A Three-Level Model of Comparative Visual Search

Marc Pomplun (marc@psych.utoronto.ca)
University of Toronto, Department of Psychology
100 St. George Street, Toronto, Ontario, Canada M5S 3G3

Helge Ritter (helge@techfak.uni-bielefeld.de)
University of Bielefeld, Collaborative Research Center 360
P.O.Box 10 01 31, 33501 Bielefeld, Germany

Abstract

In the experiments of comparative visual search reported here, each half of a display contains simple geometrical objects of three different colors and forms. The two hemifields are identical except for one mismatch either in color or form. The subject's task is to find this difference. Eye-movement recording yields insight into the interaction of mental processes involved in the completion of this demanding task. We present a hierarchical model of comparative visual search and its implementation as a computer simulation. The evaluation of simulation data shows that this Three-Level Model is able to explain about 98% of the empirical data collected in six different experiments.

Comparative Visual Search

Comparative visual search can be considered a complex variant of the picture-matching paradigm (Humphrey & Lupker, 1993). In picture-matching experiments, subjects are typically presented with pairs of images and have to indicate whether or not they show the same object. In comparative visual search, however, pairs of almost identical item *distributions* are to be compared, requiring subjects to switch between the two images several times before detecting a possible mismatch.

The stimuli in the experiments reported here showed patterns of simple geometrical items on a black background. The items appeared in three different forms (triangles, squares, and circles) and three different colors (fully saturated blue, green, and yellow), each of them covering about 0.7 degrees of visual angle in diameter. The item locations were randomly generated, but avoiding item contiguity as well as item overlap. Each stimulus picture consisted of two hemifields (size 11x16 degrees each) separated by a vertical white line. There were 30 items in each hemifield, which were equally balanced for color and form. The hemifields were translationally identical in the color, the form, and the spatial arrangement of the 30 items - with one exception: There was always a single item that differed from its "twin" in the other hemifield, either in color or in form. The subjects' task was to find this single mismatch. They were to press a mouse button as soon as they detected the mismatch. Eye movements during comparative visual search were measured with the OMNITRACK1 system, which has a temporal resolution of 60 Hz and a spatial precision of about 0.6 degrees.

Sixteen subjects participated in Experiment 1, each of them viewing 50 pictures. Subjects knew that the critical mismatch would be either in form or in color, they did not know, however, when to expect what kind of mismatch. In fact, 25 of the 50 trials contained a difference in form and 25 contained a difference in color. Experiments 2 to 6 differed from Experiment 1 in specific aspects (see Table 1) in order to provide comprehensive data on comparative visual search (cf. Pomplun, 1998; Pomplun et al., to appear).

Table 1: Six different experiments of comparative visual search

Experiment	Subjects	Trials per subject	Description
1	16	50	No information about dimension of mismatch
2	20	60	Subjects know dimension of mismatch in advance
3	16	60	No entropy in irrelevant dimension
4	14	60	Search for a match instead of a mismatch
5	16	50	Mirror symmetry between hemifields
6	16	60	Comparison of item groups of varying size

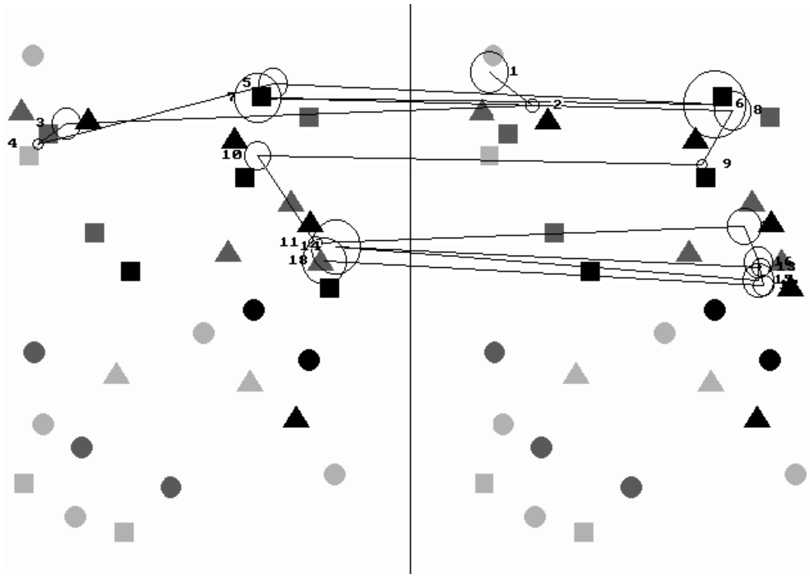

Figure 1: Example stimulus with the plotted visual scan path chosen by one of the subjects.
Fixations are numbered; circle size signifies fixation duration.

Figure 1 shows an example stimulus for Experiment 1 with a subject's gaze trajectory superimposed on it. As the example suggests, subjects switch between the hemifields very often and tend to fixate groups of items rather than single items. Moreover, they prefer to use exhaustive, self-avoiding scan paths for optimal search efficiency. For the quantitative analysis of eye movements, the independent variables local item density, local color entropy, and local form entropy were introduced. While local density indicates how closely packed the items are in a certain region of the stimulus, local entropy tells us to what extent different item features are mixed.

There are nine different dependent variables, for example fixation duration (FD) and number of successive fixations within the same hemifield (SF). These variables make it possible to investigate the influence of the local information content on a subject's eye movements. FD, for instance, increases with the local item density at the fixation point, but not with the local entropy values, indicating that short processes like single fixations are controlled by localization processes rather than by identification processes. SF, however, depends on both density and entropy: It decreases with increasing density or entropy, i.e. increasing amount of information, at the first fixation point after switching between hemifields. The quantity of this effect yields data about the capacity of visual working memory.

Taken together, eye-movement analysis in comparative visual search allows us to investigate the interaction of several perceptive and cognitive processes during the completion of a demanding task. In order to test the hypotheses derived from empirical results, a comprehensive model and its computer simulation are required.

The Three-Level Model

The Three-Level Model is not the first attempt to reproduce eye-movement patterns in comparative visual search. A simpler predecessor, the Random-Walk Model (Pomplun, 1998), directly incorporated several empirical eye-movement parameters (e.g. FD and saccade length) and their dependence on local stimulus features. The main shortcoming of the Random-Walk Model turned out to be the exclusion of higher cognitive levels, leading to unstructured search behavior. In contrast, subjects tend to structure their search, e.g. by favoring self-avoiding global scan paths.

Another problem was the direct implementation of statistical properties of empirical eye-movement variables into the model. Although this can tell us to what extent these variables determine the subjects' gaze trajectories, it does not allow us to test our interpretations of empirical findings. It is clearly more comprehensible to use a model that incorporates only these interpretations, i.e. the assumed interaction of several perceptive and cognitive components, instead of the raw empirical data. This model should generate fixations and saccades on the basis of assumed mental processes and their parameters derived from the empirical research. If the model is able to replicate the empirical eye-movement patterns, it supports our interpretations.

The Three-Level Model is a phenomenological approach meeting these requirements. Its structure is essentially motivated by the inadequacy of its predecessor, showing that different levels of processing during comparative visual search have to be distinguished. In addition to the rather schematic processes of perception, memorization, comparison etc., a higher cognitive level must be taken into account, which is responsible for global planning processes.

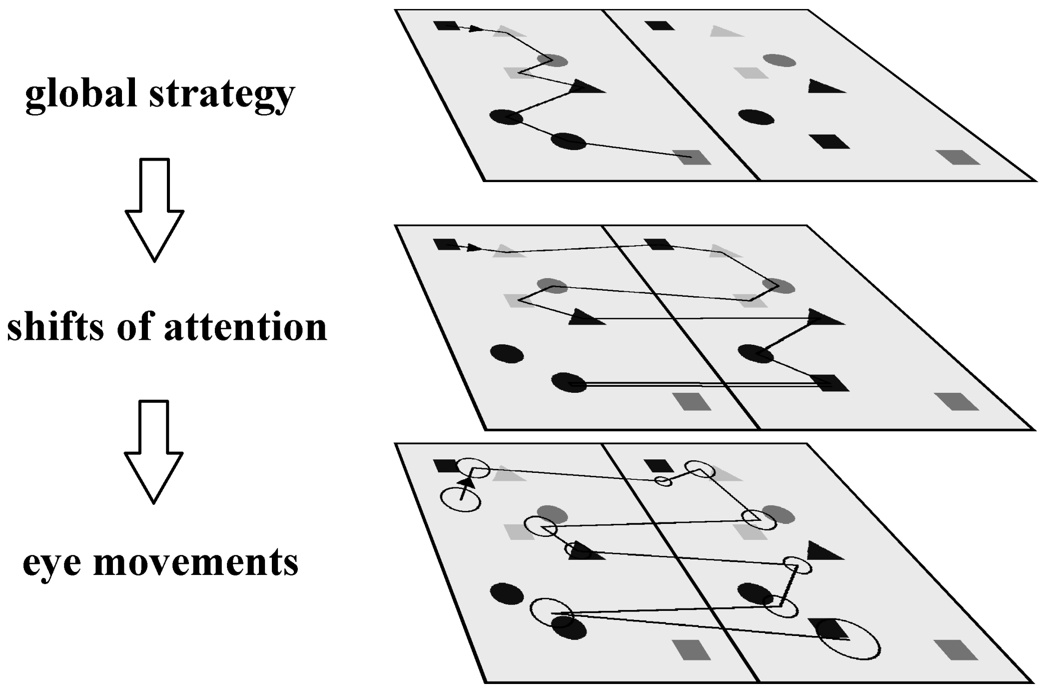

Figure 2: Scheme of the Three-Level Model. The example stimulus contains only eight items per hemifield for the sake of clarity.

Consequently, the Three-Level Model incorporates a vertical organization of mental processes, i.e. a hierarchical scheme of functional modules, better in line with current views about human brain architecture (see, e.g., Velichkovsky, 1990; Gazzaniga, 1997).

A further aspect of the model's vertical organization is the dissociation of eye movements and attention. It is a well-known fact that shifts of attention can be performed without moving the eyes (Wright & Ward, 1994; Tsal, 1983). Accordingly, the finding that subjects fixate groups of items rather than single items might be due to "invisible" shifts of attention: While attention is successively spent to all items in the display, it is not necessary to readjust the foveal gaze position for each of these steps to provide sufficient acuity. This assumption is supported by the results of studies (Pomplun, 1998) investigating the discriminability of the items used in comparative visual search: Reaction time and error rate for detecting color and form features do not vary significantly with retinal eccentricity between 0 and 10 degrees.

Figure 2 presents the structure of the Three-Level Model. On the upper level, the global strategy is planned and realized. Presumably, one of the hemifields is used as a reference with respect to this purpose; hence, the global scan path is plotted only in the left hemifield. The intermediate level is concerned with shifts of attention and processes of memorization and comparison. While the global course of processing is determined by the upper level, the local attentional shifts within and between the hemifields, needed for memorization and comparison of item features, are conducted at this intermediate level. Finally, the lower level is responsible for actually executing eye movements. The gaze position follows the attentional focus, if necessary, to provide appropriate visual acuity for the processing of information. Fixation points are chosen in such a way that the next group of items to be inspected can be memorized or compared employing as few fixations as possible. The integration of the three individual levels into a single model is described in the following sections.

The Upper Level: Global Strategy

The model's global scanning strategy is based on the Color TSP Scanning Model developed in previous research (Pomplun, 1998). It was found that subjects tend to scan a display of randomly distributed items in a "traveling salesman" fashion, i.e. using scan paths of minimal length. Moreover, subjects can take advantage of the items' colors. If the colors are clustered within the display, i.e. if there are separate areas of blue, yellow, and green items, subjects tend to completely scan each of these areas before proceeding to the next one. This strategy reduces their memory load, because keeping in memory the distinction between the items already visited and those still to be processed is easier to achieve on the basis of item clusters than on the basis of single items. No such influence was found for the items' forms, at least for the geometrical shapes used in the present context.

The Color TSP Scanning Model accounts for the influence of both the items' locations and colors on human scanning strategies, and is able to predict approximately 67.5% of gaze transitions between items, if the task is just to look once at each item. This performance is acceptable, given the fact that the maximal predictability within the analyzed scan paths was found to be 71.2% due to the observed, high variability of the data.

Therefore, the Three-Level Model's global strategy for a particular stimulus is determined as the item-by-item path yielded by the Color TSP Scanning Model for the left hemifield. Since most subjects tend to start their search at the top of the display, the items with the uppermost positions in the left hemifield are possible starting points for the path. Among them, the one that allows the construction of the shortest scan path is chosen.

It is implausible, however, that subjects plan a complete item-by-item scan path in advance. They are likely to start searching with a very coarse strategy in mind and to locally refine it to an item-based scan path during task completion. The resulting scan path, however, might be the same in both cases. We do not completely understand the dynamic development of global scan paths so far, but we know some features of the static results. This knowledge constitutes the basis for the Color TSP Scanning Model and we assume it to be transferable to the global strategy level of comparative search.

The Intermediate Level: Shifts of Attention

Attention is modeled in such a way that the sequence of items specified by the global strategy is strictly followed. Starting in a randomly chosen hemifield, attention is shifted between the hemifields during search in order to reproduce processes of memorization and comparison (see below). When the focus of attention reaches the last item in this sequence without the target items being detected, the strategy level calculates a new global scan path starting at the item in focus. A new search cycle begins, guided by the new scan path. This procedure is repeated until the detection of the target.

First, the model memorizes a number of items. This number is limited by the capacity of working memory. As suggested by empirical data, the maximum number of objects to be memorized at a time is set to three for Experiment 1. Moreover, the data show that subjects generally memorize spatially small groups of items at a time. Thus, the Three-Level Model assumes a specific radius of attention. All items to be memorized must fit into a circle of this radius. According to an estimation based on the empirical distance between neighboring fixations, the radius was set to 1.5 degrees of visual angle.

After memorization, the model ideally shifts its attention to the other hemifield, compares the stored information with the corresponding items in the same order, and starts memorizing a new group of items unless the target has been detected (see below). In most cases, however, more than one saccade between the hemifields is necessary to accomplish the comparison of two corresponding sets of items. The results of Experiment 6 indicate that the number of required between-hemifield saccades (BS) strongly depends on the number of memorized features (MF), i.e. the number of different colors plus the number of different forms contained in the set of items that are currently stored in memory. The following equation is a good linear approximation of the empirical findings and is thus incorporated into the model: $BS = 0.23 + 0.39 \cdot MF$.

In accordance with Tsal's (1983) results, the speed of the model's attentional shifts was set to one degree of visual angle per eight milliseconds. As to the dwell time of attention on the items, empirical data suggest that the processing of color may be accomplished faster than the processing of form. Consequently, we can assume attention to be focused on an item for a shorter span during specific color search than during form search or unspecified search. We adjusted the model's span in such a way that the resulting average FD corresponds to the empirical one, which is about 200 ms for all six experiments and can thus be considered a "landmark" of comparative visual search. According to this adjustment, the processing of an item's color requires 70 ms, while the processing of an item's form - or its form and color at the same time - requires 85 ms.

The attentional level is also responsible for target detection. As indicated by the empirical results, the probability of target detection is inversely proportional to working memory load, i.e. the number of item features that are memorized at the same time. In the model, the probability of target detection is defined as an experiment-specific detectability constant divided by the number of memorized features at the moment of comparing the target items to each other. For Experiment 1, the detectability constant equals 2.7.

The Lower Level: Eye Movements

As explained above, a subject's saccades are assumed to follow the shifts of attention in order to provide sufficient visual acuity in the currently attended region of the display. In the simulation, the maximum distance between the gaze position and an item to be processed is given by the radius of attention (see above). If the model directs its attention to an item outside this radius, a saccade is initiated. The target of a saccade is the next item to be inspected, if the next item but one cannot be processed within the same fixation due to a long distance separating the two items. Otherwise, the center point between these two items is chosen as the target of the saccade. Such a behavior is qualitatively indicated by the empirical eye-movement patterns; it is a reasonable way to increase search efficiency.

The model also reproduces saccadic error, i.e. a certain imprecision of saccades, mainly depending on saccade length. Its implementation follows the values reported in literature (e.g. Boff, Kaufman & Thomas, 1986). If, due to saccadic error, a saccade generated by the model does not move the intended item or any of the items to be compared into the radius of attention, another, corrective saccade is executed aiming at the same target position. Between the saccades of this kind, fixations with random durations between 90 and 110 ms are executed. The duration of saccades is modeled as a function of the respective saccade length as it can be found in literature (e.g. Boff, Kaufman & Thomas, 1986).

Finally, the empirical error in the spatial eye-movement measurement is simulated as well. The simulated eye-movement data are randomly shifted in accordance with the average error of the eye-tracker system. This feature of the Three-Level Model improves the comparability between empirical and simulated gaze trajectories.

Results and Discussion

Figure 3a shows an example scan path generated by the model on a stimulus of Experiment 1. A qualitative resemblance to empirical scan paths is obvious: Both a structured search strategy (top-down scanning) and grouping processes (e.g. within fixations number 14 to 16) can be observed. In order to quantitatively compare the simulated with the empirical data, the model was "presented" with 10.000 randomly generated stimuli for each of the six experiments.

Figure 3b presents the distribution of reaction time (RT) in Experiment 1 for both the subjects and the Three-Level Model. As can clearly be seen, the model correctly replicates a conspicuous plateau in the data between three and ten seconds. While an unstructured (random) search would lead to an exponential decay law, the plateau is indicative of a more structured search strategy. Structured search on a self-avoiding scan path leads to an RT plateau because the number of item pairs to be processed during a search cycle before encountering the target varies homogeneously randomly between 1 and 30. Neither the empirical nor the simulated data show a second plateau corresponding to the second search cycle, which has to be initiated after an unsuccessful first cycle. This result might be due to the variable duration of search cycles.

The diagram in Figure 3c presents FD as a function of local item density at the fixation point in Experiment 1. Both the empirical and the simulated FD increase approximately linearly with item density.

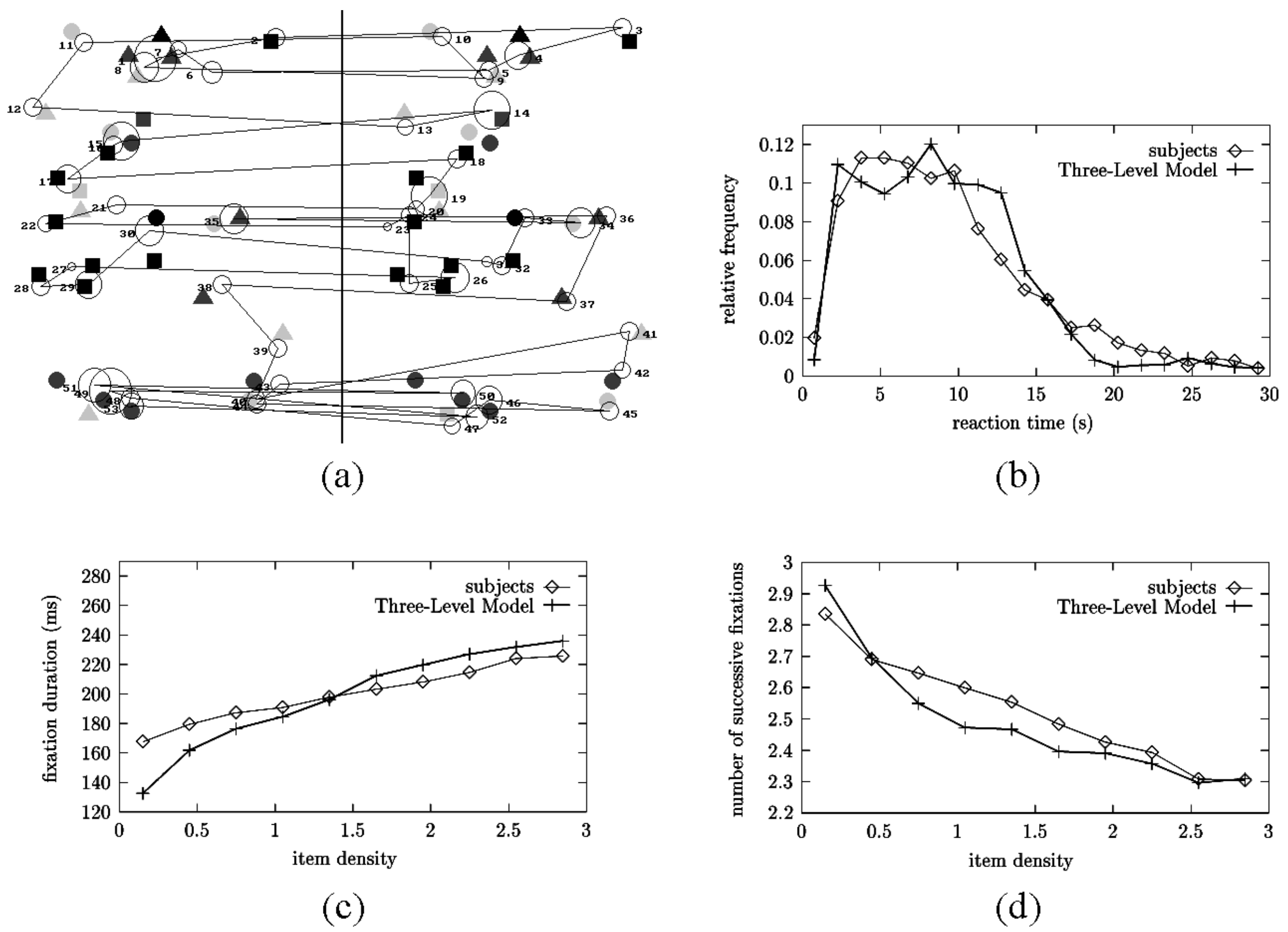

Figure 3: Results of the computer simulation: (a) example scan path, (b) RT histogram, (c) FD as a function of item density, (d) SF as a function of item density

While the two functions match well for density values above 1.5, the model's curve is steeper than the empirical one for lower density. Hence, we can assume a certain minimum duration for empirical fixations in comparative visual search even in stimulus areas providing only little information. This would explain why the duration of empirical fixations does not diminish as strongly at low local information content as could be expected regarding the results of the Three-Level Model.

Figure 3d shows SF for different item densities, measured in Experiment 1. Again, item density was calculated at the first fixation point after every transition between the hemifields. The effect of item density on the model's SF is slightly weaker than on the empirical SF for density values above 0.5, but the situation is reversed for lower density: Here, the effect on the model data is stronger. As observed in the analysis of FD, the empirical effect of low item density is weaker than it could be expected on the basis of the Three-Level Model.

All in all, the data of the Three-Level Model present for most variables a remarkably good correspondence to the empirical data. The model achieves this without directly accessing empirical eye-movement data or using freely adjustable parameters. There are only a few adjustable parameters that were set in accordance with empirical data. The analysis included three independent and nine dependent variables for six different experiments, leading to a total of 54 data distributions and 162 functional relationships to be compared, three of which are outlined above. Only 4 out of these 216 empirical data are not qualitatively reproduced by the model.

In some of the eye-movement variables, however, there are slight deviations of simulated from empirical data for regions of very low local item density. Here, item density exerts a stronger influence on the simulated than on the empirical data. On the one hand, this finding suggests the participation of inhibitory processes in comparative search, which prevent subjects from losing their search strategy, e.g. confusing processed items with unprocessed ones. On the other hand, this discrepancy could be caused by high variability in the empirical eye-movement data, which reduces the strength of measured effects. Future research will investigate to what extent the addition of inhibitory processes and higher variability at different levels can further improve the model.

Summarizing, the results are well in line with the assumed vertical organization of mental processes involved in the completion of comparative visual search. The conclusions drawn from the empirical data are strongly supported, since it is possible to build a model incorporating these conclusions and producing eye-movement patterns that are remarkably similar to the empirical ones. Thus, the Three-Level Model successfully manages to integrate a considerable number of distinct aspects investigated in different experiments into a coherent framework of mental processes and factors underlying comparative visual search.

Acknowledgments

The authors would like to thank Elena Carbone, Hendrik Koesling and Lorenz Sichelschmidt for their contribution to the present work. This research was funded by a grant from the German Science Foundation (DFG CRC 360).

References

Boff, K.R., Kaufman, L., & Thomas, J.P. (Eds.) (1986). *Handbook of Perception and Human Performance, Vol.1: Sensory Processes and Perception.* New York: John Wiley and Sons.

Gazzaniga, M.S. (1997). *The Cognitive Neurosciences.* Cambridge, Mass.: MIT Press.

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

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

Pomplun, M., Sichelschmidt, L., Wagner, K., Clermont, T., Rickheit, G., & Ritter, H. (to appear). Comparative visual search: A difference that makes a difference. *Cognitive Science.*

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

Velichkovsky, B.M. (1990). The vertical dimension of mental functioning. *Psychological Research,* 52, 282-289.

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