distance $\bar{d}_{f,dif}$ between all pairs of objects with different forms and the mean distance $\bar{d}_{f,id}$ between those with identical forms:

$$\alpha_f = \frac{\bar{d}_{f,dif}}{\bar{d}_{f,id}} \tag{2}$$

$$\bar{d}_{f,dif} = \frac{\sum\limits_{n_1=1}^{N} \sum\limits_{n_2=n_1+1, o_f^{(n_1)} \neq o_f^{(n_2)}}^{N} d(n_1, n_2)}{\sum\limits_{n_1=1}^{N} \sum\limits_{n_2=n_1+1, o_f^{(n_1)} \neq o_f^{(n_2)}}^{N} 1} \tag{3}$$

$$\bar{d}_{f,id} = \frac{\sum\limits_{n_1=1}^{N} \sum\limits_{n_2=n_1+1, o_f^{(n_1)} = o_f^{(n_2)}}^{N} d(n_1, n_2)}{\sum\limits_{n_1=1}^{N} \sum\limits_{n_2=n_1+1, o_f^{(n_1)} = o_f^{(n_2)}}^{N} 1} \tag{4}$$

$$d(n_1, n_2) = \sqrt{(o_x^{(n_1)} - o_x^{(n_2)})^2 + (o_y^{(n_1)} - o_y^{(n_2)})^2} \tag{5}$$

For example, $\alpha_f = 2$ means that, on average, objects of different forms are twice as distant from each other than objects of the same form. In our setting of 30 objects with three different forms, this would correspond to a strongly segregated distribution containing large uniform areas. $\alpha_f = 1$ means that there is no segregation at all. Fig. 13 illustrates the correspondence between α_f and the distribution of forms at three different levels.

We define the parameter *color segregation* (α_c) analogously. An iterative algorithm for generating color and form distributions with given parameters of form and color segregation can easily be implemented. Starting with a random distribution, this algorithm randomly selects pairs of objects and exchanges their colors or forms if this exchange shifts the distribution's segregation levels towards the given parameters. The algorithm terminates as

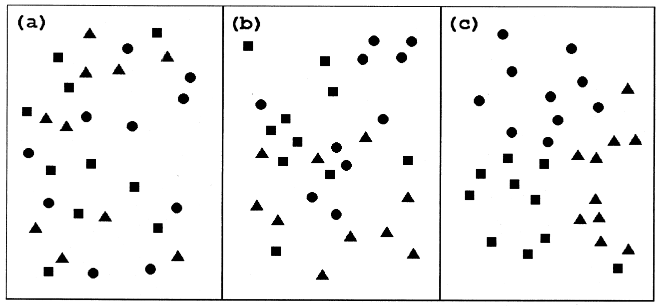

Fig. 13. Examples of object distributions at a form segregation α_f of (a) 1.0, (b) 1.3, and (c) 1.8.

soon as the difference between the actual and the desired α_f and α_c falls below a certain threshold.

In our experiment we used such an algorithm to compute the stimuli. For each scene, α_f and α_c were set to random values ranging from 1.0 to 1.3.

Appendix B. How to define appropriate independent variables

It is reasonable to assume that only *local* information in the vicinity of the gaze position is processed during comparative visual search. Although there might be an initial parallel stimulus inspection phase as is assumed in standard visual search tasks, target detection itself cannot be accomplished in parallel. Accordingly, we can suppose gaze trajectories to be controlled by local stimulus features. This assumption motivates the application of a small set of local stimulus parameters as independent variables in order to melt down the high dimensionality of the search displays into their essential features.

This is realized by defining continuous scalar functions on the two-dimensional display plane, yielding local values for each parameter. Altogether, three functions of this kind have to be defined for every point $\mathbf{p} = (p_x, p_y)^T$ on the screen:

Object density $\varrho(\mathbf{p})$: This function tells us how closely objects are packed at the location $\mathbf{p}$. The value of $\varrho(\mathbf{p})$ increases with the concentration of objects around this point, regardless of their color or form.

Color entropy $S_c(\mathbf{p})$: Color entropy is a measure of local color "disorder." If objects of all three colors are equally represented in the range of $\mathbf{p}$, the function $S_c(\mathbf{p})$ will reach its maximum value. For example, if there are only green objects around $\mathbf{p}$ without exception, $S_c(\mathbf{p})$ will yield a value near its absolute minimum. It is important to define this function in such a way that it is completely uncorrelated with object density.

Form entropy $S_f(\mathbf{p})$: This function is the equivalent to color entropy, but with respect to the objects' *forms*. We constrained form entropy to neither correlate with object density nor with color entropy.

The following two subsections give precise mathematical definitions of the local parameter functions we used in analyzing our eye-movement data.

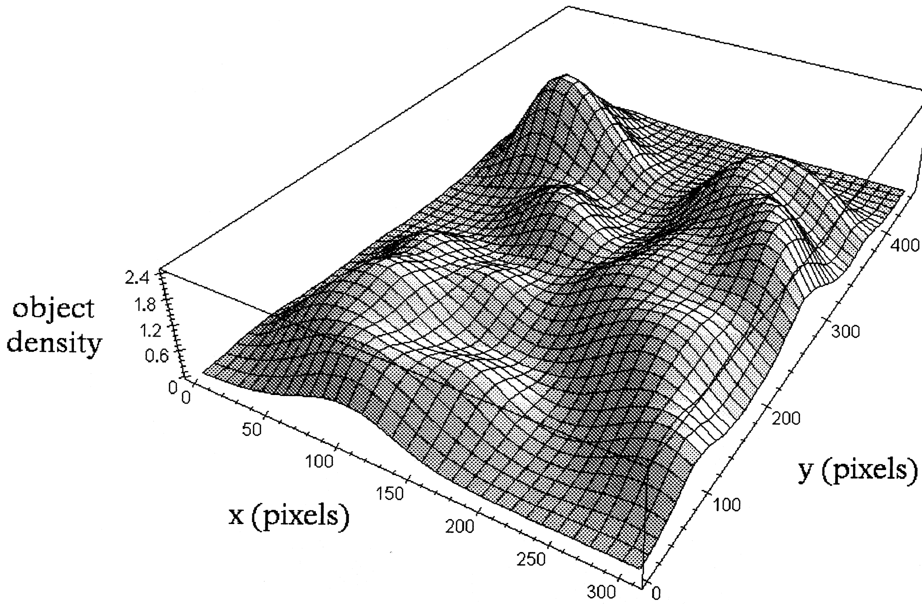

Fig. 14. Object density $\varrho(\mathbf{p})$ in the right hemifield of Fig. 1.

B.1. Local object density

How do we get a smooth continuous density function from a distribution of discrete objects as defined in Appendix A? A viable solution to this problem is to define a weight function $w(n, \mathbf{p})$ which yields a value for the influence of object $\mathbf{o}^{(n)}$ on the object density at a reference point $\mathbf{p} = (p_x, p_y)^T$. Obviously, this influence has to decrease with growing distance between the object and point $\mathbf{p}$. An appropriate choice of $w(\cdot, \cdot)$ is a Gaussian function applied to that distance:

$$w(n, \mathbf{p}) = \exp\left(-\frac{(o_x^{(n)} - p_x)^2 + (o_y^{(n)} - p_y)^2}{2\sigma^2}\right) \tag{6}$$

The standardization coefficient does not appear in this equation, because the standard deviation σ remains constant throughout the evaluation. The choice of σ cannot be perfectly calculated since there is no invariant "focus size" in human vision. We decided to use a value of 25 pixels on the screen, which corresponds to one degree of visual angle—the idealized size of the human fovea. To calculate the local object density $\varrho(\mathbf{p})$, the weight functions for all objects are summed up as follows:

$$\varrho(\mathbf{p}) = \sum_{n=1}^{N} w(n, \mathbf{p}) \tag{7}$$

How plausible is this calculation? Fig. 14 illustrates the object density values of the right hemifield of the example display (Fig. 1 on page 7). By comparing the underlying object distribution with its density "landscape", a clear and plausible correspondence can be seen. For example, the object accumulation in the upper left corner of the display causes the highest density "hill". Naturally, the display's *left* hemifield would show an identical landscape, because its objects are located at the same positions. In the present experiment, $\varrho(\mathbf{p})$ ranged between 10^{-3} and 3.9, but less than 1% of its values exceeded 3.0.

B.2. Local form and color entropy

Is it possible to define a measure of entropy in a similar way as for density? Here, we not only have to take into account the *location* of objects, but their *identity* as well. This can be achieved by calculating separate densities $\varrho_1(\mathbf{p})$, $\varrho_2(\mathbf{p})$, and $\varrho_3(\mathbf{p})$ for the presence of squares, triangles, and circles respectively:

$$\varrho_i(\mathbf{p}) = \sum_{n=1, o_f^{(n)}=i}^{N} w^*(n, \mathbf{p}),\ i = 1, 2, 3 \tag{8}$$

Then, form entropy $S_f(\mathbf{p})$ can be computed in analogy to information entropy of a probability distribution:

$$S_f(\mathbf{p}) = -\sum_{i=1}^{3} \frac{\varrho_i(\mathbf{p})}{\varrho^*(\mathbf{p})} \ln \frac{\varrho_i(\mathbf{p})}{\varrho^*(\mathbf{p})}, \quad \text{where } \varrho^*(\mathbf{p}) = \sum_{n=1}^{N} w^*(n, \mathbf{p}) \tag{9}$$

If we used the function $w(\cdot, \cdot)$ from Equation (6) as $w^*(\cdot, \cdot)$, $S_f(\mathbf{p})$ would yield the aspired value of form entropy, but also exhibit the undesirable feature of being correlated with the local object density $\varrho^*(\mathbf{p})$. This would be caused by the fact that in regions of low object density there are large areas being dominated by the influence of a single object and thus having low entropy.

Obviously, a different choice of $w^*(\cdot, \cdot)$ is required, implying some standardization to compensate for object density. We could vary the standard deviation σ of the Gaussian weight function (6), in such a way that $\varrho^*(\mathbf{p})$ is constant for all $\mathbf{p}$. In other words, the implemented "human focus size" would be inversely related to local density, which is a plausible solution.

However, simply adjusting σ would create an overcompensation in the peripheral areas. In a region of low density, for example, σ would increase or, from another point of view, σ would remain constant and the distances between all objects and $\mathbf{p}$ would shrink by the same factor. Although this induces a standardized object density in the proximity of $\mathbf{p}$, the approached peripheral objects now have a disproportionately high influence. Putting more objects into the calculation tends to produce higher entropy values, leading to a correlation between entropy and density.

A possible solution is found by shifting the distances between the objects and $\mathbf{p}$ by the same *offset* Δd, with distances lower than zero being rounded to zero. Using an offset ensures a peripheral object influence which is independent of $\varrho(\mathbf{p})$. The resulting equation reads as follows:

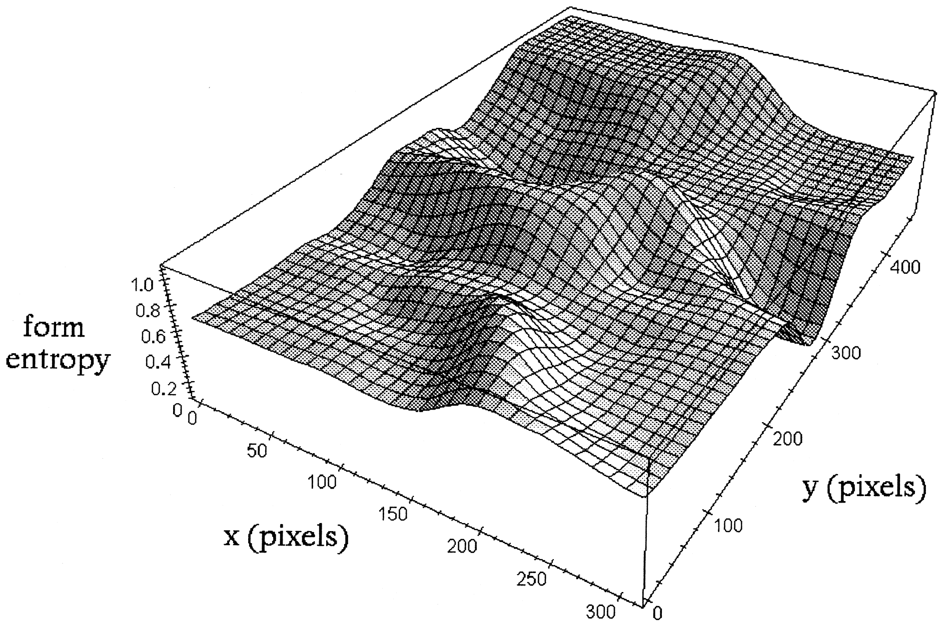

Fig. 15. Form entropy $S_f(\mathbf{p})$ in the right hemifield of Fig. 1.

$$w^*(n, \mathbf{p}) = \exp\left(- \frac{\max(\sqrt{(o_x^{(n)} - p_x)^2 + (o_y^{(n)} - p_y)^2} + \Delta d),\ 0)^2}{2\sigma^2} \right) \tag{10}$$

The offset Δd has to be chosen in such a way that

$$\sum_{n=1}^{N} w^*(n, \mathbf{p}) = c_s = \text{const}, \tag{11}$$

which can easily be achieved by an iterative procedure. In all experiments we set $c_s = 3.0$.

Fig. 15 shows a graph of the resulting function $S_f(\mathbf{p})$, again, referring to the right hemifield of the sample display in Fig. 1. Once more, we find a plausible correspondence between the parameter function and the underlying stimulus. The group of four triangles at the right side induces a steep “valley” of form entropy, and the mixed accumulation of different forms in the upper left corner causes a high plateau.

Color entropy $S_c(\mathbf{p})$ can be defined analogously. Entropy values in the current study ranged from 10^{-3} to $\ln 3 \approx 1.1$, where values below 0.3 had a frequency of less than 3%.

Appendix C: Measurement of dependent variables

What are the dependent variables in our experiment that yield most information about important features of search and comparison processes? Altogether, we found four basic and three derived variables that appear to carry essential information. In the following, the methods of measuring these variables are described in full detail in order to render the somewhat complicated analysis of eye-movement data transparent.

As mentioned in the Results section, a distinction between two successive phases in comparative search (*search and comparison* vs. *detection and verification*) was introduced. The two phases were analyzed separately in order to distinguish the processes involved in search and comparison from those involved in the ascertaining of a mismatch.

Variables marked by an asterisk were not measured during the first second after a stimulus was presented, in order not to let a possible "phase of initial orientation" influence the results. A plus mark indicates a variable being measured separately for the *search* and for the *verification* phase.

Reaction time (RT): This is the total search time measured from the presentation of the stimulus to the subject's manual reaction.

Number of fixations per trial^{+} (NF): This is the total number of fixations per trial, accounting for fixations in both hemifields.

Fixation duration^{*+} (FD): Fixation duration is simply the value in milliseconds registered by the eye tracker for each fixation. Its temporal resolution is 16.7 ms and the minimum duration for a fixation to be recorded is 83 ms.

Saccade length^{*+} (SL): Saccade length was measured as the distance in pixels between two successive fixations in the same hemifield. When the dependence of SL on local stimulus parameters was investigated, this local parameter was measured and analyzed separately at the saccade's starting or landing point. Pre-tests indicated that this method yields more valid and detailed information than the analysis of the arithmetic mean of these values or of the integral of the parameter along the saccade.

Number of successive fixations within the same hemifield^{*+} (FW): This is the number of successive fixations produced by a subject without changing to the other hemifield. Since we want to investigate the effect of local parameters on this variable, we have to answer the following question: How should one measure a local parameter that corresponds to a series of fixations? Using the *arithmetic mean* of the parameters measured at each fixation point would result in a bias, because the number of fixations that enter into the calculation is not constant. With a growing number of fixations the probability for the arithmetic mean to yield extremely high or low values decreases. Consequently, an analysis of FW as a function of any local parameter would be superimposed by a Gaussian distribution if performed in the way described above. Another method of determining the parameter value could be to measure the local parameter at the *center* of the fixations in question. However, the fact that subjects prefer to fixate areas with object density above average would then cause a problem: The more fixations are to be accounted for, the more likely their center is to be in a region of medium density. Consequently, an analysis of this kind would yield an artificial negative correlation between FW and local object density. A simple solution is found by evaluating just the local parameters of the *first* fixation that occurs after a saccade between the hemifields. In other words, we pose the following question: How many successive fixations occur within the same hemifield when the first one is located in an area with specific local parameter values?

Probability of missing the target (PM): If the subject's gaze position successively

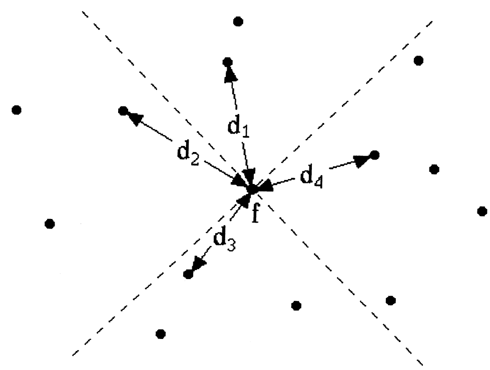

Fig. 16. Fixation point *f* and distances $d_1, \ldots, d_4$ to its nearest neighboring variant.

entered the target area in both hemifields (within a radius of 50 pixels), a "target passage" was counted, whether or not the subject pressed the button. If the gaze-position left the target area and the subject did not press the button within the following 2 seconds, a "target missing" is counted. It was possible (indeed almost certain) that another target passage was registered during the same search process. FW is the quotient of the target missing over the target passage counter.

Area coverage per fixation* (AC): During a single search process, the subject produces a fixation *pattern* on the stimulus display. The average area in the display covered by one fixation is what we were trying to measure, so an appropriate method of determining it was needed.

Fig. 16 shows an outline of such a fixation pattern with a marked fixation *f* in its center. The area covered by this fixation was defined by introducing four quadrants with the origin *f*. These quadrants were rotated by an angle of 45° with respect to the screen coordinate system, because saccade directions were disproportionately horizontal and vertical. Then four distances $d_1, \ldots, d_4$ between *f* and the nearest neighboring fixation in each quadrant were measured and their arithmetic mean μ was taken. In the case of no fixations in one or more quadrants, μ was calculated regarding only the observed distances. We defined the area AC covered by *f* as

$$AC = \pi\left(\frac{\mu}{2}\right)^2 \tag{12}$$

Only fixations in the same hemifield as *f* were considered, and, naturally, only search processes with more than one fixation per hemifield were evaluated. Pre-studies indicated that virtually all gaze trajectories start at either the top or the bottom of the display and then move downwards to the bottom or upwards to the top respectively. Thus, fixations registered after the first turn in the vertical direction of a scan path were not evaluated in order to avoid an overlapping of two or more successive passages of the search process through the same region.

References

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience, 7*, 66–80.

Bridgeman, B., van der Heijden, A. H. C., & Velichkovsky, B. M. (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences, 17*, 247–292.

Brogan, D., Gale, A., & Carr, K. (1993). *Visual search 2*. London: Taylor & Francis.

Cooper, E. E., Biederman, I., & Hummel, J. E. (1992). Metric invariance in object recognition: A review and further evidence. *Canadian Journal of Psychology, 46*, 191–214.

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

Duncan, J., & Humphreys, G. W. (1992). Beyond the search surface: Visual search and attentional engagement. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 578–588.

Ellis, R., Allport, D. A., Humphreys, G. W., & Collis, J. (1989). Varieties of object constancy. *Quarterly Journal of Experimental Psychology, 41A*, 775–796.

Goldstone, R. L. (1994). Similarity, interactive activation, and mapping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 3–28.

Humphrey, G. K., & Lupker, S. J. (1993). Codes and operations in picture matching. *Psychological Research, 55*, 237–247.

Jacobs, A. M. (1991). Eye movements in visual search: A test of the limited cognitive effort hypothesis and an analysis of the search operating characteristic. In R. Schmid & D. Zambarbieri (Eds.), *Oculomotor control and cognitive processes* (pp. 397–410). Amsterdam: Elsevier North-Holland.

Just, M. A., & Carpenter, P. A. (1987). *The psychology of reading and language comprehension*. Boston: Allyn & Bacon.

Kelter, S., Grötzbach, H., Freiheit, R., Höhle, B., Wutzig, S., & Diesch, E. (1984). Object identification: The mental representation of physical and conceptual attributes. *Memory and Cognition, 12*, 123–133.

Klatzky, R. L., & Stoy, A. M. (1974). Using visual codes for comparisons of pictures. *Memory and Cognition, 2*, 727–736.

Kowler, E. (1990). *Reviews of oculomotor research*. Vol. 4: Eye movements and their role in visual and cognitive processes. New York: Elsevier.

Lawson, R., & Humphreys, G. W. (1996). View specifity in object processing: Evidence from picture matching. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 395–416.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory and Cognition, 24*, 235–249.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology, 26*, 3–67.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences, 6*, 414–417.

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research, 38*, 1805–1815.

Nodine, C. F., Kungel, H. L., Toto, L. C., & Krupinsky, E. A. (1992). Recording and analysis of eye-position data using a microcomputer workstation. *Behavior Research Methods, Instruments, and Computers, 24*, 465–475.

O'Regan, J. K. (1989). Visual acuity, lexical structure, and eye movements in word recognition. In B. A. Elsendoorn & H. Bouma (Eds.), *Working models of human perception* (pp. 261–292). London: Academic Press.

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135–151). Hillsdale: Erlbaum.

Pashler, H. (1987). Detecting conjunctions of color and form: Reassessing the serial search hypothesis. *Perception and Psychophysics, 41*, 191–201.

Pomplun, M., Velichkovsky, B. M., & Ritter, H. (1994). An artificial neural network for high precision eye movement tracking. In B. Nebel & L. Dreschler-Fischer (Eds.), *Lecture notes in artificial intelligence: Proceedings KI-94* (pp. 63–69). Berlin: Springer.

Rao, R., & Ballard, D. H. (1995). Learning saccadic eye movements using multiscale spatial filters. In G. Tesauro, D. Touretzky, & T. Leen (Eds.), *Advances in Neural Information Processing Systems* (pp. 893–900). Cambridge: MIT Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rensink, R., O'Regan, J. K., & Clark, J. J. (1996). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.

Rock, I., Wheeler, D., & Tudor, L. (1989). Can we imagine how objects look from other viewpoints? *Cognitive Psychology*, *21*, 185–210.

Scialfa, C. T., & Joffe, K. M. (1998). Response times and eye movements in feature and conjunction search as a function of target eccentricity. *Perception & Psychophysics*, *60*, 1067–1082.

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

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*, 261–267.

Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, and Computers*, *25*, 137–142.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology*, *21*, 233–283.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 459–478.

Trevarthen, C. B. (1968). Two mechanisms of vision in primates. *Psychologische Forschung*, *31*, 299–337.

Vecera, S. P., & Farah, M. J. (1994). Does visual attention select objects or locations? *Journal of Experimental Psychology: General*, *123*, 146–160.

Velichkovsky, B. M. (1982). Visual cognition and its spatial-temporal context. In F. Klix, J. Hoffmann, & E. v. d. Meer (Eds.), *Cognitive Research in Psychology*. Amsterdam: North Holland.

Velichkovsky, B. M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics and Cognition*, *3*, 199–222.

Williams, D. E., & Reingold, E. M. (submitted). Preattentive guidance of eye movements during triple conjunction search tasks. *Psychonomic Bulletin & Review*.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 419–433.