Modeling Serial Scanning Strategies in Two-Dimensional Object Distributions

Marc Pomplun[1], Elena Carbone[2], Hendrik Koesling[3], Lorenz Sichelschmidt[4], and Helge Ritter[3]

[1]University of Toronto, Department of Psychology,

Toronto, Ontario, Canada M5S 3G3

e-mail: marc@psych.utoronto.ca

phone: (416) 978-3990

[2]Faculty of Psychology; [3]Faculty of Technology; [4]Faculty of Linguistics

University of Bielefeld, P.O.Box 10 01 31, 33501 Bielefeld, Germany

Abstract. Two experiments on visual scanning strategies are reported in this article. In Experiment 1, subjects were presented with random distributions of identical dots. The task was to look exactly once at each dot, with a starting dot specified. This setting allowed a quantitative analysis of scan-path structures and hence made it possible to compare empirical scan paths to computer-generated ones. Five different scan-path models were implemented as computer simulations and the similarity of their scan paths to the empirical ones was measured. Experiment 2 was identical to Experiment 1 with the exception that it used items with different color and form attributes instead of identical dots. Here, the influence of the distribution of colors and forms on empirical scan paths was investigated. The most promising scan-path models of Experiment 1 were adapted to the stimuli of Experiment 2. The results of both experiments indicate that a simple, scan path minimizing algorithm ("Traveling Salesman Strategy (TSS)") is most effective at reproducing human scan paths. We also found an influence of color information on empirical scan paths and successfully adapted the TSS-based model to this finding.

1 Introduction

When we are viewing a complex scene, we can very easily extract all relevant information from it. In some cases, we might hardly be aware of the fact that such scene perception is a *serial* process involving eye movements. The high efficiency of this process is not only based on the high speed of human eye movements, but also on our *strategies* to direct them. These visual scanning strategies have been optimized during a long period of evolution. They are crucial for both our understanding of the human visual system and the construction of technical vision systems. The two experiments reported here focus on a fundamental question: What factors determine the sequence in which we inspect a given set of items?

There are numerous approaches that have tried to provide at least partial answers to this question. Most experiments in the "classic" paradigm of *visual search* use simple, abstract stimuli (e.g., Treisman & Sato, 1990; Wolfe, Cave & Franzel, 1989). Subjects are typically presented with a set of abstract items, such as letters or geometrical objects, and have to decide whether a designated target item is among them. While most studies rely on reaction times and error rates as the only indicators for search performance, several researchers have also investigated the visual scan paths taken during search. Williams & Reingold (in press), for example, used a triple conjunction search task in which the presented items varied in the three dimensions color, form, and orientation. The authors analyzed the proportion of fixations on each distractor type. They found that the highest proportion of fixations was directed towards those distractors that were of the same color as the target. This finding suggests that it is possible to use color information for choosing an efficient scan path: Only the subset of items with the appropriate color has to be searched.

Eye-movement patterns during visual search and viewing images have been used as a basis for modeling visual scanning strategies. Several investigations were conducted by computer scientists intending to "teach" artificial vision systems to behave like the human visual system. Some models of human eye movements in realistic scenes use spatial filters in order to determine an image's most salient points - the ones that are most likely to attract fixations.

These filters may be sensitive to contour features like sharp angles (Kattner, 1994) or to local symmetries (Locher & Nodine, 1987; Menkhaus, 1997; Nattkemper, 1997). Rao and Ballard (1995) proposed a model of parallel search employing time-dependent filters. The location of the first fixation is determined by a coarse analysis (low spatial frequencies) of the given scene, and the following fixations are based on an analysis of increasingly higher spatial frequencies. Another approach (Rimey & Brown, 1991) uses a Hidden Markov Model that is capable of learning efficient eye-movement behavior. It optimizes its scan paths iteratively towards highest efficiency of gathering information in a given scene.

To date, however, even the best attempts at computer vision are far from reaching the performance of the human visual system. One important reason for this fact might be that we do not completely understand the fundamental cognitive mechanisms which guide our attention so efficiently during the exploration of a scene. It seems that the scenes used in the modeling studies mentioned above are too complex to yield insight into these mechanisms. In real-world scenes, a viewer's attention is guided by high-level factors, for instance, by the functional or conceptual relationships between items or the relevance of items to the viewer. It is almost impossible to parametrize such high-level factors and to obtain quantitative, clearly interpretable results from this kind of experiments.

Another problem is that neither the search tasks nor the viewing tasks described above are particularly well-suited to investigate scanning strategies. Gaze trajectories in these tasks yield only coarse information about the exact structure of scan paths, i.e. the sequence of items that receive attention. This is because visual attention can be shifted without employing eye movements. During rapid processes of scanning, small "covert" shifts of attention are likely to occur (for a review, see e.g. Posner, 1980; Wright & Ward, 1994). Therefore, gaze trajectories in search or viewing tasks do not indicate the whole sequence of attended items but – depending on task complexity and item density – only a small subset of it.

In order to obtain more comprehensive information about visual scan paths, we measured subjects' eye movements in a simplified scenario and with a simplified task: Subjects viewed a random distribution of identical dots with one exception: One of these dots – the starting

dot – was conspicuously brighter than the others. The task was to look exactly once at each dot in the display, starting at the specified dot. This task is similar to the one used by Beckwith & Restle (1966), who asked subjects to count large sets of objects. By analyzing reaction times for different types of object configurations, the authors found that subjects grouped the objects into subsets in order to count them efficiently and to avoid mistakes. In our experiments, however, we eliminated any possible interference of a concurrent counting task with the scanning process. Furthermore, we used eye tracking to measure the exact sequence of dots attended to, instead of only global reaction times.

On the one hand, the chosen task is rather artificial. In everyday life we are not used to strictly avoiding repeated attention to the same object, because the "cost" of a redundant eye movement is small (see Ballard, Hayhoe & Pelz, 1995). Although there is evidence for an attentional mechanism called *inhibition of return* (Posner & Cohen, 1984; Klein, 1988; Tipper, Weaver, Jerreat & Burak, 1994), this mechanism alone is not sufficient to generate self-avoiding and complete scan paths as demanded in our task. Therefore, subjects' scan paths are likely to be influenced by cognitive processes operating at a higher level than those being usually involved in natural situations, e.g., free exploration of surroundings. In particular, *path planning processes* are expected to take place, because subjects have to hold in memory which dots they have already visited during task completion (Beckwith & Restle, 1966).

On the other hand, our task enabled us to investigate scan paths purely based on the stimulus geometry, i.e. on the locations of the dots. Neither item features nor relations between them (other than geometrical relations) biased the observed strategies. Moreover, the demand of attending exactly once to each item brought about an enhanced comparability of scan paths taken on the same stimulus. Restricting the analysis to those paths that met this demand, made it easy to define a measure of similarity: The degree of similarity of a path A to another path B was calculated as the number of "jumps" (edges) between dots that appear in path A as well as in path B.

Experiment 1 investigated geometrical regularities of scan paths with the aim of find-

ing possible "mechanisms" that control human scanning strategies. Several models of such mechanisms were developed and implemented as computer simulations. The simulated scan paths were then compared to the empirical ones in order to evaluate the plausibility of the proposed mechanisms. Another important question was: Are there any preferred directions of scan paths? In other words, would the rotation of the stimuli exert an influence on the scan paths?

Experiment 2 went one step further towards a more naturalistic setting: While the subjects' task remained the same as in Experiment 1, the displayed items were given different color and form attributes. Beckwith & Restle (1966) showed that the distribution of color and form attributes influenced the time needed for counting a set of objects, with color having a substantially stronger effect than form. With the help of eye tracking, Experiment 2 directly investigated the influence of color and form on empirical scan paths. In addition, the most successful models of Experiment 1 were refined in such a way as to account for this additional influence.

2 Experiment 1: Geometrical Factors

2.1 Method

2.1.1 Subjects

Twelve subjects from different faculties of the University of Bielefeld took part in Experiment 1. All of them had normal or corrected-to-normal vision, none of them was color-blind or had pupil anomalies. The subjects were paid for their participation.

2.1.2 Apparatus

Stimuli were presented on a 17" ViewSonic 7 monitor. The subjects' eye movements were measured with the OMNITRACK 1 system (see Stampe, 1993). It uses two video cameras as inputs of information about the position of the head relative to the environment and the

position of the pupil relative to the head. This technique allows the subjects to move their head from the straight-ahead position up to 15^o in all directions, and therefore provides natural viewing conditions. Gaze positions are recorded at a frequency of 60 Hz. Fixations are calculated using a speed threshold in a 5-cycle time window, which means that only fixations with a duration of at least 83 ms are detected. The absolute spatial precision of the gaze-position measurement ranges from 0.7^o to 1.0^o. By using a new calibration interface based on artificial neural networks (Parametrized Self-Organizing Maps), we improved the system's precision to 0.5^o. This made it possible to even recruit subjects wearing spectacles (see Pomplun, Velichkovsky & Ritter, 1994).

2.1.3 Stimuli

Subjects were presented with sets of 30 dots (diameter of 0.5 degrees of visual angle) randomly distributed within a square area (18 degrees per side) on a black background. The dots were of the same color (blue), with the *starting dot* being clearly brighter than the others (see Figure 1, left, for a sample stimulus).

———— insert figure 1 about here ————

Five different dot configurations (stimuli) were randomly generated. In order to investigate directional effects on the scan paths, for instance top-to-bottom or left-to-right strategies corresponding to the subjects' direction of reading, each configuration was shown in four different orientations (rotated by 0^o, 90^o, 180^o, and 270^o). This resulted in a set of 20 stimuli used in this study.

2.1.4 Procedure

A written instruction informed the subjects about their task. They had to look at each dot in the display once, beginning with the starting dot. They were told not to miss any dots

or to look at any of them more than once. Furthermore, subjects had to attend to each dot for at least half a second to make sure that they actually performed a saccade rather than a covert shift of attention towards the dot (see above). After task completion subjects were to press a mouse button. The experiment started with two practice trials followed by the eye tracker calibration procedure and the 20 recording trials in random order. Each trial was preceded by a short calibration for drift correction using a single target at the center of the screen.

2.2 Results

The recorded gaze trajectories were converted to item-based scan paths. In other words, the temporal order of attended dots had to be reconstructed, because our analysis was intended to refer to these rather than to fixation points. It turned out that this could not be done automatically. The occurrence of additional fixations (conceivably used by the subjects to get their bearings), imprecise saccades as well as errors in measurement required human postprocessing. Consequently, an assistant – who was naive as to the purpose of the study – did the allocation of fixations to dots manually, on the basis of the individual trajectories with sequentially numbered fixations superimposed on the stimuli. As a result of this semi-interpretative analysis, only 139 of the 240 converted paths (57.9%) turned out to be *acceptable* in terms of the task, i.e. they visited each dot exactly once. The further analyses were restricted to these acceptable paths.

Figure 1 presents a visualization of accumulated data (right) for a sample stimulus (left). Thicker lines between dots indicate transitions (edges) used by a larger number of subjects. The lines are bisected due to the two possible directions to move along these edges. Each half refers to those transitions that started at the dot next to it. Halves representing fewer than three transitions are not displayed for the sake of clarity. Figure 1 illustrates that in the absence of any conspicuous order (upper left part of the display) we find high variability of chosen edges, whereas the linearly arranged dots on the right and at the bottom of the

display were almost always scanned in the same order.

In addition, the quantitative analysis of the data allowed us to investigate the effect of rotating the stimuli: Were there directional influences on the scan paths, for example according to the subjects' reading direction? This was analyzed by comparing similarities (as defined above) between the scan paths of different subjects. If the scan paths for the same stimuli shown in the same orientation were more similar to each other than the ones for different orientations of the same stimuli, this would indicate that the rotation exerted an effect. Actually, the average similarity value for the same orientation turned out to be 19.43 edges per path, while the value between different orientations was 19.42, constituting no significant difference, $t < 1$. Consequently, it was justified not to assume any directional influence. So we averaged the data for each of the five original stimuli over its four different orientations for subsequent analyses.

2.3 Evaluation of Scan-Path Models

Which scanning strategies are promising to model? We developed and tested five different models. Since the empirical data showed no significant dependence on the orientation of the stimuli, none of the models developed below include this factor.

In order to obtain baseline data for the evaluation of the models, we calculated a composite path with maximal similarity to the observed paths ("optimum fit") for each stimulus. An iterative algorithm determined this path within the huge set of all possible acceptable paths, regardless of whether the path actually appeared in any one subject's data. The average similarity of optimum-fit paths to empirical paths turned out to be 21.89, which exceeded the similarity of empirical ones to each other (19.43, cf. above). The calculation of optimum-fit paths also shows that no simulation can produce paths of higher similarity to the empirical data than 21.89, which is considerably lower than the "perfect similarity" value 29 (all acceptable paths consist of 29 edges). This discrepancy demonstrates the high intrinsic variability of scan paths.

Serving as a second baseline, the similarity of completely randomly generated scan paths to the subjects' paths was computed, yielding a value of as low as 1.75. An example of the optimum-fit path as well as the scan paths computed by the models are given in Figure 2, referring to the sample stimulus in Figure 1. The five models that were analyzed are described below.

———— insert figure 2 about here ————

2.3.1 The "Greedy" Heuristic

One model that suggests itself for analysis is based on what can be termed the "Greedy" heuristic. The "Greedy" algorithm always jumps to the dot which is geometrically nearest to the actual "gaze" position and which is still to be visited. Although it produces plausible, locally optimized sections of scan paths, the Greedy strategy has one drawback: On its way through the stimulus, it leaves aside items of high eccentricity. As a consequence, these items have to be "collected" later, which leads to unnaturally long saccades at the end of the scan path. The lack of memory constitutes a fundamental difference from empirically observed strategies. Nevertheless, even this simple model achieves a similarity value of 17.36, indicating that its strategy of always choosing the nearest item, i.e. the local minimization of scan paths, is already tremendously better than the pure random strategy.

2.3.2 The "Traveling Salesman" Algorithm

The shortcoming of the Greedy heuristic motivates the implementation of a TSS ("Traveling Salesman Strategy") algorithm. The Traveling Salesman Problem is a basic paradigm in computer science. A salesman who has to successively visit a certain number of cities wants to save time and energy, so his problem is to find the shortest path connecting all the cities. In the present context, this means that the TSS Model algorithmically minimizes the *global*

length of scan paths rather than just the length of the next jump. However, unlike standard TSS, the paths of this algorithm do *not* return to the starting dot. In this formulation, only the first dot is constrained. The results show that this simulation gets much closer to the actual human strategies than the Greedy heuristic: The similarity value is 20.87, which is fairly close to the optimum-fit value of 21.89. This finding suggests that not only the local optimization of scan paths – as operationalized in the Greedy algorithm – plays an important role in human scan path selection, but also their global optimization.

2.3.3 The Clustering Model

The fact that the TSS Model has yielded the best result so far motivates the investigation of a refined variant of it. This so-called *Clustering Model* is based on the assumption that human scan paths are generated by clusterwise processing of items (cf. Beckwith & Restle, 1966).

The model divides the process of scan-path computation into two steps. In the first step, the configuration of items is divided into clusters. A clustering algorithm maximizes the between-cluster distances and minimizes the within-cluster distances with the help of a cost function. We set the parameters of this iterative procedure in such a way that it generates clusters that may have either "compact" or linear shape. Five to seven clusters with four to seven items each are calculated, which is perceptually plausible (Atkinson, Campbell & Francis, 1976; Miller, 1956).

The second step consists in a TSS algorithm calculating local scan paths of minimal length connecting the dots within each cluster, as well as a global scan path of minimal length connecting all clusters. Afterwards, the within-cluster scan paths are linked together in the sequence specified by the between-cluster scan path. Thus, this model processes all dots within a cluster before proceeding to the next one ("hierarchical TSS").

A similarity analysis showed that the Clustering Model selects paths slightly more similar to the empirically observed ones (21.12) than does the TSS Model. This may suggest that clustering is a component of human scanning strategies.

2.3.4 Using a Self-Organizing Map

When simulating cognitive processes we should also consider *neural network* approaches, because their functional structure is biologically motivated. An appropriate neural paradigm is provided by Kohonen's *self-organizing maps (SOMs)*, which are capable of projecting a high-dimensional data space onto a lower-dimensional one (see Kohonen, 1990; Ritter, Martinetz & Schulten, 1992; Wieners, 1995). SOMs are networks of simulated neurons, usually a one-dimensional chain or a two-dimensional layer. They learn in an unsupervised way to partition a given *feature* or *input space* into disjoint classes or areas and to represent their class by a "typical" feature vector.

The feature space is a region of a classical vector space, where each vector $(v_1, v_2, \ldots, v_n)^T$ shows n different features or input signals. These vectors are presented to the network in random order, and a neuron "fires" if its stored feature, i.e. position vector, is the best approximation to the active input position to the network. Thus we create a map – the neural network – in which each mapped point – each neuron – represents a region of input patterns. If we also ensure that the topology of the input space is preserved, i.e. that neighboring feature vectors are mapped to neighboring neurons, or neighboring neurons stand for similar features, we get a low-dimensional structure representing a possibly high-dimensional input. This is done as follows:

1. Choose a random input vector v from feature space.

2. Select a neuron j with $|v - w_j| \leq |v - w_i|,\ \forall i \neq j$, i.e. the neuron with the best representation w_j of v; this is called the *winner*.

3. Change all neuron weights w_i towards the input vector v, with an adaptive step size h_{ij} that is a decay function of the network distance between neuron i and the winner j. Here, ϵ is an additional global adaptive step size parameter.

$$w_i^{\text{new}} = w_i^{\text{old}} + \epsilon \cdot h_{ij} \cdot (v - w_i)\,,\ \epsilon \in [0, 1] \tag{1}$$

The change of neuron weights adjusts w_i^{old} towards a better representation vector and the smooth distribution of change around the winner produces the desired topology preservation. In our case, we are only interested in a mapping from discrete 2D points onto a linear chain representing fixation order. Hence, the feature space is only the discrete set of dot positions in $\mathbb{R}^2$, one of them labeled as starting dot. Since the chain must begin at the starting dot, the first neuron is defined to be the winner if the starting dot is presented, irrespective of the actual feature-vector difference. In order to make sure that all dots are represented by neurons after the learning process, the network contains a number of additional nodes. Now, the probability to skip a dot is very low, but more than one neuron may become mapped to the same position. This must be resolved by a post-processing step to extract the simulated scan path from the chain of neurons.

The paths generated by this model look quite natural at first sight. Their similarity to the human ones, however, is substantially lower (19.45) than the results obtained by the TSS-based models.

2.3.5 A Scan-Path Model on the Basis of Receptive Fields

Another model uses neurons with a special type of *receptive fields* which are assumed to exist in the visual cortex. In a neural network, natural or artificial, the term *receptive field* stands for the region of input space that affects a particular neuron (see, e.g., Hubel & Wiesel, 1962; Lennie, Trevarthen, van Essen & Wässle, 1990). Further, the influence of stimuli in this region is not necessarily homogeneous, but dependent on variables such as the distance of the input vector from the center of the region. There may also be excitatory and inhibitory subregions, where a stimulus will respectively increase or decrease the activation of the neuron.

In our model, the receptive fields consist of an inhibitory axis and two laterally located, excitatory areas of circular shape (see Figure 3). We use 100 000 receptive fields that are randomly distributed over the input space. Their sizes vary randomly between 80% and 120% of the relevant input space, i.e. the whole area in which dots are presented. There are

eight possible orientations which are randomly assigned to the receptive fields. It is obvious from this description that the receptive fields are closely packed and overlap each other.

———— insert figure 3 about here ————

The activation of a neuron is highest if no dot is in the inhibitory region of the neuron's receptive field and as many dots as possible are in the lateral excitatory regions. The neuron with the highest activation (the "winner" neuron) thus indicates the "clearest" linear gap between two laterally located accumulations of dots. Therefore, the inhibitory axis of this neuron's receptive field can be considered to indicate a perceptually plausible bisection of the stimulus.

This first bisection separates the set of dots into two subsets. Each subset serves as the input to a new group of neurons with smaller receptive fields, calculating further bisections. This procedure is repeated until none of the sections contains more than four dots, since the number four is a plausible estimate of the number of dots that can be perceived at the same time (cf. Atkinson, Campbell & Francis, 1976; Miller, 1956). Figure 4 (left) presents the model's hierarchical partitioning of the sample stimulus previously shown in Figures 1 and 2. The bisections are visualized by straight lines with numbers indicating their level in the hierarchy. The calculation of this structure – a binary tree structure – is our attempt to simulate a subject's perceptual processing of the visual scene.

———— insert figure 4 about here ————

Finally, the scan path is derived by a TSS algorithm calculating the shortest scan path that begins at the starting dot. In the present context, however, it is not the geometrical distance that is minimized, but a linear combination of the geometrical distance and the *tree distance* between the dots. The tree distance between two dots A and B is the number of steps

that have to be taken within the tree structure to get from the subset (leaf) containing A to the one containing B. If we choose the coefficients of the linear combination in such a way that the tree distance is more relevant than the geometrical distance, the model generates the scan path shown in Figure 4 (right). It strictly follows the hierarchical tree structure, which leads to geometrical deviations.

As long as the model's linear coefficients are chosen such that the tree distance exerts a significant effect, neither the appearance of the simulated scan paths nor their calculated similarity to the empirical paths is convincing. When balancing the weights of the tree distances and the geometrical distances, we obtained scan paths with a similarity to the human paths of 18.73. This approach, at least in this rather simple form, does not yield more plausible scan paths than does the TSS Model. The hierarchical partitioning does not seem to correspond to human strategies.

———— insert figure 5 about here ————

Figure 5 displays a summary of the accuracies with which the various models simulate human scanning patterns, plus the optimum-fit value. A one-way analysis of variance (ANOVA) was conducted, excluding the optimum-fit value, which was a global value that did not vary across subjects. The ANOVA revealed a significant main effect showing differences between the five similarity values, $F(4, 44) = 32.338, p < 0.001$. Pairwise t-tests with Bonferroni-adjusted probability values were conducted to examine these differences more closely. It turned out that all of the models reached significantly higher similarity than the Greedy heuristic, all $t(11) > 3.620, p < 0.005$. The Receptive Fields Model did not significantly differ in results from the Kohonen Model. These two models, in turn, were outperformed by both the TSS Model and the Clustering Model, all $t(11) > 4.842, p < 0.006$. Finally, the TSS Model showed no significant difference from the Clustering Model.

2.4 Discussion

Basically, the results of Experiment 1 show that the simple TSS Model and Clustering Model yield better scan paths than the neural models, and that even the "primitive" Greedy algorithm is not far behind. This finding should not be interpreted as evidence for a general incapability of neural models to explain scan-path mechanisms. The neural models tested in Experiment 1 were of a very primitive nature. Multi-layered and/or hierarchical networks might be able to generate scan paths more similar to the empirically observed ones. Moreover, discretion is advisable in the interpretation of these data, since they are based on only *five* different dot configurations. Nevertheless, from the results above we can conclude that it is difficult to generate better simulations of human scan paths than those created by the simple TSS-based models. Thus the minimization of scan-path length seems to be a basic principle in human scanning strategies.

Another important result of Experiment 1 is the independence of scan paths from rotations of the stimuli. In other words, the order in which a subject scans a set of dots does not seem to change when the display is rotated by 90, 180, or 270 degrees. It is well-known from visual search experiments (e.g. Zelinsky, 1996; Pomplun, 1998) that subjects prefer to scan a display according to their reading direction, if they are allowed to freely choose the starting point. However, this was not observed in the present study. A possible reason is that the specified starting point induced rotation-invariant scanning strategies.

3 Experiment 2: Color and Form Attributes

The objective of Experiment 2 was to investigate the influence of color and form attributes on scan paths. Subjects were presented with distributions of geometrical objects (squares, triangles, and circles) in different colors (yellow, blue, and green). We might expect color and form to influence the structure of chosen scan paths, because subjects are likely to take advantage of the additional structural information. As their main concern is to remember

which of the items they have already visited, the introduction of color and form features might allow them to use perceptual groups of identical attributes as "scan-path units" which need less effort to remember than single items. This assumption is supported by the results of Beckwith & Restle's (1966) counting task. They found shorter reaction times when object colors were clustered, i.e. different colors were spatially segregated. They also found an analogous – but weaker – effect for clustering the objects by form. To examine potential corresponding effects on scan-path structure, the stimuli in Experiment 2 had three different levels of color and form clustering.

If subjects make use of the color and/or form information, these effects should be integrated into the models. It is plausible to assume that the attributes lead to a reduction in scan-path variability, which could enable the models to yield better results than in Experiment 1. Here we took advantage of the findings of Experiment 1: Since the paths generated by the TSS and the Clustering Models were most similar to the empirical data, we focused on the adaptation of these two approaches to the stimuli used in Experiment 2.

In order to make the two experiments easier to compare, the design and procedure of Experiment 2 corresponded to Experiment 1. Based on the results of Experiment 1, however, we did not further investigate the effect of stimulus rotation. In addition, the introduction of color and form attributes required to change the way of indicating the starting item. In Experiment 2, we used a *dynamic* cue, namely a flashing red circle around the starting item, appearing for a short period after stimulus onset. This method of marking the starting item did not alter its color or form attributes. The subjects' task was the same as in Experiment 1, namely to look once, and only once, at each item.

3.1 Method

3.1.1 Subjects

Twenty new subjects from different faculties of the University of Bielefeld took part in Experiment 2. All of them had normal or corrected-to-normal vision, none of them was

color-blind or had pupil anomalies. They were paid for their participation.

3.1.2 Apparatus

The apparatus was the same as in Experiment 1.

3.1.3 Stimuli

The stimuli consisted of 30 simple geometrical items (diameter of about 0.7^o) of three different colors (fully saturated blue, green, and yellow) and three different forms (triangle, square, and circle) on a black background. Their spatial distribution was randomly generated within a display of 18^o by 18^o with a minimum distance of 1.5^o between the centers of neighboring items in order to avoid item overlap or contiguity (see Figure 6).

In each stimulus array, there were a balanced number of items with each color and form. The distribution of colors and forms was not always homogeneously random, as they were clustered to varying degrees in most trials. To explain the clustering algorithm, a formalized description of the stimulus patterns is necessary: A pattern is a set of N items (objects)

$$\mathbf{o}^{(n)} = \begin{pmatrix} o_x^{(n)} \\ o_y^{(n)} \\ o_c^{(n)} \\ o_f^{(n)} \end{pmatrix} \quad , \quad n = 1, \ldots, N, \tag{2}$$

where $(o_x^{(n)}, o_y^{(n)})$ is the pixel position of the item's center on the screen, $o_c^{(n)}$ is the item's color (1 = blue, 2 = green, 3 = yellow), and $o_f^{(n)}$ is the item's form (1 = square, 2 = triangle, 3 = circle).

Now the variable *color clustering* α_c is introduced. It can be defined as the ratio between the mean distance $\bar{d}_{f,dif}$ between all pairs of items with different colors and the mean distance $\bar{d}_{f,id}$ between those with identical colors:

$$\alpha_f = \frac{\bar{d}_{f,dif}}{\bar{d}_{f,id}} \tag{3}$$

$$\bar{d}_{f,dif} = \frac{\sum_{n_1=1}^{N} \sum_{n_2=n_1+1, o_c^{(n_1)} \neq o_c^{(n_2)}}^{N} d(n_1, n_2)}{\sum_{n_1=1}^{N} \sum_{n_2=n_1+1, o_c^{(n_1)} \neq o_c^{(n_2)}}^{N} 1} \tag{4}$$

$$\bar{d}_{f,id} = \frac{\sum_{n_1=1}^{N} \sum_{n_2=n_1+1, o_c^{(n_1)} = o_c^{(n_2)}}^{N} d(n_1, n_2)}{\sum_{n_1=1}^{N} \sum_{n_2=n_1+1, o_c^{(n_1)} = o_c^{(n_2)}}^{N} 1} \tag{5}$$

$$d(n_1, n_2) = \sqrt{(o_x^{(n_1)} - o_x^{(n_2)})^2 + (o_y^{(n_1)} - o_y^{(n_2)})^2} \tag{6}$$

For example, $\alpha_c = 2$ would mean that, on average, items of different colors are twice as distant from each other than items of the same color. In our setting of 30 items and three different colors this would correspond to a strongly segregated distribution containing large single-colored areas. $\alpha_c = 1$ would mean that there is no clustering at all. We define the parameter form clustering (α_f) analogously.

———— insert figure 6 about here ————

Figure 6 illustrates the correspondence between α_c, α_f, and the distribution of colors and forms in four different sample stimuli. While the images (a) to (c) display stimuli with increasing color clustering and no form clustering, picture (d) shows a stimulus with high color *and* high form clustering. These examples demonstrate an important feature of α_c and α_f for the present experiment: Color and form clustering can be varied independently from each other. Even in an array with both high color and form clustering, the separate concentrations of colors and forms usually do not correspond.

An iterative algorithm for generating color and form distributions with given parameters of color clustering α_c and form clustering α_f can easily be implemented. Starting with a random distribution, this algorithm randomly selects pairs of items and exchanges their color or form attributes – if this exchange shifts the distribution's clustering levels towards the given parameters. The algorithm terminates as soon as the difference between the actual and the desired α_c and α_f falls below a certain threshold, which was set to 0.05 in the present experiment.

Three different levels of color and form clustering were used, namely "no clustering" (1.0), "weak clustering" (1.3), and "strong clustering" (1.7). Examples of stimuli at these levels can be seen in Figure 6. The nine possible combinations of different levels of color and form clustering constituted the stimulus categories of Experiment 2. Five stimuli of each category were used, leading to a total of 45 different stimuli. For two seconds after stimulus onset, a flashing red circle was shown around one of the items, signifying the starting item which was always the same across subjects for each given stimulus.

3.1.4 Procedure

The procedure was the same as in Experiment 1, except that 45 trials were conducted in random order.

3.2 Results

As in Experiment 1, an assistant converted the recorded fixations into scan paths connecting the items in the display. The assistant was only shown the *locations* of the items, but not their color or form *attributes*. Just as in Experiment 1, the superimposed visualization of the subject's fixations and their temporal order allowed the assistant to mark the individual scan path item by item.

The proportion of acceptable paths was 93.3%, which was substantially higher than in Experiment 1 (57.9%). Apparently, the additional color and form information helped the subjects not to "get lost" during task completion. The individual features of the items seemed to facilitate reliable memorization and recognition. The incorrect paths were approximately equally distributed among the nine categories of stimuli, and so were excluded from the analysis.

For a qualitative analysis, we can inspect the calculated scan paths of maximal similarity to the empirical ones (optimum fit). The upper row of Figure 7 presents these paths for an unclustered, a strongly color-clustered, and a strongly form-clustered stimulus. There is

no obvious evidence for the influence of color or form attributes on the subjects' strategy. Although there are some longer sections of scan paths exclusively visiting items of the same color or form, these items are always located closely together. This qualitative finding suggests that the location of items remains the most important factor to determine the structure of scan paths.

———— insert figure 7 about here ————

The quantitative investigation of the effects of color and form required a measure of color and form clustering within the empirically observed scan paths. An appropriate choice seemed to be the mean *runlength* with regard to these dimensions. In the present context, a "run" is defined as a sequence of items of the same color or form within a scan path. The runlengths ranged from one to ten, as there were always exactly ten items of each color and form in each stimulus array. In order to calculate a mean runlength across multiple paths, we employed a *weighted* mean to equally account for every single transition between items. Since longer runs comprise more transitions, we weighted each run with its runlength.

How can we test whether this measure indeed reflects the influence of item attributes rather than the geometrical structure of the stimulus? Even a subject who completely ignores color and form would generate longer runs with increasing strength of clustering in the stimulus. This is due to the fact that, according to the results of Experiment 1, subjects seem to prefer short scan paths, so neighboring items are disproportionately likely to be scanned successively. Clustering moves items with the same features closer together and thus increases the average color and form runlengths in empirical scan paths.

Fortunately, there is a "color and form blind" model, which yields paths of high similarity to the empirical ones, namely the TSS Model. We applied the TSS Model to each stimulus used in Experiment 2 to generate baseline predictions about the color and form runlengths in that stimulus. In a comparative analysis of observed scan paths, we then divided all color and form runlengths by the TSS-predicted runlengths, obtaining *relative* runlengths. Rather

than absolute runlengths, relative runlengths reveal the influence of item attributes on a subject's scan path. Relative color runlength 1, for instance, would indicate no difference to the TSS Model and thus no influence of color attributes on empirical scan paths. Longer relative runlengths would indicate increasing influence.

Figure 8 shows the subjects' relative color and form runlengths at the three levels of color and form clustering respectively. A two-way ANOVA revealed significant main effects of "dimension" (color vs. form), $F(1, 19) = 9.967, p = 0.005$, and "strength of clustering" (no vs. weak vs. strong clustering), $F(2, 38) = 4.765, p = 0.014$. There was also a significant interaction between the two factors, $F(2, 38) = 5.807, p = 0.007$, which was due to the fact that clustering had a significant effect on relative color runlength, $F(2, 38) = 5.556, p = 0.008$, but not on relative form runlength, $F(2, 38) = 2.359, p = 0.108$. For the color dimension, pairwise t-tests with Bonferroni-adjusted probabilities revealed a significant difference between "no clustering" (1.092) and "strong clustering" (1.213), $t(19) = 3.936, p = 0.003$. The differences to the "weak clustering" condition (1.131), however, were not significant, both $t(19) < 1.665, p > 0.336$. Finally, the overall relative color runlength (1.145) differed reliably from the value 1, $t(19) = 3.406, p = 0.003$, whereas overall relative form runlength (0.999) did not, $t < 1$.

———— insert figure 8 about here ————

All in all, these findings suggest that subjects use color information to guide their scan paths, because the color runlength in their scan paths is longer than predicted by the TSS Model. This effect of color guidance increases with the strength of color clustering in the stimuli. Subjects' form runlengths, however, do not exceed the predicted ones and do not depend on form clustering in the stimuli. Hence, we assume that subjects do not use form information when performing the task.

3.3 Refinement of Scan Path Models

The results of Experiment 1 motivated the adaptation of both the TSS Model and the Clustering Model to stimuli containing items with color and form attributes. Since the Clustering Model can be viewed as a refinement of the TSS Model, we started with adjusting the TSS Model.

How can we bias the TSS algorithm to react to color in the same way as the average subject does? Basically, it should still calculate scan paths of minimal length, but in doing so should weight the purely geometrical distances by the color (in)congruence ("color distance") between the neighboring items. This weighting is achieved by multiplying the distance between two items of different colors by a constant factor (*color weight*) and leaving the distance between items of the same color identical to their geometrical distance.

Understandably, the algorithm's behavior will strongly depend on the color weight. It is obvious that a color weight of 1 would lead to a standard TSS algorithm, which would not be influenced by color information at all. In contrast, a color weight of, for instance, 1000 would make the algorithm use a minimum of transitions between different colors. At first, the algorithms would visit all items of the starting item's color A, then inspect all items of color B, and finally those of color C. Within the color groups it would behave like a conventional "traveling salesman" algorithm, taking the shortest passages possible. By adjusting the color weight it is possible to control the influence of colors and hence the average color runlength produced by the TSS algorithm. Since the goal is to adapt the TSS Model to the empirical data, i.e. to produce the same runlengths as generated by the subjects, the color weight needs to be adjusted for the best match.

What is the response of the TSS algorithm to increasing the color weight? As might be expected, it reveals a tendency towards the avoidance of transitions between items of different colors, because these transitions increase the overall length of the scan path above proportion. Figure 9 shows color runlength as a function of the color weight ranging from 1.0 to 1.5. The mean runlengths are displayed separately for each of the three levels of color

clustering in the stimuli. Additionally, the empirically obtained runlengths for these levels are shown as horizontal lines.

———— insert figure 9 about here ————

We find the TSS runlengths to increase approximately linearly with increasing color weight. Higher levels of clustering lead to steeper runlength slopes. Surprisingly, we cannot determine a single value of the color weight to yield the "best" runlengths for all levels of clustering. For each level, the intersection between the runlength curve of the TSS Model and the subjects' runlength occurs at a different color weight. These are the values 1.11 for the "no clustering" condition, 1.23 for "weak clustering", and 1.33 for "strong clustering". Loosely speaking, the subjects seem to apply higher color weights with increasing color clustering in the display.

In light of these data, we must consider if the introduction of color weights, as described above, is an adequate method of modeling the observed color effects. Since the model needs different color weights depending on the strength of color clustering, we have to pose the question whether this approach is really plausible. An alternative idea would be to assign color weights for *sequences* of transitions rather than for single transitions. Starting with the value 1.0, the color weight for a whole group of successive transitions within the same color would decrease linearly with the number of items in that group. This arrangement would make the choice of longer color runs increasingly attractive to the TSS algorithm. However, testing this approach yielded a result that was in some respects inverse to the previous one: For increasing levels of color clustering, the alternative method needed *decreasing* weights for long color runs in order to produce scan paths of good similarity to the empirical ones.

To solve this problem, we could try to combine the two approaches or to use more complex functions to determine the relevant distances between items. A basic rule of modeling is, however, to use as few freely adjustable parameters as possible. The more of these parameters are integrated into a model, the easier it is for the model to fit any data, which weakens

potential conclusions drawn from the model's performance. Therefore, we kept our desired model – the *Color TSS Model* – as simple as possible by extending our initial approach. Figure 9 suggests a linear dependence of the required color weight on the strength of color clustering. Recall that the three levels of color clustering correspond respectively to the values 1.0, 1.3, and 1.7 on the cluster measure (α_c), with a maximum deviation of 0.05. We determined the parameters of the linear function to yield runlengths most similar to the empirical ones:

$$\text{color weight} = 0.264 \ \alpha_c + 0.799 \tag{7}$$

Three sample paths generated by the resulting Color TSS Model are shown in the lower row of Figure 7. In fact, some subtle differences to the TSS paths (middle row) can be found indicating that the new model better corresponds to the empirically observed strategies (upper row). A similarity analysis showed that the scan paths generated by the Color TSS Model were indeed more similar to the observed patterns (similarity value 19.51) than those produced by the unadjusted TSS Model (19.18).

Finally, we adapted the *Clustering Model* to the stimuli of Experiment 2. This was achieved analogously to the adaptation of the TSS Model. We implemented the stimulus-dependent color weight for both the first step (calculation of clusters) and the second step (cluster-based TSS) performed by the Clustering Model. The same functional relationship between color weight and color clustering in the stimulus which was calculated for the Color TSS Model (Equation 7) led to optimal runlength values for the Clustering Model as well.

The improvement of the Clustering Model achieved by its adjustment to color attributes turned out to be considerably smaller than for the TSS Model. We measured the similarity to the subjects' scan paths in Experiment 2 for both the unadjusted Clustering Model and the new *Color Clustering Model*. While the Color Clustering Model produced results slightly more similar to the empirical paths (19.03) than those generated by the original Clustering Model (18.95), it could neither compete with the TSS Model nor with the Color TSS Model.

Figure 10 shows a survey of similarities between the models' paths and the empirical ones, in ascending order. Additionally, the values for the Greedy Model (17.25) and the optimum fit (20.65) are presented. A one-way analysis of variance showed a significant main effect, i.e. differences between the five models, $F(4, 76) = 65.743, p < 0.001$. Pairwise t-tests with Bonferroni-adjusted probabilities revealed that, as in Experiment 1, the Greedy heuristic yielded a significantly lower value than all other models, all $t(19) > 9.432, p < 0.001$. While there were no reliable differences between the Clustering Model, the Color Clustering Model, and the TSS Model, the Color TSS Model produced a significantly higher value than all its competitors, all $t(19) > 3.508, p < 0.024$.

——————— insert figure 10 about here ———————

4 General Discussion

Experiment 1 provided us with some fundamental insights into visual scanning strategies. First, the results suggest that the present scanning task does not induce any preferred direction for scanning, e.g. top to bottom or left to right. The reason might be that using a random distribution of items and a specified starting point makes this kind of schematic strategy rather inefficient. Second, the five scan-path models differ substantially in their abilities to reproduce empirical scan paths. The TSS Model and the closely related Clustering Model yield clearly better results than their competitors, showing that the minimization of overall scan-path length might be an important determinant of human gaze trajectories. This does not imply that artificial neural networks are unable to generate human-like scan paths. Further research is necessary to determine adequate structures of neural networks for modeling human scanning behavior.

Experiment 2 confirmed the results of Experiment 1. Moreover, it yielded information about the influence of color and form attributes on empirical scan paths. While subjects

seem to ignore the items' forms, they use the items' colors in the scanning process, as demonstrated by disproportionately long color runs in their scan paths. The influence of color grows with increasing strength of color clustering in the stimulus. This color guidance is possibly employed to reduce memory load for generating self-avoiding scan paths. It requires less effort to keep in memory the clusters already visited and the items visited within the current cluster than to keep in memory the visited area of the display on the basis of single items, especially if suitably large clusters are available. The perceptual grouping by form, however, does not seem to be strong enough to significantly influence the subjects' scanning strategies.

These results are in line with those obtained by Beckwith & Restle (1966), who found that clustering items by color or form reduced the time needed to count them, with color having a substantially stronger effect than form. Our findings are also compatible with eye-movement studies investigating saccadic selectivity in visual search tasks (e.g. Williams & Reingold, in press). Distractor items that are identical to the target in any dimension attract more fixations than others. Again, this effect is disproportionately large for the color dimension.

Conclusions concerning differences across dimensions, however, may not generalize beyond the set of items used in the experiment. In Experiment 2, other item sets, e.g. bars in different orientations, might have led to form-biased scan paths. Reducing the discriminability between colors would at some point have eliminated the influence of color on the scan paths. From the present data we can only confidently conclude that fully saturated colors affect scanning strategies, whereas regular geometrical forms do not.

Disproving our assumption, the effect of color on scan paths did not reduce their variability. The optimum-fit value was actually lower in Experiment 2 (20.65) than in Experiment 1 (21.89), indicating higher differences between individual paths in Experiment 2 than in Experiment 1. This is probably due to the fact that, in Experiment 2, the effect of color varies considerably between subjects, which increases the range of applied strategies. The large standard error for relative color runlengths (see Figure 8) illustrates these individual differences.

Based on the empirically obtained color effect, the TSS and Clustering Models have been adapted to colored items. When using a weight for transitions between items of different colors to achieve this adaptation, this weight has to increase linearly with the strength of color clustering in the stimuli. Loosely speaking, the effect of color attributes on empirical scan paths seems to vary linearly with the amout of color clustering in the stimulus. We found the adaptation of the TSS Model – the Color TSS Model – to be a small but clear improvement over the standard TSS Model. The Color TSS Model is also superior to the Clustering Model and its refined variant, the Color Clustering Model, and hence can be considered the "winner" of our competition.

Neither Experiment 1 nor Experiment 2 showed a significant difference in performance between the "color-blind" TSS and Clustering Models. Only the adaptation to colored items was achieved more effectively for the TSS Model. This does not mean that human subjects do not apply clustering strategies. In fact, the "winning" Color TSS Model performs clustering itself, since it fits its scan paths to the color clusters given in the stimulus. While this method of clustering could to some extent be adapted to human strategies, this could not be done with the more complex and less flexible algorithm used by the Clustering Model.

Altogether, the difficulties encountered in surpassing the plain TSS Model indicate that the geometrical optimization of scan paths, i.e. the minimization of their length, is the main common principle of human scanning strategies under the given task, even when additional color and form information is provided. Further research is needed to verify the applicability of the findings to real-world situations. For this purpose, stimuli could be photographs of real-world scenes – like the breakfast scenes used by Rao & Ballard (1995) – and the task could be to memorize the scene or to detect a certain item. Will scan-path minimization still be the dominant factor to determine the scan-path structure? Will the scanning strategies be influenced by the distribution of color and form attributes or by figural or functional interpretation? Answering these questions will be an important step towards understanding the principles our visual system employs when creating gaze trajectories. In this context, the present work can be considered a starting point for a promising line of research.

Acknowledgements. We are grateful to Thomas Clermont, Peter Munsche and Karl-Hermann Wieners for their assistance in conducting the experiments. Moreover, we would like to thank Boris M. Velichkovsky, Alex Gunz, and Jiye Shen for their helpful comments on earlier drafts of this paper. The research reported here was funded by the Deutsche Forschungsgemeinschaft (CRC 360, project B4).

References

Atkinson, J., Campbell, F.W. & Francis, M.R. (1976). The magical number 4 ± 0: A new look at visual numerosity judgements. *Perception*, 5, 327–334.

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

Beckwith, M. & Restle, F. (1966). Process of enumeration. *Psychological Review*, 73, 437–444.

Hubel, D.H. & Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology (London)*, 160, 106–154.

Kattner, H. (1994). Using attention as a link between low-level and high-level vision. *Technical report*, Department of Mathematics and Computer Science, Technical University of Munich.

Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature*, 334, 430–431.

Kohonen, T. (1990). The Self-Organizing Map. *Proceedings of IEEE*, 78, 1464–1480.

Lennie, P., Trevarthen, C., van Essen, D. & Wässle, H. (1990). Parallel processing of visual information. In L. Spillmann & J.S. Werner (Eds.), *Visual Perception: The Neurophysiological Foundations*, 103–128. San Diego: Academic Press.

Locher, P. & Nodine, C.F. (1987). Symmetry catches the eye. In A. Levy-Schoen & J. O'Reagan (Eds.), *Eye Movements: From Physiology to Cognition*, 353–361. North Holland: Elsevier Science Publishers.

Menkhaus, G. (1997). Präattentive Fokussierungspunktgenerierung zur Objekterkennung mit künstlichen neuronalen Netzen. *Diploma thesis*, Technical Faculty, University of Bielefeld, Germany.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.

Nattkemper, T.W. (1997). Untersuchung und Erweiterung eines Ansatzes zur modellfreien Aufmerksamkeitssteuerung durch lokale Symmetrien in einem Computer Vision System. *Diploma thesis*, Technical Faculty, University of Bielefeld, Germany.

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*, 63–69. Berlin: Springer.

Pomplun, M. (1998). *Analysis and Models of Eye Movements in Comparative Visual Search.* Göttingen: Cuvillier.

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

Posner, M.I. & Cohen, Y.A. (1984). Components of visual orienting. In H. Bouma & D.G. Bouwhuis (Eds.), *Attention and Performance*, 10, 531–554. Hillsdale, NJ: Erlbaum.

Rao, R.P.N. & 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*, 893–900. Cambridge: MIT Press.

Rimey, R.D. & Brown, C.M. (1991). Controlling Eye Movements with Hidden Markov Models. *International Journal of Computer Vision*, 7 (1), 47–65.

Ritter, H., Martinetz, T. & Schulten, K. (1992). *Neural Computation and Self-Organizing Maps.* Reading, MA: Addison-Wesley.

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.

Tipper, S.P., Weaver, B., Jerreat, L.M. & Burak, A.L. (1994). Object-based and environment-based inhibition of return of visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, 20 (3), 478–499.

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

Wieners, K.H. (1995). Implementierung und Evaluation eines Modells zur ontogenetischen Entwicklung von Merkmalskarten im Visuellen Kortex bei Primaten. *Diploma thesis*, Technical Faculty, University of Bielefeld, Germany.

Williams, D.E. & Reingold, E.M. (in press). 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.

Wright, R.D. & Ward, L.M. (1994). Shifts of visual attention: An historical and methodological overview. *Canadian Journal of Experimental Psychology*, 48 (2), 151–166.

Zelinsky, G.J. (1996). Using eye saccades to assess the selectivity of search movements. *Vision Research*, 36, 2177–2187.

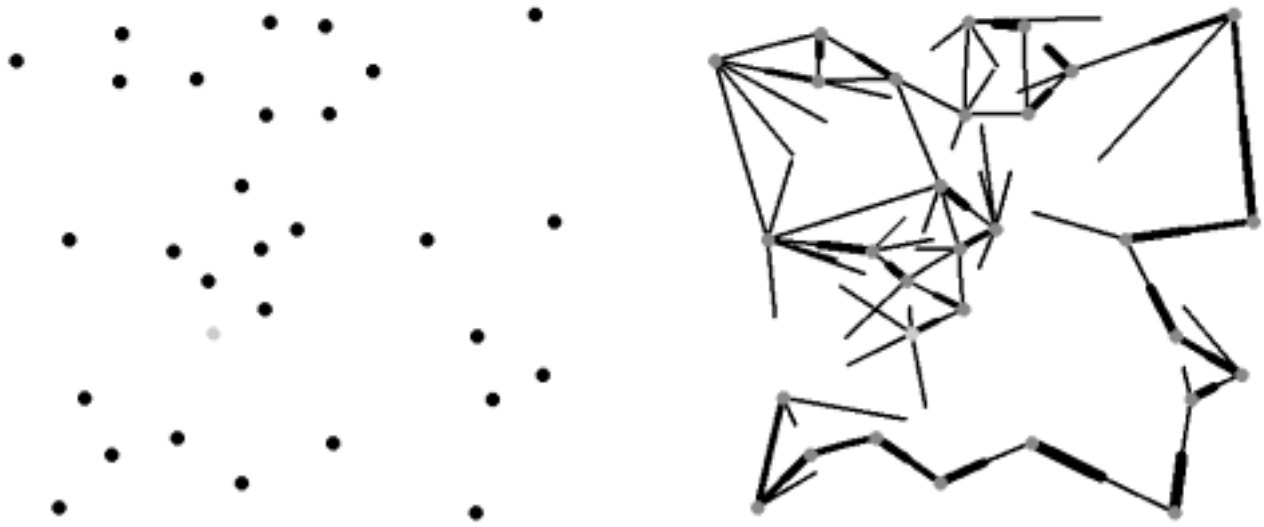

Figure 1: Sample stimulus (left) and corresponding visualized results (right)

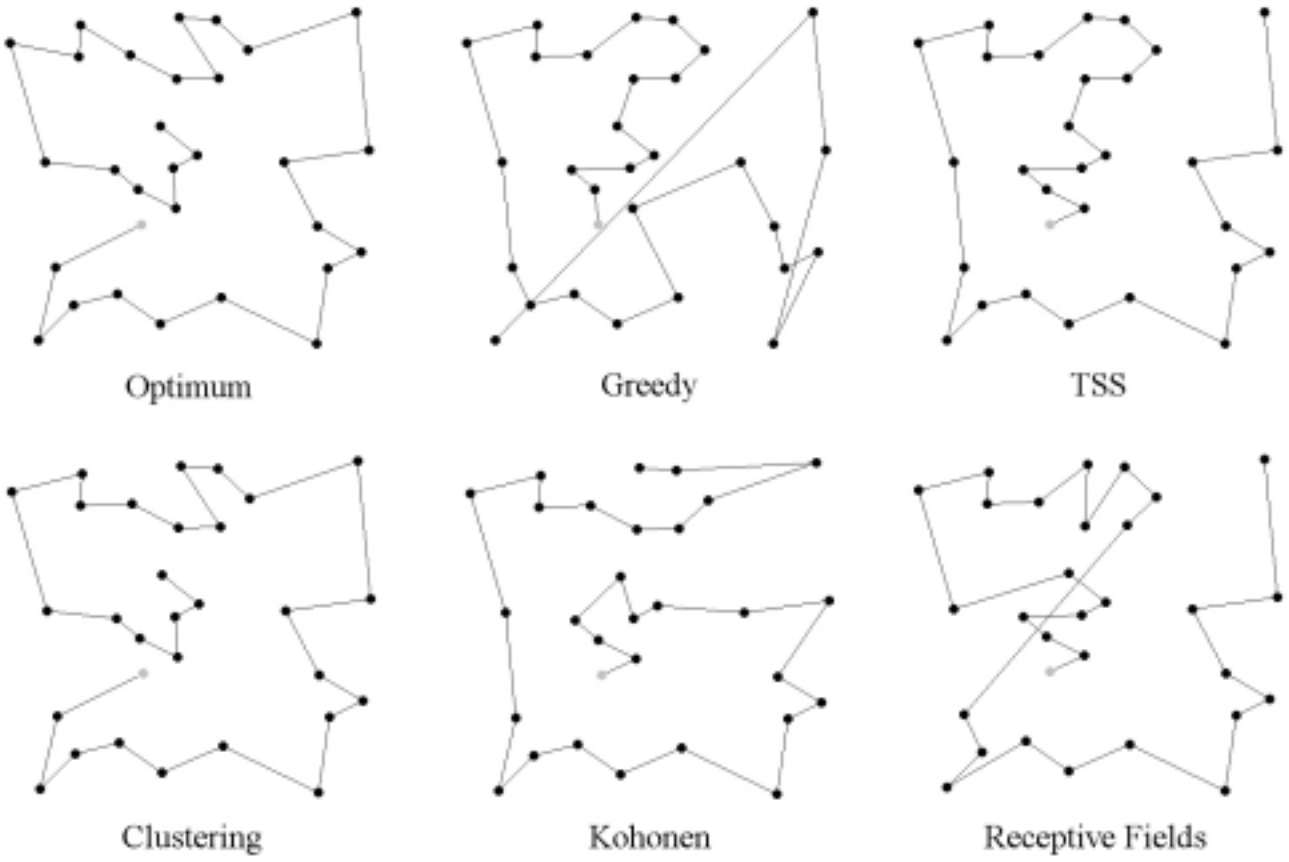

Figure 2: Scan paths generated by the different models, plus the optimum-fit path, for the sample stimulus shown in Figure 1.

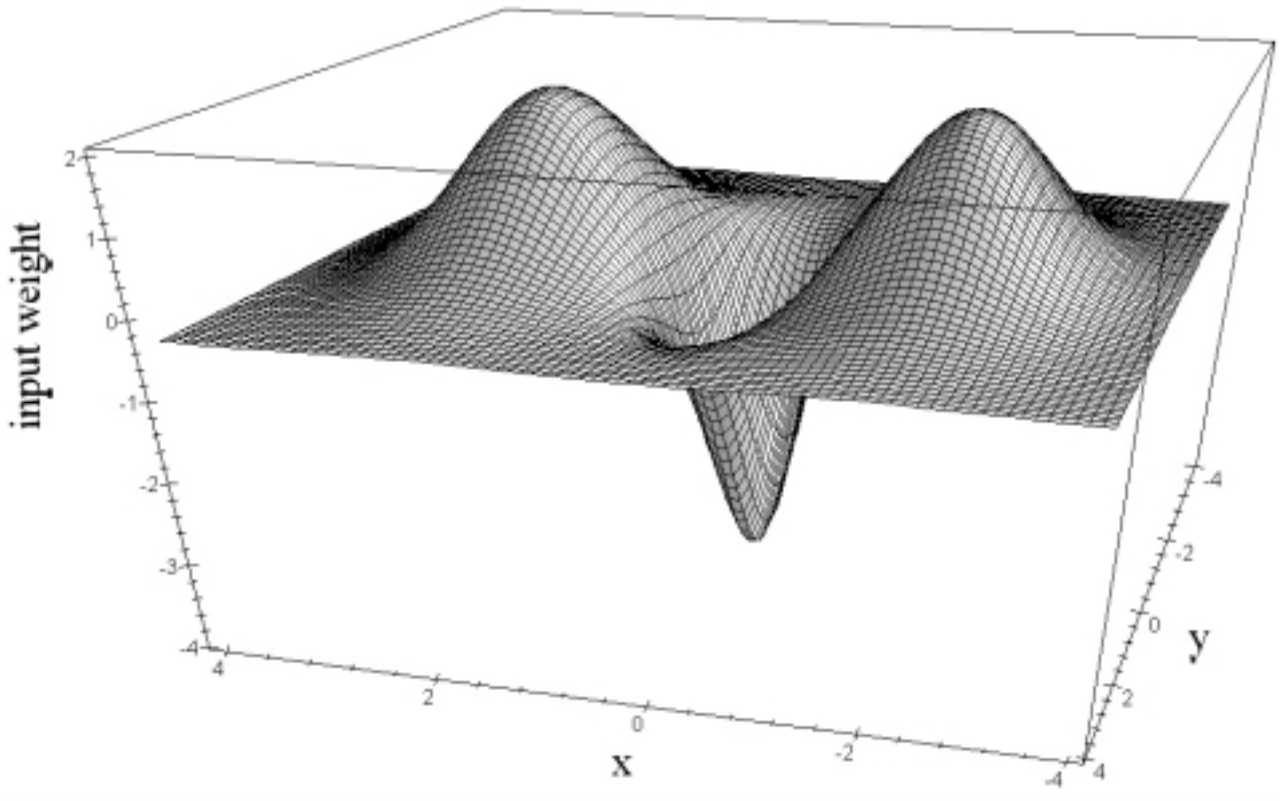

Figure 3: Illustration of the simulated receptive fields. The planar input space is represented by the dimensions x and y; positive values of input weight signify excitatory connections, negative values signify inhibitory connections.

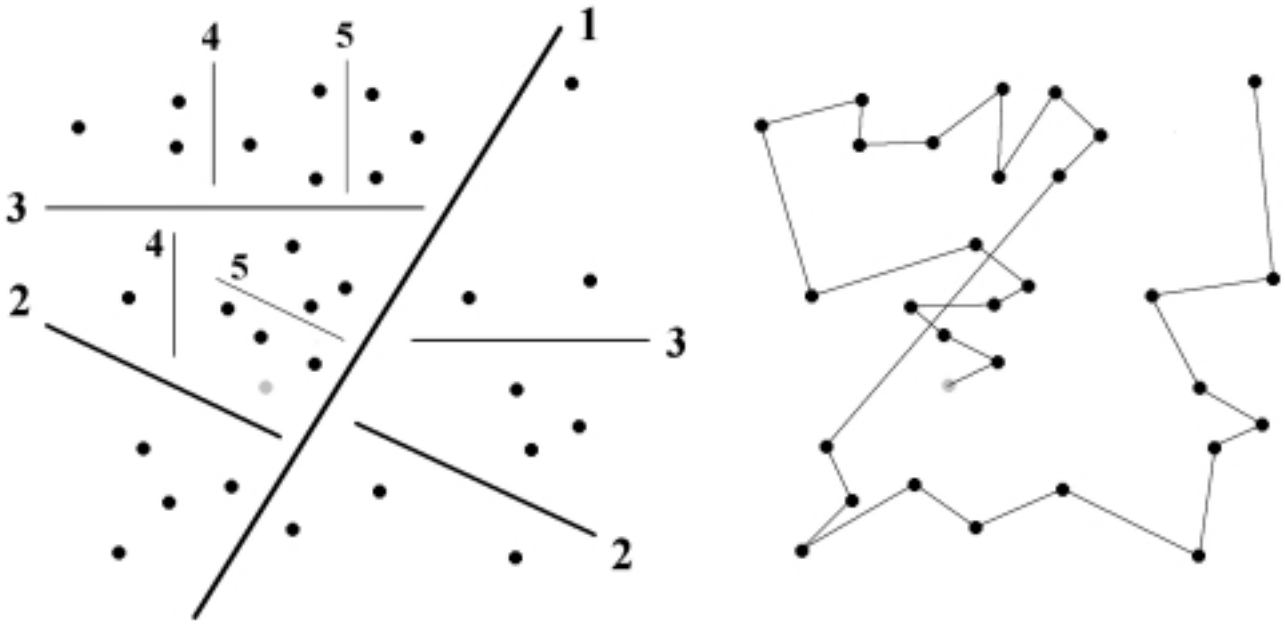

Figure 4: The model's hierarchical bisections (left) and the resulting scan path (right), for the sample stimulus shown in Figure 1

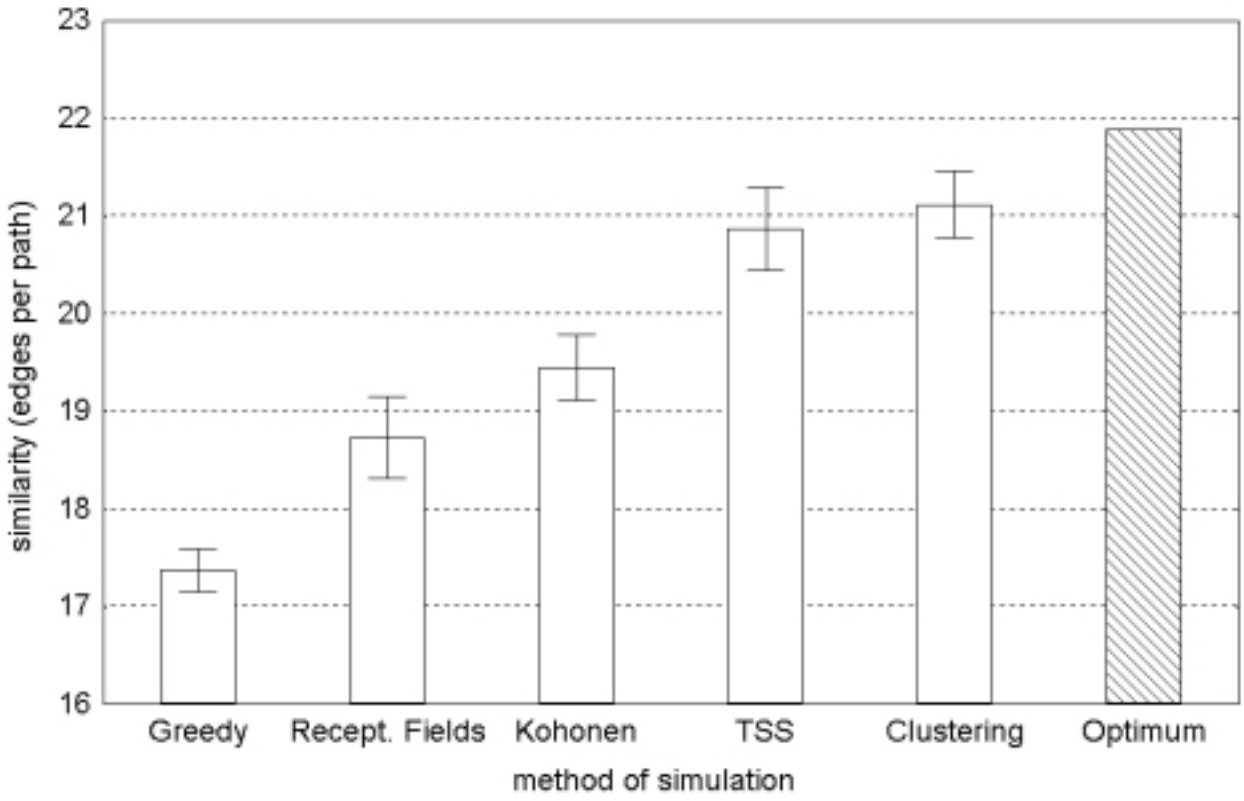

Figure 5: Similarity between the paths generated by the different models and the empirical scan paths, shown in ascending order, plus the optimum-fit value

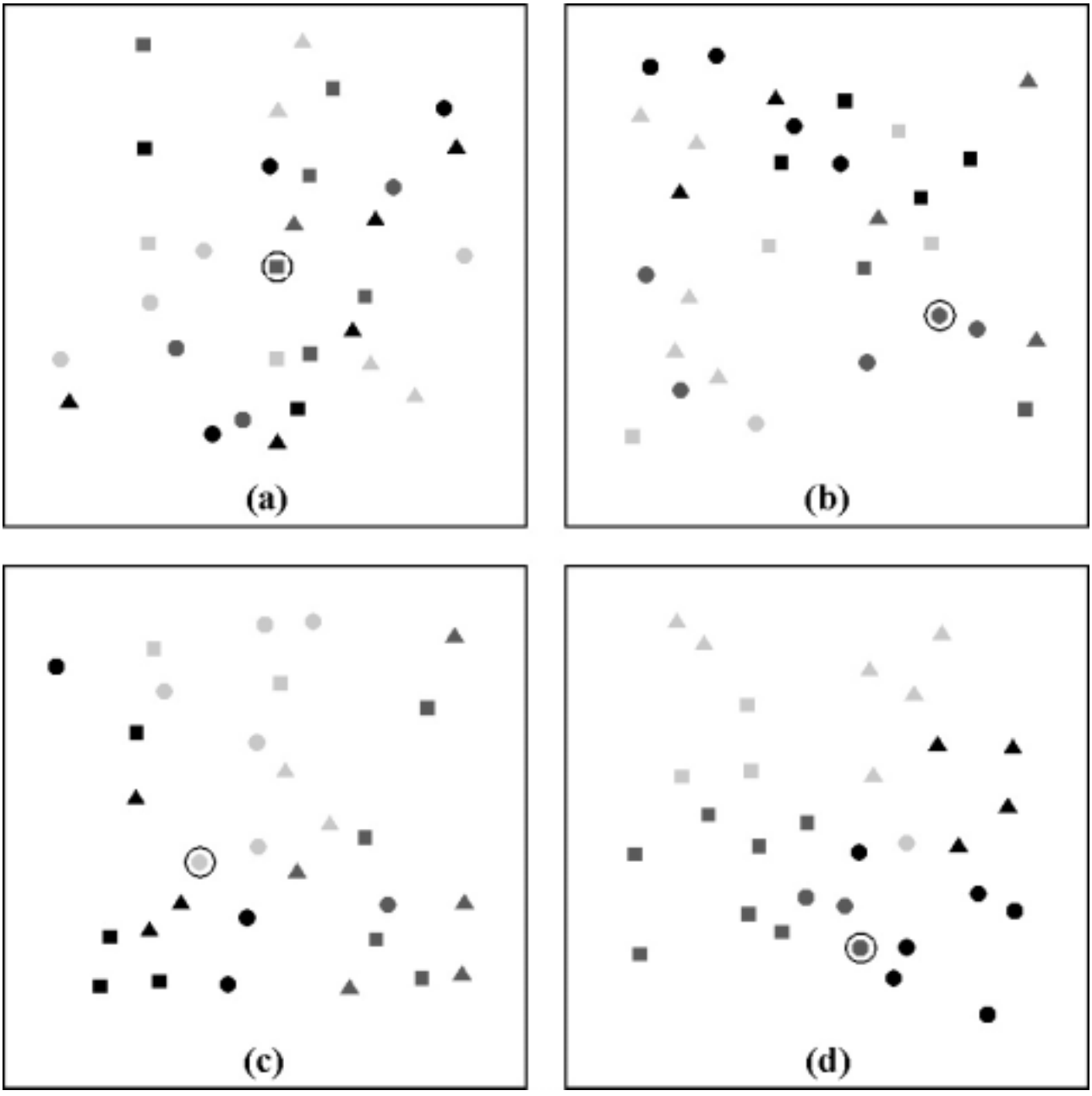

Figure 6: Examples of item distributions with different levels of color/form clustering: (a) no color and form clustering (1.0/1.0), (b) weak color and no form clustering (1.3/1.0), (c) strong color and no form clustering (1.7/1.0), and (d) strong color and form clustering (1.7/1.7). Circles indicate the starting items.

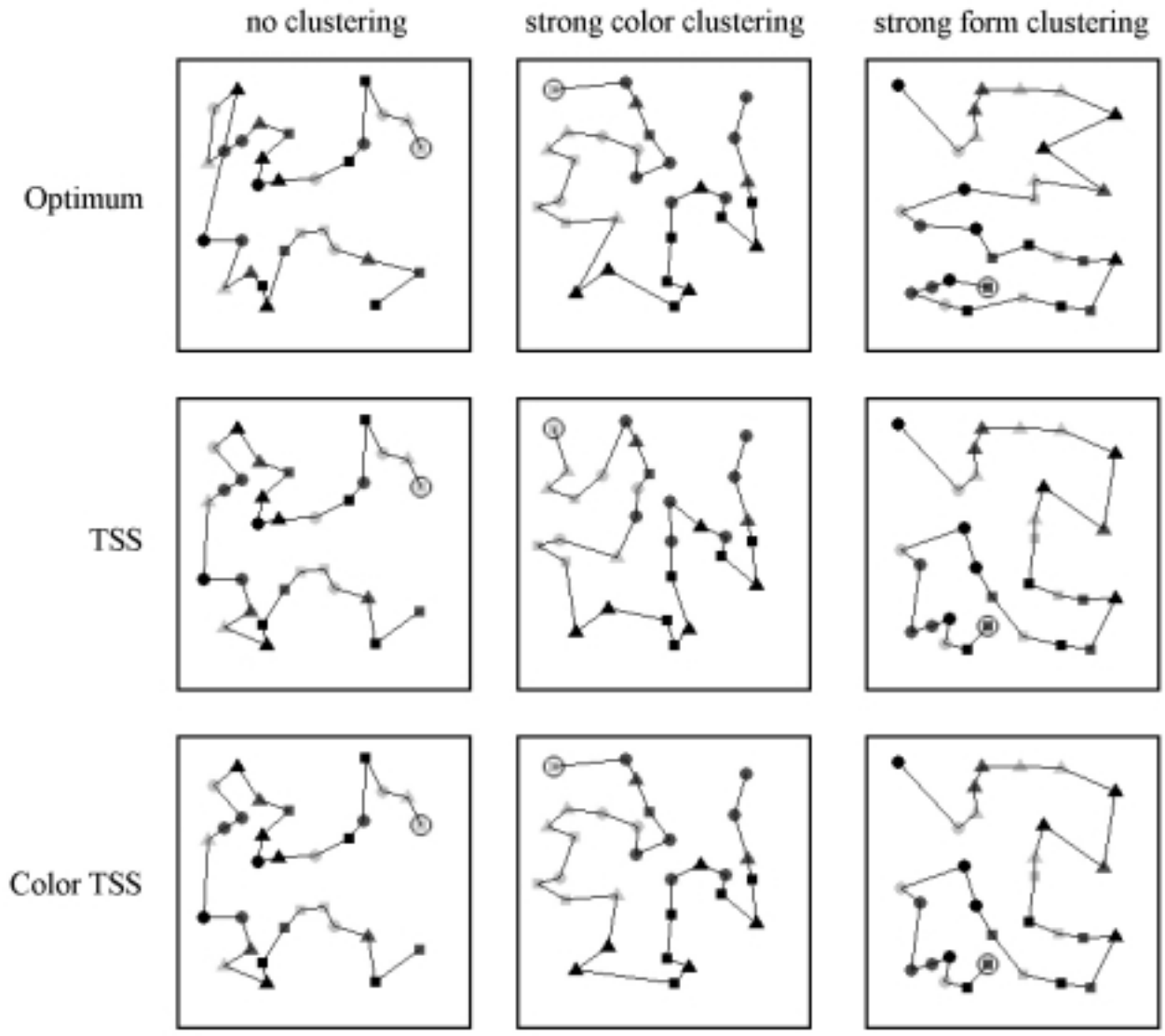

Figure 7: Scan paths generated by subjects (optimum-fit paths), the TSS Model, and the Color TSS Model. Circles indicate the starting items.

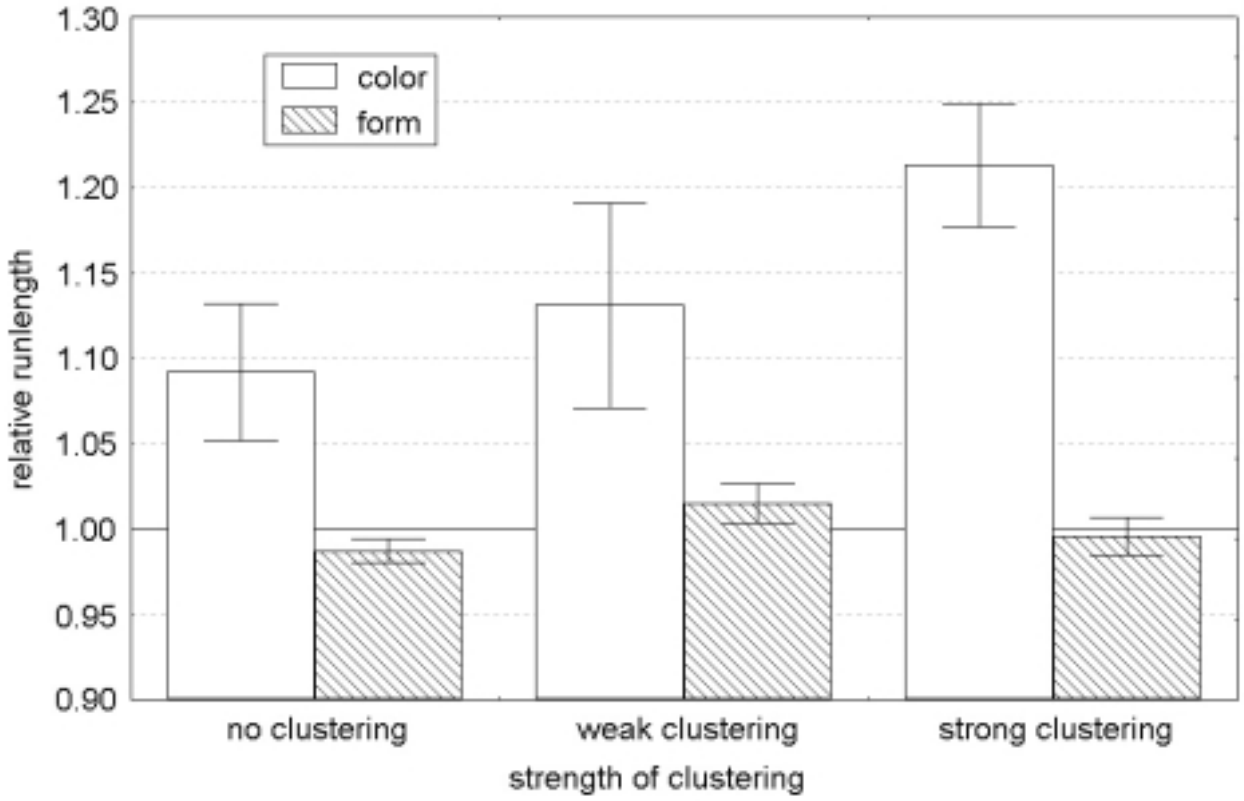

Figure 8: Mean relative color and form runlengths as functions of the strength of color and form clustering respectively.

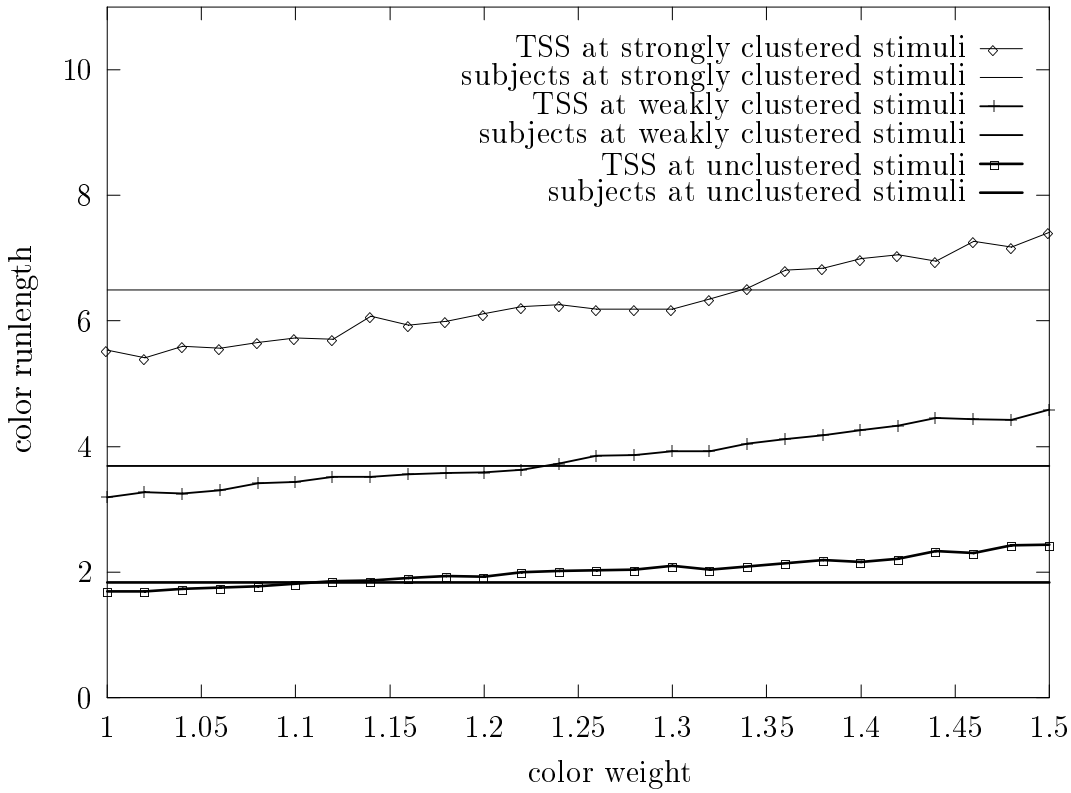

Figure 9: Color runlength generated by the TSS Model as a function of the strength of color clustering and the introduced color weight. Horizontal lines indicate empirical runlengths.

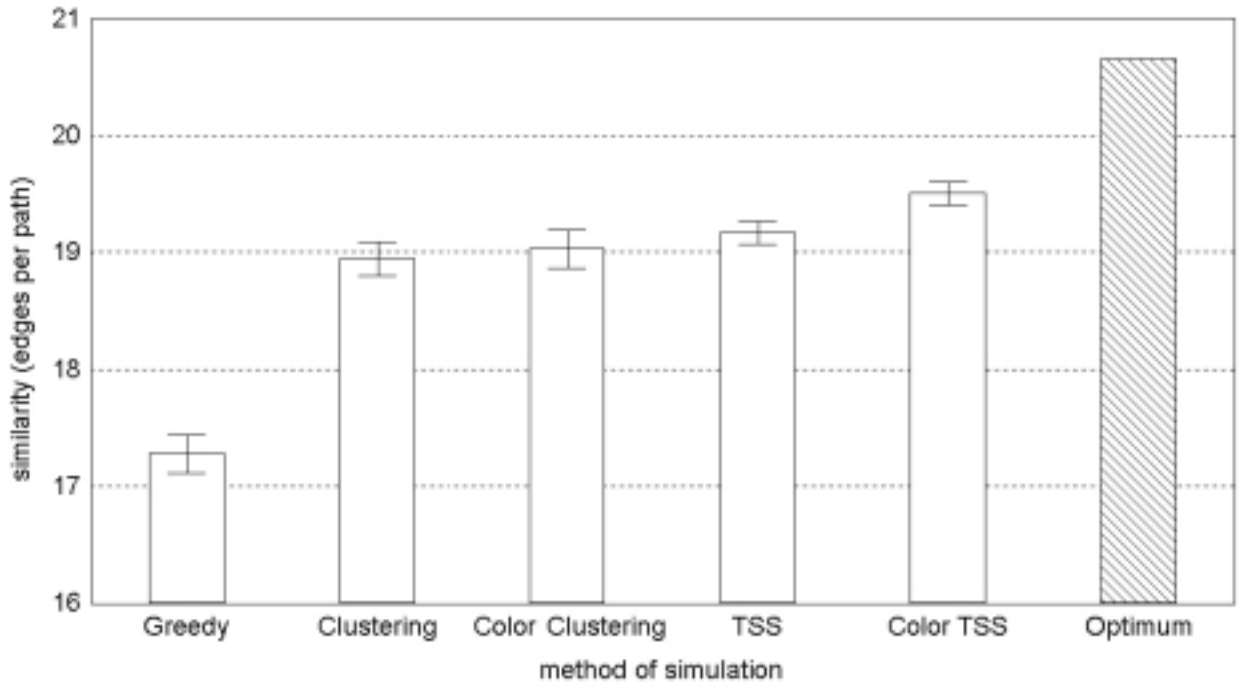

Figure 10: Similarity between the empirical scan paths of Experiment 2 and those yielded by the different models, plus the optimum-fit path