Analysis and Models of Eye Movements in Comparative Visual Search

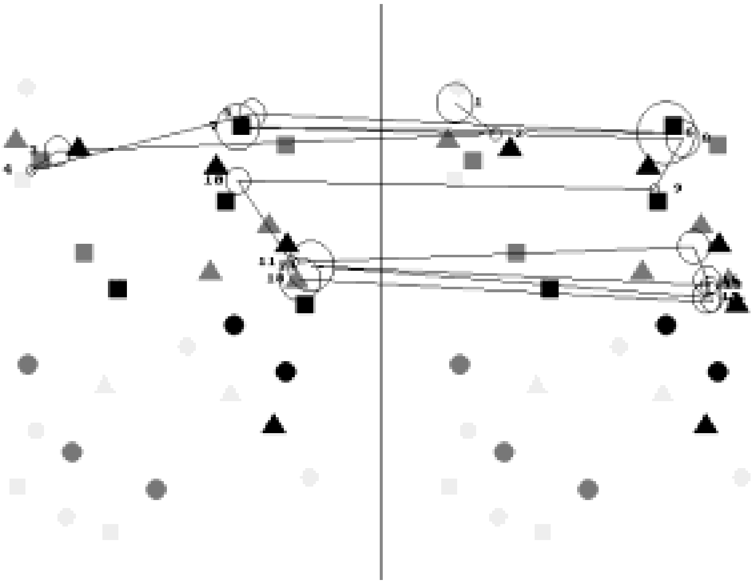
Dissertation zur Erlangung des Grades eines
Doktors der Naturwissenschaften

vorgelegt von

Marc Pomplun

bei der Technischen Fakultät an der
Universität Bielefeld

im Mai 1998



Foreword

What do eye movements tell us about the cognitive processes that underlie visual perception? The present dissertation "Analysis and Models of Eye Movements in Comparative Visual Search" by Marc Pomplun approaches this fascinating question for the situation of comparative visual search in two almost identical images. Employing a modern eyetracker system in conjunction with computer-generated images presented on a monitor, the author shows how many properties of human gaze trajectories can be illuminated and how these insights can then be used to derive quantitative computer models capturing concisely many of the observed behaviors. As a final result, the author manages to integrate the data from a remarkable number of different experimental settings in a comprehensive, three-layer model of visual search that reproduces a wide range of the experimental findings.

On the way towards this impressive result the author explains with great diligence how the analysis of fixation durations, saccade distributions as well as further statistical parameters of measured gaze trajectories can lead to intriguing and multifaceted insights into the rôle of color-, shape-, depth- and structural information for visual search processes, and he discusses the implications for visual short term memory and for the concept of separate processing of information about object location and identity.

This successful synthesis of experimental approaches and mathematical and computer-based modeling, demonstrated in the fascinating field of visual cognition, makes the present work a highly recommendable and rewarding reading for both the non-specialist as well as for the expert in the field.

in June 1998

Helge Ritter, Bielefeld, Boris Velichkowsky, Dresden

Preface

My first encounter with comparative visual search took place rather early in my scientific career: As a four-year-old, I was leafing through a TV journal. Suddenly I stopped because I had noticed a strange kind of figure. It presented two perfectly identical pictures side by side. What on earth had been the reason for this peculiar arrangement?

Fortunately, my assistant (who happened to be my mother) knew the answer to this question. Disproving my initial impression of identity, she pointed out several subtle differences between the two images. The figure turned out to be a type of puzzle called *"original and fake"* put into the journal for the sake of entertainment. The reader's task was to detect all the differences between the pictures. I started solving these puzzles with enthusiasm, not having even the faintest idea, of course, that they would provide the basis for my doctoral thesis some day.

Twenty years later, as an undergraduate student of Computer Science, I joined the Neuroinformatics Group at the University of Bielefeld in order to do the research for my master's thesis. Especially one of the various "toys" in this department fascinated me immediately, namely the *eye tracker.* I understood that the measurement of eye movements yielded a wide range of information about a person's perceptive and cognitive processes during task completion. So I began implementing new software for this system and wrote my master's thesis about the distribution of attention in ambiguous figures (Pomplun, 1994; Pomplun, Ritter & Velichkovsky, 1996).

Afterwards I got the generous offer to stay with the Neuroinformatics Group and the eye tracker as a research assistant in the Collaborative Research Center 360 "Situated Artificial Communicators". As soon as I had accepted, I started participating in the project meetings of section B4 (the eye tracker group) where we made plans for future research. The structure of this project seemed to be slightly "unfair" since it consisted of four managers plus myself as the only assistant who had to manage the practical work. Fortunately, the project began to extend, and in course of time I was accompanied by two colleagues.

Our interdisciplinary team of computer scientists, psycholinguists, and psychologists proved to be rather creative and effective. In our discussions we came across the idea to investigate the paradigm of comparative visual search ("original and fake") with the help of eye-movement recording. I translated our thoughts into action and performed a series of experiments in order to analyze comparative visual search under varying conditions. Based on the huge amount of interesting results, I began trying to model gaze trajectories by programming computer simulations. Both the adequate and the inadequate models provided additional insight into the underlying mental processes.

Most of the experiments and models are described in this doctoral thesis which is mainly addressed to scientists of related fields, but also to interested laymen with basic knowledge of psychology and mathematics. Thus motivated, I have attempted to write a clearly understandable and stimulating text that contains profitable information for everyone. I hope that all readers will enjoy this expedition from the seemingly simple "TV journal" puzzle to the level-structured information processing in our brains.

I would like to thank my supervisor Helge Ritter especially for his extensive creative support and for allowing me freedom in my research. Moreover, I am grateful to Boris M. Velichkovsky for his advice in the field of psychology and his continuous long-distance cooperation between Dresden and Bielefeld. I wish to express my thanks to my colleagues Elena Carbone, Hendrik Koesling, and Lorenz "Max" Sichelschmidt (the latter being a manager and a colleague at the same time) who helped me a lot and provided a pleasant working atmosphere. The students who have been working in our eye-tracking group during the last three years contributed to my experiments a lot by maintaining the computer hardware, recruiting subjects, searching for literature, and doing many other helpful things: Marcus Becker, Thomas Clermont (see Figure 7.1 on page 128), Kai Essig (see Figure 1.5 on page 10), Peter Munsche, Karin Wagner, and Karl Hermann Wieners. Furthermore, Dan Chen and Martin Herold supported me by proofreading this thesis, making it readable as well as understandable. My special thanks are due to Eyal Reingold and Dave Stampe for the development and hard- and software support of the eye trackers used in my experiments. This work was made possible by a grant from the German Science Foundation (DFG CRC 360/B4).

in May 1998

Marc Pomplun, Bielefeld

Abstract

The manuscript in hand is concerned with the paradigm of *comparative visual search.* A number of experiments were conducted in which subjects simultaneously viewed two nearly identical images showing sets of simple geometrical objects. Their task was to detect the only mismatch between these images which consisted in either the color or the form of one of the objects. The analysis of *eye movements* during task completion yielded diverse findings on the underlying perceptive and cognitive processes and provided the basis for two different *models* of the search process.

Chapter 1 gives an introduction to eye-movement recording and visual search in order to motivate the research work being described in this paper. The stimuli for the subsequent experiments are motivated and discussed in Chapter 2 and empirical pre-studies using these stimuli are conducted. Moreover, appropriate independent and dependent variables for the investigation of comparative visual search are defined. In Chapter 3, the Basis Experiment and its results are presented and discussed. These results provide a baseline for the subsequent experiments which introduce various changes to the stimuli and the task. Chapters 4 and 5 focus on four distinct variants of the scenario which bring about additional results on the relationship between color and form processing as well as the role of working memory. Chapter 6 describes an experiment using additional line elements to create perceptual object groups with the aim of investigating human strategies of grouping during comparative visual search. The transition from two- to three-dimensional stimuli is accomplished in Chapter 7. With the help of anaglyphs ("red-green pictures"), comparative search is studied in virtual three-dimensional space.

Chapter 8 gives a first simple approach to an eye-movement model, namely a "random-walk" simulation. The obvious drawbacks of this method motivate the conduction of further experiments intending to collect baseline data for a more plausible model. These studies, presented in Chapters 9 and 10, examine the human strategies of generating scanpaths and their dependence on various stimulus parameters. Based on the results obtained, an enhanced simulation is developed which is capable of satisfactorily explaining the empirical data (Chapter 11). Finally, Chapter 12 contains a summarizing discussion and an outlook on future research motivated by the present work.

Zusammenfassung

Die vorliegende Arbeit befaßt sich mit dem Paradigma der *vergleichenden visuellen Suche*. Hierzu wird eine Reihe von Experimenten durchgeführt, in denen den Versuchspersonen zwei nahezu identische Bilder von Mengen einfacher geometrischer Objekte dargeboten werden. Die Aufgabe der Versuchspersonen besteht darin, den jeweils einzigen Unterschied zwischen den Bildern zu finden, der entweder in der Farbe oder der Form eines der Objekte liegt. Die Analyse der *Augenbewegungen* während der Aufgabenlösung liefert vielfältige Erkenntnisse über die zugrundeliegenden perzeptiven und kognitiven Prozesse und bildet den Ausgangspunkt für zwei verschiedene *Modellierungsansätze* des Suchprozesses.

Das erste Kapitel bietet eine Einführung in die Augenbewegungsmessung und die visuelle Suche, um darauf aufbauend die im folgenden dargestellten Studien zu motivieren. Kapitel 2 erörtert die Auswahl geeigneter Stimuli und berichtet die Ergebnisse von Vorstudien für die folgenden Experimente. Desweiteren werden geeignete unabhängige und abhängige Variablen für die Untersuchung vergleichender visueller Suche definiert. In Kapitel 3 werden das Basisexperiment und seine Ergebnisse dokumentiert und interpretiert. Diese Ergebnisse bilden die Basis für die Auswertung der nachfolgenden Experimente, die veränderte Stimuli und Aufgaben einführen. Die Kapitel 4 und 5 behandeln vier verschiedene Varianten des Szenarios, mit deren Hilfe zusätzliche Erkenntnisse über das Verhältnis von Farb- und Formverarbeitung sowie die Rolle des Arbeitsgedächtnisses gewonnen werden. Kapitel 6 beschreibt ein Experiment, in dem Objekte mittels Linien gruppiert werden, um die menschlichen Gruppierungsstrategien während des Suchvorgangs zu untersuchen. Den Übergang von zwei- zu dreidimensionalen Stimuli liefert Kapitel 7. Mittels Anaglyphen ("Rot-Grün-Bildern") wird die vergleichende visuelle Suche im virtuell dreidimensionalen Raum untersucht.

Kapitel 8 beschäftigt sich mit einem ersten einfachen Modellierungsversuch, einer "Random-Walk"-Simulation. Die klaren Defizite dieses ersten Ansatzes motivieren die Durchführung weiterer Experimente zur Gewinnung von Daten für ein plausibleres Modell. Diese in Kapitel 9 und 10 dargestellten Studien untersuchen die menschliche Vorgehensweise beim seriellen Erschließen von Information in Abhängigkeit von verschiedenartigen Stimulusparametern. Die daraus resultierende, verbesserte Simulation, welche die empirischen Daten zufriedenstellend erklären kann, wird in Kapitel 11 beschrieben. Das abschließende Kapitel 12 liefert eine zusammenfassende Diskussion und einen Ausblick auf die geplante Weiterführung der hier vorgestellten Untersuchungen.

Contents

Chapter 1

Introduction and Motivation

1.1 The Brain as a Computer

Our human brain is the most complex structure on earth. In all probability it has emerged from an evolutionary process that has taken longer than a billion years. At a certain stage in evolution the onset of abstract thinking and language led to a cultural development of mankind which threw its genetic development into the eclipse. Today we can state that our genetic evolution has practically reached a deadlock and has been replaced with a cultural "explosion" especially with regard to science and technology. Meanwhile, our scientific curiosity has brought about a strange situation: The human brain wants to find out how it works *itself*.

There are several different motives for this pursuit of knowledge, for example:

Philosophy: Are human beings "only" biological supercomputers? What is consciousness and under which circumstances can it arise?

Psychology: How do individuals gather, store and share information about themselves and their environment?

Medicine: Getting more information about the brain's functional structure will result in more patients with brain injuries or abnormalities being cured.

Computer Science: What can we learn from the brain in order to improve our "Artificial Intelligence" systems? The better we understand the way our brain works, the better human-computer interfaces can be constructed.

These and other fundamental interests launched a novel research discipline being called *Cognitive Science.* It resulted from the reasonable view that the brain cannot be understood unless scientists of different faculties overcome cross-disciplinary barriers to work together in interdisciplinary teams. These teams may comprise psychologists, computer scientists, physicists, mathematicians, linguists, biologists, and physicians who try to combine their knowledge and abilities in order to explore the brain's "mechanisms".

What do cognitive scientists do? Which methods do they apply to learn about the processes occuring in our brain? One of the best-established, "traditional" methods to investigate brain processes uses electrodes to study the neurons, i.e. those nerve cells that constitute the functional part of the brain. Since neurons communicate with each other by electric signals, electrodes can be used both to stimulate neurons and to measure the neurons' activity. A large body of results has been found with this technique. For example, David H. Hubel and Torsten N. Wiesel, who won a Nobel prize later on, discovered basic properties of neurons in the visual cortex (a brain area where visual information is analyzed) of cats and rhesus monkeys (see Hubel & Wiesel, 1962). For obvious reasons, experiments of this kind can exclusively be performed on animals. Therefore, higher cognitive functions, in particular those related to language and to symbolic thought, cannot be investigated with this method.

So if we want to perform research on humans instead of animals, what are the possible methods? More recent experimental techniques are able to form images of brain activity in a non-invasive manner. One such method is positron-emission-tomography (PET). This method yields information about the spatial pattern of brain activation in high resolution (about one millimeter). Unfortunately, the temporal resolution is the range of one minute, which means that only *regions* of activation can be determined, but not the activation *dynamics*.

A simpler approach uses a number of electrodes (ranging from 16 to 64) on the subjects' scalp measuring their electroencephalogram (EEG), i.e. the oscillation of electric potentials caused by the activity of neurons. This technique provides high temporal resolution (better than one millisecond), but unsatisfactory spatial accuracy. Conceivably, only the potentials in the brain's outermost layers can be measured this way, and it is not clear to what extent these data interfere with potentials in the inner brain regions.

More favorable is magnetoencephalography (MEG) which measures the magnetic signals of the brain using a large number of magnetic sensors that are arranged in a kind of "helmet" instead of being attached to the subject's scalp. This method yields excellent data with regard to both spatial resolution (about one millimeter) and temporal resolution (in the range of one millisecond). However, only two third of the cortical currents are tangential to the skull and can thus be detected by the sensors; one third of the currents remains invisible to the technique of MEG. Another drawback is that the necessary equipment is extremely expensive.

Alternatively, there are various methods of *indirect* investigation of mental processes. It may not be necessary to measure neural activities directly. What is the idea behind this kind of methods?

Technically speaking, the brain is connected with the outside world by a large number of diverse "interfaces". Most parts of our body can be considered to be either input, output, or bidirectional "interfaces" of our brain. In order to study brain functions, a chosen "input device" can be stimulated and the corresponding reaction of a suitable "output device" can be recorded. To put it in psychologically adequate words, cognitive scientists measure parameters of subjects' *behavior* in specific experimental situations. From the resulting parameters, they draw conclusions about the underlying mental processes.

In a typical psychological experiment, the subjects' eyes receive visual stimulation and the subjects are instructed to press one of two possible buttons in response to the stimulus. For instance, subjects are presented with a monitor picture showing a set of items and they have to decide whether the picture contains a specific target item. In such experiments, reaction time and error rate can be measured as a function of item number and item distribution, yielding information about the underlying perceptual processes.

Indirect methods are by far the most widely applied ones; since the early stages they present a standard in cognitive psychology. The reasons for this dominance are obvious: First, the application of indirect methods is cheaper and less demanding in comparsion to direct ones. Second, the interpretation of directly measured data is more difficult. Correspondences between patterns of neural activity and specific mental processes, especially with respect to high-level functions, are hard to establish.

However, it must be stated that the information gained by standard indirect methods – as in the example described above – is rather sparse: Reaction times and error rates. This drawback motivates the choice of a different human "output device" to be observed. Remarkably, the perhaps most promising choice is the same as that for the "input device": *our eyes.*

1.2 Eyes as a Window into the Brain

Why use our eyes as indicators for mental processes? Our eyes are rather sophisticated instruments. Let us compare them to the simpler compound eyes of a fly which can be viewed as arrays of light receptors. They yield a roughly homogeneous spatial resolution that enables the fly to avoid "flight accidents" and to react to sudden threats from any direction. The fly does not need to visually analyze the detailed structure of its surroundings. Its moderate powers of vision combined with its well-developed sense of smell are sufficient for it to survive and to reproduce.

In contrast, the evolution of mammals (and many other classes of animals) took a different direction. More powerful brains and "interfaces" emerged. This brought about the drawback of a long developmental stage and thus slower reproduction. On the other hand, the enhanced capabilities of interaction with the environment led to strong evolutionary advantages over other species. The new prototype of eyes played an important part in the development of these capabilities. Today, our human eyes are still based on this prototype.

Figure 1.1 (left) shows a sectional view of the advanced eye conception, namely the *eyeball*. Similar to the process taking place in a photo camera, light from the outside world falls through a lens in such a way that it produces a sharp upside-down projection on a light-sensitive surface. This surface, the so-called *retina*, contains millions of photo receptors (light-sensitive neurons) that transform the image on the retina into a pattern of neural activity. Before leaving the retina through the *optic nerve*, this information is processed by more than a hundred

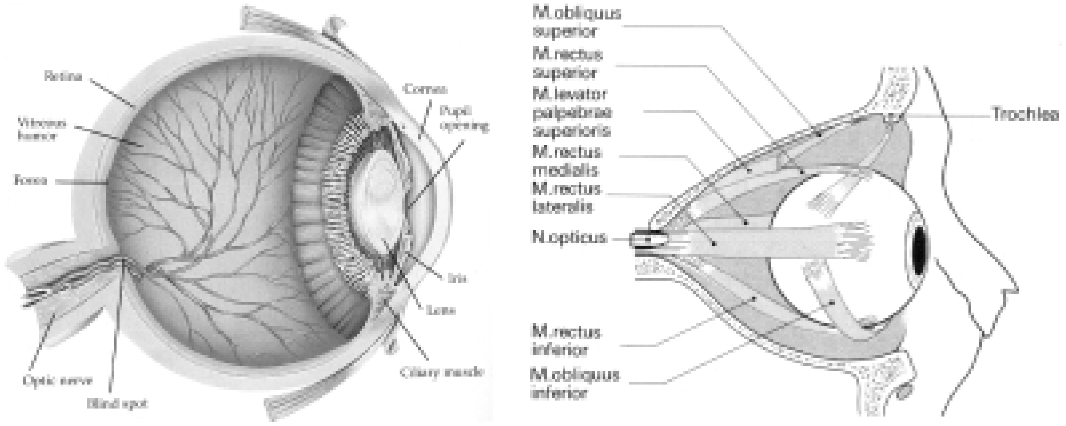


Figure 1.1: Left: Scheme of the eyeball; right: Ocular muscles (taken from Saccuzzo, 1987, and Schmidt, 1985)

different types of neurons, resulting in a substantial data compression. The retinal pre-processing is indispensible, because it would be extremely difficult to connect all receptors directly to the relevant brain areas. Furthermore, the compression has to be performed anyway, since the capacity of the human brain is limited.

What is the advantage of this design in comparison with the compound eye? There are two basic tricks our eyes use in order to provide high resolution and a wide field of view at the same time:

Gradient of receptor density: The receptor cells are not homogeneously distributed over the retina. There is a region called *fovea* of high receptor density and therefore high spatial resolution in the center of the retina. Outside this area with a radius of about one degree of visual angle, the density decreases exponentially with growing excentricity. This means that we possess very detailed vision in the center of our visual field and only coarse perception in the peripheral regions (Figure 1.2).

Eye movements: The eyeball can be quickly rotated by three antagonistic pairs of muscles (see Figure 1.1, right side), reaching a maximum speed of about 700 degrees per second. This enables us to direct our fovea successively to all regions in our field of view which we want to inspect in more detail after their rudimentary extrafoveal perception.

The fly can perceive its complete surroundings in parallel, which enables its straightforward brain to convert visual input into motor reactions extremely fast. We all know that it is nearly impossible to catch a fly with our bare hands. In the human brain, vision works in parallel as well, in order to examine things more closely, to read written language, to recognize faces etc., however, we have to direct our fovea to the relevant location. Thus, our high-level perception is a serial process.

Figure 1.2: Left: Photograph of the Holstentor with surroundings; right: Illustration of visual resolution while attention is being paid to the Holstentor

When looking at static scenes or pictures, our eye movements are never performed continuously. In fact, the gaze position “jumps” between inspected locations. These quick jumps are called *saccades* and the motionless phases in between are called *fixations*. Visual information can only be perceived during fixations, which is beneficial to the brain’s job of constructing a stable image of the surroundings.

Referring to the abovementioned idea of the eyes as being indicators for the brain’s performance, this means that *fixation duration* can be considered as a measure of the effort of information processing. The longer our attention rests at a certain location, the longer it presumably takes us to deal with the visual information presented there. This relationship has been coined the “eye-mind” hypothesis (Just & Carpenter, 1987). It is strongly supported by results from reading research. The fixation duration of a subject reading written text clearly depends on the length of the currently fixated word and its frequency in the relevant language. Since fixation duration is *not* influenced by the characteristics of the previously read word, the syntactic and semantic analysis of a word is evidently performed during its fixation. It is plausible to assume that the eye-mind hypothesis holds for other kinds of stimuli, e.g. pictures and video sequences, as well.

Furthermore, *saccade length* reveals how thoroughly a certain region of the visual field is scanned. Short saccades indicate that the fovea is directed to positions close to each other, signifying a “high resolution” scanning process. Long saccades imply that the local scene is only roughly perceived or that its information content is low. Fixation duration and saccade length are the basic eye-movement variables. Additionally, eye movements yield data about where and in which temporal order a subject acquires visual information, i.e. eye movements reveal a subject’s distribution and dynamics of visual *attention*.

What is attention? Attention is defined as a state of focused awareness coinciding with the readiness to react to stimulation. Since we are not able to consciously process the huge amount of input to our brain (e.g. visual, auditive, and tactile input) at the same time, we have to *select* a part of it to enter our consciousness. With regard to visual attention, this means that we always tend to look at those regions in our field of view that we consider to be the most relevant for our current interest. Consequently, our awareness of the other regions is reduced.

The relationship between visual attention and eye movements, however, is not perfectly determined by this connection. As everybody knows, it is possible to stare into space while thinking of something completely different from the visual scene, for example during daydreams. In this case, eye movements do not tell anything about visual attention. If subjects have to solve a particular visual task, however, they must direct their attention towards the stimuli on the display such that gaze position and attention are correlated.

A different aspect is that, during a fixation, we are able to focus our attention on different points, i.e. shifts of attention can occur independently of eye movements. Since our highest visual acuity is located in the fovea, we are likely to keep our attention within the foveal region and to employ eye movements if we want to shift attention. Only if these shifts are sufficiently small or if there is not enough time for extensive inspection, “covert” shifts of attention may occur even during fixations.

Finally, another attentional variable cannot be derived from eye movements directly, namely the *size of visual focus*. We are capable of using a wide focus in order to survey a large area in our field of view at the cost of a detailed perception. It is also possible for us to apply a narrow focus in order to examine particular features more precisely and to ignore subtleties in the periphery (see Figure 1.2, right picture). Nevertheless, the focus size can be estimated on the basis of distribution of fixations and local stimulus features. If there are, for example, only a few fixations spent on an area showing low complexity, we can assume that a wide focus is applied. This idea demonstrates the general usefulness of defining additional eye-movement variables (e.g. “local density of fixations”) and appropriate stimulus variables (e.g. “local stimulus complexity”).

Obviously, eye movements can give considerably more insight into mental processes than sheer manual responses. Serving as “a windows into the brain”, eye movements have been studied in many different fields of research:

Reading research: While reading written text, a subject’s eye movements tell us the duration needed for processing a particular word. These data enable scientists to draw conclusions about the structure of language information stored in our brain.

Medical research: Eye-movement measurement can help physicians to diagnose certain diseases of the nervous system, e.g. schizophrenia or Parkinson’s disease, because these diseases lead to characteristic distortions of eye-movement parameters. Moreover, eye-movement analysis can provide information about the state of a patient’s healing process during his/her therapy.

Traffic research: A car driver's eye movements tell scientists which factors distract the driver's attention and are thus likely to cause traffic accidents. The arrangement of instruments, for example, can be optimized with the help of these investigations.

Consumer research: It is important for advertising agencies to test the visual appeal of their commercial spots or brochures before launching a publicity campaign. Subjects' eye movements can indicate which parts of the spot or brochure attract most of the subjects' attention. In particular, it can be investigated whether the name of the boosted product is shown in an adequate position to be recognized.

As a matter of course, we have to pose the following question: How can eye movements be measured? This will be the subject of the following section.

1.3 Tracking Eye Movements

When talking about the common phrase *eye tracking*, it should always be kept in mind that the pure measurement ("tracking") of the eyeball's movement is not satisfactory for most research purposes. Understandably, scientists are rather interested in a subject's *gaze position* measured in the coordinate system of the presented two- or three-dimensional stimulus. This demand implies the determination of not only the subject's eyeball orientation, but also the subject's head position and head orientation to facilitate the computation of gaze position. Moreover, *temporal aspects* of eye movements are important as well. This means that *gaze trajectories*, i.e. spatio-temporal scanpaths, are the optimal data to be obtained from eye-tracking experiments.

The pioneers of eye tracking avoided some of these difficulties by imposing a severe restriction on the subject's freedom of action: The subject's head was being fixed. In most cases this was accomplished by an individually made bite bar on which the subject had to bite throughout the trial. Evidently, this experimental setting did not provide completely natural conditions, but at least it did lead to the first eye tracking results like e.g. those by the Russian psychologist Yarbus (1967). He used a fairly unpleasant method: A tiny mirror was fastened on one of the subject's eyeballs by a suction-cup. During the experiment, a narrow light ray was directed towards this mirror and reflected onto a photo-sensitive paper. Eye movements induced shifts in the angle of reflection and thus in the location of the light spot on that paper. Thanks to the linear correlation between gaze and light spot positions, the resulting pattern on the paper could be viewed as the recorded gaze trajectory. Figure 1.3 shows an example stimulus and the accordant experimental result. It clearly points out the most attended regions of the picture, namely the eyes, the nose, and the mouth. However, it does not yield any information about the *temporal* aspects of eye movements.

The tools for temporal analysis came with the development of digital cameras and powerful computers for image processing. It was not only to the subjects'

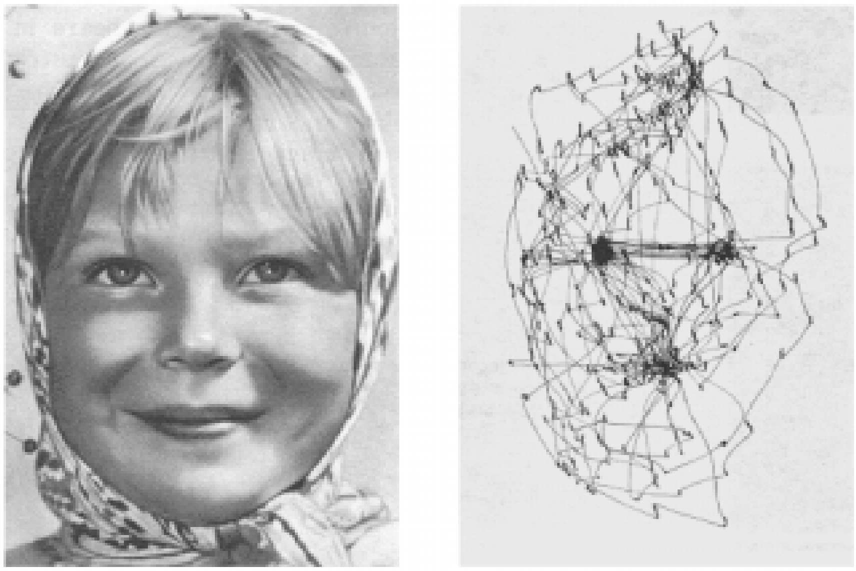


Figure 1.3: Stimulus and corresponding gaze trajectory in an early experiment conducted by Yarbus (1967)

advantage that the eye was being observed by a remote camera instead of attaching items to the eyeball now. What is more, the continuous computation of the pupil position in the digital camera image made it possible to record the temporal as well as the spatial course of eye movements. For small angles the correlation between pupil and gaze position (on a screen) was approximately linear such that the data could easily be superimposed onto the corresponding stimulus. Another improvement compared to Yarbus' experiments was achieved by the digital form of the data which enabled researchers to perform various quantitative analyses of gaze trajectories. One big handicap, however, could not be eliminated: Head movements were still forbidden.

A completely different approach to eye tracking has to be mentioned as well, namely the *Purkinje* method. Here, a laser beam is pointed to the subject's eye. It is refracted and reflected by the cornea and the lens (see Figure 1.1). The resulting light pattern falls, for example, on an array of photo-sensitive electronic receptors which facilitates a computer-based analysis. Although there is no linear dependence, it is possible to calculate the gaze position from the light pattern. Furthermore, the accomodation of the pupil can be derived from these data. This information is especially interesting in the context of experiments using three-dimensional stimuli, because it reveals the distance between the eye and the object in focus. Nevertheless, the Purkinje method does not tolerate head movements either.

There is no doubt that fixing the subject's head causes unnatural and unpleasant experimental conditions that are likely to blur the eye-movement data. Besides, if we want to investigate issues like the interaction of eye movements and

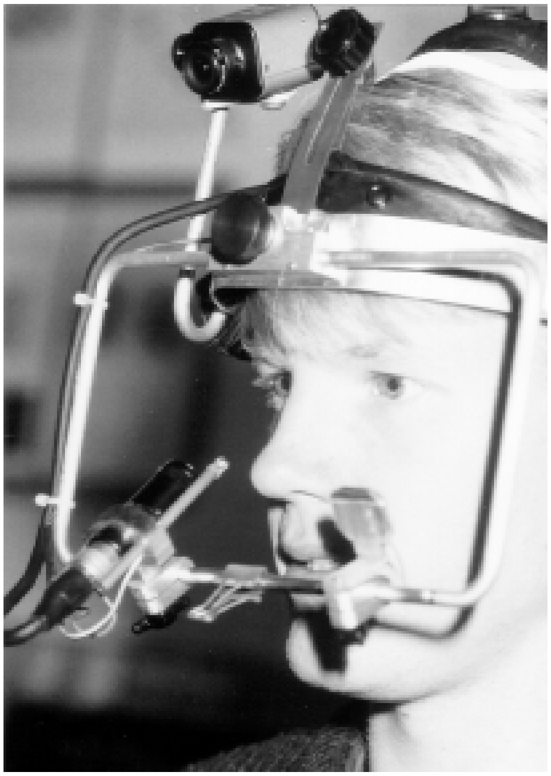


Figure 1.4: The author wearing the OMNITRACK1 head-set

speech production, head movements are absolutely inevitable. So how can this serious problem be solved?

Recent advances in eye-tracking technology have indeed solved this problem. The Neuroinformatics Group at the University of Bielefeld owns two modern eye trackers that do not require the subject's head to be fixed. There are the systems *OMNITRACK1* and its successor *SMI EyeLink* which use basically the same technique (for further details see Stampe, 1993). As shown in Figure 1.4, the OMNITRACK1 system utilizes a head-set to be worn by the subject. Two small digital infra-red cameras are fastened to this head-set, an *eye camera* pointing at the subject's right eye and a *head camera* which "looks" roughly in the same direction as the subject. The subject is placed in front of a computer monitor for stimulus presentation.

Four infra-red LEDs (*light emitting diodes*) near the corners of the monitor screen are essential for the "trick" of head movement compensation. The head camera yields an infra-red image of the subject's field of view, and of course the four LEDs cause the only four light spots in the otherwise black image. The locations of these spots in the image and thus the relative position of the subject's head with respect to the screen can effortlessly be calculated.

As in early eye tracking experiments using digital cameras (see above), the pupil position is derived from the eye-camera image. Proceeding from the head

Figure 1.5: SMI EyeLink: the binocular successor of OMNITRACK1

and pupil positions, a non-linear projection leads to the gaze position coordinates on the screen. The parameters of this projection are gained from a calibration procedure which has to be performed before the start of any experiment. In this procedure a dot is shown sequentially at nine different positions on the screen and the subject has to track it visually. This comparably quick calibration leads to a spatial accuracy which is more than sufficient for most purposes.

All technical data are documented in Section 2.1.1. At this point, it is only important to state that the OMNITRACK1 system provides both natural conditions (freedom of head movements) and high accuracy of measurement.

However, some of the system's features are not optimal. For a start, the headset is rather heavy and tends to slide on the subject's head, which is a potential source of large errors in measurement. Moreover, only the gaze position of the right eye is recorded, yielding only partial information about the subject's eye movements. Finally, the system's relatively low temporal resolution and technical problems concerning stimulus presentation and data access leave much to be desired.

It has been possible to overcome these problems of OMNITRACK1 with the development of the system *SMI EyeLink*, without any change in the mode of operation. When looking at Figure 1.5, one feature of progress immediately strikes the eye: Now there are *two* eye cameras, one for each eye, facilitating binocular eye tracking. Furthermore, the head-set is lighter which reduces the probability of shifts and brings about more pleasant conditions for the subject. Other major technical problems have been solved as well (see Section 2.1.3).

Now we have seen that eye movements yield interesting information about perceptive and cognitive processes in our brain and that we are able to measure them with highly qualified devices. However, just measuring eye movements would not provide any significant results yet. The first thing we have to do is to choose a research paradigm which is promising with regard to eye tracking. In particular, we have to answer the following questions:

- Which type of stimuli is shown to the subjects?
- What is the subjects' task?
- Which hypotheses are to be tested?
- Which variables are investigated in order to test these hypotheses?

The only indisputable points are the use of visual stimuli and the measurement of eye movement variables including those mentioned in the previous section. All other aspects of experimental design are open to our choice, and that seems to be a hard one. How can we find our way through the infinitely many possibilities that emerge in front of us?

A good strategy of orientation is to find an interesting standard paradigm and to use it as a point of reference. When looking at the research literature on this paradigm, we will come across open questions and unchecked hypotheses. Maybe the introduction of slight changes to the paradigm and/or the use of enhanced methods, e.g. eye tracking, will allow us to answer these questions and to extend this field of research in a sensible way.

A standard paradigm based on visual stimuli is *visual search*. Does it constitute a suitable basis for our eye-movement research?

1.4 A "Classical" Paradigm: Visual Search

The paradigm of visual search is a well-explored field of psychological investigation. Surprisingly, the subject's task in most of the innumerable studies that have been published so far is exactly the same. It consists in finding out as fast as possible whether a shown set of items includes an "odd one" that is different from the others in a certain way. The subjects have to report their decision by pressing one of two buttons labeled "yes" and "no". It is obvious that the only variables that can be measured this way are reaction times and error rates. Nearly all analyses refer exclusively to reaction times as indicators for mental processes.

What are the theories about visual search that have been developed on this basis? As a starting point, there is a distinction between two fundamental types of search processes:

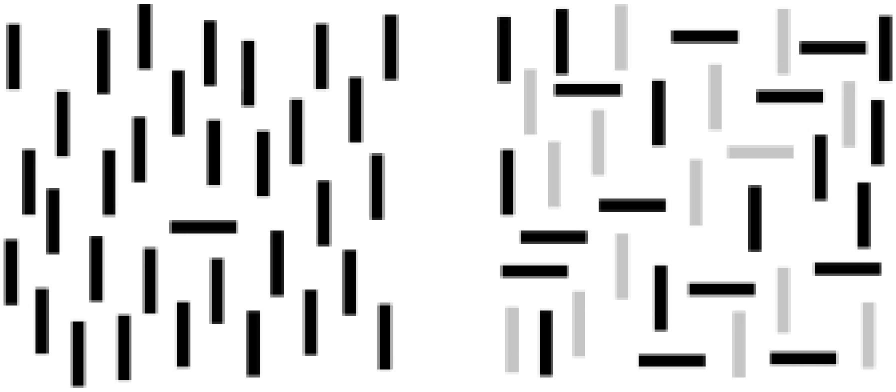


Figure 1.6: Stimuli for visual search. Left: Disjunctive search allowing parallel processing; right: Conjunctive search requiring serial scanning

Parallel search: If the target item differs from all distractor items with respect to only one dimension (*disjunctive search*) it “pops out”. Figure 1.6 (left) presents an example with *orientation* being the relevant dimension. In this case, the search process is very short, because in most cases we detect the target immediately.

Serial search: The task becomes more challenging if the differences occur in more than one dimension and the target item is specified by a unique combination of them (*conjunctive search*). As shown in Figure 1.6 (right), the differing item is hard to detect at first sight, hence the display must be scanned in a sequential fashion.

This parallel-serial dichotomy was proposed by the *Feature Integration Theory* (Treisman & Gelade, 1980). The dichotomy is empirically based on reaction time slopes: Whereas the time needed for disjunctive search is more or less independent of the number of distractor items, the reaction time for conjunctive search increases linearly with the item number in the display. The Feature Intergration Theory assumes the existence of specific “feature maps” in the brain, enabling us to detect differences in single dimensions even without attention being involved, resulting in parallel search. In conjunctive search, however, attention is needed to analyze local combinations of different feature maps. Attention is inevitable for detecting the dissimilar item. Since the focus of attention can cover only a part of the display at a time, serial scanning has to be performed.

Several years later, these assumptions were questioned by other researchers (Wolfe, Cave & Franzel, 1989). In their experiments, the authors found parallel search for conjunctive color-form targets which could not be explained in terms of the Feature Integration Theory. The precondition for parallel conjunctive search was reported to be sufficient feature saliency. Strong color contrasts both between different items and between items and background may facilitate parallel processing. Accordingly, reaction time slopes can be considered functions of feature saliency and they thus constitute a continuum rather than two distinct groups. The authors proposed the hypothesis that parallel conjunctive search is the "default" case which can be prevented by insufficient feature saliency.

In the same work (Wolfe, Cave & Franzel, 1989), they pointed out that the assumption of autonomous parallel and serial processes, as given by the Feature Integration Theory, is unplausible. Why does the serial search process not benefit from the information gathered during the preceding parallel process? As a consequence, they proposed the *Guided Search Model*. According to this model, an initial parallel process can determine areas in a display that are likely to contain the target item. This information is accessible for a subsequent serial search process.

Although the idea of an initial, preattentive (i.e. without utilizing attention) stage followed by a serial analysis is shared by most theories (Treisman, 1992), they diverge in their explanation of how the parallel stage can support the subsequent serial stage. Treisman & Sato (1990), for example, revised the Feature Integration Theory, reacting to Wolfe, Cave & Franzel (1989). This new version of the theory is rather similar to Guided Search, however, it assumes an *inhibition* of unlikely target features instead of an *activation* of promising locations.

Other theories account for *perceptual grouping* as a relevant factor in visual search. For instance, Pashler (1987) proposed that the display is scanned cluster by cluster using parallel processing within the clusters and sequential shifts of attention between the clusters. A refined theory (Duncan & Humphreys, 1989; 1992) proposed that the whole search process is controlled by the global similarities between the items. Following this approach, visual search is not based on single features but on subsets of items which are separated from each other by perceptual grouping during preattentive processing. This assumption of multi-item grouping is supported by a variety of experimental results. Humphreys, Quinlan & Riddoch (1989), for example, presented subjects with T-shaped targets among upside-down Ts serving as distractors. They found clearly higher search effciency in the case of items forming a coherent global shape. Following their interpretation, if subsets of items can be perceptually combined to form a well-known structure, subjects are able to process them as a whole. Hence, the *perceptual units* of visual search are groups of items instead of single items, which facilitates target detection. In the theories accounting for perceptual grouping, features are considered as a result of mental processes rather than as a precondition for them, as supposed by other theories. Obviously, the dichotomy between disjunctive and conjunctive search loses its relevance when viewed from this perspective.

Current research focuses on the *transition* between different stages of processing (e.g. Grossberg, Mingolla & Ross, 1994). How do the stimulus parameters control the time course of search patterns applied by subjects? Here, it is important to ask whether the measurement of reaction times is sufficient to answer questions on this level.

As mentioned before, reaction times and error rates give only restricted insight into the perceptive and cognitive processes that are involved in visual search. We must be aware of the fact that conclusions about these processes are based on the assumption of *additive* latency components (Meyer, Osman, Irwin & Yantis, 1988). Only a strictly sequential cognitive structure of task completion would justify inference from reaction times. However, the models of visual search mentioned above propose parallel processes as well. Should this be considered a contradiction in terms?

Without doubt, the measurement of *absolute* reaction times can only provide rough information about mental processes participating in visual search. Therefore, most studies are based on the analysis of reaction time *slopes* rather than on absolute reaction times. As discussed in the context of the parallel-serial dichotomy in visual search, these slopes can indicate basic differences in processing depending on the characteristics of the stimuli. The slopes also serve as reference data for formal models of visual search (Grossberg, Mingolla & Ross, 1994; Humphreys & Müller, 1993). On the other hand, it is obvious that information about *hybrid* processes is not accessible by investigating reaction times only.

Referring to our initial question, is eye tracking an appropriate methodological extension which could prove able to overcome these restrictions? In fact, the analysis of eye movements during visual search has been successfully conducted, resulting in different models of gaze trajectories (e.g. Rao & Ballard, 1995; Wolfe, 1994). The completion of standard visual search tasks, however, does not require subjects to apply complex scanpaths. For instance, disjunctive search within an area of several degrees of visual angle is likely to be solved without any eye movements at all. If no or only a few eye movements are necessary for target detection, it is obvious that the recording of eye movements does not constitute a substantial improvement over traditional methods.

Consequently, it seems reasonable to use a more complex task for our investigations. It should make the subjects employ eye movements which yield insightful information about the perceptive and cognitive processes going on during task completion. The increased complexity of mental activities should not be problematic, since the information content and detailedness of the data obtained are immensely enhanced due to the additional eye tracking. In the light of these considerations, more intricate tasks can even considered to be advantageous, because they promise to yield more insight into the interactions between different mental processes.

Apart from the validity of available data (*construct validity*), another matter of general importance should be taken into account, namely the problem of *ecological validity*. The human brain has not been evolutionarily designed for solving abstract tasks in laboratory experiments, but rather for interacting with its natural environment. Hence, we have to make the subjects' situation in our experiments ecologically adequate or *valid*, because otherwise the results would be as artificial as the experimental conditions.

As to the subjects' task, the standard visual search paradigm corresponds to the everyday situation of looking for a well-known item. During task completion, the subjects have to remember the relevant features of the target item as well as information about the actual scene, especially the distinction between inspected and not yet inspected items. Unfortunately, this class of tasks is not an ideal background for the investigation of working memory in visual perception, because the target item and thus the respective memory content are constant during the search process. In most experiments the target item does not even vary systematically between the trials. Therefore, standard experiments in visual search do not provide insight into processes of memorization, but only into processes of retrieval and comparison.

Concerning stimulus materials, almost all experiments in visual search utilize displays of simple, abstract, and well-defined items. They enable the researchers to explicitly control the dimensions to be investigated (e.g. color or form) and to obtain clearly interpretable results.

To sum up, standard tasks of visual search are not perfectly suited for eye-movement research. They do not involve many eye movements and they are not designed to yield information about an important cognitive component, namely working memory. Therefore, we have to introduce appropriate changes to the paradigm of visual search. A promising variant of visual search in this respect is presented in the following section.

1.5 Extending the Classical Paradigm: Comparative Visual Search

What does “comparative visual search” mean? Surprisingly, this search task is rather famous, but not among scientists so far. It often appears in magazines and is known as “original and fake”. Two almost identical pictures are presented to the reader, and the task consists in finding several subtle differences between them which were introduced on purpose.

Comparative visual search is closely related to picture matching (e.g. Humphrey & Lupker, 1993). In a typical picture matching experiment, subjects are presented with pairs of pictures, which can occur either simultaneously or sequentially. The subjects’ task is to indicate whether the pictures show the same item. A well-known result from many picture-matching studies is 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). Moreover, it was established that response times are a sinusoid function of the difference in the angle between the two presented items, which is known as *mental rotation* (Rock, Wheeler & Tudor, 1989; Shepard & Cooper, 1982). Finally, there are specific objects having “canonical views” such as frontal or profile view, which are particularly easy to process (Cooper, Biederman & Hummel, 1992; Palmer, Rosch & Chase, 1981).

The scenario used by Bruce & Morgan (1975) can be considered a first step from picture matching towards comparative visual search. The authors presented subjects with left-right mirror symmetric or horizontally translated line patterns, some of which contained small symmetry violations. The subjects’ task was to manually report whether the symmetry was complete or not. Both the subjects’

Figure 1.7: Typical "original and fake" task: Find the two local mismatches between the left and the right image.

reaction times and error rates were higher with mirror symmetry than with repetition. The advantage of mirror symmetry decreased with the enlargement of the horizontal distance of the violation from the center of the stimulus, since the two mismatching locations were farther apart from each other. Thus, the advantage of mirror symmetry exposed an eccentricity effect.

The task of comparative visual search is more complex than those used in picture-matching studies including the work of Bruce & Morgan (1975). Loosely speaking, comparative visual search combines visual search and picture matching. Figure 1.7 shows an example stimulus for comparative visual search. Two local differences are to be detected.

The completion of this relatively simple example task does not take much effort: From the left to the right image, a sphere in the upper left corner turns into a cube and the color of the girl's turned-up sleeve gets considerably darker. During task completion, we notice that a number of different mental processes are necessarily involved in our actions. Locally, *memorization* and *comparison* have to be performed. In one of the pictures, a local set of items or structures is memorized first. After "filling up" our working memory this way, we proceed to the corresponding region in the other picture and compare the information given there with the content of our memory in order to detect potential mismatches. Globally, we have to *search* the whole picture, and this cannot be done efficiently unless *strategy planning* is carried out. No area within the image should be missed or visited more than once.

Obviously, many eye movements have to be performed during task completion. They indicate shifts of attention within and between the hemifields, telling us where and in which order subjects process information. The investigation of the resulting gaze trajectories by means of eye tracking is clearly more attractive than in standard visual search tasks.

What can be stated about the ecological validity of comparative visual search? The situation is substantially different from standard visual search. In comparative visual search, the setting is equivalent to real-world circumstances being caused by a discrepancy between an actual state and a target state. If we want to construct an aeroplane model from single pieces on the basis of a drawing showing the completed plane, for instance, we have to continuously compare (and adjust) the actual real-world situation to the target state. Here, processes of memorization and comparison of local information can be observed, leading to a deeper understanding of the role of working memory during the efficient completion of complex tasks. While ecological relevance is taken care of in both visual search and in comparative visual search tasks, the comparative variant clearly provides a more profound insight into the interaction of different factors, especially with regard to working memory. By the analysis of eye movements we can obtain information about basic parameters of working memory, e.g. capacity and access time, and how these parameters control the subjects' strategies of solving the task.

As to stimulus materials, the situation is similar to the one encountered in standard visual search. Although stimuli showing realistic scenes like the example in Figure 1.7 are preferable with regard to ecological validity, they are not adequate for basic studies of comparative visual search because they involve factors operating on a high semantic level. For instance, we all know a wide variety of faces, buildings, streets etc. and thus have certain concepts of how to perceive them. This knowledge influences our eye-movement patterns as well as our capability of mismatch detection. Experiments with ambiguous pictures (Pomplun, Ritter & Velichkovsky, 1996), for example, showed that the distribution of attention depends not only on the geometrical structure of the inspected stimulus, but also on its semantic interpretation by the subjects. Evidently, factors cannot conveniently be parametrized on this level and hence they are not accessible to quantitative analysis.

The discussion above shows that there is a kind of trade-off between ecological validity and experimental control, for both standard and comparative visual search. At least for our initial investigations it seems useful to emphasize experimental control at the cost of ecological adequacy. Based on the results obtained, further research with more realistic scenes will be carried out to overcome this shortcoming.

All in all, what are the most important aspects to be investigated? With the help of eye tracking, the comparison scenario allows us to study several important questions:

- Which information-related features influence visual search?
- How are these reflected in the spatial and temporal aspects of the search path?
- What can we infer about memory and recognition processes?
- Is it possible to build a model capable of generating statistically similar search paths?

This last point is especially relevant because the integration of empirical findings into a model is a well-suited way to test their consistency and validity. If a computer simulation of eye movements, incorporating this model, produces scanpaths that resemble the empirical ones, the consistency and validity of the underlying results and interpretations are strongly supported. This topic will be thoroughly discussed in Chapters 8 and 11.

First of all, the following Chapter 2 is dedicated to the preparation of empirical research based on comparative visual search. Its technical realization as well as the selection of appropriate stimuli for experiments of comparative visual search are described. Furthermore, the basic characteristics of the stimuli with regard to human perception are investigated, resulting in important baseline data for the subsequent experiments.

Chapter 2

Methodological Preliminaries

2.1 Technical Equipment

What are the technical facilities for our research? As mentioned in the previous chapter, we currently use two eye trackers in our group, namely the monocular OMNITRACK1 device and its successor, the binocular SMI EyeLink system. Since the EyeLink system did not become available before 1997, the first experiments of comparative visual search were conducted using the OMNITRACK1 device.

2.1.1 The Basic OMNITRACK1 System

Basically, the OMNITRACK1 apparatus uses a single computer (486/DX2-66) called "Operator PC" for eye tracking as well as for stimulus presentation. It is connected with three external devices:

Head-set with infra-red eye and head cameras: The eye camera yields digital images of the subject's right eye, while the head camera observes the four infra-red LEDs at the corners of the monitor screen in front of the subject. These camera data are processed by ISCAN RK-416PC pupil-tracking boards in the Operator PC which are capable of detecting the pupil and LED positions respectively in the pictures using simple image-processing algorithms. Based on these coordinates, special software running on the Operator PC calculates the subject's gaze position on the screen. Furthermore, saccades and fixations are detected (see Stampe, 1993).

Computer monitor for the subject: All stimuli are presented to the subject on a 17" ViewSonic 7 monitor, including those for the calibration procedure. This calibration procedure, which has to be performed before the start of every experiment, demands the subject to look at a dot that is sequentially presented at nine different positions on the screen. The calibration procedure yields required reference data for the subsequent computation of gaze positions. During experiments, a video-data switch makes it possible to display stimuli from either the video card of the Operator PC or from an external device, for example a video recorder. Between trials (or any other parts of experiments) a quick re-calibration featuring only one target dot can be applied in order to compensate for shifts of the head-set.

Computer monitor for the operator: The operator is watching a separate 17" ViewSonic 7 monitor in order to supervise the experiment. This monitor is also controlled by the Operator PC. Experiments can be programmed in a special command language which includes commands for calibration, image presentation, timing, data formatting, message output etc. There is also a set of commands to display messages for the operator on his/her monitor, like for example: "Press <Enter> to start next trial". During eye-movement recording, the operator monitor shows the same scene as the subject monitor plus a moving gaze cursor indicating the subject's current gaze position. This enables the operator to get a visual impression of the eye movements before their statistical analysis and to decide whether a re-calibration is necessary due to low accuracy of measurement.

Although this configuration is perfect for psychological standard research, the command language for programming experiments turns out to be too limited for advanced purposes. For example, only static bitmap images can be shown, and it is impossible to present computer-controlled dynamic stimuli or stimuli created on-line from data files. In order to conduct experiments of comparative visual search in a sensible way, the eye-tracker system had to be extended.

2.1.2 The Extended OMNITRACK1 System

Some of the system's limitations are eliminated by the integration of a second computer into the system which generates and displays the stimuli for comparative visual search. Thanks to the video data switch, the presentation of video signals from an additional computer ("Subject PC") is no problem at all. The calibration procedure and the control of experiments is still performed by the Operator PC, but during eye-movement recording, the video card of the Subject PC is in charge of the subject monitor.

However, the main problem consists in the synchronization of the Subject PC with the Operator PC. The data recording has still to be carried out by the Operator PC, whereas the possibly dynamic stimuli are produced by the Subject PC. Therefore, the Subject PC has to "know" when the eye-movement recording begins in order to start presenting the dynamic scene. If there is no synchronization between the PCs, the correlation between the recorded data and the state of the shown stimuli is unclear.

Unfortunately, the eye-tracker software is not designed for communication with other computers, and since no source code is provided, it is not possible to add any software interfaces. Hence, we had to extended the system in a slightly peculiar way, as shown in Figure 2.1.

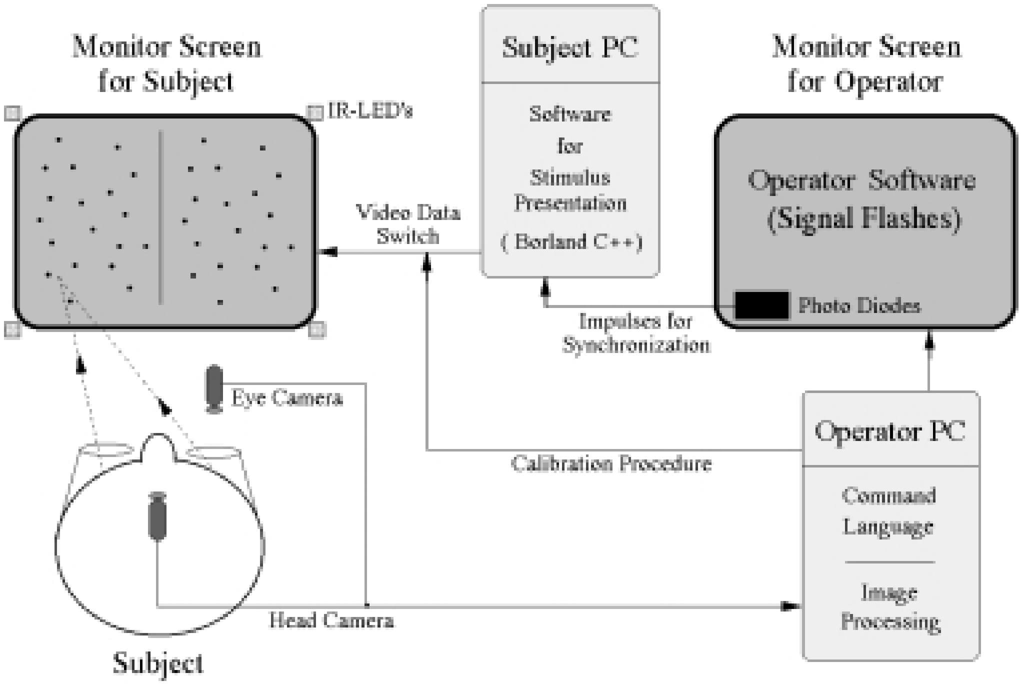


Figure 2.1: Scheme of the extended OMNITRACK1 system

As mentioned above, the OMNITRACK1 command language allows us to present messages on the operator monitor. In the new configuration, this facility is used to send data to the Subject PC. It works as follows: Immediately before the start of any data-recording trial, "signal flashes" are displayed at eight certain positions near a corner of the operator screen. At each position, either a white block or the black screen background can be shown. A set of eight photo diodes receives this eight-bit information which is sent to the parallel port of the Subject PC via a specific interface. Thus, synchronization and even data transmission can be achieved.

Of course, this method does not work without delay. The blocks are not displayed before the electron beam of the monitor has reached their position, which it does 60 times per second. Furthermore, the signal is received by the Subject PC using the parallel-port interrupt which leads to further delay. These factors cause a total temporal uncertainty of ±50 ms. Obviously, the new arrangement is far from being optimal, but nevertheless it fundamentally increases the capabilities of OMNITRACK1. At least for the experiments presented in this work, the temporal "jitter" does not have any negative effect on the experimental data.

The advantages of this system compared to early technologies are undeniable, though: The subject's head is free with permitted deviation from the straight-ahead position up to 15^o and the field of view is practically unrestricted (80^o in the horizontal dimension and 60^o in the vertical). Moreover, experiments can be run under natural illumination conditions.

The average absolute maximal precision of the gaze-position measurement lies within the range from 0.7^o to 0.9^o. By using a new calibration interface based on parametrized artificial neural networks, the precision of measurement is improved to 0.4^o. This makes it possible to recruit even subjects wearing spectacles (see Pomplun, Velichkovsky & Ritter, 1994). The system's temporal resolution is 16.7 ms due to the camera frequency of 60 Hz. In order to detect a fixation, the system needs to analyze five successive frames, which means that fixations shorter than 83 ms cannot be registrated. Unfortunately, the result of an experiment consists only in a data file listing all fixations, their spatial and temporal positions, their durations, and the pupil size while fixating. No online-access to eye-movement data or analysis of gaze positions instead of detected fixations is possible. As far as these aspects are concerned, the advanced SMI EyeLink system offers relief.

2.1.3 The SMI EyeLink System

The SMI EyeLink can be viewed as the successor to OMNITRACK1. Accordingly, the drawbacks of OMNITRACK1 have been eliminated as far as possible:

- The standard EyeLink system includes both an Operator PC *and* a Subject PC. They are connected with each other by an Ethernet-link that does not only achieve synchronization, but also allows eye-movement data and other information to be transferred.

- The gaze position data are already available to the Subject PC about 6 ms after their registration. In contrast to OMNITRACK1, these data can now be used on-line even from self-written C or C++ programs to control the display. This allows us to display stimuli that are contingent on the gaze trajectory. Nevertheless, a simple command language for programming the experiments can be used instead.

- The temporal resolution is improved to 250 Hz. Every four milliseconds, actual gaze positions are accessible to the users who can decide if they want to detect saccades and fixations with the help of individual software or to use a built-in detection algorithm.

- Binocular eye tracking is feasible due to the introduction of a second eye camera. In spite of that, the weight of the head-set has decreased to 600 g which is a reduction of about 50% in comparison to OMNITRACK1. This is appreciated by the subjects, and what is more, its lower inertia makes head-set shifts and thus loss of data accuracy less probable.

Figure 2.2 demonstrates that the arrangement of the EyeLink system strongly resembles the extended version of OMNITRACK1, the interface between the PCs and the software flexibility have been substantially improved, though. The hardware has been upgraded as well: The Operator PC is a Pentium 133 using a 17" ViewSonic 7 operator monitor, and the Subject PC contains a Pentium 166 and

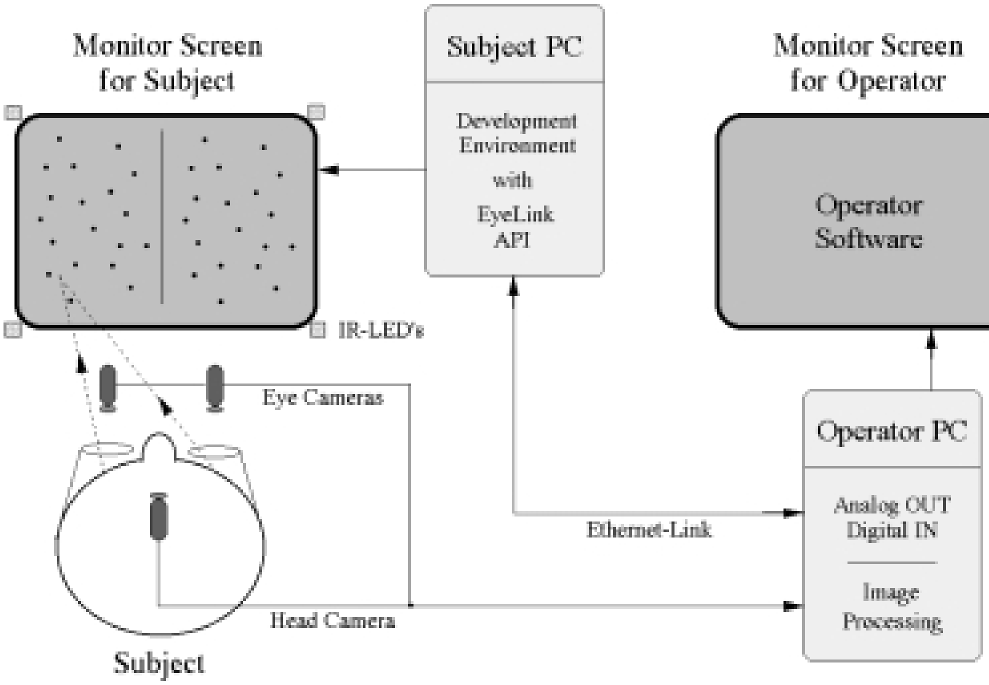


Figure 2.2: Scheme of the SMI EyeLink system

displays stimuli on a 20” Trinitron subject monitor. With respect to the subject’s field of view and the spatial accuracy of measurement, no changes were introduced.

2.2 Designing Stimuli for Comparative Visual Search

As shown in the previous section, the technical equipment enables us to implement C programs which generate individual stimuli for each subject and provide thus an excellent statistical basis for the analysis of gaze trajectories. The most important point to be clarified, then, is how to define the appearance of the stimuli. The following two sections concentrate on this topic.

The choice of stimuli is closely connected with the definition of independent and dependent variables. Which combination of stimuli, tasks, and methods of analysis is most promising to yield interesting information about comparative visual search? Appropriate definitions of variables are developed in Section 2.4.

2.2.1 Choice of Elementary Stimulus Items

When exploring a new paradigm, such as comparative visual search, it is usually a good strategy to first focus on a restricted set of stimulus dimensions. Consequently, we eliminated high-level factors in order to achieve explicit control of image parameters and their statistical distribution. Following the majority of studies in visual search, we decided to focus on three dimensions, namely *color, form,* and *location.* Instead of realistic items we introduced random distributions of simple geometrical items (squares, triangles, and circles) in three different colors (green, blue, and yellow). In every stimulus, exactly one mismatch was present which consisted in either the color or the form of one of the items.

The main advantages of this design are obvious: No high-level information, i.e. semantic content, interferes with the basic task of comparative search, and the stimuli are completely determined by a small set of quantitative parameters, i.e. the color, form, and location of items and the position and dimension (color or form) of the mismatch. Since there are identical numbers of different colors and forms, comparative analyses between these two dimensions can be performed. We can investigate, for instance, differences between the detection of color mismatches and the detection of form mismatches, or we can study distinct influences of color and form information on the subjects' eye-movement patterns.

The item forms were dimensioned in such a way that their subjective size appeared identical. For this purpose, a pre-test with five subjects was conducted. Each subject was successively shown six different scenes, each of them presenting a square, a circle, and a triangle in random positions. In each scene, the subjects had to adjust the size of the circle and the triangle in such a way that they appeared to have the same size as the square. The results showed that circles were adjusted to cover almost the same average area (221 pixels) as did the squares (225 pixels), while the triangles were adjusted to cover a substantially smaller area (187 pixels). Obviously, the less "compact" shape of triangles seems to make them look "bigger" than both squares and circles covering the same number of pixels. Since the subjective item size is a more adequate cognitive variable than is the number of covered pixels, we decided to use the empirically adjusted item sizes for our experiments.

For technical reasons, the choice of colors was restricted. The OMNITRACK1 system requires all stimuli to be converted into video format, because the implemented Subject PC replaces the system's video recorder (see Section 2.1.1). The resulting monitor output presents a slight fuzziness caused by high values in the red component of the video data. Therefore, the boundaries of red items exhibit a noticeably stronger blur than, for instance, blue items. Since the distinctness of specific boundaries is an important factor for object recognition, we decided to exclude the color red from the stimuli. Instead, the colors green, blue, and yellow were chosen.

An important requirement for the elementary items is that all possible pairings of colors and forms should appear equiluminant. While the colors' hue and saturation were set to adequate values (fully saturated green, blue, and yellow), their brightness could be adjusted to meet the requirement of equiluminance. The optimal brightness value for each pairing cannot be determined mathematically since the subjective perception of item intensity might be different from physical quantities. It might also depend on the form of an item and on perceptual inter-

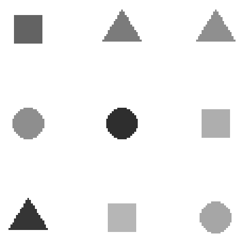


Figure 2.3: Simultaneous presentation of all possible pairings of colors and forms in a task requiring the adjustment of their brightness

actions between its color and its form. In order to account for these subjective factors, we conducted a pre-test investigating the subjects' perception of item luminance.

In this study, subjects were presented with arrangements of different items as shown in Figure 2.3. The nine possible pairings of colors and forms were shown as a 3×3 array with the brightness of each array item chosen randomly from the range between 50% and 100%. Both their hue and saturation, however, were identical within the color groups. The subjects could adjust the initially randomly chosen brightness values with the keyboard. The task for each subject was to adjust the individual brightness of each object in such a way that they perceived all of them as equiluminant.

This task was repeated six times for each subject, using a different random initialization of brightness values in each trial. Five subjects participated in this study, resulting in a total of 30 brightness values obtained for each combination of color and form.

Since there was no "reference luminance" given to the subjects, the absolute values strongly varied between subjects and trials. The relative values, however, exhibited almost the same significant pattern for all subjects: The brightness of green items was consistently adjusted about 8% higher than the brightness of yellow items, and the brightness of blue items was consistently chosen about 14% higher than the brightness of yellow items. No significant differences were found between distinct forms. Finally, there was no evidence for the hypothesized interaction between color and form.

The brightness of the elementary items for comparative visual search was adapted to these results. This was the final step in the "design" of appropriate items.

Before starting the experiments of comparative visual search, it is important to establish the perceptual features of the chosen items with regard to *standard* visual search. Undoubtedly, this knowledge is helpful for both the interpretation of findings and the construction of models in *comparative* visual search. With this aim in view, the following section is concerned with the investigation of possible "pop-out" effects among the chosen items.

2.2.2 Do the Chosen Features “Pop Out”?

In Section 1.4, we discussed the differences between parallel and serial search. If a target item differs from all distractor items along a specific dimension, it possibly “pops out” from the background. Such “pop-out” effect is characterized by the reaction time which is independent of the number of distractors and indicating that we are able to perform parallel search on the items. In contrast, reaction time slopes indicate serial search. It was stated that there is a continuum of search processes rather than a parallel-serial dichotomy.

This characterization of search efficiency might be relevant for comparative visual search as well. During task completion, local areas of the display are compared to each other. If, for example, such an area in the left hemifield contains only yellow items, while the corresponding area in the right hemifield shows yellow items plus the blue target item, it might be possible that “local pop-out” takes place.

It is important for us to know whether parallel detection of color or form attributes is possible in principle. If color targets, for example, can be detected in parallel, whereas form targets require serial processing, we expect comparative search for color mismatches to proceed faster than search for form mismatches, at least if subjects know the dimension of the mismatch.

Moreover, the *absolute* difference in reaction time between color and form targets in standard visual search should be reflected in the results of comparative visual search as well: At least the latency before the subjects’ manual response should exhibit a corresponding dependence on the relevant dimension. Finally, it is important to know whether reaction time and error rate vary with individual colors, forms, and specific target-distractor combinations. If there are striking differences in the detectability of target-distractor combinations, for example, subjects are likely to adapt their distribution of attention to this situation. They might pay more attention to those stimulus regions that are more “difficult” in this respect, impairing the analysis of eye movements as a function of the stimulus structure.

To investigate these issues, we first conducted a standard visual search experiment using the chosen items. In the experiment, subjects were shown random distributions of these items. In each trial, all items had the same color and form except for a possibly appearing target item that differed from them in either its color or in its form. The subjects had to press the left button on a computer mouse if a target was shown, and to press the right button otherwise. They were to react “as quickly and as precisely as possible”.

The experimental design incorporated five different independent variables:

Target presentation: In half the trials, a target was presented, in the other trials, it was not. It was important to integrate this factor into the design, because the presentation of targets in every trial or most of the trials would lead to “mechanical” manual reactions by the subjects.

Knowledge of the relevant dimension: The experiment was divided into three phases which came in random order and which were announced to the subjects: A *color* phase presenting a possible color difference between target and distractors, a *form* phase presenting a possible form difference, and a *color-or-form* phase randomly presenting either a possible color or a possible form difference. As can clearly be seen, both the color and form phases provide knowledge of the relevant dimension, whereas the color-or-form phase does not.

Dimension of the target-distractor difference: As mentioned above, the difference between target and distractors could be either in color or in form. The other dimension presented an identical value for all items which was randomly chosen.

Number of items: The displays could consist of three different numbers of items: 10, 20, or 30. As a matter of fact, their order of occurrence was randomized. The display size, however, remained constant (square-shaped with a lateral extent of about six degrees of visual angle), which allowed us to interpret the resulting reaction time slopes in the context of the distinction between serial and parallel search.

Target-distractor combination: If we define non-target stimuli as presenting targets of the same color and form as the distractors, there are nine possible target-distractor combinations for color and form each. The differentiation between these enables us to figure out effects of individual features (color and form) and their combination.

The definition of these variables led to 144 different types of stimuli, i.e. 144 cells in the experimental design. For each type, five correct reactions per subject were obtained, while incorrect judgements were disregarded in the analysis. Following the tradition of visual search, the dependent variables were chosen as reaction time and error rate. Seven subjects were tested; all of them were able to distinguish between colors.

As to the results, a five-factorial analysis was calculated for both reaction time and error rate. On average, an error rate of 1.75% was found which is rather a low value. The error rate did not depend significantly on any of the five independent variables.

Reaction time, however, presented a significant dependence on the dimension of the target-distractor difference ($F(1;6) = 29.91; p = 0.002$). It was shorter for color differences (601.59 ms) than for form differences (700.43 ms). The number of items in the display did not exert a significant effect on reaction time. As shown in Figure 2.4, neither color search nor form search induced reaction time slopes, indicating strong "pop-out" effects for both dimensions. As in every bar chart in this work, the errorbars indicate the respective *standard error*.

Furthermore, the knowledge of the relevant dimension significantly influenced reaction time ($F(1;6) = 10.69; p = 0.017$). Reaction time was shorter when subjects knew the relevant dimension (621.85 ms) than when they did not (680.17

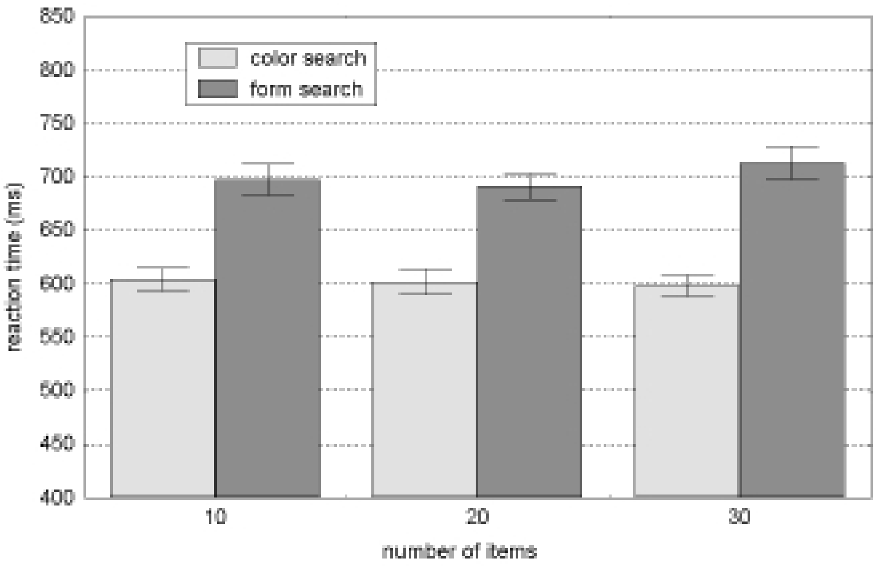


Figure 2.4: Reaction time as a function of the number of displayed items

ms). Also, this variable presented a significant interaction with the dimension of the target-distractor difference ($F(1;6) = 11.66; p = 0.014$). As illustrated in Figure 2.5 (left), the knowledge of the relevant dimension (specific search) facilitated color search more strongly (553.20 ms versus 649.99 ms) than form search (690.50 ms versus 710.35 ms). This finding suggests a fundamental asymmetry between the processing of color and the processing of form. It motivates a closer investigation in the context of comparative visual search (cf. Chapter 4).

Target presentation exhibited a reliable effect on reaction time as well ($F(1;6) = 10.04; p = 0.019$). The presentation of a target led to shorter reaction time (621.30 ms) than did the non-target situation (680.72 ms). Moreover, a significant interaction with the dimension of the target-distractor difference was established ($F(1;6) = 7.79$; $p = 0.032$). Figure 2.5 (right) shows that this interaction corresponds to the previous one: Color search benefits more strongly from the presentation of a target item (561.66 ms versus 641.52 ms) than does form search (680.94 ms versus 719.91 ms). At least a part of this effect can be attributed to the faster reaction to color targets than to form targets even if the relevant dimension is unknown (see above). If no target is shown, the situation of unspecific search leads to completely identical stimuli for color and form search and thus to identical reaction times. Since half the trials provide unspecific search, the reaction time difference in the "no target" condition diminishes.

Do individual color or form combinations exert specific effects on reaction time? Figure 2.6 shows the reaction times for different feature combinations between target and distractors with regard to the relevant dimension. The light columns indicate reaction time for the target-distractor combination specified at the bottom of the respective column, while the dark columns refer to the inverse

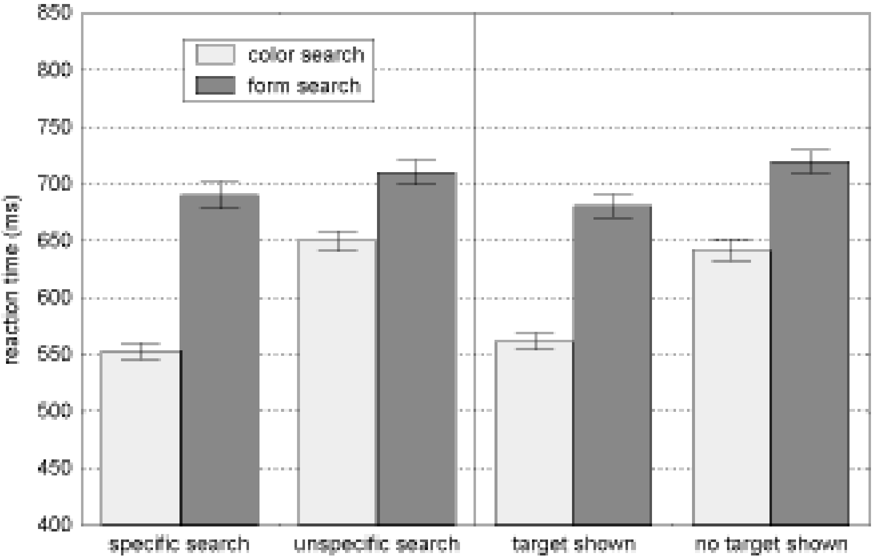


Figure 2.5: Interactions of the relevant dimension (color or form) with specification or target presentation respectively

situation, e.g. a green target and yellow distractors for the leftmost dark column. Only those trials with target presentation are taken into account.

With regard to color combinations, a significant influence on reaction time was found ($F(5; 30) = 6.23; p < 0.001$). The combinations yellow-green and green-yellow induced significantly longer reaction times than all other combinations except for blue-green. Form combinations revealed a reliable effect as well ($F(5; 30) = 5.36; p = 0.001$). The combination square-triangle led to reliably longer reaction time than the other combinations apart from the inverse combination triangle-square. The combination triangle-square, in turn, caused significantly longer reaction time than both square-circle and triangle-circle.

With regard to the “no target” condition, do the item features still affect reaction time? No such influence was found in the case of color search (see Figure 2.7). Although the mean value for blue items was lower than for both yellow and green ones, this difference did not reach significance. Form search, however, showed a significant effect ($F(5; 30) = 3.05; p = 0.024$). Stimuli consisting of triangles led to significantly longer reaction time than those consisting of squares or circles. Some of the subjects reported distributions of triangles to constitute especially “complex” structures which impair the detection of a possible target.

All in all, the contrasts between yellow and green as well as between squares and triangles exhibit a tendency towards longer reaction time in comparison to other contrasts in the respective dimension. Non-target displays consisting of triangles induce longer reaction time than others. These relatively small differences are compatible with our demand for ecological validity. They are unlikely to exert an influence on the subjects’ gaze trajectories which could decrease the validity

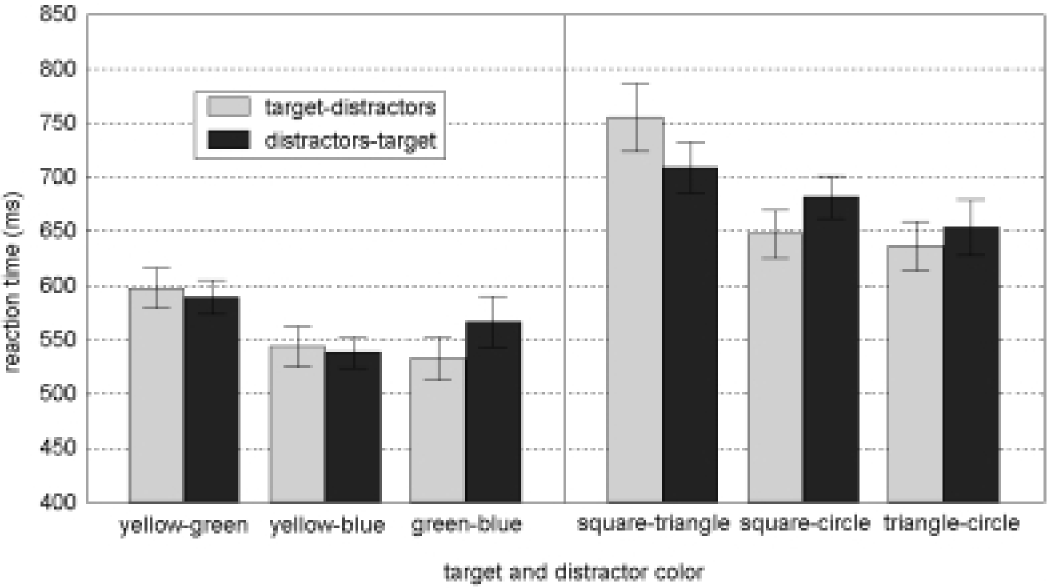


Figure 2.6: Reaction time for different combinations of target-distractor features

of results. *Perfectly* equal detectability of different items is neither ecologically adequate nor strictly required for experimental control.

As mentioned in Section 1.5, comparative visual search can be considered a composition of visual search and picture matching. Since we have investigated the properties of the chosen items with regard to visual search now, it is near at hand to take a look at picture matching as well.

2.2.3 Picture Matching and Retinal Eccentricity

A task of picture matching, using the introduced items, was performed in our group. In his experiment, Marcus Becker combined a serial same-different task with the investigation of the influence by retinal eccentricity.

As discussed in Section 1.2, visual attention can be shifted without employing eye movements. A divergence between the foveal gaze position and the attended region of the visual field, however, leads to reduced visual acuity with regard to the attended information. Therefore, shifts of attention are usually accompanied by eye movements in order to provide optimal visual acuity. During rapid processes of scanning like in comparative visual search, small "covert" shifts of attention might occur within fixations. These covert shifts have been investigated in numerous studies. Typically, in these studies subjects are asked to constantly fixate a marker displayed in the center of the screen. They are presented with target items appearing at different positions on the screen for variable durations. Their task is to press a specific button corresponding to the type of target, i.e. they have to recognize the type of target without moving their gaze position away from the marker. Experiments of this kind are well-suited for the investigation

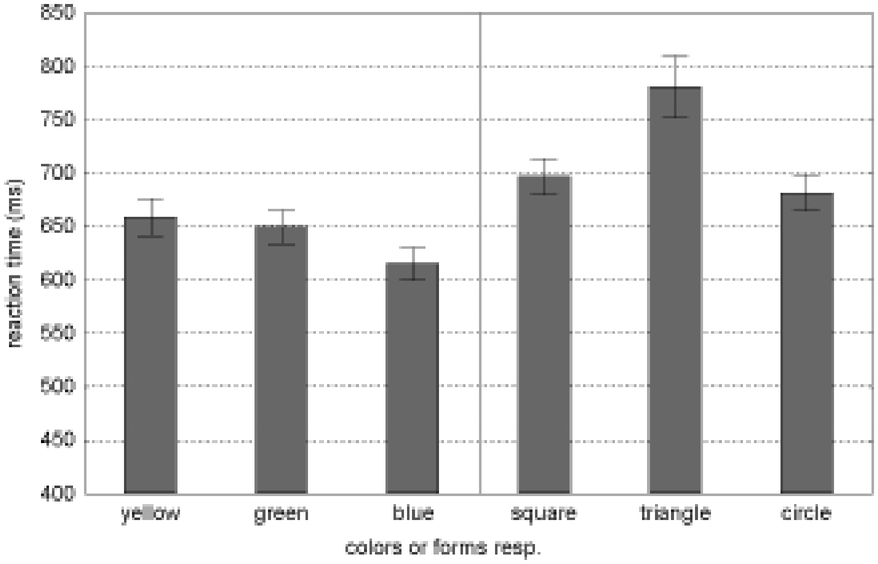


Figure 2.7: Reaction time depending on distractor features in trials without target presentation

of the nature of attentional shifts, since the time needed to direct attention to specific locations can be estimated (for a review, see e.g. Posner, 1980; Wright & Ward, 1994). There is evidence for these shifts to have an approximately constant maximum velocity. According to the findings of Tsal (1983), this velocity is about one degree of visual angle per eight milliseconds. However, these results must be viewed with discretion, because they are based on rather rough measurements.

With regard to our intended research, it would be interesting to know the "costs" of covert attentional shifts needed for the identification of the elementary items. This knowledge would enable us to estimate the amount of covert shifts being applied during comparative visual search, which is important for both the interpretation of eye-movement patterns and the construction of computational models. The results of Becker's experiment provide some information in this respect.

In his experiment, subjects were presented with an item of random color and form in the center of the screen. The possible colors and forms were identical to those chosen for comparative visual search. After one second, the presented item was replaced by a white dot which served as a marker. Subjects were to fixate on this point all the time, while a second item was shown either above, below, to the left of, or to the right of the marker. Corresponding to the experiment of visual search, subjects had to press the left button on a computer mouse if the two items differed in the relevant dimension (which was known), and to press the right button otherwise. Three different independent variables were systematically varied:

Relevant dimension: The relevant dimension was either color or form. Subjects

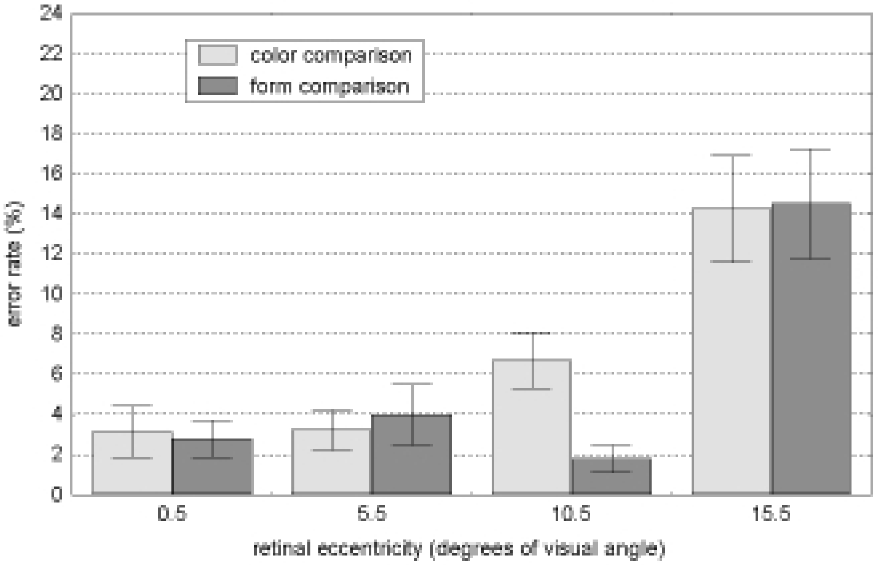


Figure 2.8: Error rate depending on retinal eccentricity

always knew this dimension in advance.

Retinal eccentricity: The retinal eccentricity of the second item was randomly varied among four levels: 0.5 degrees (as a baseline), 5.5 degrees, 10.5 degrees, and 15.5 degrees of visual angle (the size of the fovea is about 1 degree of visual angle). It should be noted that eccentricities of 10.5 degrees and especially 15.5 degrees are unlikely to occur in our scenario of comparative visual search.

Direction of the shift: As mentioned above, there were four possible, randomized directions of shifts between the two items: Up, down, left, right.

Each of the ten participating subjects made ten judgements for each of the possible combinations, resulting in a total of 320 trials per subject. The empirical error rates and reaction times were entered into a three-factorial analysis of variance.

Retinal eccentricity revealed a significant effect on error rate ($F(3;27) = 4.58; p = 0.010$). Error rate was significantly higher for retinal eccentricity of 15.5 degrees (14.38%) in comparison to all other eccentricities (2.97% for 0.5 degrees, 3.60% for 5.5 degrees, and 4.23% for 10.5 degrees); no other contrasts reached significance. These data are shown in Figure 2.8. Neither of the factors relevant dimension and direction of the shift exerted a significant effect on the error rate.

The analysis of reaction time led to a similar pattern of results. The main effect of retinal eccentricity reached significance ($F(3;27) = 11.10; p < 0.001$). An eccentricity of 15.5 degrees induced longer reaction time (819.1 ms) than the

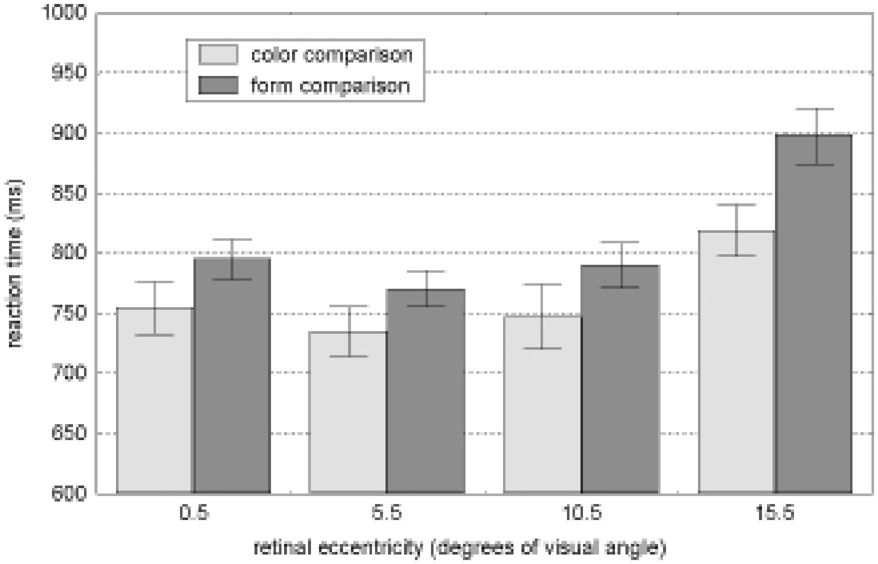


Figure 2.9: Reaction time as a function of the second item's retinal eccentricity

other levels (775.0 ms for 0.5 degrees, 752.8 ms for 5.5 degrees, and 768.4 ms for 10.5 degrees). The other contrasts remained below significance. Although the diagram shown in Figure 2.9 seems to suggest that the comparison of colors can be accomplished faster than the comparison of forms, this difference is not reliable. Finally, the direction of the shift was not significant either.

These results indicate that shifts of attention with an amplitude of 5.5 degrees or 10.5 degrees of visual angle do not induce noticeable "costs" of reaction time or error rate. Only the condition of 15.5 degrees reveals a significant effect. This value of retinal eccentricity, however, is not relevant for our intended experiments (see above). If subjects employed attentional shifts of this amplitude they would not need to move their eyes at all during search, which is absolutely implausible.

We must take into account, however, that the situation in comparative visual search is more complex than in the experiment described here. During comparative search, the information of other items shown in the display *interferes* with the currently relevant information. Therefore, stimulus areas presenting accumulations of objects can be expected to induce stronger effects of retinal eccentricity than those measured in Becker's experiment. This point is supported by numerous studies in visual search (e.g. Mackworth, 1976). Nevertheless, the results suggest that shifts of attention are very efficient in our scenario. It is reasonable to assume that subjects apply many covert shifts of attention in addition to eye movements in order to reach optimal search efficiency. This finding constitutes an important baseline for the interpretation and modelling of results.

The error rates and reaction times for individual feature combinations were analyzed as well. A specific combination of colors, namely yellow and green, turned out to induce substantially higher error rates than other combinations in situations of high retinal eccentricity (15.5 degrees). As mentioned before, possible retinal eccentricities during comparative visual search are clearly lower, hence this finding is irrelevant for the judgement on our choice of item colors.

All in all, the chosen items have proven to be well-suited for the intended research. The encountered differences between individual colors and forms are unlikely to interfere with experimental control. Furthermore, the basic properties of the items with regard to visual search and picture matching have been determined, resulting in valuable baseline data. These data will be referred to in the context of comparative visual search.

Logically, the next step consists in the composition of these elementary items to stimulus pictures. The following section explains how appropriate distributions of items can be achieved.

2.3 Algorithmic Generation of Item Distributions

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

- Items may overlap and thus constitute new, “hybrid” forms which are likely to induce distinct recognition processes and to violate the simple parametric structure of the stimuli. Hence, there has to be a minimum distance of 20 pixels between neighboring items. Apart from this constraint, items are to be placed homogeneously randomly within a rectangle of 260×400 pixels in each hemifield.

- It may happen that in some stimuli specific colors or forms are strongly overrepresented. Inhomogeneous sets of items can lead to deviating search strategies, biasing the experimental results. Hence, we demand that each of the three colors and the three forms must be equally represented, i.e. there have to be 10 triangles, 10 blue items etc. No specific combinations of color and form may appear more frequently than others.

- A homogeneous random function is unlikely to yield large regions of same colors or forms within a stimulus. The occurrence of such regions, however, is an important precondition for investigating the effects of item uniformity versus item variety on eye movements. Thus, we should use a random function that tends to create regions of same attributes more likely than a homogeneous random distribution would do.

Obviously, the first two points can easily be achieved, whereas the third one requires the mathematical definition of color and form *segregation* respectively and an algorithm being capable of generating color and form distributions according to these two 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: A pattern is a set of N items (objects)

$$\mathbf{o}^{(n)} = \begin{pmatrix} o_x^{(n)} \\ o_y^{(n)} \\ o_f^{(n)} \\ o_c^{(n)} \end{pmatrix}, \quad n = 1, \ldots, N, \tag{2.1}$$

where $(o_x^{(n)}, o_y^{(n)})$ is the pixel position of the item's center on the screen, $o_f^{(n)}$ is the item's form (1 = square, 2 = triangle, 3 = circle), and $o_c^{(n)}$ is the item's color (1 = blue, 2 = green, 3 = yellow). In our experiments, every pattern consists of two almost identical parts and in the following these two hemifields are described separately, so $N = 30$.

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 items with different forms and the mean distance $\bar{d}_{f,id}$ between those with identical forms:

$$\alpha_f = \frac{\bar{d}_{f,dif}}{\bar{d}_{f,id}} \tag{2.2}$$

$$\bar{d}_{f,dif} = \frac{\sum_{n_1=1}^{N} \sum_{n_2=n_1+1, o_f^{(n_1)} \neq o_f^{(n_2)}}^{N} d(n_1, n_2)}{\sum_{n_1=1}^{N} \sum_{n_2=n_1+1, o_f^{(n_1)} \neq o_f^{(n_2)}}^{N} 1} \tag{2.3}$$

$$\bar{d}_{f,id} = \frac{\sum_{n_1=1}^{N} \sum_{n_2=n_1+1, o_f^{(n_1)} = o_f^{(n_2)}}^{N} d(n_1, n_2)}{\sum_{n_1=1}^{N} \sum_{n_2=n_1+1, o_f^{(n_1)} = o_f^{(n_2)}}^{N} 1} \tag{2.4}$$

$$d(n_1, n_2) = \sqrt{(o_x^{(n_1)} - o_x^{(n_2)})^2 + (o_y^{(n_1)} - o_y^{(n_2)})^2} \tag{2.5}$$

For example, $\alpha_f = 2$ means that, on average, items of different forms are twice as distant from each other than items of the same form. In our setting of 30 items and three different forms this would correspond to a strongly segregated distribution containing large uniform areas. $\alpha_f = 1$ means that there is no effect of segregation at all. Fig. 2.10 illustrates the correspondence between α_f and the distribution of forms at three different levels.

Analogously, we define the parameter α_c of color segregation. An iterative algorithm for generating color and form distributions to given parameters of form segregation α_f and color segregation α_c can easily be implemented. Starting at a random distribution, this algorithm randomly selects pairs of items and exchanges their colors or forms if this exchange shifts the distribution's segregation levels

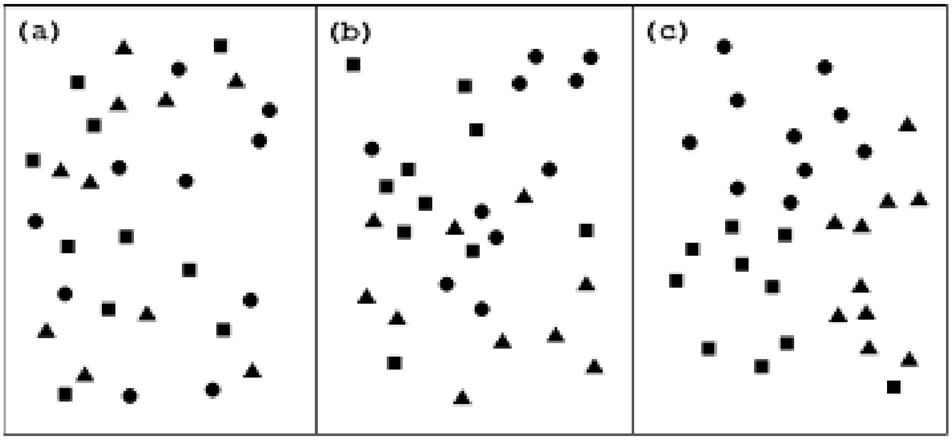


Figure 2.10: Examples of item distributions at a form segregation α_f of (a) 1.0, (b) 1.3, and (c) 1.8.

towards the given parameters. The algorithm terminates as soon as the difference between the actual and the desired α_f and α_s falls below a certain threshold.

In our experiment we used this algorithm to compute the stimuli. For each scene, α_f and α_s were set to random values ranging from 1.0 to 1.3. Figure 2.11 presents an example for the resulting type of "abstract" stimuli. The right hemifield is a copy of the left one which is horizontally shifted by 360 pixels. A form mismatch is integrated into this stimulus: One of the squares in the left hemifield is replaced by a triangle in the right hemifield. For technical reasons, in the following chapters the search pictures will not be shown in their original colors, but as items of different grey levels on a white background.

Stimuli of this kind will be used in our experiments of comparative visual search. These stimuli can be considered as the "abstract variant" of typical original-and-fake pictures (see Figure 1.7). All high-level semantic information has been eliminated for the sake of experimental control.

Speaking of experimental control, we must define adequate independent and dependent variables in order to make use of the excellent experimental conditions provided by the stimuli. The fact that the stimuli are determined by a set of quantitative parameters should ideally lead to quantitative results with regard to the influence of stimulus parameters on eye-movement parameters. Therefore, we need suitable mathematical definitions of variables for the analysis of eye-movement data.

2.4 Analysis of Empirical Data

2.4.1 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

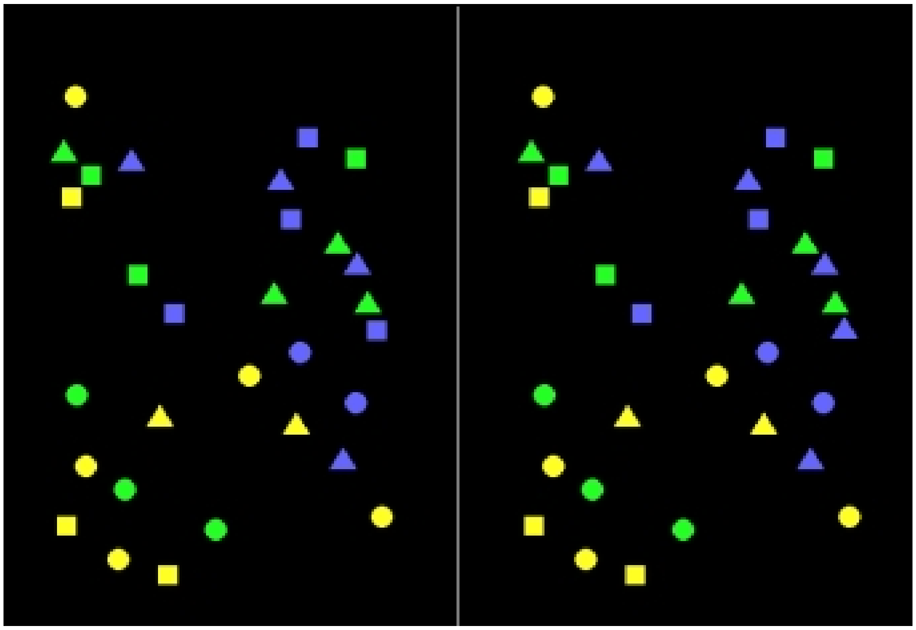


Figure 2.11: Example of a randomly generated stimulus picture

an initial parallel phase of stimulus inspection as in standard visual search, target detection itself cannot be accomplished in parallel. Thus, we can assume 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 pictures into their essential features.

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

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

Color entropy $S_c(\mathbf{p})$: Color entropy is a measure of local color “disorder”. If items 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 items around $\mathbf{p}$ without exception, $S_c(\mathbf{p})$ will yield a value near its absolute minimum. It is important that this function has to be defined in such a way that it is completely uncorrelated with item density.

Form entropy $S_f(\mathbf{p})$: This function is the equivalent to color entropy with respect to the items' *forms*. We demand that form entropy is neither correlated with item density nor with color entropy.

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

Item Density

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

$$w(n, \mathbf{p}) = \exp\left(-\frac{(o_x^{(n)} - p_x)^2 + (o_y^{(n)} - p_y)^2}{2\sigma^2}\right) \tag{2.6}$$

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 accomplished since there is no invariant "focus size" in human vision. We decided to use a value of 25 pixels on the screen or one degree of visual angle respectively, corresponding to the size of the idealized human fovea. To calculate the local item density $\varrho(\mathbf{p})$, the weight functions for all items are summed up as follows:

$$\varrho(\mathbf{p}) = \sum_{n=1}^{N} w(n, \mathbf{p}) \tag{2.7}$$

How plausible are the results yielded by this density function? Fig. 2.12 demonstrates its effect on the right hemifield of the example picture previously shown in the introduction (Figure 2.11). By comparing the underlying item distribution with its density "landscape", a clear and plausible correspondence can be seen. For example, the item accumulation in the upper left corner of the picture causes the highest density "hill". Naturally, the picture's *left* hemifield would show an identical landscape, because its items are located at the same positions. $\varrho(\mathbf{p})$ ranges approximately between 10^{-3} and 3.9, but less than 1% of its values exceed 3.0.

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 items, 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:

$$\varrho_i(\mathbf{p}) = \sum_{n=1, o_f^{(n)}=i}^{N} w^*(n, \mathbf{p}) \;\;, \;\; i = 1, 2, 3 \tag{2.8}$$

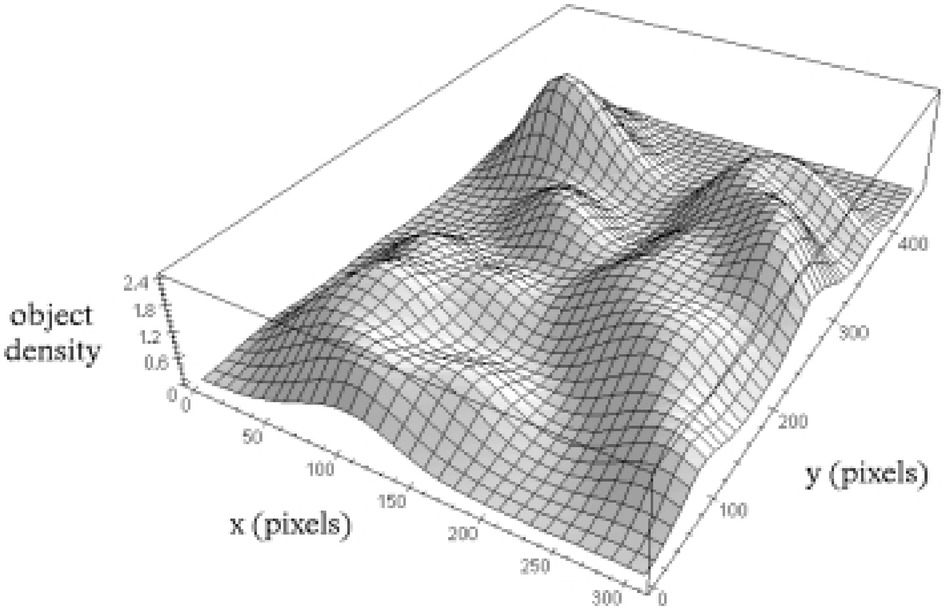


Figure 2.12: Item density $\varrho(\mathbf{p})$ in the example picture's right hemifield

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

$$S_f(\mathbf{p}) = -\sum_{i=1}^{3} \frac{\varrho_i(\mathbf{p})}{\varrho^*(\mathbf{p})} \ln \frac{\varrho_i(\mathbf{p})}{\varrho^*(\mathbf{p})} \quad , \text{ where } \varrho^*(\mathbf{p}) = \sum_{n=1}^{N} w^*(n, \mathbf{p}) \qquad (2.9)$$

If we used the function $w(\cdot,\cdot)$ from Equation (2.6) as $w^*(\cdot,\cdot)$, $S_f(\mathbf{p})$ would yield the aspired values of form entropy, but also exhibit the undesirable feature of being correlated with the local item density $\varrho^*(\mathbf{p})$. This is caused by the fact that in regions of low item density there are large areas being dominated by the influence of a single item and thus presenting low entropy.

Obviously, a different choice of $w^*(\cdot,\cdot)$ is required, implying some standardization to compensate for item density. We could vary the standard deviation σ of the Gaussian weight function (2.6), in such a way that $\varrho^*(\mathbf{p})$ is constant for all $\mathbf{p}$. In other words, the implemented "human focus" would expand at lower and narrow at higher local density, a plausible solution.

However, simply adjusting σ would create an effect of 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 items and $\mathbf{p}$ would shrink by the same factor. Although this induces a standardized item density in the proximity of $\mathbf{p}$, there is now a disproportionate high influence of the approached peripheral items. In a tendency, a significant influence of more items leads to higher entropy values and thus to a correlation between entropy and density.

A possible solution is found by varying the distances between the items and $\mathbf{p}$ by the same *offset* Δd instead of the same *factor*. Occuring distances lower than zero are rounded to zero. Using an offset ensures a peripheral item influence which is independent of $\varrho(\mathbf{p})$. The resulting equation reads as follows:

$$w^*(n,\mathbf{p}) = \exp\left(-\frac{\max(\sqrt{(o_x^{(n)}-p_x)^2+(o_y^{(n)}-p_y)^2}+\Delta d),0)^2}{2\sigma^2}\right) \qquad (2.10)$$

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

$$\sum_{n=1}^{N} w^*(n,\mathbf{p}) = c_s = \text{const} \ , \qquad (2.11)$$

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

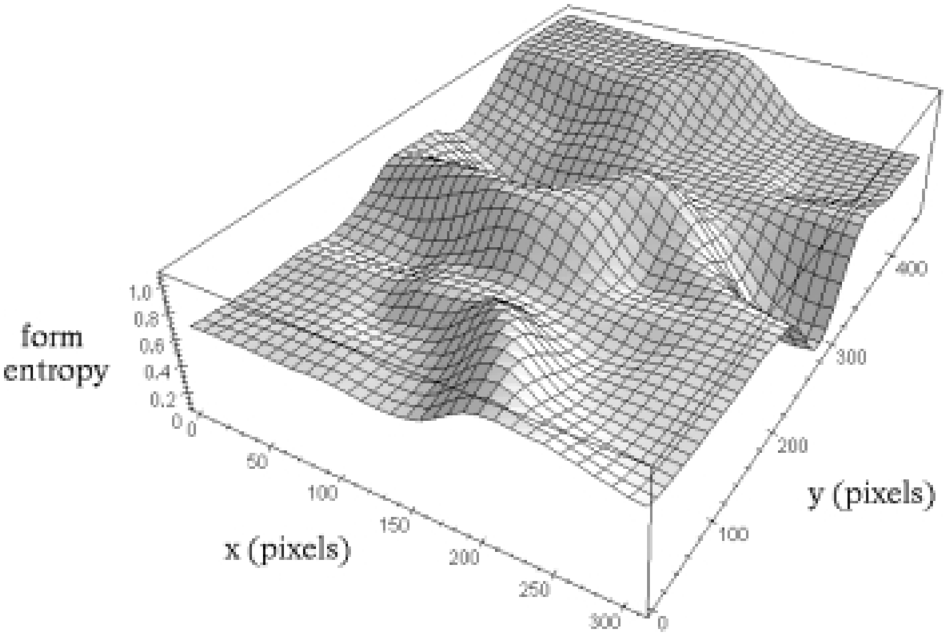


Figure 2.13: Form entropy $S_f(\mathbf{p})$ in the example picture's right hemifield

Fig. 2.13 shows a graph of the resulting function $S_f(\mathbf{p})$, again referring to the right hemifield of the example picture in Figure 2.11. Once more, we find a plausible correspondence between the parameter function and the underlying picture. 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 the highest plateau.

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

2.4.2 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 five derived variables that carry essential information. In the following, the methods of measuring these variables are described in full detail in order to make the partially complicated analysis of the gaze-position data transparent.

Motivated by the subjects' 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*, subjects are likely 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 defined to occur when the following two conditions are met: (a) the subject's gaze position gets closer than 50 pixels (i.e. two degrees of visual angle) to any of the target items and (b) a manual reaction is registered within the next two seconds. The two phases are analyzed separately in order to properly distinguish search and comparison from cognitive processes involved in the ascertaining of a mismatch.

Variables marked by an asterisk are not measured during the first second after a stimulus is presented, in order not to let the subject's "phase of 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 picture^{+} (NF): This is the total number of fixations per search process, accounting for fixations in both hemifields.

Duration of fixation^{*+} (FD): Duration of fixation is simply the value in milliseconds registrated by the eye tracker for each fixation. Its temporal resolution is 16.7 ms with regard to the OMNITRACK1 system and 4 ms with regard to the EyeLink system.

Saccade length^{*+} (SL): Saccade length is measured as the distance in pixels between two successive fixations in the same hemifield. If the dependence of SL on local stimulus parameters is investigated, this local parameter is measured and analyzed separately at the saccade's starting or landing point respectively. This method turned out to yield more valid and detailed information than the analysis of the arithmetic mean of these values or of the parameter integral 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 into the other hemifield. How should a local parameter that corresponds to a series of fixations be measured? 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 item density above average causes a problem: The more fixations are to be accounted for, the more likely their center is to be in a region of medium density. Accordingly, an analysis of this kind would yield an artificial negative correlation between FW and local item 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 if 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 enters both target areas (radius = 50 pixels) with only one saccade between visiting each area, a target passage is counted. If, 1.5 seconds later, the gaze-position has left the target area again and the subject does not press the button within the following 2 seconds, a missing of the target is counted. It is possible (and almost certain) that another target passage is registrated during the same search process. FW is the quotient of the target missing and the target passage counter.

Additional search time (AT): AT is the amount of time that is "wasted" by missing the target. It is measured from the start of the first target passage of a trial to its final one. This means that if the target is not missed at all, AT has the value 0.

Area coverage per fixation* (AC): During a single search process, the subject produces a fixation *pattern* on the stimulus picture. The average area in the picture covered by one fixation is the desired value, but we need an appropriate method of determining it.

Fig. 2.14 shows an outline of such a fixation pattern with a marked fixation f in its center. The area covered by this fixation can be defined by introducing four quadrants with the origin f. These quadrants are rotated by an angle of 45^o with respect to the screen coordinate system, because saccade directions were found to tend to both the horizontal and to the vertical axis. Now four distances $d_1, \ldots, d_4$ between f and the nearest neighboring fixation in each quadrant are measured and their arithmetic mean μ is taken. In the case of no fixation in one or more quadrants, μ is calculated regarding only the

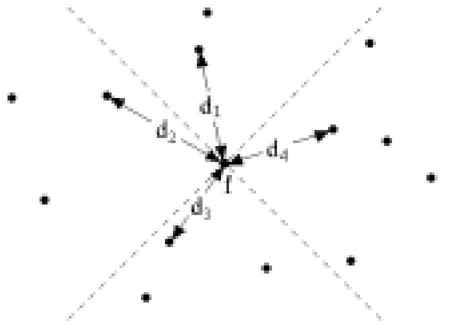


Figure 2.14: Fixation point f and distances $d_1, \ldots, d_4$ to its nearest neighboring fixation in each quadrant

remaining distances. We define the area AC covered by f as

$$AC = \pi(\frac{\mu}{2})^2 \tag{2.12}$$

Only fixations in the same hemifield as f are considered, and, naturally, only search processes with more than one fixation per hemifield are evaluated. Fixations registrated later than 7 seconds after the presentation of the stimuli are not evaluated in order to avoid an overlapping of two or more successive passages of the search process through the same region.

Speed of processing* (SP): Due to the observation that in almost all cases subjects scan the display from the top to the bottom or vice versa, SP accounts only for the vertical component (y-component) of eye movements. It is defined as the distance between the y-coordinate y_{start} of the first fixation and the y-coordinate y_{end} of the last fixation within the measuring range divided by the difference in time between the starts of these two fixations ($t_{end} - t_{start}$). Consequently, it is measured in pixels per second. Figure 2.15 illustrates the determination of the measuring range. The first fixation to be recorded at least one second after stimulus onset is always taken as the starting point of the measuring range (y_{start}, t_{start}). The following fixations are successively processed; their y-distance to y_{start} is calculated and the current maximum y-distance derived. As soon as any y-distance is shorter than the maximun y-distance by more than 50 pixels (caused by y_{return} in the example), the algorithm stops. If the verification phase is initiated (see above) the algorithm stops as well. The fixation corresponding to the maximum y-distance is taken as the end point of the measuring range (y_{end}, t_{end}). This definition of SP accounts only for the first, straight part of a gaze trajectory, which is an appropriate standardization providing comparability between trials, subjects, and experiments relatively independently of both the starting points of scanpaths and of the probability of missing the target.

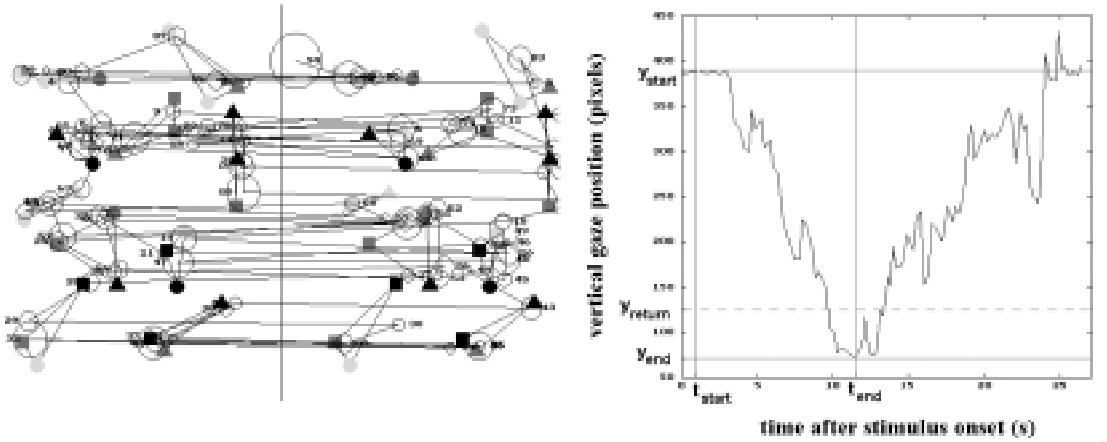


Figure 2.15: Measurement of SP in a gaze trajectory (left side). The diagram (right side) shows the vertical component of the gaze position as a function of the time after stimulus onset.

2.5 Summary

In this chapter, we have established the basic scenario for comparative visual search. We chose three simple geometrical forms (square, triangle, and circle) and three fully saturated colors (blue, green, and yellow) for the elementary items. In a pre-test, the *size* of these items was adjusted in such a way that they appeared to have identical size. While the *hue* of the items' colors was determined by technical restrictions, their *brightness* was empirically adjusted for equiluminance between all items.

In a visual search experiment, both the colors and the forms proved their capability to create strong "pop-out" effects. Furthermore, the results of the study revealed color targets to be detected faster than form targets. If subjects were told the relevant dimension (color or form) in advance, their average reaction time was shorter than without this knowledge. These findings are particularly interesting, because we can expect to observe corresponding effects in *comparative* search.

Another study investigated the influence of retinal eccentricity on the recognition of item features. The results suggest that even eccentricities of ten degrees of visual angle do not severely impair recognition, since neither reaction time nor error rate increases significantly. Therefore, we should expect a large proportion of covert shifts of attention during comparative visual search which are – as a matter of course – invisible to eye-movement analysis.

We decided on two – except for the mismatch – identical random distributions of these items to constitute the stimuli for comparative visual search. Three constraints are applied to the random function: First, there has to be a minimum distance between neighboring items; second, the items' colors and forms have to be counterbalanced within the stimuli; and third, there is a tendency towards the creation of uniform areas for the investigation of homogeneous versus heterogeneous local item configurations.

The independent variables to be investigated were chosen as local item density, local color entropy, and local form entropy. While density measures the accumulation of objects at a specific location, entropy is a measure for the local "disorder", i.e. the heterogeneity of the respective feature in a region of the stimulus. Dependent variables are basic eye-movement variables like fixation duration and saccade length, but also additional variables that were specifically derived for comparative visual search, e.g. speed of processing.

Thus, we have prepared the "ground" for our research. The following Chapter 3 will describe the first experiment conducted on this basis.

Chapter 3

Experiment A: The Basis Scenario

3.1 Variables and Hypotheses

At this point, all necessary "tools" for our research are prepared: First, we have defined suitable stimuli for comparative visual search and tested the basic properties of elementary items. Second, we have developed promising methods, i.e. independent and dependent variables, for the analysis of empirical data. So, what is the best plan for the conduction of experiments?

First of all, the basis scenario, as defined in the previous chapter, has to be thoroughly investigated. All possible effects of the independent variables on the dependent ones have to be tested in order to establish basic findings about mental processes and their interaction during comparative visual search. These findings, in turn, are likely to motivate the examination of further, possibly more specific aspects.

As a consequence, the stimuli and/or the task will be modified in order to provide a basis for the investigation of emerging questions. Since the results of further experiments raise further questions, an "explorative iteration" of experiments can be conducted. Proceeding this way, the understanding of comparative visual search can be extended in a flexible and sensible manner.

At a certain stage of knowledge gained about this paradigm, we should start to develop computational eye-movement models incorporating this knowledge. Starting with a sophisticated model, however, is probably an inappropriate strategy, because the detection of essential mistakes in complex approaches can be difficult. Instead, it might be useful to start with a very simple model. The comparison of scanpaths generated by this model with empirical gaze trajectories could indicate the requirements for a more adequate model.

A plausible model, yielding scanpaths that resemble the empirical ones, can substantially support the results of our research. The integration of empirical research and computer simulation is a powerful approach to form an extensive understanding of complex processes. Therefore, the development of computational models constitutes the final part of this work.

A central issue of our first, basic investigations of comparative visual search is the following question: How does comparative search proceed? Since the number of items in each hemifield is relatively large, comparative search must proceed in several steps. Accordingly, saccades within one hemifield should be distinguishable from longer saccades between hemifields. The oculomotor parameters should, to some extent, depend on the *local* structure of the display, in particular on the attributes that hold for the fixation point in question. As a matter of course, the *global* gaze trajectories should be influenced by the global structure of the stimulus.

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 genuine form; however, they may serve to illustrate the range of possibilities and provide three useful points of reference for a discussion.

3.1.1 The Travelling Salesman Strategy

A first strategy could be termed the "travelling salesman" strategy. Its name is derived from the travelling 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 round-trip that is possible. If subjects use an analogous strategy for their scanpaths, they would be expected to prefer object-to-object paths of minimal 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 the items in the display. Second, the task of comparing corresponding items influences the 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 to deviate from the "optimal" scanpath in terms of the TSP. Hence, there is a "trade-off" between memory usage and scanpath minimization. Let us assume – for the benefit of a simple model – that subjects memorize only one item at a time. What could we expect their scanpaths to look like?

After memorizing the first item in any of the two hemifields (A), it is compared to its counterpart in the other hemifield (B). In order to minimize the length of their scanpaths, subjects are likely to memorize the next item in hemifield B rather than to switch back to hemifield A and memorize the next item there. This means that after the memorization of the first item, subjects are likely to process successively two items in the same hemifield: The item 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 scanpath – the one that touches every item pair with as short saccades as possible. On the assumption that two fixations fall to each item 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. However, if fixations cover more than one item, the travelling salesman strategy would become indistiguishable from the strategies below.

3.1.2 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 items covered would depend on the density of the items in that particular subregion. Thus, topology now determines which particular items 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 travelling 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. Anyway, the viewpoint here is a spatial one – to scan the display with a minimal number of fixations so that every item up to the target item is visited at least once (and, at best, only once), while keeping saccades as short as possible. Similarly as for the travelling salesman strategy, two successive fixations per "visit" of one of the hemifields would be expected. With the searchlight strategy, the spatial distribution of items in the display should be an important determinant of fixation duration and saccade length; one would expect longer fixations on subareas with a high item density, which is also where saccades should be preferentially aimed at.

3.1.3 The Clustering Strategy

A third strategy could be to proceed cluster by cluster: With each fixation, subjects might process a certain number of items. These items would be assumed to be the maximal subset of items 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 items closest to the fixation point.

(2) Definition by distance: A cluster consists of all items within a circle of radius r around the fixation point.

(3) Definition by item attributes: A cluster consists of those items 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 items' features, i.e. items of the same color or form could be processed with the same effort as items of mixed colors and forms, which is rather implausible. Furthermore, the distances between the items are not considered; for example, two items with a distance of 15 degrees of visual angle between them could be processed as efficiently as two items with a distance of 2 degrees between them. Definition (2) is more adequate with respect to visual attention, 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 item features on processing effort. It is completely inadequate, however, in its representation of memory capacity: Any number of items 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 items, with their eccentricity, and with the entropy of their attributes. If the local item 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 less and fixation duration to be longer than with the travelling salesman strategy; to some extent, fixation behavior should depend on the information content, or entropy, of the items 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.

3.1.4 Strategy Issues

A common viewpoint to characterize the three candidate strategies would be in terms of the minimization of a cost function. While the travelling 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 the 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. This is achieved by grouping the items 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 items into as few clusters as necessary which comprise as many items as possible that are maximally alike. So, for one, the economical principles featured in the three strategies are different.

Another difference lies in the segmentation of the set of items to be compared. First of all, the strategies differ in the amount of information processed with each fixation. The number of items covered per fixation is one for the travelling salesman strategy. In contrast, both the searchlight strategy and the clustering strategy maintain that with each fixation a variable number of items is processed. According to the searchlight strategy, that number is determined by the item 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 items in terms of location, color, and form.

In addition, there are qualitative differences. In order to keep the length of the scanpath to a minimum, it would suffice to roughly analyze the spatial layout of the items to determine their location. Thus, information about the spatial location of the items is necessary and sufficient for optimization of the scanpath in accordance with the travelling salesman strategy. Employing the searchlight strategy, however, is somewhat more exacting. The searchlight strategy presupposes that subregions have been coarsely analyzed as to the density of items, in order to fixate subregions with many items by preference and to avoid fixating empty ones. Thus, the information necessary and sufficient for the searchlight strategy is spatial location plus local density. An even more comprehensive analysis is necessary to license the clustering strategy. A cluster can be defined as a group of items that is similar with respect to dimensions such as location, color, or form. It follows that various feature dimensions of the items must be analyzed in conjunction in order to achieve perceptual grouping. The clustering strategy must take into account not only where the items are located and how far they are apart (density) but also whether or not neighboring items are alike in color or form (entropy). This raises the question to what extent we have to consider also the costs for performing the cost minimization itself.

From the above considerations it follows that the travelling salesman strategy could be characterized by the occurrence of approximately as many fixations as there are items in each hemifield: On average, a mismatch can be detected after scanning 50% of the item 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 item can be processed with each fixation. A summary is given in the first row of Table 3.1.

Predictions can also be made for fixation duration. In the case of the travelling 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 item density. In the case of the clustering strategy, item density should also take an effect; in addition, however, variance in item features like color or form (entropy) will be taken into account. Since it only makes sense to talk of variance when a fixation covers more than one item, we expect an interaction effect of item density and entropy. The second row of Table 3.1 provides a summary of these expectations.

Also, saccades should be relatively long when assuming a travelling salesman strategy as compared to either a searchlight or a clustering strategy. More specifically, saccade length should be a function of the spatial layout of the display, that is, of the local density of the items, for any of the three candidate strategies. Be-

Table 3.1: Hypotheses about expected values and factors for the candidate strategies (by dependent variables)

Variable	"Travelling Salesman"	"Searchlight"	"Clustering"
Number of fixations	$= n$ (items 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)

sides, for the clustering strategy, saccade length could be affected by interactions: as a cue to clustering, similarity in location (item density) may be diminished by variance in item features (entropy). These expectations are summarized in the third row of Table 3.1.

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 3.1.

How should the above hypotheses be tested in this first study? Technically speaking, the design of the study was a factorial with repeated measures on all independent variables. Independent variables were type of mismatch (color versus form) and three local stimulus parameters – as defined in Section 2.4.1 – for each fixation point registered: Item density $\varrho(\mathbf{p})$, color entropy $S_c(\mathbf{p})$, and form entropy $S_f(\mathbf{p})$ at the point $\mathbf{p}$ in question. An important feature of these local parameters is that they are completely uncorrelated, that is, the local item density in a particular area does not influence the probability of finding high or low color entropy or high or low form entropy respectively, at the same location.

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 for the specific purpose of this study; they included measures that might prove essential for the understanding of comparative search: Number of successive fixations within the same hemifield (FW), probability of missing the target (PM), additional search time (AT), area coverage per fixation (AC), and speed of processing (SP). For details concerning the calculation of dependent variables see Section 2.4.2.

3.2 Method

3.2.1 Subjects

The subjects ($N = 16$) were students of various fields at the University of Bielefeld. Each subject was paid 7 DM for the participation. All subjects had normal or corrected-to-normal vision, none had pupil anomalies and all were able to distinguish between colors.

3.2.2 Materials

The stimulus pictures were presented on a computer screen with a spatial resolution of 640×480 pixels. The pictures 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), as explained in Section 2.2. The size of the items was about 0.7 degrees of visual angle in diameter. The item locations were randomly generated, but avoiding item contiguity as well as item overlap. The random distribution was chosen such that it slightly tended to create regions of similar items in order to make it possible to study eye movements in homogeneous versus heterogeneous item distributions (cf. Section 2.3).

Each stimulus picture consisted of two hemifields with 30 items each. The items in each hemifield were equally balanced for form and color. The hemifields were translationally identical in the form, the color, and the spatial distribution 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.

3.2.3 Apparatus

Eye movements during comparative visual search were measured with the OMNITRACK 1 system (see Section 2.1.1). Subjects were seated about 60 cm away from a 17” color monitor. Prior to experimentation, a calibration procedure was to be performed by making the subject fixate specified points on the screen.

3.2.4 Procedure

Subjects were tested individually. Their task was to find the only difference between the two hemifields of each picture. Subjects were to press a mouse key placed in front of them as soon as they had detected the mismatch. Each subject viewed 50 pictures. After every tenth picture, the eye tracker was recalibrated in order to compensate for the possible sliding of the head-set due to head movements by the subjects. The stimulus pictures were newly generated for every subject so that none of the patterns occurred twice. 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.

3.3 Results and Discussion

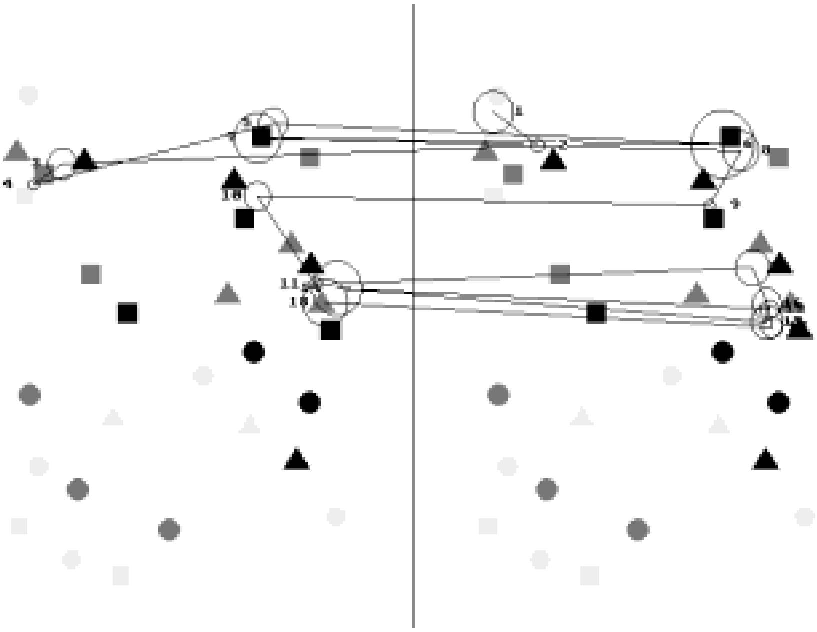


Figure 3.1: Example picture of Figure 2.11 with the plotted visual scanpath chosen by one of the subjects. Fixations are numbered; circle size signifies fixation duration.

The data recorded during the search and comparison phase were subjected to analyses of variance. The repeated measures analyses were performed using conservative adjustment of degrees of freedom. Level cutpoints for density and entropy were set at the lower and upper third (low versus medium versus high). The density range was split up 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”; they were thus entered into a one-factorial analysis of variance each. Each of the variables FD, SL, FW, and AC was entered into a four-factorial analysis of variance (type of mismatch, item 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 the subjects did not know the type of difference before detecting it. Therefore, data were collapsed over "type of mismatch". It was thus possible to reduce 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 one-factorial 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 is explained below.

3.3.1 Basic Dependent Variables

Reaction Time

On average, subjects needed 10 950 ms to process a picture. A histogram of reaction times in the experiment is shown in Figure 3.2; 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 effects of local parameters. Also, the distinction between search and comparison on the one hand and detection and verification on the other does not make sense here because the latter phase is defined on the basis of overall reaction time.

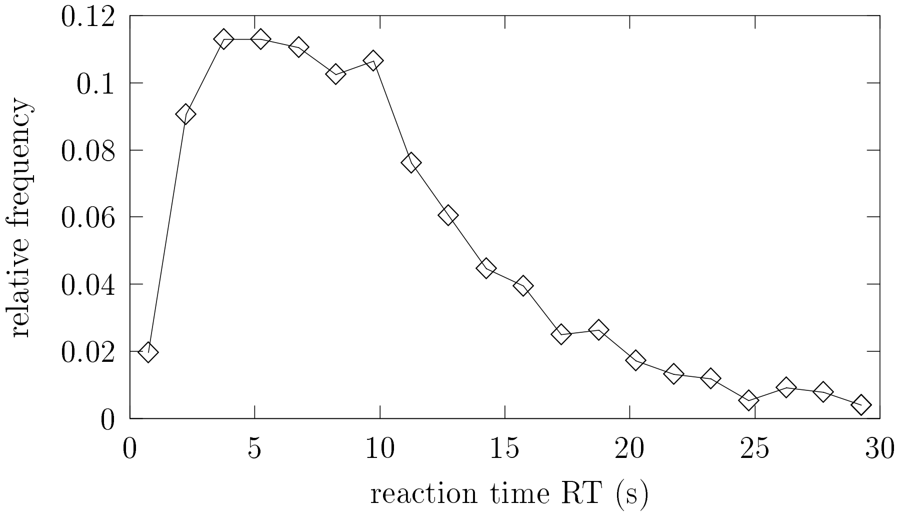


Figure 3.2: Histogram of reaction times

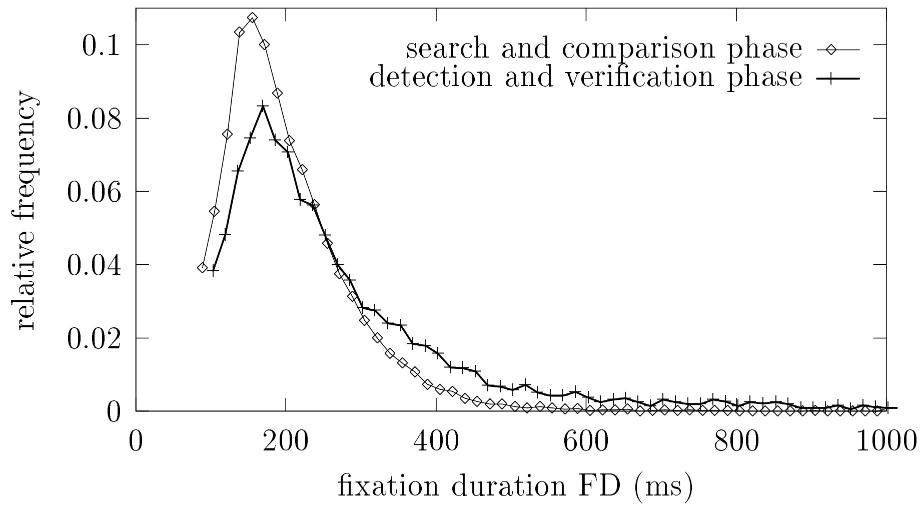


Figure 3.3: Histogram of fixation durations

Number of Fixations

Averaged over all subjects and all types of difference, subjects needed 39.6 fixations to finish a trial. Processing of a picture that had the difference in color generally took fewer fixations (35.7) than processing of a picture with 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 up 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 4.5 fixations, regardless of the type of difference ($F(1; 15) = 0.07; p = 0.788$).

Fixation Duration

In comparative visual search, 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. 3.3 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"). Overall, one in 113 fixations was a long one; the proportion of long fixations, however, was about 15 times higher in the detection and

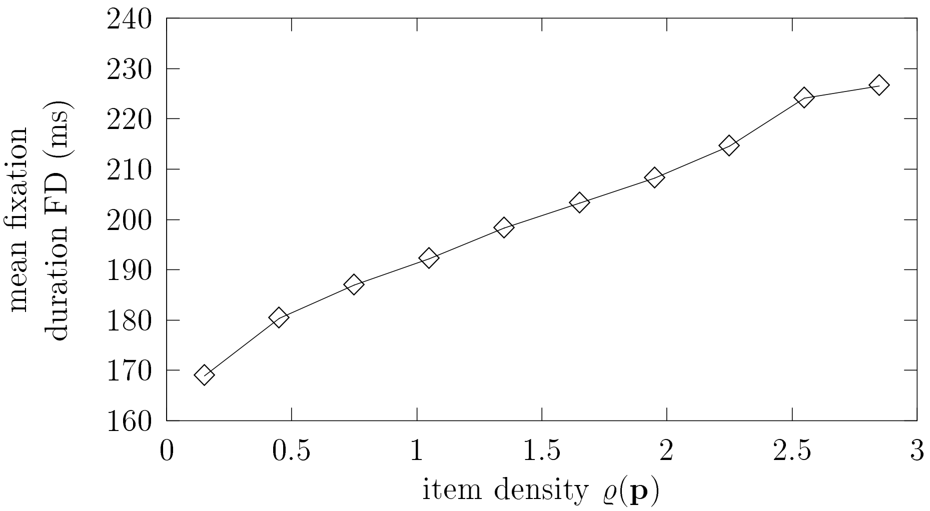


Figure 3.4: Mean fixation duration as a function of local item density at the fixation point

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 item density ($F(2;30) = 8.77; p = 0.001$); no other effects were significant. FD is plotted against $\varrho(\mathbf{p})$ in Figure 3.4. The subjects' fixations in high density regions (213.6 ms) were larger 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) ($F(1;15) = 9.88; p = 0.007$).

Saccade Length

The histogram of saccade lengths (SL), given in Figure 3.5, suggests to distinguish between two types of saccades: saccades linking fixations within the same hemifield (about 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 items 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

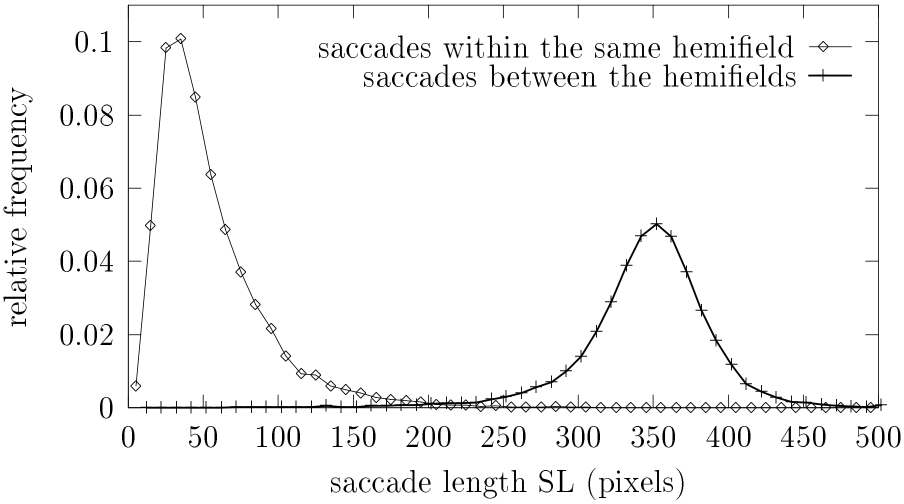


Figure 3.5: Histogram of saccade lengths

information, we decided to perform two 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 item density at the starting points ($F(2;30) = 35.79; p < 0.001$). 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 of form entropy ($F(2;30) = 6.40; p = 0.005$). Saccades starting in regions of low form entropy (59.3 pixels) and medium form entropy (57.6 pixels) were longer than those starting 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 item density on SL ($F(2;30) = 72.31; p < 0.001$): Saccades ending in medium density regions were longer (54.9 pixels) than those ending in high density regions (46.4 pixels) ($F(1;15) = 58.59; p < 0.001$). Even longer saccades (72.3 pixels) were those ending in low density regions ($F(1;15) = 52.91; p < 0.001$). However, analysis (b) did not reveal any significant effects of color or form entropy. Figure 3.6 illustrates SL as a function of local item density, whereas Figure 3.7 shows SL as a function of local color and form entropy respectively.

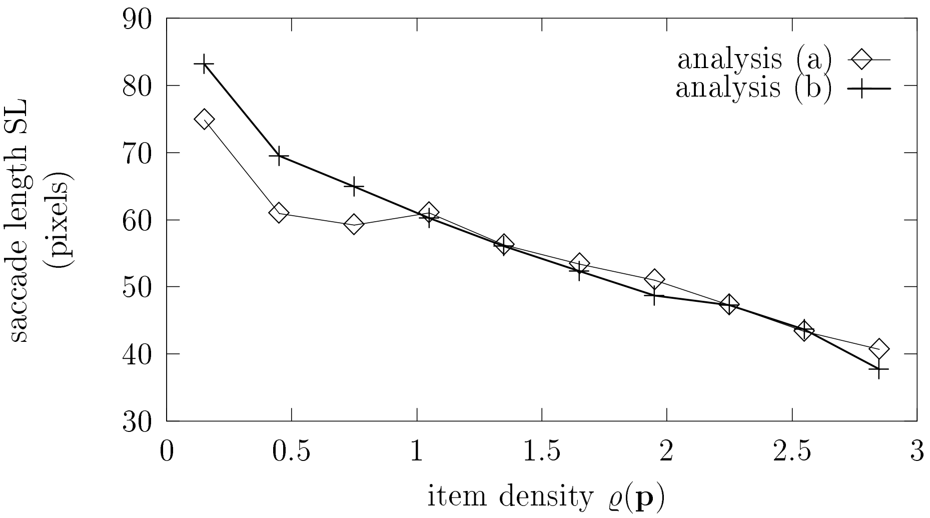


Figure 3.6: Mean saccade length as a function of local item density at its starting point (analysis a) and at its target point (analysis b)

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 difference to be detected made a difference in reaction time and the number of fixations. 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 items in each hemifield. This observation can be interpreted that people occasionally failed to detect a mismatch at first sight, so that parts of the display had to be scanned twice.

As to local parameters, both fixation duration and saccade length were affected by item density. The observation that fixations took longer in high density regions is not compatible with a travelling 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 items; representing these should take longer. It is in line with the clustering strategy as well: In high density regions more items can be memorized or compared per fixation: Accordingly, such fixations should be longer.

The fact that saccades were shorter in high density regions agrees with the predictions of all three strategies since the distance to the next suitable fixation point – be it the next item on the “salesman’s” path, in a neighboring area of fixed size, or in the neighboring cluster – should be rather short. The observed

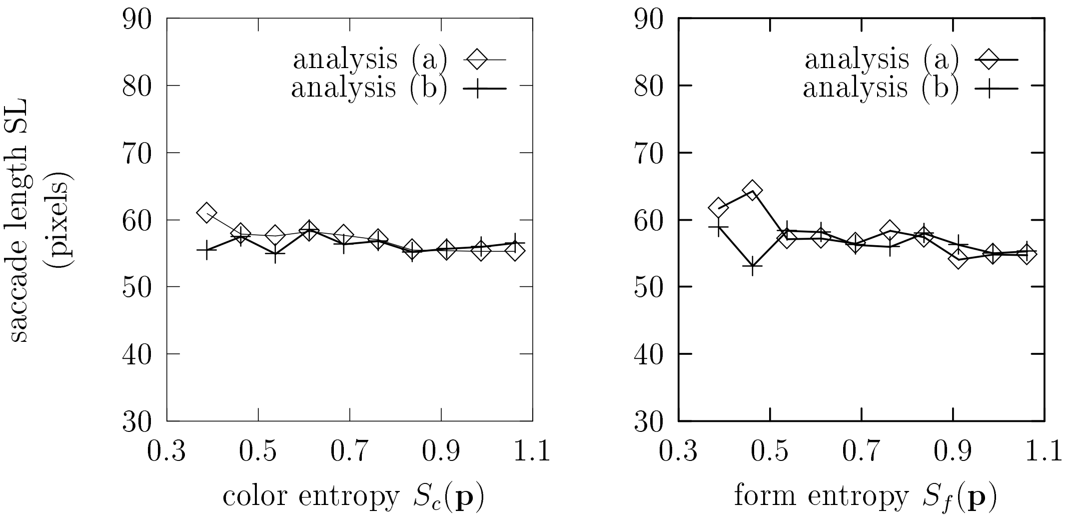


Figure 3.7: Mean saccade length as a function of color entropy (left) or form entropy (right) at its starting point (analysis a) and at its target point (analysis b)

effect of form and – as a tendency – color entropy at the starting point of a saccade indicates that the number of items processed at a time also depends on their content of information. Conceivably, processing the items in uniform areas is hardly demanding so that subjects tend to focus on areas of higher entropy within the same hemifield of the display.

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

3.3.2 Derived Dependent Variables

Successive Fixations within Hemifields

Eye-movement measurement does not only allow to distinguish search and comparison from detection and verification, but it also enables researchers to reconstruct the scanpaths that lead to the detection of mismatches in comparative search. 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 currently attended items. Consecutive saccades within the same hemifield link those fixations which pertain to the items to be compared in that particular step. Thus, the number of successive fixations within the same hemifield (FW) is of particular relevance for a detailed

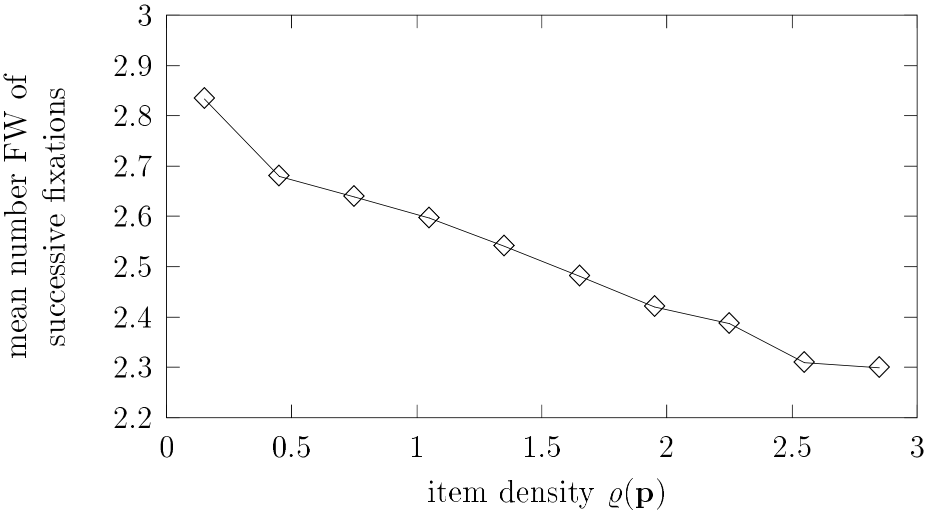


Figure 3.8: Mean number of successive fixations within the same hemifield as a function of the local item density at the first fixation point

investigation of the strategies that people follow in comparative search.

On average, subjects needed 2.45 fixations before shifting over to the other hemifield. The analysis of variance showed that subjects switched between hemifields more readily 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 item density ($F(2;30) = 13.99; p < 0.001$): When items were far away from each other, 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 item density on FW is plotted in Figure 3.8.

Also, FW was affected by form entropy ($F(2;30) = 3.46; p = 0.045$): The more the items in the vicinity of the fixation point varied in form, the more fixations people needed 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. Figure 3.9 illustrates FW as a function of the local entropy values.

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 items or their vicinity. This was the case

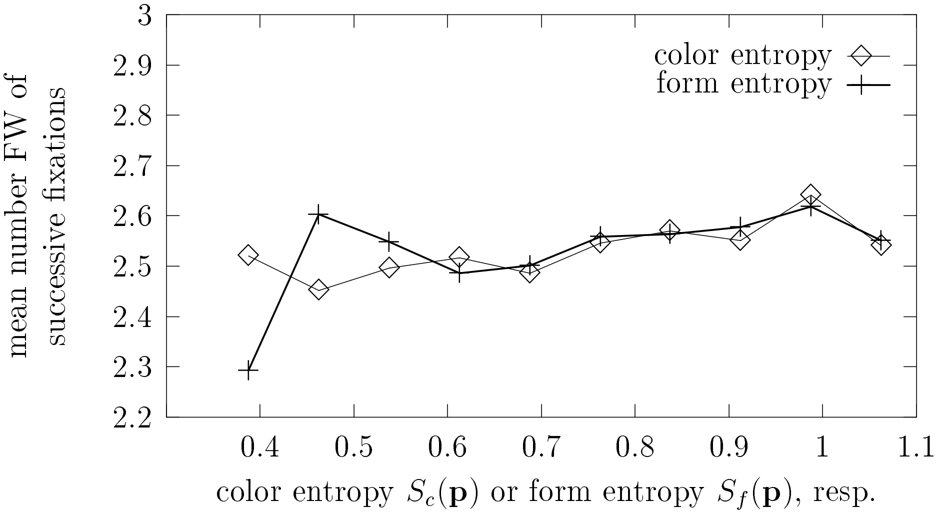


Figure 3.9: Mean number of successive fixations within the same hemifield as a function of local color or form entropy at the first fixation point

in 12.9% of all trials. Complementarily to this probability of missing the target (PM), the probability of detecting the mismatch was about 87.1% once each of the target regions was fixated (maximum distance between gaze position and target: 50 pixels) with only one saccade from one hemifield to the other between fixating the first and the second one.

Which factors influence PM? We checked four factors: Local item density, local color entropy, local form entropy, and type of mismatch (color or form). Each local factor was divided into two levels ("low" versus "high"), the cutpoint set at its arithmetic mean value at the center of each target item. 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 entropy cutpoint was set at 0.85, the density cutpoint at 1.5.

The four-factorial 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 item density, however, had a significant effect on PM ($F(1;15) = 9.52; p = 0.008$). High item density increased the rate of missing the target (15.3%) compared to low item density (10.5%). Furthermore, the analysis revealed an effect of local form entropy: Targets were more frequently missed (14.6%) at high form entropy in the target area than at 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 items led to a higher probability of missing the mismatch, regardless of whether it was in color or in form.

Additional Search Time

With regard to color targets, the additional search time (AT) reached a value of 1248 ms, while it was 2855 ms for form targets. This effect reached significance ($F(1; 15) = 9.26; p = 0.008$). Accordingly, subjects "wasted" less time by missing color mismatches than by missing form mismatches, which is compatible with the results of PM.

Area Coverage per Fixation

This derived variable provides a measure of the average area covered by one fixation (AC). It is defined as a circular area with a diameter calculated on the basis of the average distance to four neighboring fixation points. 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 of all "local" factors: item 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 to item density, AC was large (3516 pixels) in low density areas, but smaller (2751 pixels) in medium density areas ($F(1; 15) = 44.99; p < 0.001$). An even smaller AC was observed 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 $S_c(\mathbf{p})$ 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 at medium form entropy (2870 pixels) compared to high form entropy ($F(1; 15) = 8.51; p = 0.011$). As illustrated in Figures 3.10 and 3.11, which show AC as a function of density and of entropy respectively, the area covered per fixation decreased with growing local complexity of the display.

Speed of Processing

Speed of processing (SP) is a global measure – as RT and AT – that cannot be related to local stimulus features. Therefore, we can only obtain two SP values, namely one for color mismatches (45.59 pixels/s) and another for form mismatches (45.30 pixels/s). The difference between these two values does not reach significance. This is plausible since SP is defined in such a way that it is independent of PM (see Section 2.4.2).

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: Item density and entropy played a

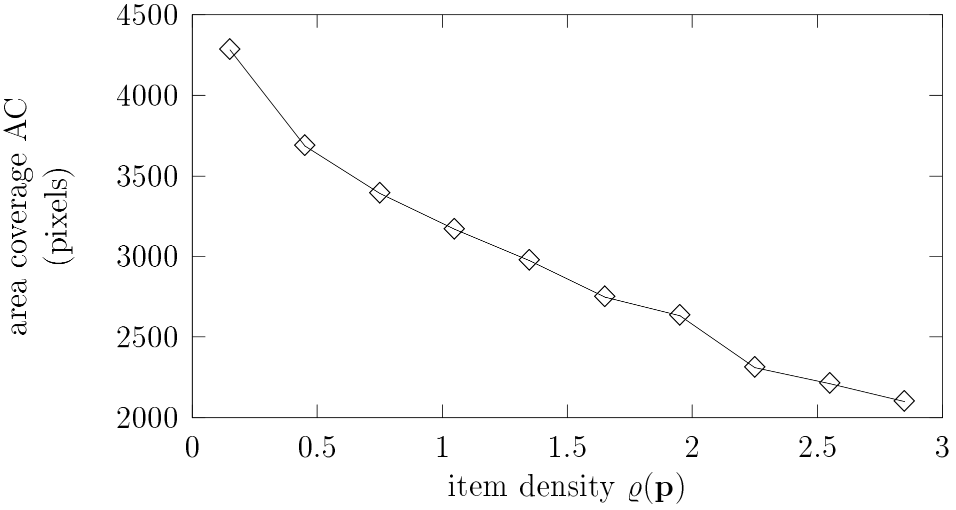


Figure 3.10: Mean covered area per fixation as a function of the local item density at the fixation point

role for the number of consecutive fixations within one hemifield, for the average area covered per fixation, as well as for the probability of missing the target. Second, people's actual gaze trajectories can be reconstructed more precisely if taking 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. Moreover, 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 people, in determining their eye movements in comparative search, 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.

The varying probabilities of missing the target again 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. These findings might give rise to a more thorough investigation of the strategies in comparative search for color in contrast to comparative search

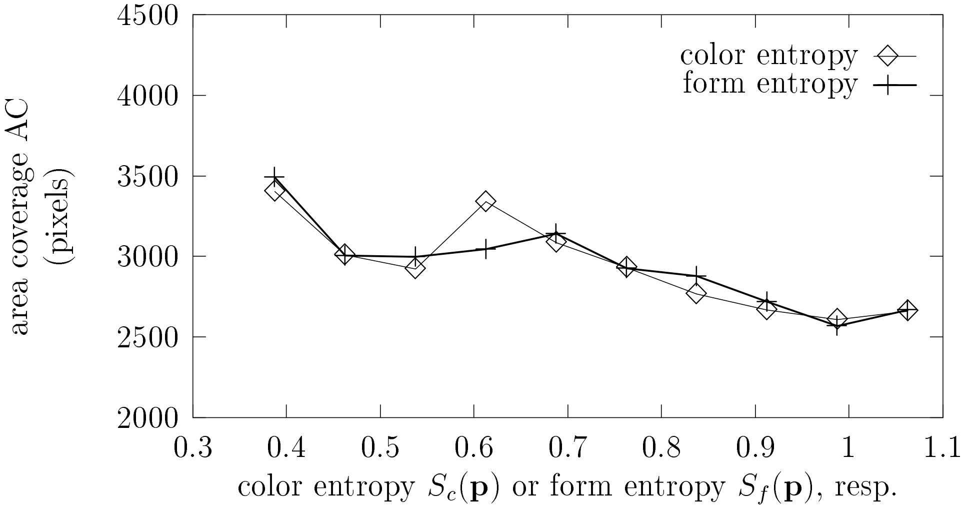


Figure 3.11: Mean covered area per fixation as a function of the local entropy values at the fixation point

for form.

3.4 General Discussion

The experiment reported in this chapter raises two questions:

- Which new insights about the process of visual search does eye-movement measurement yield?
- What does the comparative visual search paradigm tell us about the cognitive system?

If the experiment had restricted itself to the standard method of analyzing subjects' reaction times we would have found that detecting a difference in color is about two seconds faster than detecting a difference in form. Without analyzing subjects' eye movements, however, we hardly would have been able to figure out the reason for this. Eye-movement measurement has rendered it possible to split up 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, is increased by a factor of 15. This suggests that during the second phase, people concentrate on identifying and processing the mismatching items rather than quickly scanning the hemifields in search of a difference as they do during the first phase. Moreover, average saccade lengths are shorter in the second phase, indicating that the subjects' attention is now focused on a single "suspicious" item 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.

As for reaction time, we have found 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 there is 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. This assumption is supported by the results of the pre-test with regard to standard visual search (see Section 2.2.2): Subjects exhibited significantly longer reaction time for the detection of form targets than for the detection of color targets.

To sum up, it may be said that the analysis of eye-movement behavior with respect to the type of mismatch (color versus form) has yielded considerable insight into the cognitive process of comparative visual search. We have found that differences in form take more processing time than differences in color, and we can attribute this mainly to the fact that subjects fail to detect form mismatches more likely than they fail to detect color mismatches. We are, however, interested in a more detailed investigation of comparative search, focusing on how viewers' eye movements are determined by local display characteristics, namely by *local item density*, *local color entropy*, and *local form entropy*.

The average length of saccades has proven to be inversely proportional to local item density. This is intuitively plausible because people tend to direct their saccades at items; in low density subregions, items are located relatively far from each other which makes people produce relatively long saccades. The fact that the length of a saccade is codetermined by the degree of form entropy and – as a tendency – color entropy in its starting point, is also in line with the idea that processing economy is a highly important principle in comparative search. With respect to its end point, the length of a saccade depends on local item 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; since the end point is usually located in a subject's parafoveal region, only the local density can coarsely be estimated in advance but the entropy cannot.

These findings extend the conclusions reached by other researchers who have presented evidence that saccade length changes along 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 item 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 item density, shows saccade length to be primarily determined by the spatial distribution of the items, regardless of any other features. This pattern of results suggests that it might be useful to distinguish between parameters related to spatial location ("where?") and parameters related to item features ("what?").

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

Again, this observation goes beyond the results of previous research. Although the relationship between fixation duration and the spatial layout of the items to be processed has been studied before, even to the extent that global density measures, such as average distance, predict fixation duration less exactly than does minimum item 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. Conceivably, the time spent on a fixation varies along with the number of items covered by it. This implies, however, that the specific features of the items 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 items one by one but in chunks (the size of which probably varies). On the other hand, in determining relevant item configurations, subjects appear to merely localize the items – regardless of their color or their form – by chunks, gathering only coarse information required for memorization or comparison respectively.

Thus, the basic dependent variables, saccade length and fixation duration, mainly yield information about low-level processes involved in comparative visual search, in particular about the perception of item 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 item density increases. This observation can be attributed to economical principles in the usage of capacity-limited working memory. People appear to be able to process more items 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 item *density*, a local parameter which merely reflects the spatial location of the items to be scanned.

In addition, however, eye movements in comparative search are co-determined by item *entropy*, a local parameter which presupposes that the identity of the items has been established, that is, that their features have been identified and the appropriate color and form values have been computed. As to the number of successive fixations within the same hemifield, form entropy, but not color entropy, has an effect. This pattern of results does not only suggest that people, when gathering information from one hemifield, analyze the items thoroughly with regard to particular features, but also that the processing of color and form follows different principles – a point already made in the discussion of the different number of "misses" of color and form mismatches.

The influence of local entropy on search and comparison becomes more transparent when considering the area covered per fixation. Not unlike the notion of "grain size" – operationally defined via the minimum distance between neighboring items (cf. Jacobs, 1991) –, this measure gives an impression 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 item density, color and form entropy. While the strong dependence of area coverage on item 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 the establishment of the identity of the items. The more complex a particular region of the display is with respect to color and form of its items, the more fixations are required to gather a sufficient amount of information for detecting a mismatch.

Altogether, the perceptual and cognitive processes involved in comparative visual search can be viewed in terms of the working of two hypothesized subsystems in visual perception (cf. Velichkovsky, 1982; Bridgeman, van der Heijden & Velichkovsky, 1994): A fast system that serves the purpose of localizing items ("where?") and another, somewhat slower system for identifying items by their features ("what?"). The "where?" system might be taken to control search tactics; operating rather fast, it might guide short-term processes such as the maintenance of working memory and the chunking of items into clusters. In contrast, the "what?" system might be taken to control the overall search strategy; being somewhat slower than the "where?" system, it might guide mid- or long-term processes such as the planning of scanpaths and the extraction and memorization of specific item features for comparison.

The distinction between "where?" and "what?" systems in visual search corresponds to pertinent neurobehavioral evidence. Trevarthen (1968), for instance, argued that spatial vision and the perception of items are based on anatomically separate cortical mechanisms. One of these is ambient; it is applied to localize events over a wide field. The other is focal; it serves to obtain detailed vision with a restricted scope and is associated with the occurrence of small saccades. In a similar vein, Mishkin, Ungerleider & Macko (1983) distinguish two separate multisynaptic cortical pathways in primates: One specialized for "spatial" vision, another specialized for "object" vision.

The distinction between "where?" and "what?" systems also corresponds to current models of visual search which emphasize the role of perceptual grouping (e.g. Duncan & Humphreys, 1992; Humphreys & Müller, 1993). The claim that preattentive processing involves a reorganization of the set of stimulus items on the basis of global similarities which eventually leads to disjoint clusters is close in spirit to the above conception of the functioning of the "where?" system: The most obvious measure of global similarity is spatial distance, so that the resulting clusters are disjoint with respect to the localization of the items. Likewise, the claim that processing between clusters is sequential in nature and thus comparatively slow largely matches the way the "what?" system is supposed to work: Comparison of specific features proceeds by clusters; it requires that both the features of the items in that particular cluster and the clusters already processed be kept in mind.

The findings from the experiment have several implications. First, they pose a number of interesting research questions. How does the distinction between "where?" and "what?" relate to the *levels of processing* notion in Cognitive Science? How exactly do the subsystems interact in the course of visual search? Will the "what?" subsystem have to be subdivided into "what color?" and "what form?"? What about the effects of color or form entropy if subjects know in advance what kind of mismatch (color or form) they are to detect? How is oculomotor behavior affected by the detection of a mismatch? Proceeding from the basis experiment, questions like these will be addressed in the experiments reported subsequently in this work.

Second, the 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 pictured in the example scanpath in Fig. 3.1. The scanpaths frequently start in the upper part of the left hemifield. To begin with, the item set is subdivided into disjoint clusters on the basis of local item density. Then, with a few fixations connected by relatively short saccades, the first cluster is analyzed in more detail as to the color and the form of the items. The actual number of fixations conceivably depends on working memory capacity and on the number of items processed per fixation. That number, again, is determined by the local stimulus parameters, item 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 item 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 item density), and so forth. If, however, search and comparison yields a "suspicion", that phase will turn into a detection and verification phase which is characterized by few, but extraordinarily long fixations. It should be noted that people are more likely to miss the mismatch upon encountering the target cluster when the difference is in form than when it is in color.

The picture of comparative visual search that emerges from the experiment does not provide any support at all for a travelling salesman strategy which proceeds item by item: In contrast to the corresponding predictions, saccade length and fixation duration do depend on local stimulus parameters. To some extent, the results are rather compatible in part with the clustering and the searchlight strategy: With each fixation, a limited number of items is processed before proceeding to the next cluster. As for saccade length and fixation duration, the pattern of results corresponds to the predictions made on the basis of the searchlight strategy. These basic measures of oculomotor behavior are affected by the local parameter relating to the spatial configuration of the items 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 item density, color entropy, and form entropy suggest that economical aspects play a part in determining the search path. Within the limit set by the capacity of working memory, the number of items that can be processed with each fixation depends on item similarity in terms of location, color, and form. The main effect of local item density signifies that a cluster comprises the more items, the closer they are to each other. The main effects of color and form entropy signify that a cluster comprises the more items the more homogeneous they are in terms of color and form.

When discussing strategy issues, however, we must take into account that shifts of attention may occur relatively independently of eye movements. There might be local strategies of scanning that are invisible to the analysis of eye movements, at least with regard to the present scenario. This point is addressed in further investigations described in this work, especially in the context of scanpath modelling.

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 item density. Conceivably, the maximum distances between neighboring items have not been sufficient to reduce entropy effects in regions of low density in our experiments. According to the investigation of the influence of retinal eccentricity on feature recognition (Section 2.2.3), this finding is plausible: The features of single items can be recognized even at high eccentricities from the gaze position without causing significant cognitive "costs". 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 item density. Due to this flexibility, the entropy of item features plays an important part even if the items 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 modelled best. It will be taken up in Chapter 6 below.

As stated in the summary above, one fundamental aspect of comparative visual search was only touched so far, namely the question of differences between color and form search. Since subjects did not know the type of mismatch in advance, their eye-movement patterns were not influenced by the dimension of the mismatch. The only difference was found to be a higher probability of missing the target in the case of form targets, causing increased additional search time and, consequently, increased reaction time. There might be further differences between color and form search. The setting of Experiment A, however, does not allow us to investigate them.

In order to overcome this restriction, Experiments B and C were conducted. These experiments provided situations in which subjects were able to perform *specific search* for color or form targets respectively. In this context, the results of Experiment A serve as a baseline for comparisons between these experiments.

Chapter 4

Experiments B and C: How to Disregard Irrelevant Stimulus Dimensions

4.1 Introduction

The processing of different dimensions of visual items has been a central issue in perceptual studies at all times. One question within this field of research concerns the integrality of stimulus dimensions: Do stimulus dimensions such as color and form capture a viewer's attention automatically even if they are irrelevant with regard to a certain task? Or is the observer able to process particular stimulus dimensions and to disregard others because of his knowledge of what is relevant and what is irrelevant in a given situation? In other words, should control of attention be conceived of in a bottom-up fashion or in a top-down fashion?

This issue has been investigated within the scope of various paradigms. One of these paradigms is *visual search* (Pashler, 1988; Theeuwes, 1992; Bacon & Egeth, 1994; Friedman-Hill & Wolfe, 1995). The basic concepts of visual search have been outlined in Section 1.4. In a typical experiment investigating the question posed above (Bacon & Egeth, 1994), subjects had to search a display for a target (circle) among distractors (diamonds) and to state whether a line segment inside the target was horizontal or vertical. When all items were of the same color, this task was very easy to accomplish since the discrepant item appeared to "pop out" among the distractor items. In effect, reaction times were short and virtually independent of the number of distractors. However, when one of the distractors differed in color from both the target and the other distractors, reaction times increased. The additional discrepant item seemed to compete for attention with the actual target. The same results were obtained for a color target and form as the irrelevant dimension (Theeuwes, 1991). All in all, the observations from visual search lead to the conclusion that subjects are not able to disregard the irrelevant color or form dimension.

Another experimental approach is the *same-different paradigm* (Dixon & Just, 1978; Howard & Kerst, 1978; Santee & Egeth, 1980; Watanabe, 1988). As mentioned in Section 1.4, in a same-different task two stimuli are presented to the subject either simultaneously or in succession, and the subject has to decide whether they are the same or different with respect to certain relevant stimulus dimensions. In a classical experiment, Egeth (1966) used stimuli that were defined by three dimensions: Color (red or blue), form (square or circle), and tilt (the figure contained an ascending or descending line). A mismatch in a specified characteristic should lead to a "different" answer. An analysis of the "same" answers showed that reaction times were faster when the stimuli were identical with respect to the irrelevant dimensions than when they were different. If, for example, the subjects had to decide whether or not two stimuli had the same color, a "same" response was faster when their forms were identical as well. Findings like these indicate that, as in visual search, subjects cannot ignore irrelevant dimensions.

These two paradigms share the characteristic that the tasks are not too demanding. The stimulus sets are easy to survey and it takes only a few processing steps to generate an appropriate response. Furthermore, the dependent variables in these experiments are typically reaction time and error rate which reflect cognitive processes in visual search or same-different decisions only coarsely. The use of rather coarse dependent variables possibly goes along with relatively simple tasks: With such rough measures, it may be necessary to use simple tasks in order to arrive at significant conclusions about the cognitive processes involved in visual search and same-different decisions.

It seems possible to get a more precise answer to the initially posed question with the help of comparative visual search. As supported by the results of Experiment A, comparative search enables us to thoroughly investigate the dependence of eye movements on local stimulus features during the completion of a complex task. Since comparative search can be considered a combination of visual search and a same-different task (see Section 1.4), it seems to be a reasonable paradigm for the continuation of the studies mentioned above.

In order to address the question of disregarding irrelevant stimulus dimensions, we complement Experiment A by two further Experiments B and C (to be described in more detail in Sections 4.3 and 4.4). Both experiments employ the comparative visual search scenario of Experiment A. However, Experiment B differs from Experiment A in that now the subject is verbally informed about the type of mismatch (i.e. either "color" or "form") in advance. This would provide any top-down processes for the control of attention with the necessary information what stimulus dimension might be ignored. Therefore, the contribution of any such processes might become visible in a comparison of the results of Experiments A and B.

Experiment C differs from Experiment B in that now the irrelevant dimension is held constant. This means that all items in the display share the same form if there is a color mismatch. Analogously, all items have the same color if a form mismatch is present. This modification allows us to investigate any facilitating effects of bottom-up processes on the subjects' performance, which would be indicated by differences between the results of Experiments B and C.

Moreover, these experiments permit us to compare color and form as irrelevant dimensions. Are subjects equally effective in filtering out color and form information? Or are top-down processes sufficient to disregard one of the two dimensions while bottom-up processes help to disregard the other one? Any disparities could be interpreted as indicating differences in the cognitive processing of color and form.

After a brief characterization of the relevant independent and dependent variables of Experiments A to C, in Section 4.2 we recapitulate the main findings for Experiment A, viewed from the perspective of our present discussion. Then, Experiment B is described (Section 4.3), followed by the comparison of its results to those of Experiment A. In Section 4.4, we report Experiment C and compare the respective findings to those of Experiment B. The chapter closes with a general discussion.

What are promising variables to investigate? Since we indend to compare the results between the experiments, the first independent variable in our design is *experiment*. As stated above, an important question to be investigated is whether the control of attention exerts specific effects on color and form search. Thus, another independent variable is the *type of mismatch*, i.e. whether the difference between corresponding items is in color or in form. In order to investigate the effects of local color and form information on the subjects' eye movements, the variables *local color entropy* and *local form entropy* are considered as well. In Chapter 3, these variables have already proven their relevance for the investigation of mental processes during comparative visual search.

The dependent variables were reaction times and selected eye-movement variables. As mentioned above, *reaction time* (RT) is a global measure of the subjects' efficiency in a visual search or same-different task. RT should decrease if the search gets more efficient, i.e. if subjects are able to disregard the irrelevant stimulus dimension.

The analysis of eye movements is based on one "classical" and two derived variables: The "classical" variable of eye-movement research is *fixation duration* (FD). FD is known to indicate quantitative as well as categorial differences in mental effort (Velichkovsky, 1995; Velichkovsky, Sprenger & Pomplun, 1997). As suggested by the results of Experiment A, FD is likely to depend on the amount of information that has to be encoded at any step; thus, it should decrease if the subjects are able to disregard the irrelevant stimulus dimension.

Furthermore, Experiment A has shown that the time required to detect a mismatch does not only depend on the speed of search and comparison, but also on the *additional search time* (AT) caused by missing the target. AT can be viewed as a measure of the subjects' target detection capability. Therefore, AT should decrease if subjects ignore one stimulus dimension.

As a measure of the efficiency of memorization and comparison which is neither influenced by AT nor by the manual reaction, the *speed of processing* (SP) is taken into account as well. Presumably, SP should increase if the subjects restrict their attention to one instead of two stimulus dimensions.

Finally, the *area coverage per fixation* (AC) is analyzed, because AC is not only a *global*, but also a *local* measure of search efficiency. The global AC value should increase with the general search efficiency. If AC increases with the local color or form entropy value, this would suggest that subjects do not completely disregard the respective dimension.

To sum up: The independent variables within each experiment are *type of mismatch, local color entropy,* and *local form entropy.* For between-experiment analyses, the independent variable *experiment* is added. The dependent variables are RT, FD, AT, SP, and AC.

4.2 Experiment A: Unspecific Comparative Search

In Experiment A we assessed the effects of color and form as the irrelevant stimulus dimension on any of the dependent variables. Subjects had to analyze both dimensions because they were not informed about the type of mismatch in advance. Under these circumstances, neither top-down nor bottom-up filtering processes are possible.

The results of Experiment A provide a baseline for further comparisons between experiments. Experiment A itself does not yield unequivocal information about processing differences because it is impossible to decide if those differences should be attributed to differences in cognitive processes such as modes of control or to characteristics of the stimuli, in particular the similarities between chosen colors and forms. For instance, by using only slightly differing shades of grey instead of fully saturated blue, green, and yellow as item colors, the detection of color mismatches could be made much harder than the detection of form mismatches.

Although the results of Experiment A have already been reported in Chapter 3, the relevant subset of these results is reported again in order to facilitate the between-experiment comparisons in the present context. Moreover, due to the modified entropy levels, some of the analyses had to be adapted to this situation.

As to the global variables, RT was significantly shorter with respect to color mismatches (9903 ms) than with respect to form mismatches (11997 ms) ($t(15) = -2.94; p = 0.010$). The type of mismatch had no significant effect on SP (45.59 pixels/s for color mismatches and 45.30 pixels/s for form mismatches), on AC (2619.7 pixels versus 2619.8 pixels), and on FD (208.75 ms versus 207.46 ms). Therefore, these variables cannot account for the RT difference. AT indicates the amount of additional search time caused by missing the target. Here, a form mismatch "costs" 2855 ms as compared to 1248 ms for a color mismatch. This difference was significant ($t(15) = -3.04; p = 0.008$). This means that AT accounts for 1607 ms of the overall RT difference of 2094 ms. As discussed in Chapter 3, it is plausible to assume AT plus divergent latencies of the subjects' manual reaction to be responsible for the difference in RT.

Referring to the local entropy values as independent variables, only the dependent variables FD and AC can be investigated, because RT, AT, and SP are global measures. Differing from the analyses of variance reported in Chapter 3, only two levels of entropy ("low" versus "high") are distinguished with the cutpoint set at 0.75. This bisection leads to an improved comparability of entropy influences between the experiments. However, this might lead to deviations in the mean values of the respective dependent variables from the mean values presented in Chapter 3.

FD was neither significantly affected by color entropy (207.58 ms for low color entropy and 208.38 ms for high color entropy) nor by form entropy (208.10 ms for low form entropy and 207.82 ms for high form entropy). AC, however, showed a significant dependence on color entropy ($F(1;15) = 62.99; p < 0.001$) as well as on form entropy ($F(1;15) = 6.37; p = 0.023$). AC was larger in areas of low color entropy (2793.7 pixels) than in areas of high color entropy (2492.6 pixels); low form entropy induced larger AC (2728.9 pixels) than did high form entropy (2557.4 pixels). As a matter of fact, these differences did not depend on the type of mismatch.

These findings suggest the following coarse characterization of the differences between color and form search in Experiment A: When the subjects are not informed about the type of mismatch, detecting a form target is more difficult. More specifically, subjects are more likely to "miss" a form mismatch than a color mismatch. Missing the target forces the subject to continue searching. This extra search takes time (AT), causing an increase in RT. However, we are not yet able to decide if these differences between color and form search depend on differences in the underlying mental processes or on differences in the similarity between colors and forms chosen. To illucidate that issue, we now turn to two new Experiments B and C.

4.3 Experiment B: Top-Down Control of Attention

Experiment B was designed to investigate whether top-down control can enable subjects to filter out the irrelevant stimulus dimension (color or form). As a straightforward way to allow subjects to attend to the relevant stimulus dimension and to disregard the irrelevant one, we used a verbal instruction that informed the subjects in advance when to expect which type of mismatch. An efficiency gain would be indicated by a significant decrease in RT, FD, and AT and an increase in SP and AC from Experiment A to Experiment B. In technical terms, these expectations correspond to a significant main effect of the factor experiment.

Another objective of Experiment B was to study whether color and form can be ignored in an equally effective way. This could be assumed if the changes in the efficiency parameters turn out to be independent of the type of mismatch. If, however, the efficiency gain is more pronounced for color search than for form search or vice versa – indicated by a significant interaction between the factors experiment and type of mismatch – this would suggest that the feasibility of top-down control of attention depends on the choice of the relevant/irrelevant dimension.

The influence of local entropy values on AC is likely to reveal the extent of which subjects are able to disregard a specific dimension. In Experiment A, AC significantly depends on both the color and the form entropy. If, for example, form entropy has no signifcant effect on AC during color search in Experiment B, we can conclude that top-down control permits subjects – at least partially – to disregard form information. Thus, asymmetrical top-down influences on color and form processing can be detected this way. If subjects are able to ignore the irrelevant dimension, we expect only the relevant dimension to exert an entropy effect on AC.

4.3.1 Method

Subjects

A new group of twenty subjects with normal or corrected-to-normal vision was recruited at the University of Bielefeld and paid 7 DM for their participation. As in Experiment A, none of them was color-blind or had pupil anomalies.

Apparatus, Stimuli, and Procedure

Apparatus, stimuli, and procedure were identical to Experiment A with one exception: The trials were arranged in six blocks, each block consisted of ten trials with either only color or only form mismatches. Between blocks, the subjects were explicitly informed which type of mismatch to expect during the next block. The information was given in written form.

4.3.2 Results and Discussion

As in Experiment A, RT was significantly shorter with respect to color mismatches (7330 ms) than with respect to form mismatches (10541 ms) ($t(19) = -5.45; p < 0.001$). The type of mismatch had no significant effect on FD; FD was 208.04 ms during color search and 203.91 ms during form search. SP was faster for color mismatches (57.92 pixels/s) than for form mismatches (48.47 pixels/s) ($t(19) = 3.28; p = 0.004$). The AC difference between color search (3170.8 pixels) and form search (2873.7 pixels) missed significance ($t(19) = 1.97; p = 0.063$). Advance information about the type of mismatch seems to enhance color search but not form search. Again, AT was shorter for color mismatches (818 ms) than for form mismatches (2500 ms) ($t(19) = -4.14; p < 0.001$). In contrast to Experiment A, the AT difference of 1682 ms did not completely account for the RT difference of 3211 ms. The two values differed significantly from each other ($t(19) = -2.59; p = 0.018$). A plausible explanation is to assume that the extra time was caused by different speeds of processing (SP).

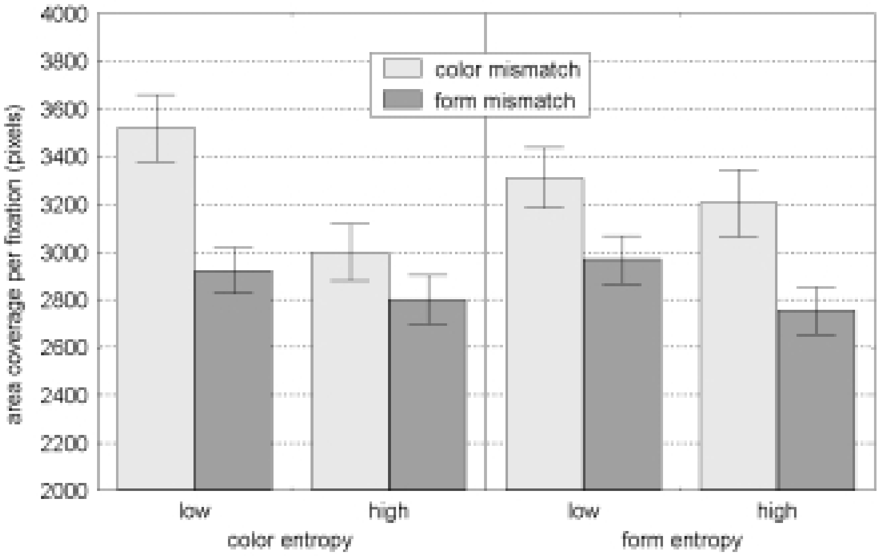


Figure 4.1: Mean area coverage as a function of local color and form entropy

More explicitly: Even if subjects know in advance which kind of mismatch to expect, form search seems to be harder to accomplish than color search, at least with these particular colors and forms. Search for a difference in form is slower than color search; also, more time is “wasted” by overlooking form mismatches in comparison to color mismatches. Thus, the increase in RT is due to longer AT as well as to higher SP.

While color and form entropy did not affect FD, they exerted specific effects on AC (see Figure 4.1). A three-factorial analysis of variance (type of mismatch, color entropy, form entropy) showed that, during color search, AC was significantly larger in areas of low color entropy (3520.6 pixels) than in areas of high color entropy (2999.3 pixels) ($F(1;19) = 19.03; p < 0.001$), whereas form entropy had no reliable effect (3312.2 pixels versus 3207.7 pixels). Correspondingly, during form search, AC was not significantly affected by color entropy (2922.9 pixels versus 2801.6 pixels), but presented a significant effect of form entropy ($F(1;19) = 10.29; p = 0.005$): AC was larger in areas of low form entropy (2969.5 pixels) than in areas of high form entropy (2755.1 pixels). There was a significant interaction between the factors type of mismatch and color entropy ($F(1;19) = 9.03; p = 0.007$), while the interaction between type of mismatch and form entropy remained below significance. These results suggest that subjects are able to disregard irrelevant color information as well as irrelevant form information. Disregarding color, however, seems to be more effective than disregarding form.

To analyze the data with respect to efficiency gains and the potential difference between color and form as irrelevant dimension, analyses of variance on RT, FD, AT, SP, and AC with the within-subjects factor type of mismatch (color versus form mismatch) and the between-subjects factor experiment (Experiment A versus Experiment B) were conducted. For the sake of completeness, both main effects and the interaction are reported for all dependent variables. Our focus, however, is on the differences *between* the experiments, because the main effects could be traced back to the particular colors and forms used in experimentation.

The type of mismatch had significant effects on RT ($F(1;34) = 33.55; p < 0.001$), on SP ($F(1;34) = 7.98; p = 0.008$), and on AT ($F(1;34) = 25.24; p < 0.001$). In detail, a color target was detected faster (8617 ms) than a form target (11269 ms), SP was faster for color mismatches (51.76 pixels/s) than for form mismatches (46.89 pixels/s), and AT was shorter when search was directed at a color mismatch (1033 ms) than when it was directed at a form mismatch (2678 ms).

As to the factor experiment: Average RTs in Experiment B (8936 ms) were faster than in Experiment A (10950 ms) ($F(1;34) = 5.26; p = 0.028$). No other variables showed an effect of the factor experiment, as shown in the comparative diagrams presented in Figures 4.2 to 4.6.

The only significant interaction effect was on SP ($F(1;34) = 7.07; p = 0.012$). Simple effects analyses indicated that the factor experiment had an effect only in color mismatch trials ($F(1;34) = 4.70; p = 0.037$), while type of mismatch had an effect only in trials in Experiment B ($F(1;34) = 16.93; p < 0.001$). In other words: SP was identical for color and form mismatches in Experiment A and for form mismatches in Experiment B. The only condition that differed from the others was search for color mismatches in Experiment B: SP is higher.

How are these results to be understood in the light of our assumptions? One efficiency measure, namely RT, is faster in Experiment A than in Experiment B. Obviously, subjects benefit from the information about the type of mismatch. They are able to disregard an irrelevant stimulus dimension – at least in part – irrespective of the type of mismatch. On the other hand, the interaction between type of mismatch and experiment indicates that color search benefits from information about the type of mismatch while form search does not. In other words: It seems easier to disregard irrelevant form information via top-down processes than irrelevant color information.

So far, we might conclude with some discretion that color and form search benefits from information about the type of mismatch and that this benefit is stronger for color search than for form search.

4.4 Experiment C: Bottom-Up Control of Attention

It must be pointed out, though, that “a-priori” information about the relevant dimension does not necessarily preclude the irrelevant dimension from being processed. Rather, the effects of a verbal instruction are restricted to those attentional mechanisms that are controlled by top-down processes. Other mechanisms may still make subjects attend to the irrelevant dimension.

Thus, Experiment C was designed to investigate the following question: Can a *constant* irrelevant stimulus dimension be disregarded more effectively than a *variable* one, i.e. do *bottom-up* processes facilitate comparative visual search? As a matter of fact, holding constant the irrelevant dimension necessarily goes along with informing subjects about the relevant dimension in advance. The comparison of results between Experiment B and Experiment C can therefore be assumed to indicate effects of “data-driven”, i.e. bottom-up control of attention. Once again, enhanced search efficiency would be signified by shorter RT, FD, and AT, and by higher SP and AC in Experiment C in comparison to Experiment B.

Another important point is to figure out whether color and form search are unequally influenced by bottom-up control of attention. This could be assumed if the potential changes in the efficiency measures depend on the type of mismatch. In technical terms, we would expect a significant interaction between the factors experiment and type of mismatch.

4.4.1 Method

Subjects

Sixteen new subjects were recruited at the University of Bielefeld. They had normal or corrected-to-normal vision and none of them was color-blind or had pupil anomalies. All subjects were paid 7 DM for their participation.

Apparatus, Stimuli, and Procedure

Apparatus, stimuli, and procedure were identical to those described in Experiment B. However, the subjects were not only informed about the relevant dimension, but additionally the irrelevant dimension was held constant, i.e. all forms were identical in color-mismatch trials and all colors were identical in form-mismatch trials. The frequencies of the three forms and the three colors that constituted the irrelevant dimension were balanced.

4.4.2 Results and Discussion

A significant difference in RT was found between color and form mismatches: Subjects needed 7422 ms to respond if color was the relevant dimension and 9279 ms if form was relevant ($t(15) = -4.28; p = 0.001$) (diagrams are shown in Figures 4.2 to 4.6). FD revealed no significant effect of type of mismatch (218.01 ms for color mismatches and 214.66 ms for form mismatches). SP was significantly higher for a difference in color (58.78 pixels/s) than for a difference in form (46.57 pixels/s) ($t(15) = 4.90; p < 0.001$). Color search induced larger AC (3671.3 pixels) than form search (2900.2 pixels) ($t(15) = 5.47; p < 0.001$). The type of mismatch affected AT as well: Subjects spent 598 ms due to missing color targets and 1272 ms due to missing form targets ($t(15) = -3.59; p = 0.003$). As in Experiment B, this AT difference of 674 ms does not sufficiently explain the overall RT difference

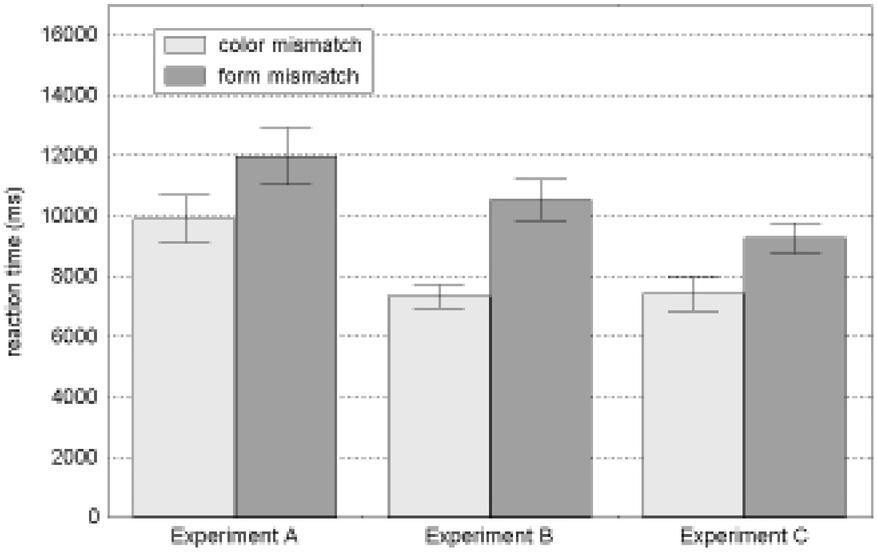


Figure 4.2: Mean reaction time for all three experiments and both types of mismatch

of 1857 ms ($t(15) = -2.73; p = 0.016$). Thus, the remaining difference could be attributed to the lower SP value for the detection of form mismatches.

Reasonably, the analysis of the independent variable color entropy was restricted to color search and the analysis of form entropy was restricted to form search. Neither of these variables significantly influenced FD, but AC was affected both by color entropy ($F(1; 15) = 11.74; p = 0.004$) and by form entropy ($F(1; 15) = 35.23; p < 0.001$). Low color entropy led to larger AC (3975.8 pixels) than did high color entropy (3483.0 pixels) and low form entropy induced larger AC (3026.5 pixels) than did high form entropy (2710.9 pixels).

All in all, holding one dimension constant does not seem to change the relation between color and form search, except for the AC difference which constitutes only a tendency in Experiment B and reaches significance in Experiment C. In Experiment C, search is still harder to accomplish for form mismatches than for color mismatches: Speed of processing and area coverage are lower and additional search time is longer, resulting in longer RT.

In analogy to the comparison between Experiment A and B, analyses of variance were calculated on RT, FD, AT, SP, and AC with the within-subjects factor type of mismatch (color versus form mismatch) and the between-subjects factor experiment (Experiment B versus Experiment C).

The type of mismatch exerted a significant influence on all dependent variables, namely on RT ($F(1; 34) = 43.82; p < 0.001$), on FD ($F(1; 34) = 4.35; p = 0.045$), on SP ($F(1; 34) = 30.52; p < 0.001$), on AC ($F(1; 34) = 25.78; p < 0.001$), and on AT ($F(1; 34) = 23.61; p < 0.001$). Subjects needed less time to detect a color target (7376 ms) than to detect a form target (9910 ms). FD was longer for

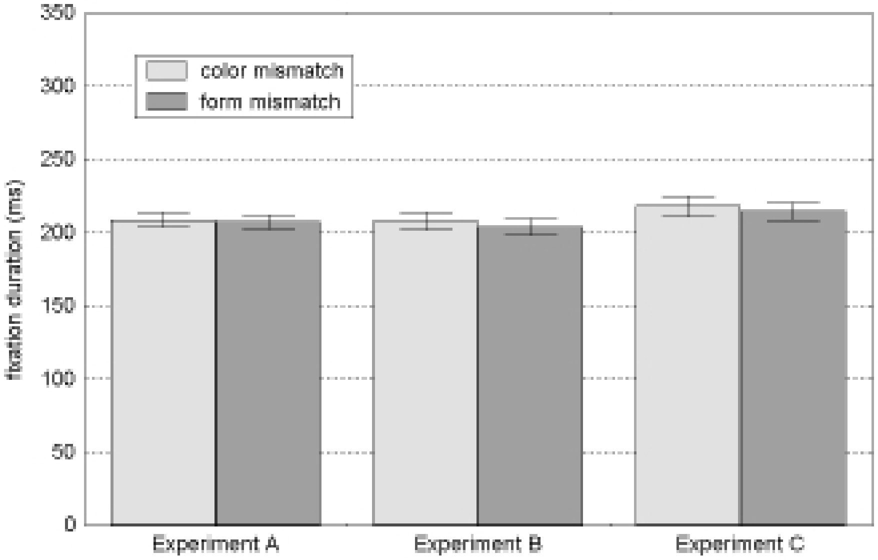


Figure 4.3: Mean fixation duration

color search (213.03 ms) than it was for form search (209.29 ms), SP was higher for color mismatches (58.35 pixels/s) than for form mismatches (47.52 pixels/s), AC was larger during color search (3393.2 pixels) than during form search (2885.1 pixels), and AT was shorter when subjects searched for a color mismatch (708 ms) than when they searched for a form mismatch (1886 ms).

As to the factor experiment, it exerted a significanct effect on AT ($F(1;34) = 5.92; p = 0.020$): AT was shorter in Experiment B (1659 ms) than in Experiment C (935 ms). The other efficiency measures were independent of the factor experiment.

The comparison of Experiments B and C revealed two significant interaction effects of the factors type of mismatch and experiment, namely an effect on AC ($F(1;34) = 5.08; p = 0.031$) and another on AT ($F(1;34) = 4.32; p = 0.045$). Simple effect analyses indicated only a tendency for the factor experiment to influence AC ($F(1;34) = 3.38; p = 0.075$). AC was significantly affected by the type of mismatch both in Experiment B ($F(1;34) = 4.49; p = 0.042$) and in Experiment C ($F(1;34) = 24.18; p < 0.001$). As to AT, the factor experiment only affected form mismatch trials ($F(1;34) = 6.44; p = 0.016$) and the type of mismatch only affected trials in Experiment B significantly ($F(1;34) = 27.07; p < 0.001$), while Experiment C presented only a tendency for an influence of the type of mismatch ($F(1;34) = 3.48; p = 0.071$).

Probably, the variation in Experiment C reconciles the AT values through the influence on form mismatch trials. The only condition that differs significantly from the others is search for form mismatches in Experiment B: In this condition, AT is higher.

What does the comparison between Experiments B and C tell us with regard

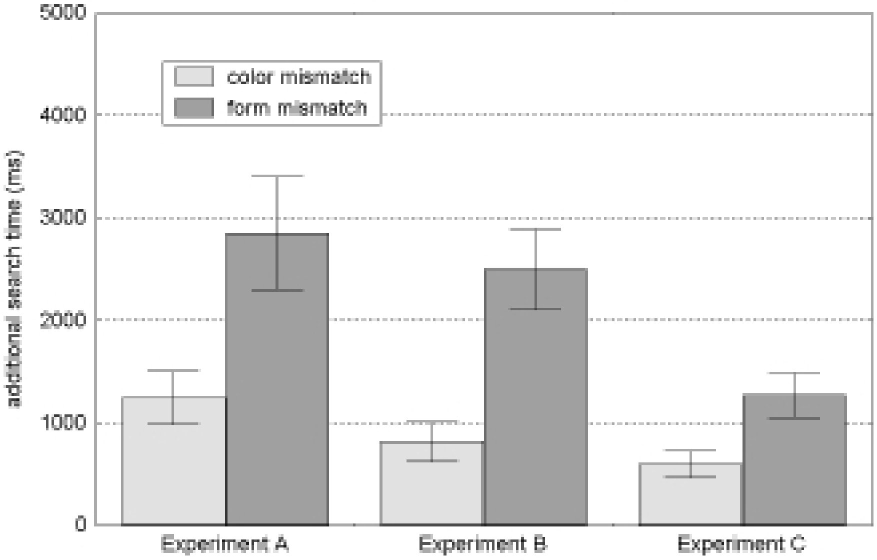


Figure 4.4: Mean additional search time caused by missing the target

to our assumptions? Again, only one of the efficiency parameters, this time AT, differs between the two experiments, but this main effect loses its relevance in the light of the significant interaction for AT between the factors type of mismatch and experiment. Keeping the irrelevant color dimension constant seems to improve the efficiency of form search, but keeping the irrelevant form dimension constant does not seem to improve color search any further than just a verbal instruction about the type of mismatch.

4.5 General Discussion

In view of our central question, we have to ask how the results are to be interpreted concerning the disregarding of irrelevant information under different conditions. As argued in Section 4.2, differences concerning color versus form search within the same experiment cannot be viewed as universally valid, because they strongly depend on the similarity of chosen colors and forms. Therefore, emphasis is put on the differences *between* the experiments. The interactions between the factors experiment and type of mismatch are especially important, because they suggest color-form asymmetries in the effects of bottom-up control of attention (comparison of Experiment A to Experiment B) and of top-down control of attention (comparison of Experiment B to Experiment C).

Before discussing the results yielded by the *global* measures, we should focus our attention on the analysis of *local* parameters, i.e. local color and form entropy. In contrast to the global variables, the effects of local entropy on the subjects' eye movements can be assumed to *directly* indicate to what extent a specific stimulus

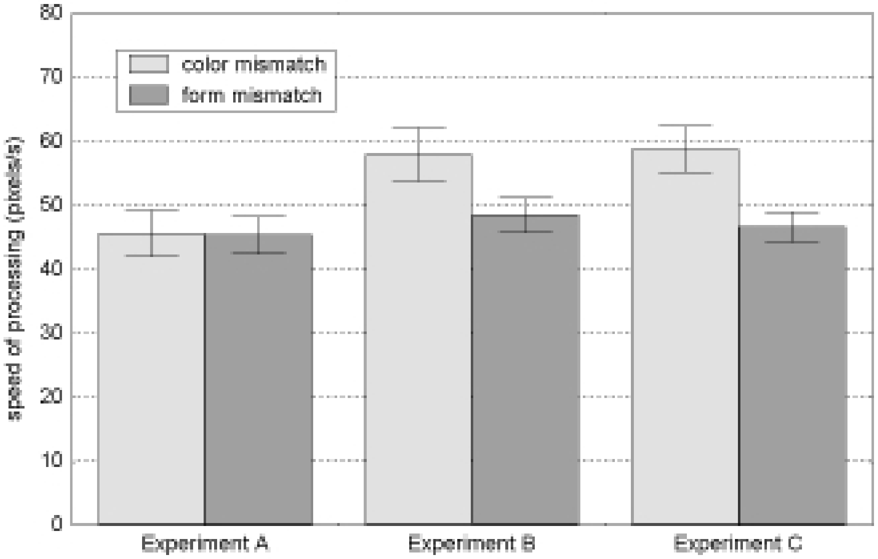


Figure 4.5: Mean speed of processing in y-direction

dimension was regarded or disregarded: If, for instance, the subjects' performance decreases in display areas of high color entropy (i.e. areas presenting much color information), this would suggest that the color information is not completely disregarded.

Fixation duration does neither depend on local color entropy nor on local form entropy in any of the experiments. Conceivably, fixation duration cannot be considered an appropriate measure of a subject's efficiency in comparative visual search. Area coverage, however, is significantly affected by both color entropy and form entropy in Experiment A. This result is plausible: Subjects do not know the dimension of the mismatch in advance, hence they have to pay attention to the colors as well as to the forms of the items. In Experiment B, the entropy effects are restricted to the relevant dimension: AC significantly depends on the entropy of the relevant dimension, but not on the entropy of the irrelevant dimension. High values of irrelevant entropy seem to decrease AC as well, but this effect does not reach significance. While the interaction of the factors type of mismatch and color entropy shows a reliable effect, the interaction of type of mismatch and form entropy remains below significance. These findings suggest that both color and form can, at least to some extent, be disregarded on the basis of bottom-up control of attention. Disregarding form information, however, seems to be more efficient than disregarding color information. The results of Experiment C are in line with the above interpretations; AC always depends on the relevant entropy.

Are the effects of local entropy reflected in the global measures of efficiency? The analysis of variance between Experiments A and B does not show an interaction between the type of mismatch and any of the variables RT, FD, AC, and AT. This finding indicates that the bottom-up control introduced in Experiment

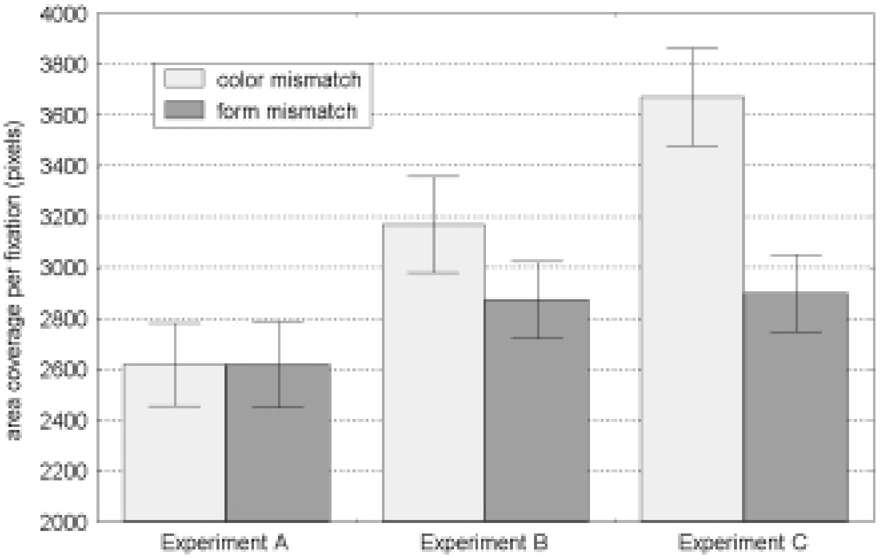


Figure 4.6: Mean area coverage per fixation

B does not affect these variables differently for color and form search. However, there *is* such an interaction for SP (see Figure 4.5). SP shows no significant shift in situations of form mismatches, but an acceleration at the detection of color mismatches.

Why is this SP effect not reflected in the overall RT? Logically, RT should decrease with increasing SP. When regarding the RT results in Figure 4.2, such a correspondence can actually be assumed: Color RT seems to decrease more strongly than does form RT between Experiments A and B. This interaction, however, does not attain significance. The reason for this missing significance is likely to be the high variance of RT, because RT is the sum of at least three different components: The time to reach the target, the additional time caused by eventually missing the target (AT), and the time needed for manual response. This “noise” handicaps the analysis of RT. Therefore, investigating eye movements in comparative visual search does not only enable us to recognize the factors that determine the overall RT, but also to increase the validity of the results.

The SP and AC data indicate, as expected, that the detection of color targets is facilitated by verbally specifying the relevant dimension. Surprisingly though, there is no analogous effect on the detection of form targets. Although it seems to be possible – to some extent – to disregard color information during form search (see above), this capability does not enhance the speed of processing. Hence, we might assume the maximum search speed to be higher for color targets than for form targets.

According to the results of SP and AC, simultaneous search for color and form in Experiment A is not significantly less efficient than specific search for form in Experiment B. Thus, it seems that form processing might imply color processing.

In the light of these considerations, the effects between Experiments B and C are especially insightful: How does the additional bottom-up help of eliminating the irrelevant stimulus dimension interact with the search for color and form targets? This time, the variables RT, FD, and SP do not signify any interactions between the factors experiment and type of mismatch, whereas the variable AC is significantly influenced: AC benefits from the bottom-up control only with regard to color search, supporting our view that color search is not completely independent of the form information given in the display.

AT, however, remains almost constant with regard to color search and clearly decreases for form search. Irrelevant color information during form search seems to be more "confusing" for the subjects, leading to longer additional search time after missing the target than does irrelevant form information during color search. The improving effect of bottom-up control on form search is restricted to target detection itself; search efficiency is not affected, as indicated by SP and AC. Again, no significant effect is established for RT, although the RT difference corresponds well to the divergence of AT between Experiment B and Experiment C.

All in all, the results enable us to outline a rough picture of the processing and disregarding of color and form during comparative search. As a matter of fact, there is no evidence supporting a dichotomy of "processed" versus "non-processed" (disregarded) information. The irrelevant dimension seems to influence the subjects' search performance, both during color and form search, but the influence of the relevant dimension is stronger in either case.

There are, however, basic differences between color and form search. The efficiency measures SP and AC indicate that form search is neither significantly facilitated by top-down nor by bottom-up control of attention. This finding supports the assumption raised above: Form processing may imply color processing. Handling form information seems to be more demanding and to be achieved at a certain maximum speed that cannot be increased by the control of attention. During form search, color information – as a "by-product" – may become accessible to the cognitive system as well.

During color search, however, it seems possible to avoid form processing to some extent. Search efficiency benefits both from top-down control (SP) and from bottom-up control (AC). The bottom-up effect on AC suggests that irrelevant form information is not completely disregarded. This finding underlines the conclusion drawn above: The capability of disregarding irrelevant information varies along a continuum rather than between the states "yes" and "no".

Chapter 5

Experiments D and E: Variations of the Basis Scenario

Motivated by the insights gained from the carefully assigned changes between Experiments A, B, and C about the factors that influence comparative visual search, we will now consider further variations and analyze changes in the eye-movement parameters in order to investigate the cognitive processes in comparative visual search in more detail. What kind of variations appear most promising in this respect?

Comparative visual search is practically an untouched field of research so far, hence countless combinations of novel stimuli and tasks are open to scientific investigation. It goes without saying that only a very small fraction of them can be covered in a single work like the one in hand. Therefore, it seems useful to take a look at substantial changes in the scenario in order to explore the range of phenomena. Before introducing new elements to the stimuli, we should consider those changes that do not affect the "old" two-dimensional, geometrical items with three colors and three forms.

Fundamental modifications of the scenario can be achieved by various kinds of its *inversion*. When thinking about inversion, one idea is likely to strike our minds immediately: So far, subjects have been searching for a mismatch among otherwise corresponding items. What about having them search for a *match* among otherwise *dissimilar* items? This idea has been taken as the basis of Experiment D, which is described in the following section. Another aspect of inversion refers to the geometry of the stimulus pictures: In the previous experiments, the right hemifield has been a *translated* copy of the left one (except for the target item). What will happen to the subjects' eye movements if the right hemifield were a *mirror image* of the left hemifield instead? Experiment E is an attempt to answer this question.

5.1 Experiment D: Searching for a Match

There are various possibilities of arranging stimuli for match detection. For instance, in the simplest situation, subjects could search for a combined color and form match among pairs of items that differ in either their color or form. However, we decided to use the *exact* inversion of Experiments A and B in order to make the experiments comparable to each other: The distractors differ in both dimensions, while the targets are of either the same color or form, as shown in Figure 5.1. Nevertheless, the question remains whether the subjects should be informed about the type of match in advance, i.e. whether the experiment should correspond to Experiment A or to Experiment B.

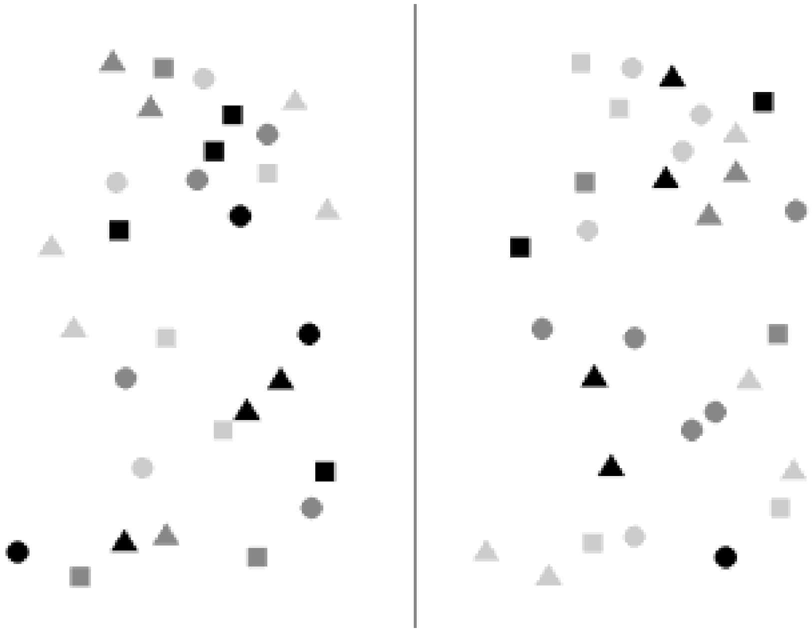


Figure 5.1: Example stimulus used in Experiment D. Subjects are to find a color or form match between items in corresponding locations.

Both alternatives were investigated in a pre-study. It revealed that keeping the subjects uninformed about the relevant dimension results in too difficult a task. The subjects' mean reaction time was more than twice as long as in Experiment A. Moreover, subjects soon got fatigued, and some of them completely lost their motivation to solve the task (solving the example task in Fig. 5.1 may help to understand this effect). Such phenomena have not been observed in any other experiment using comparative visual search. In order to ensure overall comparability to those, Experiment D was designed in such a way that subjects know the dimension of the match in advance. Consequently, Experiment D corresponds to Experiment B, hence it is useful to compare their results. Technically speaking, the independent variables are *type of match* (color versus form) and *experiment* (Experiment B versus Experiment D).

What effects can be expected to occur when subjects have to search for a match rather than a mismatch? As in Experiment B, only one of the two stimulus dimensions (color or form) is relevant for task completion, the other one is purely distracting. The comparison between Experiments A and B in Chapter 4 has shown that the subjects' knowledge about the relevant dimension increases their search efficiency only with respect to *color* search, suggesting that irrelevant form information can be disregarded more effectively than irrelevant color information. Basically, the situation in Experiment D is the same; only one stimulus dimension is relevant within each trial. If we assume that there are no fundamental differences between the detection of matches and mismatches, the pattern of results should resemble the one obtained in Experiment B. In particular, we expect an increased efficiency during color search which should be indicated by the variables *reaction time* (RT), *speed of processing* (SP), and *area coverage per fixation* (AC).

It might be possible, however, that the comparison of merely *different* color and form information between the hemifields is more demanding than the comparison of mainly *identical* information, because individual information provided in each hemifield is likely to induce higher memory load during its processing. Since we assume a constant capacity of working memory, the number of items (or the subarea of the display) memorized at a time might be smaller in Experiment D than in Experiment B. This could be indicated by lower values of the efficiency measures RT, SP, and AC. Additionally, *saccade length* (SL) is an appropriate measure: Shorter saccades could signify a more thorough search strategy caused by a smaller focus of attention.

Finally, interesting questions are whether matches are missed more frequently than mismatches and whether this frequency is affected by local entropy in the same way for matches and mismatches. It is important to investigate these questions, because the answers may provide information about the factors that determine the "visibility" of matches and mismatches. Therefore, we analyze the dependence of the *probability of missing the target* (PM) on the factors *experiment, type of mismatch, color entropy,* and *form entropy.*

5.1.1 Method

Subjects

The subjects ($N = 14$) were students of various fields at the University of Bielefeld. They were paid 7 DM for their participation. All of them had normal or corrected-to-normal vision; none had pupil anomalies, and all were able to distinguish between colors.

Materials

The parameters of the stimuli and their presentation were identical to Experiment B with one fundamental difference: This time, item "twins" neither had the same color nor the same form, except for the target items which corresponded in exactly *one* of these dimensions (see Figure 5.1 for an example). Again, the random distribution of colors and forms tended to generate areas of similar items (see Section 2.3).

Apparatus

Eye movements were recorded using the OMNITRACK1 system. The experimental set-up was the same as in the previous experiments.

Procedure

Subjects were tested individually. They were instructed to find the only color or form match between corresponding items shown in the display. They had to press a mouse key as soon as they detected the match. Each subject viewed 60 pictures which were individually generated. After every tenth picture, the eye tracker was recalibrated. Three of the six resulting blocks of stimuli contained a color match, the other three blocks contained a form match. Their order of presentation was randomized. Prior to each block, subjects were informed which kind of match to expect.

5.1.2 Results and Discussion

Since the results of Experiment D cannot be interpreted without their comparison to Experiment B, all results of Experiment D are immediately followed by their comparison to Experiment B in this section. This seems to be the most comprehensive way to report the findings.

Figure 5.2 illustrates the scanpath generated on the example stimulus by one of the subjects. Obviously, most of the fixations are located closely to items, demonstrating a thorough search strategy. All in all, the search process seems to proceed more slowly in Experiment D than in the corresponding Experiment B. This is corroborated by the results of the quantitative analyses. Saccade length (SL) is found to be significantly shorter than in Experiment B, regardless of the dimension of the match or mismatch respectively. Concerning color targets (color matches or mismatches respectively), SL decreases from 62.26 to 55.50 pixels ($t(32) = 2.28; p = 0.030$). With regard to form targets, it diminishes from 58.86 to 52.46 pixels ($t(32) = 2.53; p = 0.016$). Within Experiment D, SL is significantly higher during color search ($t(32) = 2.83; p = 0.014$) than during form search. This finding suggests that color search still induces higher efficiency of task completion than form search, unaffected by the task inversion between Experiments B and D.

Moreover, the variable search speed (SP) reveals a corresponding pattern of results. In Experiment D, SP is significantly slower than in Experiment B, which holds for color search (35.39 versus 57.92 pixels/s; $t(32) = -3.63; p < 0.001$) as well as for form search (29.52 versus 48.47 pixels/s; $t(32) = -5.11; p < 0.001$). The

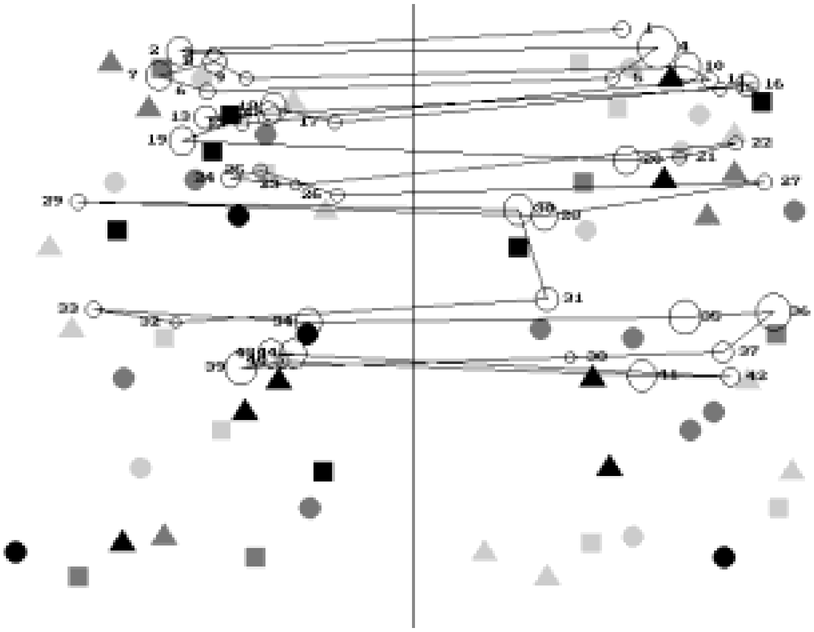


Figure 5.2: Example picture with a subject's gaze trajectory. Fixations are numbered; fixation duration is indicated by circle size.

SP difference between the detection of color and form matches in Experiment D shows a tendency towards higher speed during color search; it does not, however, reach significance.

The same effects are found with respect to the area coverage per fixation (AC). It is reliably smaller in Experiment D than in Experiment B, both during color search (2253 versus 3171 pixels; $t(32) = -3.65; p < 0.001$) and during form search (2055 versus 2874 pixels; $t(32) = -4.21; p < 0.001$). The mean AC in Experiment D is somewhat larger for color search than for form search, however, this difference closely misses significance.

Since the observations for the variables SL, SP, and AC suggest some tendency towards higher search efficiency of color detection as compared to form detection, we should expect reaction time (RT) to reveal an analogous effect. However, this is not true. Matching color targets require a mean RT of 13694 ms, while form targets require 13496 ms, which is almost the same result. In comparison to Experiment B, however, efficiency is reduced: Color search causes longer RT in Experiment D than in Experiment B (7330 ms) ($t(32) = 7.63; p < 0.001$), and the same holds true for form search (Experiment B: 10541 ms) ($t(32) = 3.12; p = 0.004$).

In the previous Chapter 4, we have argued that RT is not completely determined by the variables SL, SP, and AC. The probability of missing the target (PM) has a substantial impact on RT as well. A four-factorial analysis of variance (type of match, local item density, local color entropy, local form entropy) unveils an interesting effect: Color targets lead to significantly higher PM (22.8%) than form targets (17.8%) ($F(1;13) = 6.83; p = 0.021$), which explains the missing RT difference between color and form search. This finding is inverse to the corresponding result of Experiment B, where PM is *lower* for color mismatches (9.2%) than for form mismatches (19.4%) ($F(1;19) = 14.38; p = 0.001$). Moreover, in Experiment D, PM significantly depends on the local color entropy at the target items. It is smaller for low color entropy (12.1%) than for high color entropy (19.0%) ($F(1;13) = 16.21; p = 0.001$). Form entropy, on the other hand, does not exert a reliable effect on PM. Again, the situation is inverse in Experiment B, which exhibits a significant influence of local *form* entropy on PM. PM increases from low form entropy (8.9%) to high form entropy (12.7%) ($F(1;19) = 8.91$; $p = 0.008$), while it does not depend on color entropy.

The results obtained are not consistent with the assumption that match and mismatch detection involve the same kind of processes. As indicated by the variables SL, SP, AC, and RT, search efficiency is substantially higher during mismatch detection than during match detection. The reason for this discrepancy conceivably lies in the fact that subjects memorize and compare *groups* of items rather than single items. If subjects sequentially compared single pairs of corresponding items, search efficiency should not strongly depend on the task (match versus mismatch detection). However, subjects tend to memorize and compare about two to four items at the same time, as estimated in Chapter 3. In Experiment B, the items corresponding between the hemifields had the same *identity*, which means that they were of the same color as well as of the same form. Only the target items had differing identities. Hence, groups of items could be compared *holistically*; they matched perfectly between the hemifields unless they contained the target items. The situation in Experiment D is completely different: Neither the distractors nor the target items share the same identity between the hemifields. Since item groups always differ between the hemifields, their holistic comparison is impossible. Additionally, it is more difficult to identify the items which correspond to the actually memorized ones, because the only cue is their location, whereas in Experiment B their color and form constitute further signals.

Apart from the generally reduced efficiency in Experiment D, the increased value of PM with regard to color targets has to be explained. Conceivably, this effect is to some extent caused by the subjects' overestimation of their own capacities. As Experiments B and C have demonstrated, it is easier to perceive, memorize, and compare the color information of item groups than their form information. Since a holistic comparison of such groups is possible, PM is even smaller for color search than for form search. In Experiment D, the detection of color targets still *seems* to be easier, because the pure *perception* of color information is not impaired. Accordingly, subjects might proceed faster during color search than during form search, memorizing larger groups of items at the same time. This is supported by the variables SL, SP, and AC. The impracticability of holistic group comparisons, however, might eliminate the better detectability of color targets. Hence, subjects are likely to"pay" for their faster scanning during color search with higher PM. As a matter of fact, such an effect would increase with the complexity of color patterns, i.e. the local color entropy at the target items.

In the situation of mirror symmetry between the hemifields, holistic processing should be impaired as well. Thus, it is promising to compare the effects to be observed in Experiment E with those established in Experiment D in order to verify the considerations mentioned above.

5.2 Experiment E: Stimuli with Mirror Symmetry

The effects of translational or "repeated" symmetry versus mirror symmetry were examined in various contexts. For instance, Julesz (1969) reported that subjects were able to detect mirror symmetry in random-dot patterns faster than repetition. The study of Bruce & Morgan (1975) is more closely related to comparative visual search (see Section 1.5). The authors used left-right mirror symmetric or horizontally shifted line patterns as stimuli. Some of these stimuli contained symmetry violations which the subjects had to detect. Similar to Julesz' (1969) results, the subjects' search efficiency was significantly higher with mirror symmetry than with repetition. This advantage of mirror symmetry was found to diminish with an increasing horizontal eccentricity of the violation due to a longer distance between the two mismatching locations. Unfortunately, neither of the two studies mentioned above investigated the subjects' gaze trajectories.

In order to keep Experiment E as comparable as possible to Experiment A, the parameters of the stimuli used in Experiment A were not changed at all, with the sole exception of the right hemifield being a horizontally mirrored rather than a translated copy of the left hemifield. An example stimulus is shown in Figure 5.3. Additionally, we decided to use the same *task* as in Experiment A, hence subjects were not informed about the dimension of the mismatch they had to search for.

Since the difference between Experiments A and E is of a geometrical nature, it is especially interesting to compare the spatial structure of the respective scanpaths. As a matter of course, we expect the distribution of saccade length between the hemifields to be homogeneous rather than "peak-shaped" as in previous experiments, because the distance between corresponding items varies in Experiment E. This distance can be substantially longer than in the situation of translational symmetry. Thus, particularly long saccades are to be executed which may require subsequent corrective saccades in order to direct the gaze to the intended position. SL is likely to be affected by the occurrence of these additional saccades.

Furthermore, we expect to replicate the eccentricity effect reported by Bruce & Morgan (1975) under the conditions of Experiment E. This effect could be caused by two different factors: First, the probability of missing the target (PM) might be lower for items near the symmetry axis, and second, subjects may tend to scan the area near the axis *before* proceeding to more distant regions. If at least one of these considerations holds, there should be an effect on the subjects'

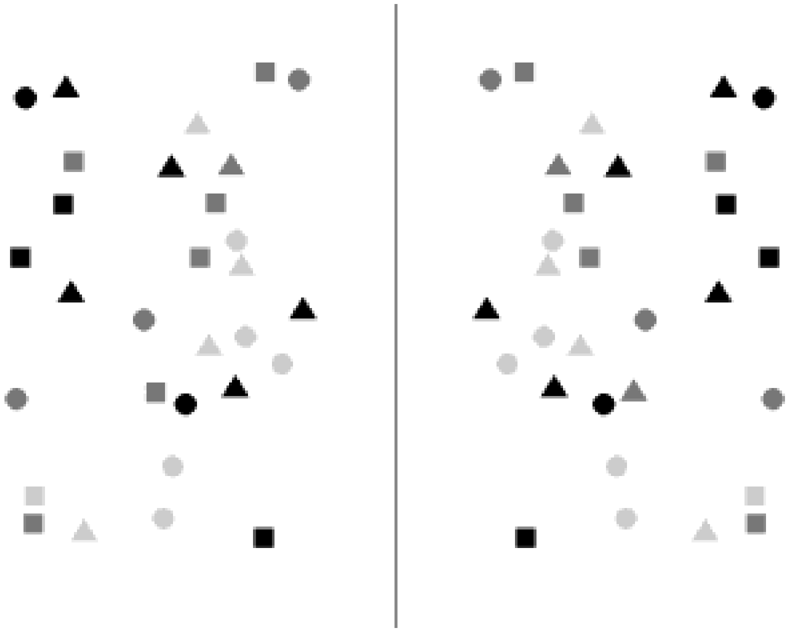


Figure 5.3: One of the stimuli used in Experiment E. The right hemifield is a mirror image of the left one, except for one mismatching pair of items.

reaction time (RT). RT should increase with the distance between the target and the symmetry axis.

But can we expect our subjects to require shorter reaction times in Experiment E than in Experiment A in replication of the Bruce & Morgan (1975) study? We must be careful, because the setting used by Bruce & Morgan (1975) is quite different from that in Experiment E. In their study, the mismatch between the hemifields consisted in missing or additional line elements. The stimuli were monochrome and all lines were of the same rectangular shape, such that no complex local features had to be compared. Moreover, the stimuli occupied only about six degrees of visual angle, allowing subjects to perceive the hemifields almost as a whole instead of conducting sequential steps of local memorization and comparison. In Experiment E, on the other hand, the global mirror symmetry is less relevant due to the necessity of local processing. It might at best be possible that targets in the vicinity of the dividing line are detected faster, because no or only small lateral eye movements have to be applied.

As a matter of course, the local information is affected by the type of symmetry as well. In the case of translational symmetry, the memorized information can be directly compared with the corresponding data in the other hemifield, while mirror symmetry requires the subjects to mentally perform the respective transformation. There is no obvious reason why this additional transformation should facilitate search performance, hence we should rather expect the opposite effect. At any rate, it is useful to compare the results of Experiments D and E in order to establish any effects of increasing task complexity.

As a basis for the comparison to Experiments A and D, we investigate the variables RT, SP, AC, SL, and PM as functions of the type of mismatch in Experiment E.

5.2.1 Method

Subjects

Sixteen paid subjects (7 DM each) participated in Experiment E. They were students of various faculties at the University of Bielefeld. They had normal or corrected-to-normal vision; none of them had pupil anomalies or was unable to distinguish between colors.

Materials

The type of stimuli and their presentation were the same as in Experiment A with one variation: The right hemifield was not a *translational copy* of the left hemifield as in the previous experiments, but its *mirror image*. The white line that separated the hemifields was taken as the symmetry axis. The random distribution of colors and forms as well as the integration of mismatches did not differ from Experiment A.

Apparatus

The OMNITRACK1 device was used to record the subjects' eye movements. The experimental set-up was the same as in Experiments A to D.

Procedure

All subjects were tested individually. Their task consisted in detecting the only color or form mismatch between corresponding, i.e. "reflected" items in the display. Subjects were to press a mouse key as soon as they found the mismatch. Each subject viewed 50 individually generated stimuli, 25 of which contained a color mismatch and 25 a form mismatch. The order of presentation of stimuli was randomly permutated. After every tenth picture, the eye tracker was recalibrated. Subjects were *not* informed when to expect which type of mismatch.

5.2.2 Results and Discussion

Figure 5.4 shows one subject's gaze trajectory on the underlying example stimulus. In fact, the subject starts scanning exclusively the targets which are located close to the symmetry axis, and indeed succeeds. Conceivably, he takes advantage of the assumed increased detectability of these "favorable" targets. This example motivates the investigation of the question raised above: Are targets near the dividing line detected faster than others? In order to assess the magnitude of the assumed horizontal target bias of RT, we compare it with the vertical one. In

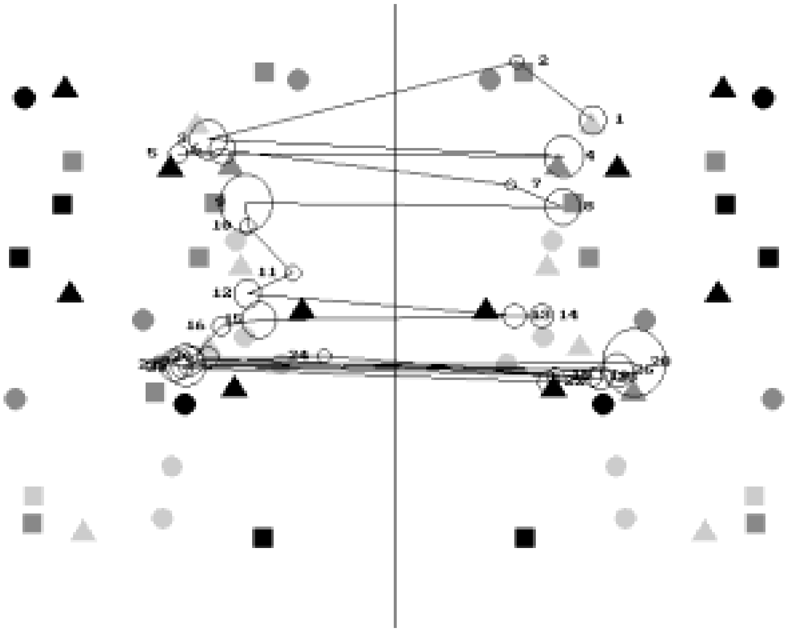


Figure 5.4: Example picture with a subject's plotted visual scanpath. Fixations are numbered; the circle size indicates fixation duration.

Experiments A to D, subjects tended to start scanning at the top of the display and then move downwards. Which is the dominant strategy in Experiment E?

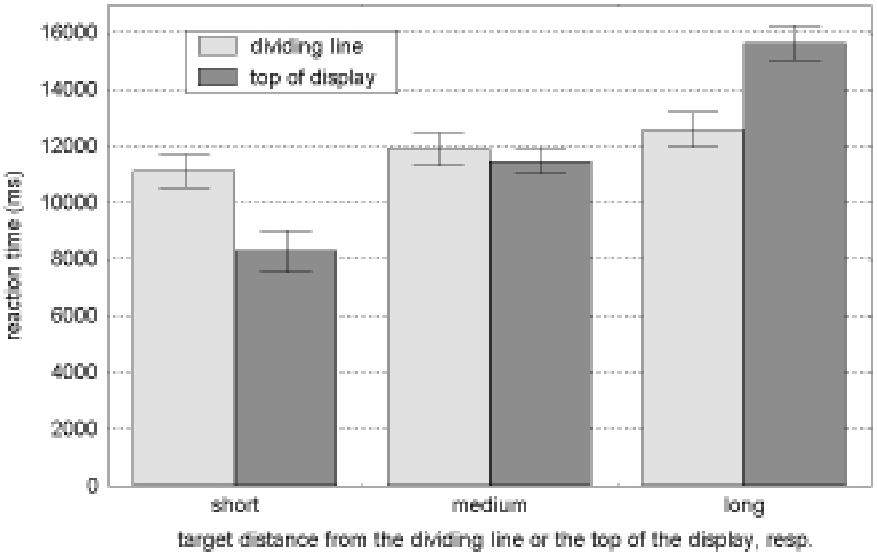


Figure 5.5: Reaction time for different target distances from the top of the display or the dividing line respectively

Figure 5.5 presents the RT values for three equidistant intervals of target distances. The light columns represent the horizontal intervals, the dark columns the vertical ones. A two-factorial analysis of variance (horizontal and vertical target distance) reveals that the vertical target position exerts a significant effect on RT ($F(2;30) = 28.91; p < 0.001$). A short distance to the top of the stimulus leads to shorter RT (8298 ms) than a medium distance (11490 ms) ($F(1;15) = 12.56; p = 0.003$). A long distance induces higher RT (15660 ms) than a short one ($F(1;15) = 40.22; p < 0.001$) or a medium one ($F(1,15) = 25.93; p < 0.001$). Increasing horizontal target distance from the dividing line exhibits a tendency to cause longer RT as well (11137 ms for a short, 11903 ms for a medium, and 12590 ms for a long distance). Just like the interaction between the two factors, however, the tendency to cause longer RT does not reach significance.

It must be stated that the influences of the horizontal and the vertical target position are not perfectly comparable, since the height of a hemifield exceeds its width by about 50%. The order of magnitude between the horizontal and the vertical target bias of RT, however, cannot be caused by the height-width ratio of the stimuli alone. Obviously, the top-down strategy still determines the subjects' scanpaths much more strongly than does the horizontal factor. Subjects prefer the plain, vertical alignment of scanpaths, although strategies as shown in Figure 5.4 are likely to be more efficient on average. Conceivably, a clear distinction between visited and not yet visited items is considered to be more important than the chance of faster target detection. This attitude is absolutely reasonable, because confounding visited and unvisited items could lead to substantially increased probability of missing the target (PM), which in turn could hardly be compensated by statistically faster detection.

With regard to the differences between color and form search, RT presents a significant dependence ($t(15) = -2.80; p = 0.014$); it is shorter for color mismatches (10903 ms) than for form mismatches (12958 ms). This difference of approximately two seconds compares well to the one obtained in the corresponding Experiment A (9903 ms versus 11997 ms). The absolute values, though somewhat higher in Experiment E, do not significantly differ from Experiment A. In all respects, the hypothesis of increased search efficiency in mirror-symmetric stimuli does not hold. This finding clearly supports our assumption of local information processing being impaired by additional mental transformation.

Accordingly, the efficiency parameters SP, AC, and SL should be higher in Experiment A than in Experiment E. Search speed (SP) is in fact substantially faster in Experiment A (45.44 pixels/s) than in Experiment E (31.65 pixels/s) ($t(30) = 5.31; p < 0.001$). Area coverage (AC) presents a similar effect between Experiment A (2619 pixels) and Experiment E (2006 pixels) ($t(30) = 3.92; p < 0.001$). Since subjects are not informed about the type of mismatch in both experiments, neither SP nor AC reveals a significant difference between color and form search. However, the third efficiency variable, namely saccade length (SL), exhibits a difference between Experiments A and E in the opposite direction. SL is significantly shorter in Experiment A (55.14 pixels) than in Experiment E (60.68 pixels) ($t(30) = -2.70; p = 0.011$). What can be the reason for this unexpected effect?

The reason conceivably lies in the occurrence of extremely long distances to be covered by single saccades. In Experiment A, the saccades between the hemifields are of approximately the same length, because there is a constant horizontal shift of 360 pixels between corresponding items. The stimuli in Experiment E, however, present horizontal distances between 100 and 620 pixels and hence require a certain proportion of very long saccades. Saccades are "programmed" before their execution rather than being continuously controlled during execution. This programming might be especially fast and often incorrect due to the subjects' lack of time. The end points of longer saccades are more likely to hit a "wrong" group of items in the opposite hemifield or sometimes even in the same hemifield, thus requiring an additional saccade for correction. If the imprecise as well as the corrective saccades are the reason for the increased value of SL, the within-hemifield saccades with starting points or end points of high horizontal eccentricity should be longer than those of low eccentricity. In fact, SL was significantly higher (69.85 pixels) for saccades with long distances between their end point and the dividing line than for those with a short distance (49.52 pixels) ($t(15) = 16.34; p < 0.001$). Obviously, attending to "wrong" item groups that have already been visited or still have to be visited does not strongly affect AC. It might, however, decrease SP, because "split" saccades go along with longer latencies than direct ones.

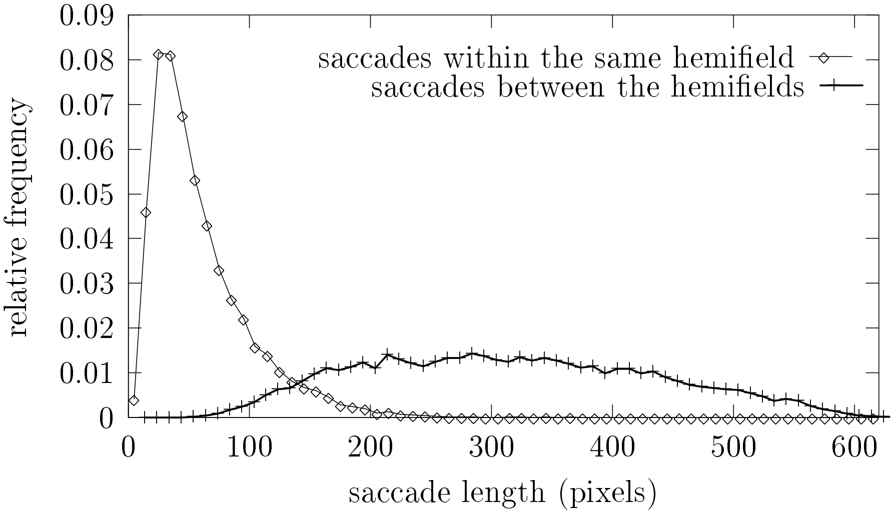


Figure 5.6: Histogram of saccade length

Besides the occurrence of corrective saccades, it is obvious that the introduction of mirror symmetry must exert an effect on the *distribution* of SL. Figure 5.6 shows SL histograms for the saccades within and between the hemifields respectively. Since the horizontal distance between corresponding items is variable, it is not surprising that the between-hemifield SL distribution in Experiment E has a much higher dispersion and does not show the "peak" found in Experiment A (see Figure 3.5). Interestingly, the maximum distance of 620 pixels between corresponding items is practically not occupied in the SL histogram. This is in line with the abovementioned idea that long distances are probable to be covered by two successive saccades instead of a single saccade. Furthermore, since subjects tend to inspect item groups rather than individual items, they prefer to conduct saccades between the innermost points of groups to be compared. As supported by the visual inspection of scanpaths, these two circumstances are responsible for an almost complete elimination of saccades longer than 580 pixels.

Generally speaking, the effects of search efficiency induced by the introduction of mirror symmetry resemble those being caused by match detection instead of mismatch detection in Experiment D. In fact, there are no significant differences in SP or AC between Experiment D and E, neither for color nor for form search. However, color search leads to smaller RT than form search in Experiment E (see above), whereas Experiment D does not show a dependence of RT on the type of match. RT for color search is significantly longer in Experiment D (13694 ms) than in Experiment E (10903 ms) ($t(28) = 3.09; p = 0.004$), while there is no such effect for form search (13496 ms versus 12958 ms). The low RT value in Experiment E is accompanied by an accordingly low PM value. With regard to color targets, PM is significantly higher in Experiment D (22.8%) than in Experiment E (15.9%) ($t(28) = 2.81; p = 0.009$). Since the subjects do not know the dimension of the mismatch, they can never ignore the form information, which seems to prevent them from “overestimating” their own capacities like e.g. in Experiment D.

5.3 Conclusions

Unexpectedly, the tasks in Experiments D and E, which were based on different types of inversions, have led to similar patterns of results. Search efficiency is lower than in the corresponding Experiments B and A respectively, indicating that inversion requires increased mental effort for the completion of the task. An advantage in efficiency of mirror symmetry over translational symmetry (e.g. Bruce & Morgan, 1975) has not been established. This effect seems to be restricted to global symmetry, and does not occur during successive processing of local information. Instead, successive processing of local information requires subjects to perform an additional mental transformation on the visual input which impairs their search efficiency.

A holistic comparison of local item groups appears to be the most efficient, because the correspondence of items between the hemifields is clearer. Furthermore, the findings suggest the dominance of “ad hoc” processing of information in tasks of comparative visual search. If holistic processing is possible, the visual information need not enter higher perceptual or cognitive levels. It is rather the case that the immediate visual impression can be memorized and matched with the corresponding region in the other hemifield. This strategy is most efficient if only a holistic same-different decision is required.

However, if the memorized information cannot be compared as a whole, its superficial processing is not sufficient for task completion. In Experiment D, neither the forms nor the colors of items correspond between the hemifields, hence not even the corresponding *position* of memorized information can be determined holistically. The locations of items might be processed independently of their color and form attributes in order to find the complementary items in the other hemifield. Subsequently, the memorized data are retrieved and compared, presumably demanding a serial process, since there is no holistic correspondence with mismatches "popping out" as in Experiments A to C. Experiment E makes similar demands on the subjects; although the items' identity between corresponding locations is the same, their relative positions are horizontally mirrored. Moreover, the mirror symmetry seems to disable subjects to perform a holistic comparison of memorized item information.

In Experiments D and E, the local information has to be represented in a rather flexible way, because transformations are to be performed and locations and attributes are to be retrieved separately. As indicated by the variables SL, AC, and SP, these increased demands on working memory, combined with a strive towards efficiency, may lead to smaller groups of items being memorized at the same time. Interestingly, fixation duration (FD) does not vary significantly between Experiments A to E, supporting our assumption of a constant capacity of working memory. Irrespective of the task, memory usage is adjusted in such a way that single steps of local memorization and comparison take about 200 ms, which seems to be the optimal duration in combination with the time "cost" of necessary saccades. Consequently, increasing qualitative demands on working memory will lead to smaller quantities of objects being memorized during each step.

A more thorough investigation of working memory usage during comparative visual search is performed in the following Chapter 6. The extension of the basis scenario allows to observe how working memory controls the search process. Furthermore, the stimulus-dependent capacity of working memory and its effects on search efficiency are investigated.

Chapter 6

Experiment F: From Items to Item Clusters

6.1 Guiding the Subjects' Attention

Despite the findings from Experiments A to E, it seems that several interesting basic questions cannot be addressed within this "class" of experiments:

- Are there general principles of scanpath selection? If so, what are these principles?
- How many items are held in working memory at the same time? What is the effect of color and form entropy on this number?
- Which perceptual characteristics of the stimuli guide the viewers' attention?

We might try to design visual displays which strongly suggest a grouping of items into clusters of a particular size. Assuming that the suggested grouping is indeed used by the subjects and determines their memory load, we might then study the resulting scanpaths and compare the findings for different sizes of suggested groupings. In this way, we might get a clue how scanpath generation and memory load (assuming the latter to be largely determined by the size of the suggested groupings) are interrelated. The dependence of various eye-movement variables on local stimulus features might become amenable to quantitative analysis. Finally, the sequential scanning of item groups can be assumed to reduce the immense variabililty of scanpaths found in the previous experiments. This situation may facilitate the revelation of general, basic characteristics of scanpaths.

What is an appropriate method of stimulus design with regard to this purpose? For a start, we could use spatially clustered items instead of homogeneously randomly distributed ones as in Experiments A to E. However, this approach has two drawbacks: First, the comparability to the former experiments would be rather low, because all items are located in regions of high density and can thus be perceived with fewer fixations and shorter saccades. Second, conclusions about the capacity of working memory are not generally valid for the paradigm of comparative visual search, since in most cases working memory can be "filled up" without employing any saccades during this process in most cases.

Recent research in our group (Koesling, Pomplun & Ritter, submitted 1997) suggests a different approach of how to create suitable stimuli. In Hendrik Koesling's experiments, subjects were presented with scenes as shown in Figure 6.1. These scenes showed two non-identical random distributions of dots with additional *connecting lines* in the left hemifield, forming groups of approximately equal size. The subjects had to decide whether there were more dots displayed in the left or in the right hemifield. As a matter of fact, they had to base their estimate on perception and were not allowed to count the dots (the reader is invited to solve the example task displayed in Figure 6.1 before reading the correct answer in the next paragraph. Which of the hemifields has more dots?).

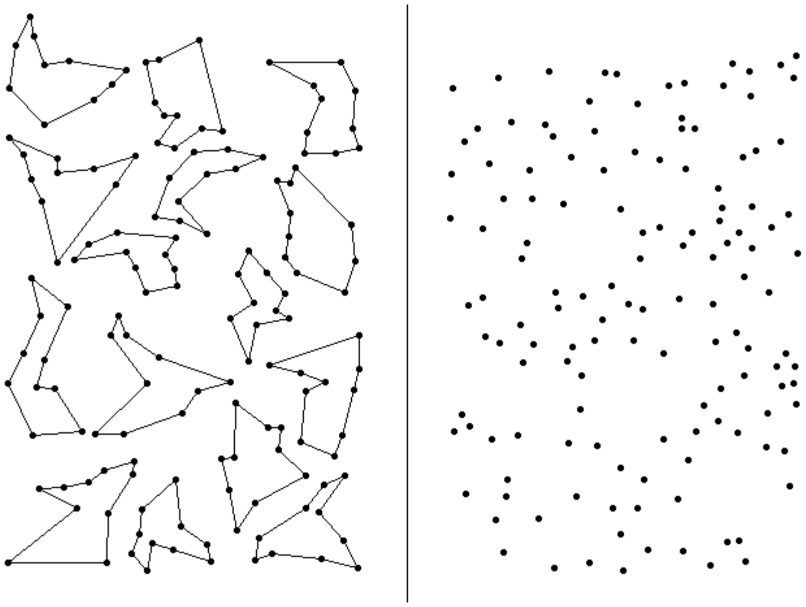


Figure 6.1: One of the stimuli used in Koesling's experiments. Which of the two hemifields contains more dots?

As the empirical results showed, people tend to believe that it is the *right* hemifield that contains more dots. In fact, there are 150 dots in the left hemifield while there are only 135 dots in the right hemifield. This perplexing effect of quantitative underestimation caused by the additional line elements (edges) was found to increase with the number of displayed dots. In another series of experiments, subjects were to spontaneously adjust the number of single dots shown in the right hemifield to match the number of connected dots in the left hemifield. The data obtained revealed a mean underestimation ranging from 20% (30 dots shown) to 44% (150 dots shown). Moreover, the studies demonstrated that the introduction of line elements had a substantially stronger impact on the subjects' estimation bias than many other methods of grouping which were tested as well: Grouping by color, contour, or pointers exerted, if any, only small effects.

The observed occurrence of underestimation can be attributed to the *perceptual grouping* of dots induced by the additional lines. In the left image of Figure 6.1, for instance, we perceive groups of dots rather than a distribution of single dots plus another distribution of lines or polygons. In fact, the lines make it hard to perceive the (existing!) homogeneous random arrangement of dots as such. The structural information given by the edges leads to a perceptual "simplification" of the left hemifield causing the estimation bias (for a detailed discussion see Koesling, Pomplun & Ritter).

With respect to our initial problem, what can we infer from these findings? The experiments described above yielded evidence for connecting lines to cause a strong effect of perceptual grouping in random distributions of dot-like items. It is important to note that no change of any item positions was necessary in order to achieve this perceptual impression. Hence, the idea near at hand is to connect the items in our scenario of comparative visual search with *straight lines*. Compared to our aforementioned suggestion to group items by moving them closer to each other ("spatial clusters"), this method has the advantage of not changing the overall random distribution of item locations. Although we have to take into account the presence of additional line elements, the comparability to previous experiments as well as the general validity of results is, therefore, significantly enhanced.

What effects of the additional connecting edges on the subjects' gaze trajectories can we expect? First of all, the stimuli are likely to be scanned in a cluster-by-cluster fashion; this was our main requirement for their creation. Consequently, *cluster size* is an important independent variable in Experiment F: In which way do the subjects' eye movements depend on the size of the perceptual units, i.e. the number of items per cluster?

Being a "classical" measure of search efficiency, reaction time (RT) is investigated as a dependent variable. Ideally, we expect RT to present a distinct minimum value for a specific cluster size, indicating the presumably optimal number of items to be memorized at a time during comparative search. As suggested by Experiments A to C (see Chapter 4), however, speed of processing (SP) and area coverage per fixation (AC) might be more reliable indicators for efficiency than RT. Maximum values of SP and AC would suggest the optimal cluster size. At any rate, RT, SP, and AC should signify the same size to be the optimal one.

As argued above, if subjects scan the display in a clusterwise pattern, we are able to precisely investigate the dependence of eye-movement parameters on the *size*, *color entropy*, and *form entropy* of a cluster being processed. For this purpose, we have to use appropriate within-cluster parameters of gaze trajectories as dependent variables: *Number of comparisons between hemifields* (NC), *number of fixations* (NF), and *duration of processing* (DP). If these variables depend proportionally on cluster size, the potential effects of cluster size on the global measures RT, SP, and AC cannot be attributed to processes *within* the clusters. In this case, the global efficiency gains for specific cluster sizes are conceivably caused by the cluster structure facilitating the construction and tracking of global scanpaths.

6.2 Method

6.2.1 Subjects

Sixteen students of various fields at the University of Bielefeld took part in Experiment F. They were paid 7 DM for their participation. All of them had normal or corrected-to-normal vision; none had pupil anomalies, and all were able to distinguish between colors.

6.2.2 Materials

Basically, the stimuli had the same appearance as in Experiment A. Colors and forms were still distributed with a tendency towards the generation of uniform areas, and the mismatch was either in color or in form.

As mentioned before, one of the key questions to be investigated is the influence of cluster size on the eye-movement parameters. In this context, cluster size means the number of items in each perceptual group (or cluster) formed by the connecting edges. This cluster size was varied from one to six with “size one” corresponding to the presentation of unclustered stimuli like those used in Experiments A and B. These unclustered stimuli were integrated into Experiment F as well, yielding baseline data for the investigation of grouping effects. It goes without saying that the total number of items per hemifield was kept constant at a value of 30, which led to a problem with stimuli of cluster size four. Since 30 cannot be divided by four, we chose stimuli of six clusters of size four plus two clusters of size three. In Experiment F, 60 different stimuli were used, ten of each cluster size. Five stimuli per cluster size contained a color mismatch, the other five contained a form mismatch.

Thin white lines were introduced to connect the centers of items and thus to form item clusters. These lines were identical for the left and the right hemifield. In analogy to the experiments of Koesling, Pomplun & Ritter, the clusters were algorithmically generated with the minimization of geometrical within-cluster distances between items being the clustering criterion. Items within the same cluster were connected with lines using a standard “travelling salesman” algorithm (see Section 3.1.1). It can be mathematically demonstrated that the connecting lines of a valid TSP path never intersect each other. Therefore, using the TSP algorithm led to a standardized and clearly arranged appearance of clusters. Figure 6.2 presents one example stimulus for each cluster size.

Since we expected a reduced variability of scanpaths due to the “guiding effect” of the suggested clusters, it did not seem useful to generate individual stimuli for each subject as performed in Experiments A to E. Presenting each subject with the same set of stimuli in randomized order instead, enabled us to

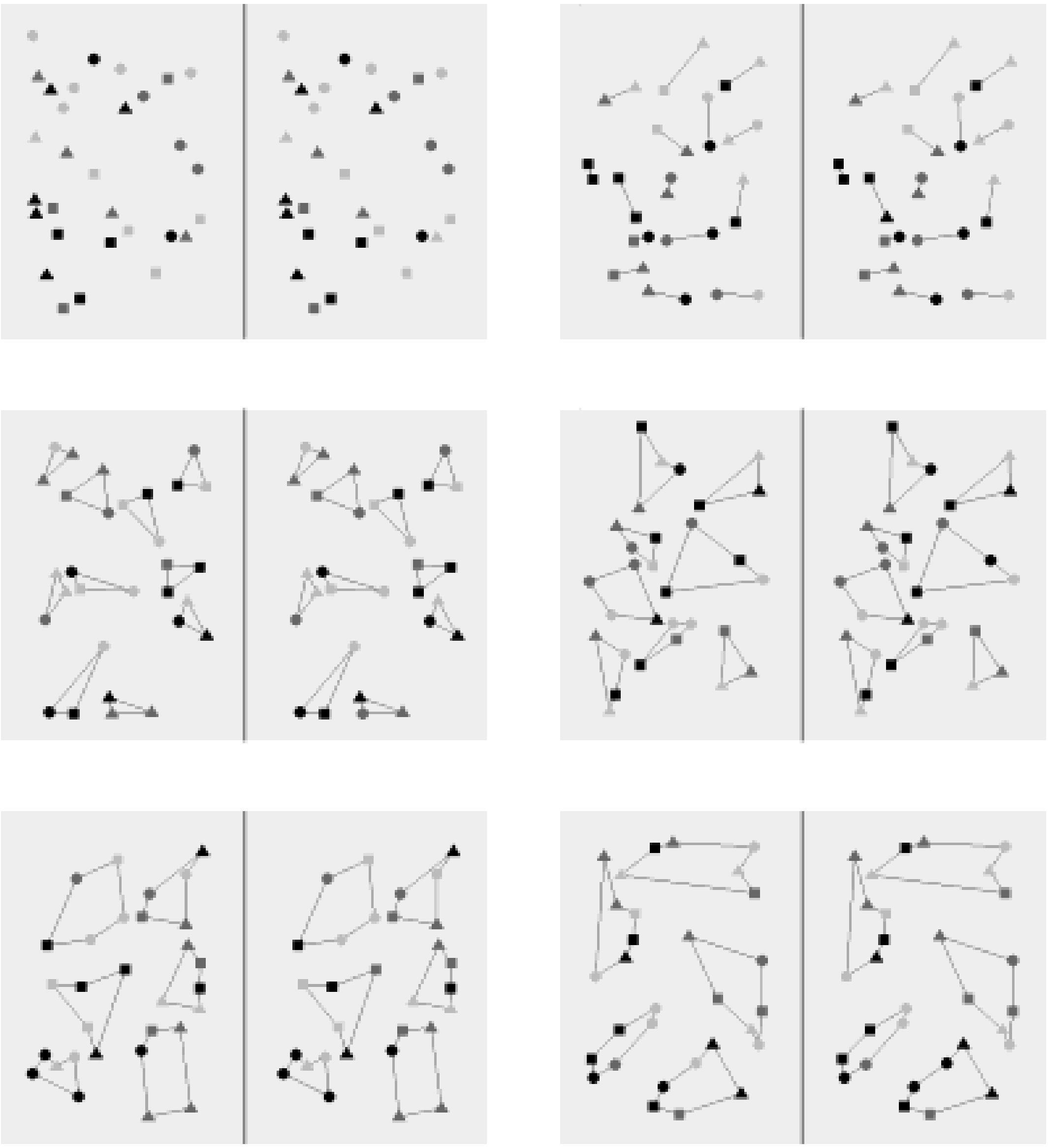


Figure 6.2: Example stimuli used in Experiment F with cluster size increasing from one to six

superimpose their scanpaths in order to discover common regularities.

Although this experimental design was favorable, the restricted stimulus set required us to take a closer look at its statistical balance. For example, the overall reaction time RT was known to depend to a high extent on the vertical target position (see Experiment E). In order to obtain valid RT results, the target position had to be counterbalanced across stimuli. For this reason, the left hemifield of every generated stimulus was divided into three times three rectangular areas of equal size. For each cluster size, the stimuli used in the experiment were chosen in such a way that the targets were located twice in the center area and once in each of the eight surrounding areas.

6.2.3 Apparatus

Eye movements were recorded with the help of the OMNITRACK1 device. The experimental setting remained the same as in the previous experiments.

6.2.4 Procedure

Subjects were tested individually. They were instructed to find the only mismatch between corresponding items without being informed about its type (color or form mismatch). They were to press a mouse key as soon as they detected the mismatch. Each subject viewed the 60 stimuli in an individually generated random order. After every tenth picture, the eye tracker was recalibrated.

6.3 Results and Discussion

6.3.1 Qualitative Analysis of Scanpath Characteristics

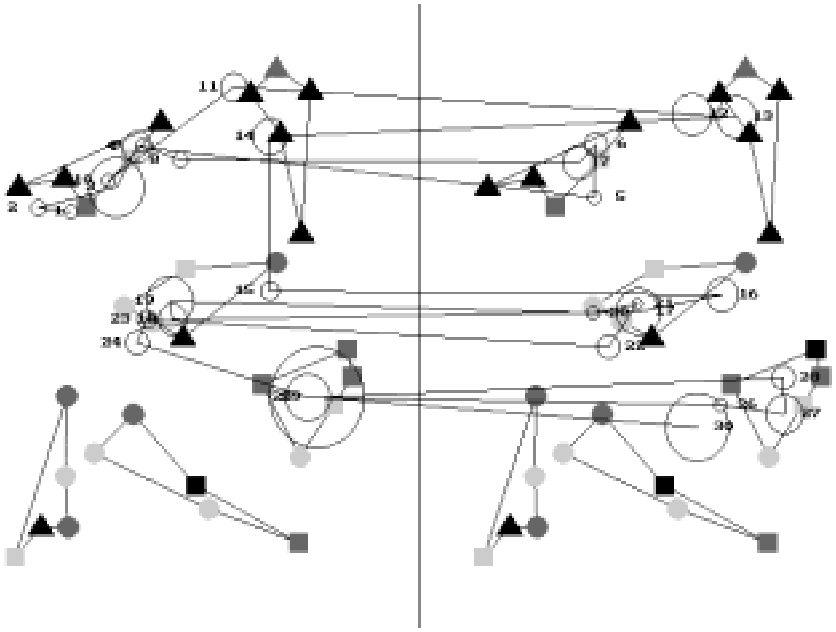


Figure 6.3: One of the stimuli presenting five items per cluster, superimposed with a subject's scanpath. Fixations are numbered; circle size indicates fixation duration.

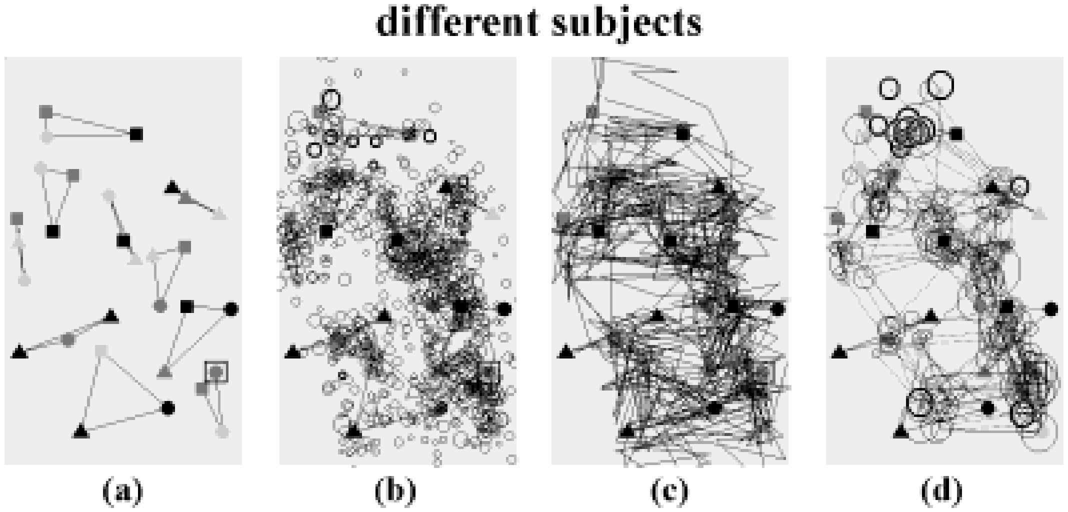
Did the additional line structure essentially influence the subjects' eye movements? For reasons of illustration, Figure 6.3 presents the scanpath generated by one of the subjects for an example stimulus of cluster size five. The effect exerted by the grouping lines is clearly visible: The subject employed a cluster-by-cluster strategy. Actually, this was the typical way displays were scanned in Experiment

F. As a rule, subjects did not proceed to a different pair of clusters before the currently visited pair had been completely inspected. This behavior was observed for all subjects and almost all stimuli with cluster sizes ranging from three to six. Cluster size two did not seem to trigger this effect, but this may be due to difficulties in assigning gaze positions to particular clusters, because clusters of two items are shaped like lines rather than areas. An alternative interpretation would be that the information content of two items may be insufficient for subjects to fill up their working memory, so that single clusters of size two cannot attract the subjects' attention completely. We will return to this point further below.

Obviously, the additional structural information served as a "guideline" for the creation of a global, self-avoiding search path. When using clusters instead of individual items as scanpath units, it is much easier for the subjects to remember which part of the display they have already visited during task completion.

The fact that our clusters control the subjects' eye movements to some extent should result in a reduced variability between scanpaths. Thus, it is promising to pose one of the initial questions once again: Are there general principles of scanpath selection? An appropriate method of investigation is to visualize the eye-movement data of all 16 subjects at the same time on the background of the underlying stimulus picture. For the benefit of a clearly arranged illustration, it is useful to disregard the between-hemifield saccades and to superimpose all data onto *one* hemifield.

The upper row of Figure 6.4 shows the results of this procedure with regard to an example stimulus of cluster size three. One hemifield of this stimulus with a square marking the target item is shown in diagram (a). The next diagram (b) presents the fixations produced by all subjects while processing the example stimulus. Each fixation is represented by a circle with its radius being proportional to fixation duration, the thick circles indicating the first fixation measured for each subject. Not surprisingly, the fixations are attracted to the items such that the fixation pattern reveals a "hole" in the item-free region to the left of the image center. Moreover, we see that most of the subjects start their search at the top of the picture, some of them at the bottom, but no one in between. Apart from these rather trivial findings, no particular characteristics of eye movements can be detected. This is also the case for diagram (c), which shows all saccades that were generated by the 16 subjects. The only available information is that there are more saccades in areas of high local item density. Finally, diagram (d) gives a more detailed insight into scanpath features by displaying the subjects' cluster-to-cluster transitions. For each scanpath, all fixations belonging to the same cluster are summed up. The position of the circle indicates their center of gravity, and its radius signifies the sum of their durations. These circles are connected by lines according to their temporal order of occurrence and the start circle is thicker than the other ones. While the centers of gravity and the sums of durations still exhibit strong variability, the structure of the cluster-to-cluster paths reveals a certain tendency. There is a "chain" of four clusters leading from the center to the lower right corner of the image. These clusters have only small distances between each other and are aligned to roughly the same direction. As the accumulation of connecting lines in diagram (d) proves, subjects clearly tend to visit these four groups in immediate succession. It seems that the *minimization of scanpath length* plays an important role in the subjects' search strategies, reminding us of the travelling salesman problem (see above).



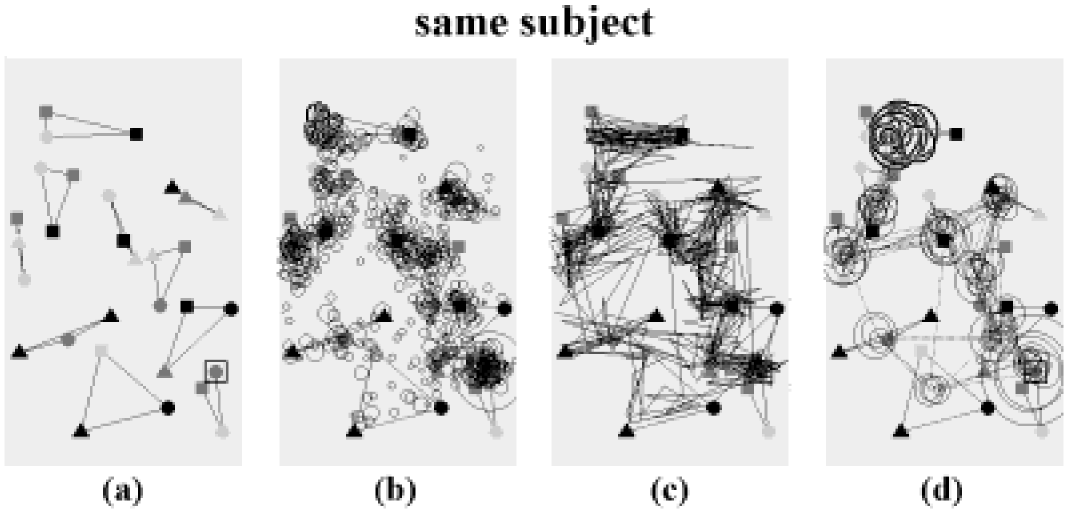


Figure 6.4: Eye-movement data accumulated over 16 different subjects (upper row) and 10 trials of the same subject (lower row) performed on an example stimulus (a). The distribution of fixations (b), the scanpaths (c) as well as the cluster-to-cluster trajectories (d) are displayed with circle size indicating duration and thick circles signifying the starting point of each path.

At this point, another interesting question arises: Is there an even more reduced variability between scanpaths of the *same* subject, yielding more precise information on strategies and their individual characteristics? To try to give an answer to this question, we had one additional subject participate in the experiment *ten times* with delays of three or four days between the sessions. Undoubtedly, the investigation of a group of subjects instead of a single subject would have led to results of higher general validity, however, the effort for the conduction of such a study would have been beyond the scope of the present work. Being aware of the missing statistical significance, we inspected the eye movements of the single subject in order to get an idea of the within-subject variability of scanpaths.

As a matter of fact, an effect of learning partially influenced his eye movements and led to shorter reaction times. This a-priori knowledge was indicated by long saccades hitting the target area within the first one or two seconds after stimulus onset. It was found by his report and the inspection of his scanpaths that in the case of only eleven of the 60 stimuli no such influence took place in any of the ten respective trials. One of those was the stimulus shown in Figure 6.4, where the lower row presents the eye-movement patterns accumulated over the subject's ten trials. Diagram (b) exhibits a stronger concentration of fixations around the items than in the between-subjects analysis, but their exact locations are still unpredictable. In all trials the subject started searching at the top of the display. His saccades, shown in diagram (c), suggest that he prefers a distinct global strategy. There seems to be less "noise" than in the 16 paths of different subjects. This is true for diagram (d) as well, which indicates only very few deviations from his favorite scanpath. Unlike most of the other subjects, he integrated a fifth group of items into the abovementioned chain of clusters, which might indicate the existence of stable individual preferences. Moreover, the centers of fixations within clusters do not shift strongly between his paths, suggesting that he paid an individually constant amount of attention to each item. This attentional pattern is not reflected in the distribution of fixations, as we have seen in diagram (b).

All in all, the qualitative analysis of both between-subject and within-subject scanpaths gives some idea of their global structure and their variability. On the level of single fixations, however, we do not find any common regularities.

6.3.2 The Influence of Cluster Size

Now that we have studied the effects of clustered stimuli in a qualitative fashion, we have to take a look at the impact of cluster *size* on the empirical data. The first idea is to analyze the dependence of the overall reaction time (RT) on cluster size in order to determine which size is most convenient for the subjects. The mean RT values separated for different cluster sizes are shown in Figure 6.5. Neither a linear dependence on cluster size nor an RT peak is visible. In fact, there is no significant effect of cluster size on RT.

What does this finding mean? Is the factor cluster size completely irrelevant to the subjects' performance? Before such conclusion, we might remember that, in Experiments B and C, the pure analysis of RT did not prove to be sufficient to establish such effects either. RT was biased by subjects' target misses, which led to the more specific performance indicators *search speed* (SP) and *area coverage per fixation* (AC) being checked for clarification. In Experiment F, the situation seems to be comparable: An even higher probability of missing the target (PM) of 18.1% adds a lot of "noise" to the RT values. Therefore, a closer analysis of

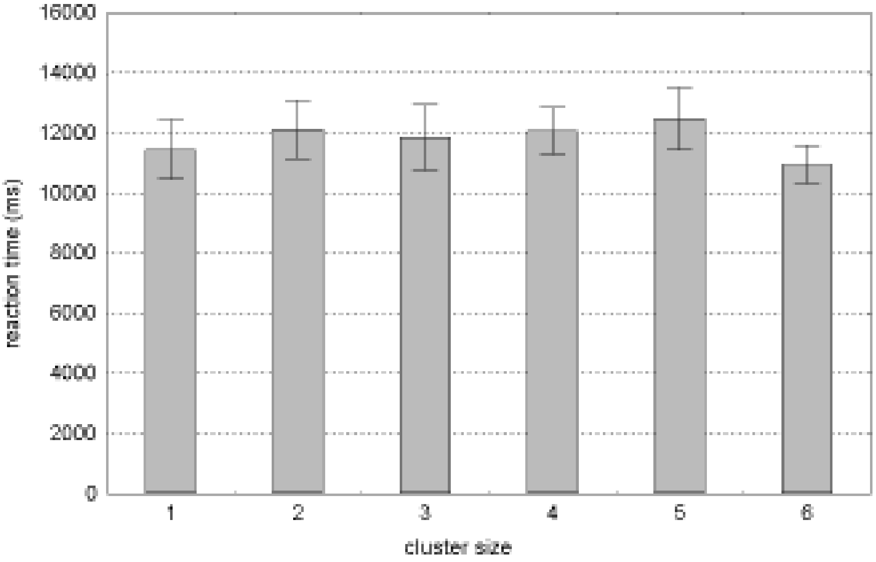


Figure 6.5: Reaction time RT for different cluster sizes

SP and AC is required before we draw any conclusions.

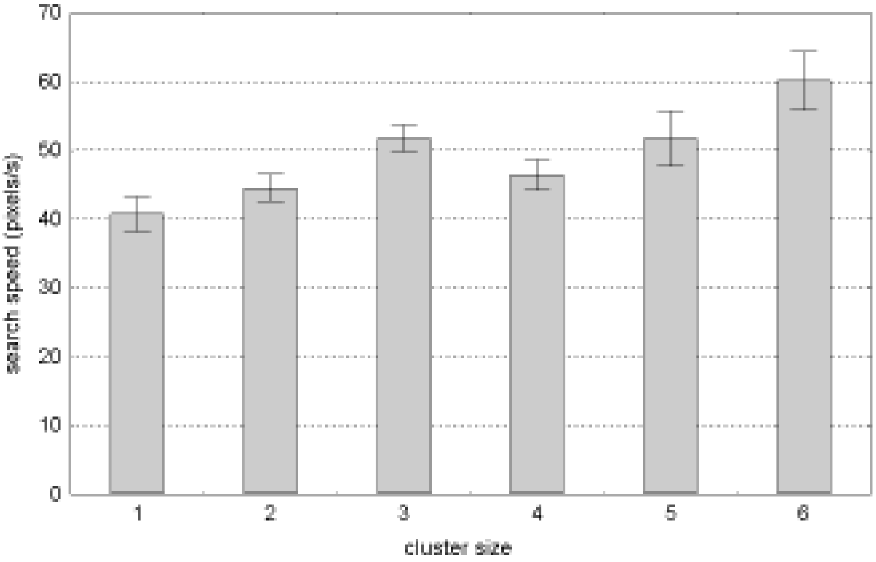


Figure 6.6: Search speed SP as a function of cluster size revealing a "double stairway" characteristic

In fact, the SP results shown in Figure 6.6 present an interesting pattern: SP increases from cluster size one to three, declines from size three to four, and increases again from size four to six, yielding a pattern that reminds us of a "double stairway". The analysis of variance reveals a significant effect of cluster size on

SP ($F(5;75) = 10.82; p < 0.001$). Also, the contrasts between different cluster sizes are analyzed: While the SP difference between size one (40.76 pixels/s) and two (44.61 pixels/s) only presents a tendency, the SP value for size three (51.82 pixels/s) is significantly higher than for size one ($F(1;15) = 33.40; p < 0.001$) as well as size two ($F(1;15) = 16.18; p = 0.001$). The "step back" between size three and four (46.43 pixels/s) also reaches significance ($F(1;15) = 9.14; p = 0.009$). The SP variation from size four to five (51.76 pixels/s) is just a tendency, whereas SP for size six (60.17 pixels/s) is significantly higher in comparison to both size four ($F(1;15) = 11.90; p = 0.004$) and size five ($F(1;15) = 5.02; p = 0.041$). Last not least, the difference between the two peaks at size three and size six is reliable as well ($F(1;15) = 5.54; p = 0.033$).

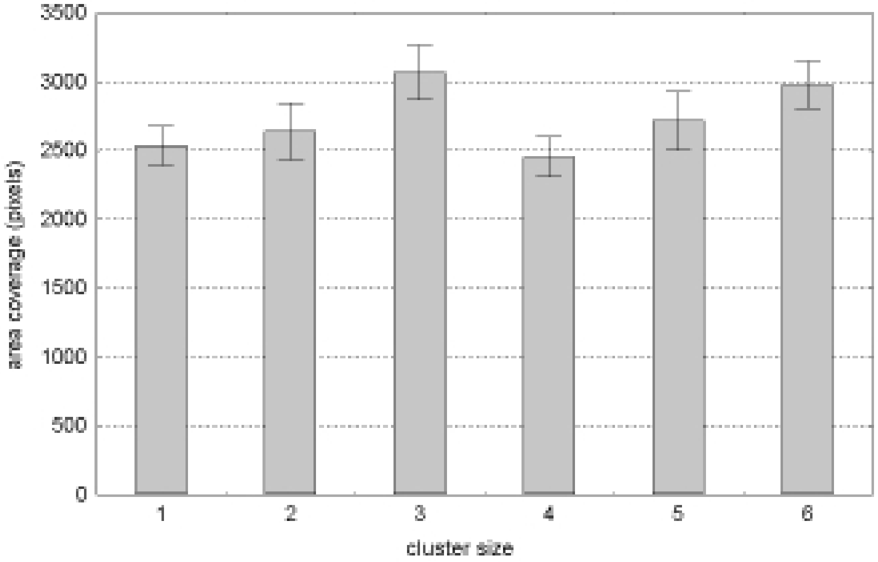


Figure 6.7: Area coverage AC in dependence on cluster size

Basically, the same "double stairway" pattern as for SP occurs when analyzing AC (see Figure 6.7). An analysis of variance exhibits a significant dependence of AC on cluster size ($F(5;75) = 3.03; p = 0.002$). The analysis of contrasts between different cluster sizes led to the following results: The increase in AC between size one (2539.3 pixels) and size three (3068.4 pixels) constitutes a significant difference ($F(1;15) = 10.00; p = 0.006$) as well as its reduction from size three to size four (2456.1 pixels) ($F(1;15) = 14.45; p = 0.002$). Furthermore, its increase from size four to size six (2978.1 pixels) reaches significance ($F(1;15) = 11.57; p = 0.004$) as well. The AC values for cluster size two (2638.5 pixels) and five (2716.4 pixels) show statistical tendencies to differ from their neighboring values, however, without attaining significance.

The basic variables FD and SL are not affected by cluster size. Moreover, it should be noted that the values of RT, SP, AC, FD, and SL for cluster size one do not differ significantly from the respective values obtained in Experiment

A. Conceivably, the connecting edges shown in most trials do not motivate the subjects to generally apply different strategies of search and comparison.

From the results of SP and AC it is obvious that the subjects' performance increases approximately linearly from cluster size one to three, falls off at size four and increases again in an approximately linear fashion up to size six. This finding suggests that the processing of stimulus subsets containing *three items* or multiples of object triplets leads to optimal search efficiency. According to this interpretation, memorizing clusters of size one or two does not completely "fill up" working memory, so that subjects have to combine them to larger ones. This effort is not necessary in a situation of three items per cluster, which increases the subjects' search efficiency. Clusters of four items appear to be processed in two steps; they are split up into two groups containing either one and three items or two items each. An alternative method, namely memorization and comparison of item triplets across cluster boundaries, does not seem to be applied, as the qualitative analysis of eye movements suggests. It could be inferred that the capacity of working memory is not exhausted and thus the optimal search efficiency cannot be reached. The same problem occurs for clusters of size five, but here the impact on efficiency is less dramatic since they can be divided into two groups of two and three items respectively. Apparently, six items per cluster allow a perfect separation into two clusters of size three and therefore do not inhibit the search progress. On the contrary, the possibility of choosing how to split up the big clusters seems to allow even faster search than when there are three items per cluster, as the SP data suggest. Another circumstance that conceivably contributes to the SP increase might be that five clusters of size six simplify the selection of scanpaths even more than do ten clusters of size three.

Although it involves some speculation, the assumption of an "optimal" cluster size of three is plausible; it is in line with the results obtained so far. This observation motivates the investigation of eye movements within the processing of single clusters in order to find out whether the SP and AC effects are based on differences during the comparison of clusters or rather during the transition between them. Moreover, the dependence of within-cluster variables on cluster size is likely to reveal information about working memory and its utilization during comparative visual search.

6.3.3 What Happens Within a Cluster?

In order to analyze a section of a scanpath that corresponds to a subject's processing of a certain cluster pair, we have to identify the first and the last fixation directed at this pair. The algorithm for the analysis of this works as follows: Each fixation is attributed to one of the cluster pairs in the display, namely, the one that contains the item with the shortest geometrical distance to the fixation point. The onset of the first fixation attributed to a certain cluster pair is taken as the starting time of processing. Now the algorithm could successively inspect the subsequent fixations until it finds one that belongs to a different cluster pair. It then could take the end of the last fixation that still belongs to the current cluster pair as the end time of processing.

However, this method would not account for the fact that saccades between hemifields sometimes go along with orientational or imprecise fixations (see Section 5.2.2). These are likely to be assigned to a different cluster pair than the one that is being processed. Thus, if one fixation of this kind occurred during a subject's inspection of a certain pair, the algorithm would interpret it as the "end signal" of the ongoing processing and the next one as the "start signal" of another processing of the same cluster pair. As a result, we would find too many and too short sections of processing, leading to improper results. In order to avoid this, the algorithm disregards those single fixations of a pair A which are preceded and followed by one or more fixations of a different pair B. A processing of a cluster has to include at least one saccade between the hemifields in order to be accepted by the algorithm. Visual inspection shows that these refinements lead to plausible scanpath intersections.

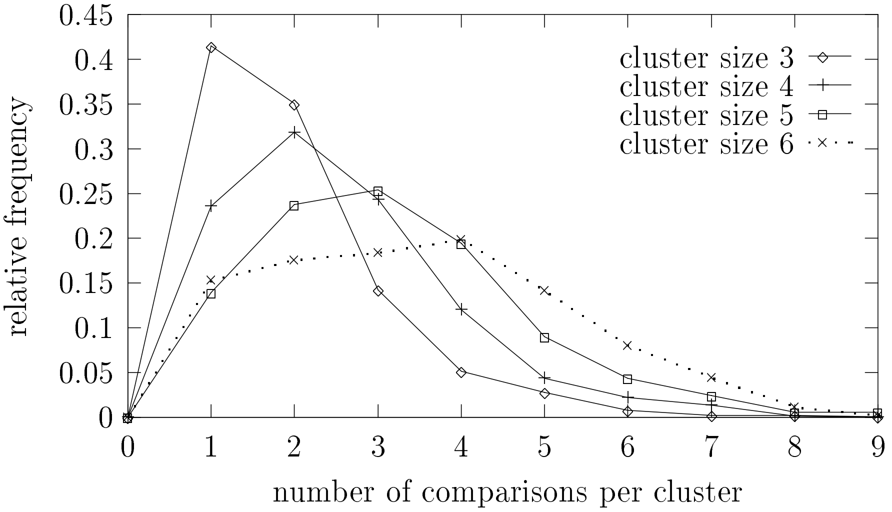


Figure 6.8: Histograms of the number of comparisons per cluster processing in dependence on cluster size

Now that the facilities for clusterwise analysis of gaze trajectories are available, what are promising variables to look at? With the investigation of working memory in view, one of the most interesting within-cluster variables is the *number of comparisons* (NC). It is defined as the number of saccades between the hemifields and tells us how many switches (comparisons) between hemifields the subjects needed to process a cluster of a certain size. Figure 6.8 presents histograms of NC for cluster sizes three to six. Sizes one and two are not considered since they do not admit a clear correspondence between fixations and clusters (see above). In stimuli of size four, the two cluster pairs of size three are excluded from the analysis.

The mean values of NC were entered into an analysis of variance with the only factor being cluster size. It revealed a significant main effect ($F(3;45) = 76.50; p < 0.001$) as well as significant contrasts between cluster sizes: NC is higher at size four (2.429) than at size three (1.788) ($F(1;15) = 77.23; p < 0.001$). Cluster size five induces greater NC (3.012) than size four ($F(1;15) = 51.43; p < 0.001$), and finally size six leads to higher NC (3.430) than size five ($F(1,15) = 16.71; p = 0.001$). As a matter of fact, all other contrasts reach significance as well.

All in all, we find that with increasing cluster size the histograms grow towards larger numbers of comparisons. This dependence of NC on cluster size is in agreement with the assumption of a constant capacity of working memory. The more items are to be compared, the more processes of memorization and retrieval are to be conducted. However, when inferring cognitive functions and their parameters from these data, we must be careful. Since all clusters consist of homogeneously randomly distributed items, the area covered by them *increases* with cluster size. This means that the dependence of NC on the number of items in a cluster could be attributed to *geometrical* cluster size. Even if the subjects scanned the display at random, the larger clusters would statistically be hit by more fixations and thus yield higher NC values. As a consequence, the results would reflect features of the stimuli rather than characteristics of specific cognitive processes.

It is important to investigate these possibilities, but how can we separate cognitive factors from stimulus factors? An idea is to do another analysis of NC, but this time based on computer-generated, pseudo-randomly located “fixations”. If NC still showed the same functional relationship to cluster size we would have to reject a cognitive interpretation of our findings.

However, a pattern of random fixations is quite different from the empirical pattern during comparative visual search, making it difficult to compare the two analyses. In order to preserve the spatiotemporal characteristics of subjects' eye movements while dissolving any relationships with geometrical characteristics of the display, the empirical fixation patterns were rotated by 180^o around the centers of the underlying stimuli.

An analysis of variance showed that NC is still significantly affected by cluster size ($F(3;45) = 22.48; p < 0.001$), but in contrast to the previous analysis its differences between size four and five as well as between size five and six do not reach significance. Moreover, the differences between the mean values of NC (2.225 for size three, 2.470 for size four, 2.674 for size five, and 2.805 for size six) have harshly diminished. The results of this “imitation” analysis are illustrated in Figure 6.9. As can easily be seen, the NC differences between cluster sizes appear to be much smaller than in the “genuine” analysis (see Figure 6.8). The histograms of cluster sizes three to six practically present the same shape. On this basis, we may safely assume the cognitive view to hold, and to push the investigation of within-cluster variables even further.

Another important indicator of cognitive processes is the *number of fixations* (NF) during the processing of a cluster pair. We can assume NF to increase even more strongly with cluster size than NC does, because longer within-cluster distances have to be bridged. The results displayed in Figure 6.10 do not support

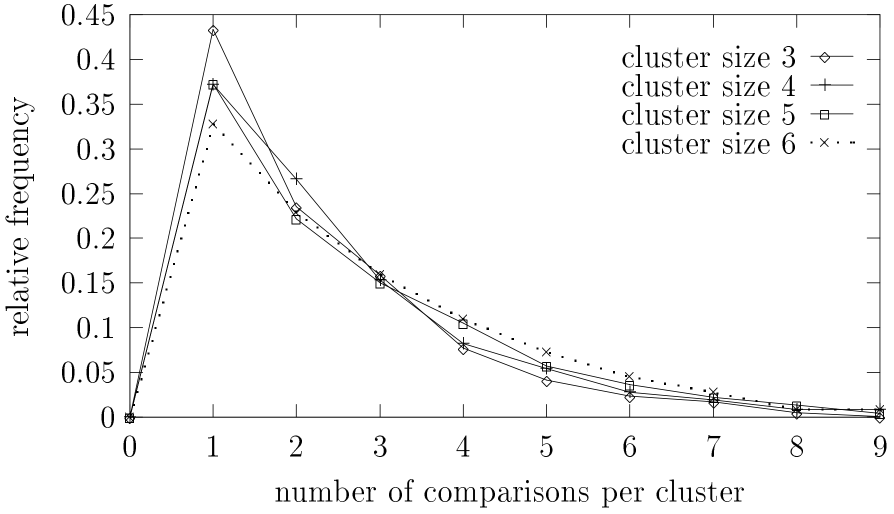


Figure 6.9: Histograms corresponding to the previous figure, but computed on the basis of dislocated fixations

this assumption: The distributions of NF resemble the ones of NC quite closely. In fact, the mean values of NF (4.740 for size three, 6.541 for size four, 7.979 for size five, and 9.101 for size six) are approximately proportional to NC. One step of comparison corresponds to 2.65 fixations. This finding suggests a close relationship between NC and NF that is independent of cluster size.

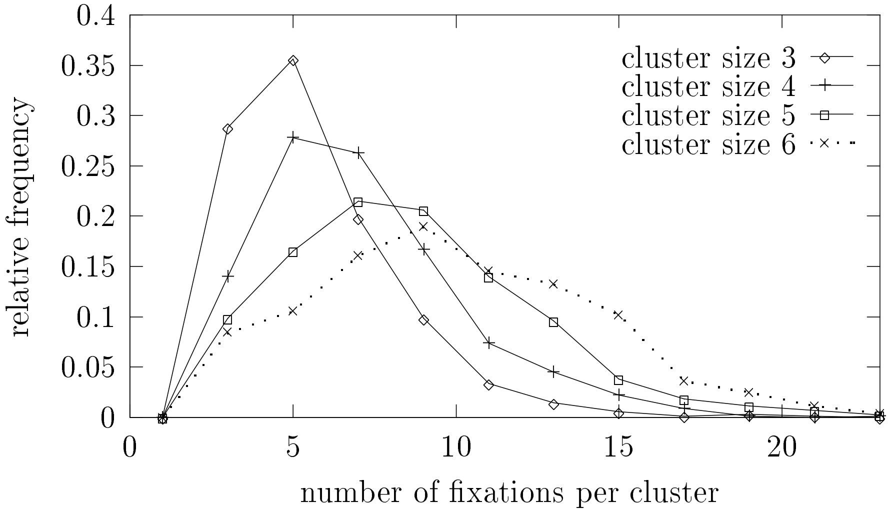


Figure 6.10: Histograms showing the number of fixations per cluster processing for different cluster sizes

Finally, we should take a look at the *duration of cluster processing* (DP). This variable is measured as the difference between the starting time and the end time of processing a cluster pair. The analysis of DP is of particular importance, because it accounts for the duration of orientational fixations which were excluded from the analyses of NC and NF. Deviations of DP from the proportionality to NC and NF would hint at the existence of cluster-size dependent effects of orientational fixations on the subjects' gaze trajectories that would have to be studied more closely. T-tests revealed that the mean DP values (915.9 ms for size three, 1269.9 ms for size four, 1492.8 ms for size five, and 1767.7 ms for size six) do not significantly deviate from proportionality to NC and NF. One step of comparison takes about 512 ms. Furthermore, the inspection of Figure 6.11 suggests that the histograms of DP do *not* differ from those of NC and NF.

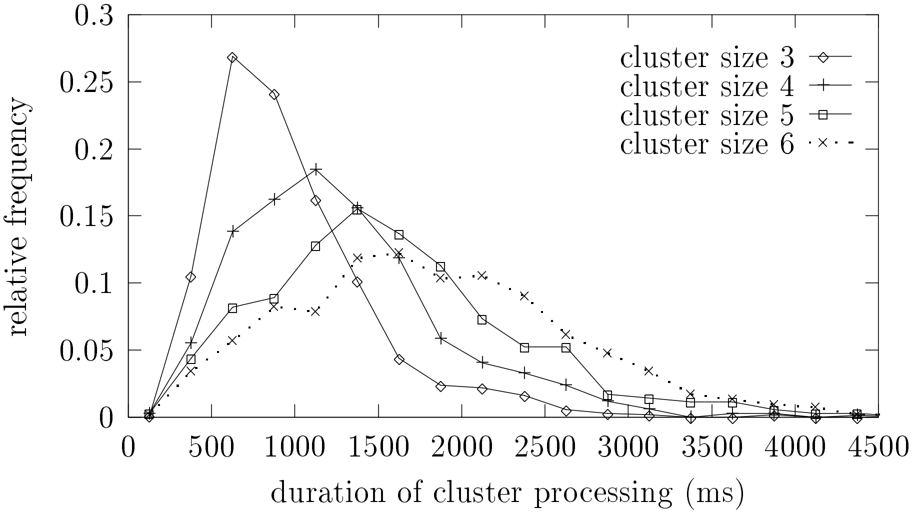


Figure 6.11: Histograms showing the duration of cluster processing in dependence on cluster size

How do the within-cluster results relate to the effects of cluster size on the global variables SP and AC? If those effects are based on differences that occur *during* the processing of clusters, we would expect the within-cluster variables to show a dependence on cluster size corresponding to SP and AC. When investigating this point, we have to take into account that SP and AC refer to the processing of a constant number of item pairs (30), whereas NC, NF, and DP refer to a number of item pairs that varies with cluster size. In order to make these two groups of variables comparable to each other, we divide the within-cluster variables by their respective cluster size so that all of them refer to the processing of exactly *one* item.

In Figure 6.12, the relative values of these variables for different cluster sizes are given. The data for size three are taken as the basis for standardization, which means that all displayed values have been divided by the respective values for size

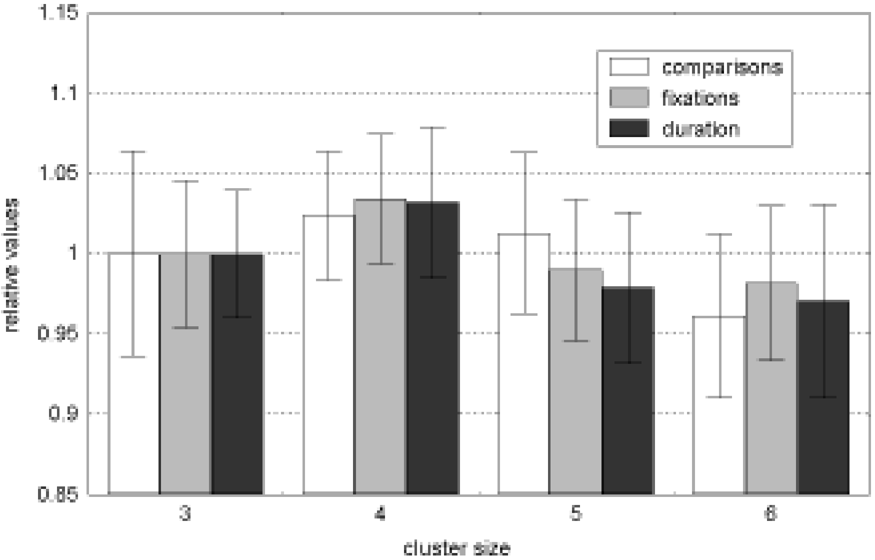


Figure 6.12: Relative values of comparisons (NC), fixations (NF), and duration (DP) for cluster sizes three to six with size three serving as the basis for standardization

three. We find a tendency of NC, NF, and DP to increase from size three to four, to decrease from size four to five, and to further decrease from size five to six. Since these three variables are measures of *inefficiency*, i.e. since they decrease with higher search efficiency, they should behave inversely proportional to SP and AC. Indeed, Figure 6.12 seems to confirm this hypothesis, but the differences are rather weak and far from being significant, as the large errorbars indicate. Therefore, the SP and AC effects are not likely to be caused by within-cluster processes alone, but between-cluster factors seem to play an important part as well.

6.3.4 Entropy and Working Memory

In the previous sections we have extensively discussed the effects of cluster size, i.e. of the *number* of items, on the subjects' eye movements and the utilization of working memory. Yet, the *identity* of items should influence the utilization of working memory as well. We do not expect a group of items with mixed colors and/or forms (high entropy) to be memorized as easily as the same number of identical items (low entropy). In Experiments A to E, our entropy considerations were based on local entropy values in the regions being fixated, but we did not know *which* of the items displayed there were currently being processed. This was a reasonable handicap, because the notion of entropy is meaningful only with regard to item *combinations* and not with regard to single items. To give an extreme example: If subjects held only one item in memory at a time, the local entropy would have no effect on their scanpaths at all.

The fact that subjects used a cluster-by-cluster strategy of comparative search in Experiment F enables us to precisely investigate the influence of color and form entropy on working memory. Our way of within-cluster analysis – as applied in the previous section – allows to take a look at scanpath sections that correspond to the processing of a known subset of items. Hence, we can define a suitable measure of entropy within an item set and study the effects of entropy on the variables NC, NF, and DP. These variables are expected to increase with entropy due to the higher effort needed for processing. The question has to be asked how entropy should be defined here.

In any case we should retain the distinction between color and form entropy, because these two dimensions are likely to exert different effects on the subjects' performance. Since clusters consist of only a small number of discrete items, it is not justified to apply a complicated formula as (2.10) given in Section 2.4.1. The interpretation of results based on such a definition of entropy would be unnecessarily difficult. Here is an easier way to arrive at reasonable entropy values: Given a cluster of three items, for example, only *three* different levels of color entropy are possible:

Level 1: All items are of the same color.

Level 2: The cluster contains two different colors, which goes along with two items of color A and one item of color B.

Level 3: There are three different colors, resulting in all items having distinct colors.

Unquestionably, this is the simplest and most transparent definition of entropy with respect to groups of three items. Though the scale of measurement is ordinal in nature, it enables us to analyze the effects of color and form entropy (with three levels each) on the within-cluster variables at cluster size three. Figure 6.13 illustrates the obtained results with the help of a two-dimensional column diagram of NC. Obviously, our hypothesis is confirmed: Higher entropy seems to cause increasing NC. This effect is especially strong between the simultaneous minimum and maximum of color and form entropy (1.000 versus 2.548 comparisons). The results were entered into a two-factorial (color and form entropy) analysis of variance. Since a within-subjects design would have had several empty cells, all 1112 measured NC values for clusters of size three were treated as independent samples. Though this way of analyzing data does not match the actual design, it is conservative in that error variances are overestimated; thus, it is a practicable way to proceed. The analysis revealed a significant main effect of color entropy ($F(2; 1103) = 3.27; p = 0.038$). The calculation of contrasts between different cluster sizes exhibited that NC is significantly higher at level three of color entropy (2.102) than at level one (1.599) ($F(1; 1103) = 4.93; p = 0.027$), and neither of these values shows a reliable deviation from level two (1.909). The main effect of form entropy reaches significance as well ($F(2; 1103) = 7.48; p < 0.001$). Here, NC is significantly higher at level two (1.955) than at level one (1.429)

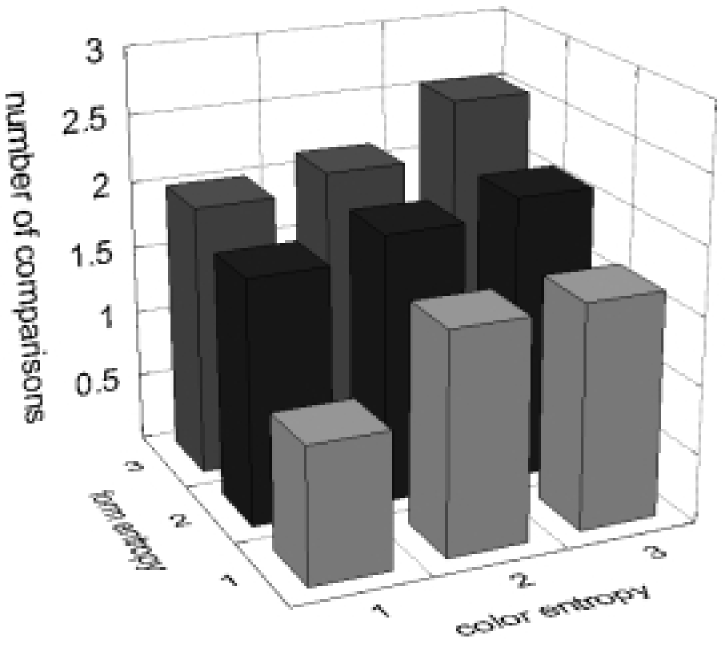


Figure 6.13: Number of comparisons in clusters of size three as a function of color and form entropy

($F(1; 1103) = 5.90; p = 0.015$). Level three (2.227) induces higher NC than level one ($F(1; 1103) = 12.66; p < 0.001$) and level two ($F(1; 1103) = 6.46; p = 0.011$). No reliable interaction between color and form entropy was found.

All in all, form entropy exerts a strong effect on NC. The NC values increase significantly with form entropy. There is an effect of color entropy as well, however, it is somewhat less pronounced.

The situation becomes more complicated if we want to study clusters of *four* items. On the one hand, we can still define, for instance, color entropy as the number of different colors that appear in the cluster, which results in three possible entropy levels again. Accordingly, the analysis of entropy effects on NC has the same design as the previous one. With regard to color entropy, NC has the value 2.273 at level one, 2.439 at level two, and 2.688 at level three. Form entropy causes NC values of 2.565 at level one, of 2.447 at level two, and of 2.501 at level three. Neither the main effects nor the interaction reach significance.

On the other hand, we could argue that there are actually *four* levels of entropy in clusters of size four, because level two comprises two sub-levels: The two colors can be distributed either equally (two items each of colors A and B) or unequally (three items of color A, one item of color B). Nevertheless, the NC difference between these sub-levels is not significant.

Both definitions of entropy ("three levels" and "all possible levels" respectively) were applied to analyze clusters of five and six items in the same way. In none of these cases any entropy effects were found.

A plausible interpretation of these findings is that in clusters containing more than three items, subjects are unlikely to memorize all items at the same time. Therefore, the subjects' behavior is determined by the entropy within the chosen subsets of items rather than by the overall entropy within the cluster. Our interpretation of the SP and AC dependence on cluster size is clearly compatible with this view: Three items can be processed as one unit, but larger groups have to be split up.

6.4 Conclusions

The results obtained in Experiment F have provided answers to the questions posed at the beginning of this chapter. First of all, the visualization of accumulated scanpaths can be viewed to be compatible with a "travelling salesman" strategy: Subjects preferred cluster-to-cluster scanpaths of minimized length. As suggested by visual inspection, these paths appeared to be even less variable within the same subject than between subjects. Interestingly, the fixations measured in within-subject scanpaths had almost constant centers of gravity for each cluster, indicating an invariant distribution of attention. On the level of single fixations, however, no correspondence between locations was discovered. This result suggests that, during local inspection of the display, eye movements and shifts of attention are only loosely coupled with each other. In Chapter 11 this point is taken up in the context of scanpath modelling.

Furthermore, there was evidence as to the number of items being held in memory at a time. The "magical number" three is compatible with an earlier estimation in Experiment A, suggesting a range from two to four items. The specific value, three, was not only indicated by the patterns of SP and AC as functions of cluster size, but also by the cluster-size dependence of entropy effects. The processing of three items per cluster is strongly influenced by color and form entropy, which indicates that all three items are memorized at the same time. Clusters of size four to six do not exhibit significant influences of entropy, which suggests that they are processed in multiple, sequential steps.

Apart from entropy, the *quantity* of items exerts significant effects on the number of comparisons, the number of fixations, and the duration of processing. These three variables vary proportionally with each other, suggesting that some local, low-level characteristics of gaze trajectories are not biased by the additional line structures. Other findings support this point of view:

- The within-cluster variables increase proportionally with the number of items per cluster. Therefore, no specific cluster size seems to facilitate comparative search *within* a cluster.

- The basic eye-movement variables fixation duration and saccade length are not influenced by cluster size.

- The results of Experiment A are replicated under the condition of cluster size one in Experiment F. Thus, we can assume that the basic search behavior remains unchanged, regardless of the connecting edges appearing in most of the stimuli.

Based on these observations, we can assume our cluster stimuli to be well-chosen for the purpose of Experiment F: The subjects "accept" the pre-defined cluster structure as a perceptual guideline for task completion without changing their low-level mechanisms of comparative visual search.

Altogether, the novel stimuli used in Experiment F enable us to discover facts beyond the range of the basis scenario. Here, the augmentation of search pictures turned out to be highly fruitful. Experiment G, described in the following Chapter 7, extends the paradigm in yet a different direction, namely into the *third dimension*.

Chapter 7

Experiment G: Introducing the Third Dimension

7.1 Using Three-Dimensional Stimuli

Although the task of comparative visual search has ecological relevance (see Chapter 1), the *stimuli* used in Experiments A to F are far from being natural. In everyday life, we are not likely to be confronted with sets of two-dimensional, abstract, geometrical items. Our environment and the objects we perceive and manipulate are rather three-dimensional. These objects have specific meaning and are characterized by their materials, textures, functions etc. Being aware of this discrepancy between experimental and natural conditions, the question has to be posed whether we can design a reasonable variant of the basis scenario that gets us closer to an ecologically plausible situation.

As shown in Chapter 2, the choice of artificial objects as elementary search items was well-considered. The high-level factors involved in the perception of realistic things and scenes would have caused problems in the analysis and interpretation of the empirical data. For the same reason, most experiments in visual search (see Section 1.4) are based on abstract items rather than on realistic ones.

However, another aspect of ecological inadequacy can be overcome without facing these difficulties and without giving up the comparability to previous experiments. It is the restriction to two dimensions. Three-dimensional scenes require perception on a higher level as well, but since this enhancement is of a purely geometrical nature, semantic factors as knowledge or interpretation do not play an important part. Consequently, the most fruitful strategy is to stick to artificial items that should be as similar to the ones used in Experiments A to F as possible, and to embed them into a three-dimensional scene. There are many possible ways of realizing this idea which deserve to be taken into consideration.

The technically simplest method consists in using patterns of two-dimensional stimuli that are *interpreted* as three-dimensional arrangements. This is what we automatically do when viewing photographs or drawings: We do not perceive them as planar distributions of colors or lines, but as spatial objects. Perspective, texture gradients, occlusions, and many other cues help us to perceptually reconstruct the inherent three-dimensional information. Various experiments of visual search investigating this effect have been conducted (for a review, see Grossberg, Mingolla & Ross, 1994). Enns & Rensink (1990), for instance, used images of cubes as items with the target item having a different shading than the distractor items, indicating that our judgement of lighting is affected by a three-dimensional interpretation. In several control experiments, the same shadings were used on items which could only be interpreted as two-dimensional. It was found that only the first experiment enabled parallel search, whereas the others required serial search (for the distinction between parallel and serial search, see Section 1.4). This result shows that three-dimensional perception can enhance efficiency in visual search, even if the stimuli are in fact two-dimensional. Moreover, it falsifies the wide-spread assumption (e.g. Treisman & Gelade, 1980) that visual search is exclusively controlled by feature-coding processes on a low perceptual level.

The findings mentioned above clearly demonstrate that the introduction of the third dimension considerably extends the horizon of visual search experiments. It motivates analogous studies in *comparative visual search*. However, the presentation of shaded two-dimensional stimuli is still not yet ecologically adequate, because it neglects an essential perceptual factor, namely “binocular disparity”. When we look at a certain point in three-dimensional space, we direct the two visual axes of our eyes to this point in such a way that its image is projected to corresponding locations on both retinae. This projection on corresponding locations is called zero binocular disparity. The projected images of items in front or behind this point cause crossed or uncrossed binocular disparity respectively on the retinae. Binocular disparity is the dominant indicator of relative spatial depth for the visual system. Items that are much nearer or much more distant than the fixation point such that they cause disparity of more than approximately one degree (lying outside the so-called *Panum's fusional area*, see e.g. Schor & Ciuffreda, 1983) cannot be fused. They are seen as doubled images, but their depth is still perceived due to their binocular disparity.

Since the EyeLink system is capable of recording binocular eye movements, it seems to be especially fruitful to use “genuine” three-dimensional stimuli involving binocular disparity. The measurement of the *vergence angle,* i.e. the angle between the visual axes, enables us to investigate three-dimensional gaze trajectories during comparative visual search. But how can spatial stimuli be presented on a standard computer monitor?

More specifically, the question should be: How can each of a subject's eyes get individual information from one and the same screen, inducing different levels of binocular disparity? There are four basic techniques for trying to solve this question:

Stereo image-pairs: Two images are presented side by side on the screen. Subjects have to fixate an (imaginary) point in front or behind the center of the screen, adjusting their visual axes to different images by crossed or parallel view respectively. After binocular fusion, a stable 3D image is perceived which does not require a specific point of fixation anymore.

Auto-stereograms: This is the famous type of "magic eye" images. The principle is identical to the stereo image-pairs with the difference that in auto-stereograms the information for both eyes is combined in a *single* image. Usually, subjects perceive a three-dimensional, textured surface.

Shutter glasses: Subjects wear special glasses that alternately "shut" (darken) their left or right side, which is synchronized with the monitor frame rate. Between frames, the monitor display changes from the right-eye to the left-eye information or vice versa, so that each eye views its individual scene exclusively. If the frame rate is sufficiently high, both scenes can be fusioned to a stable, spatial image.

Red-green images (anaglyphs): Here, the two eye-specific images have to be monochrome. They are converted in such a way that one of them shows varying shades of red and the other one varying shades of green. Afterwards, they are superimposed on the monitor screen. The subjects wear a pair of glasses with the left glass colored red and the right glass colored green. The glasses serve as filters which allow the left eye to perceive only differences in the red value, whereas the right eye is solely sensitive to green values. Thus, the two original images are separately shown to the eyes and perceptually combined to a monochrome, three-dimensional scene.

There are not many experiments in visual search that employ one of these methods. This may be due to the increased complexity of conduction and evaluation of such studies. He & Nakayama (1992) performed an experiment which is comparable to the abovementioned study by Enns & Rensink (1990) to some extent, but it was technically based on shutter glasses instead of two-dimensional images. The authors used white L-shaped target items and reversed Ls as distractors (or vice versa), which were accompanied by black squares. In one condition, the squares were shown behind the Ls, in another condition they appeared in front of them. In the second condition, the Ls were perceived as white *squares* being partially occluded by the black ones. According to this three-dimensional interpretation, the difference between target and distractor items is perceived as a horizontal shift of the white square rather than the reversal of an L-shape. The second condition leads to longer reaction times than the first one, demonstrating the impact of three-dimensional perception on visual search.

Actual three-dimensional search instead of three-dimensional item features was studied by Nakayama & Silverman (1986). Using the technique of stereo image-pairs, they simultaneously presented items distributed on two different depth planes. The distractor items had specific colors (or directions of movement) for each plane, while the target item was of a different color (or different direction). It was shown that conjunctive search for the combinations of depth and color as well as depth and motion can be performed in parallel. A similar

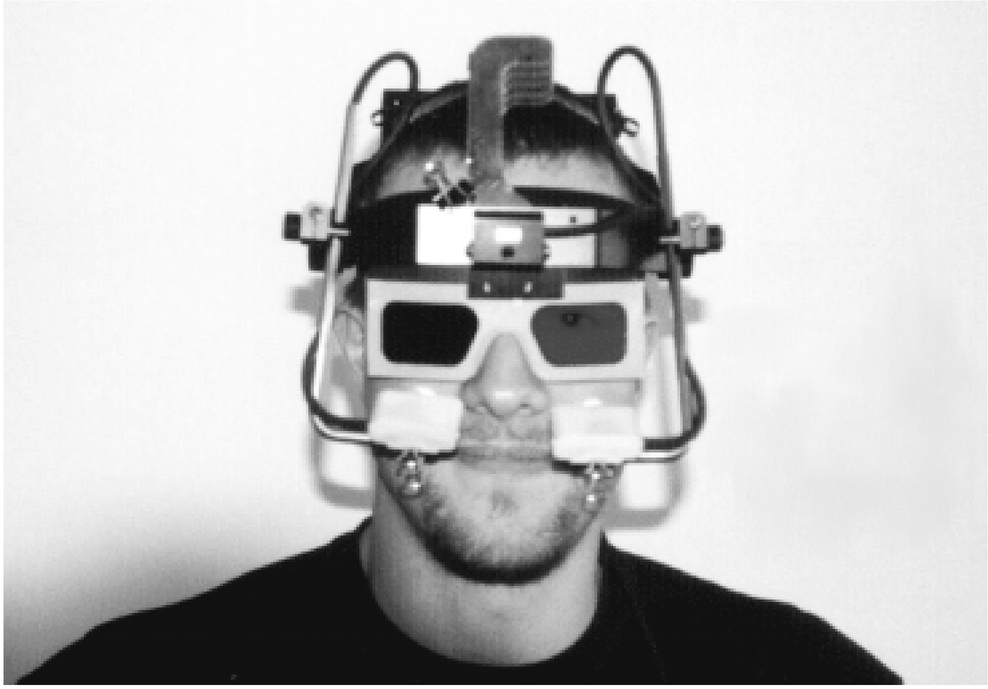


Figure 7.1: Subject simultaneously wearing the EyeLink head-set and red-green glasses

study by Steinman (1987) that used vertical lines basically confirmed these results. These findings contradict the classical assumption that the detection of feature combinations requires serial search (e.g. Treisman & Gelade, 1980), and hence suggest that stereoscopic depth plays a special role in visual perception.

O'Toole & Walker (1997) have recently argued that the experiments of Nakayama & Silverman (1986) and Steinman (1987) were inappropriate, because their targets could be detected by *monocular* cues as well. In order to achieve improvement, they used random-dot stereograms to verify those results and found conditions under which stereoscopic disparity requires serial search. In their opinion, the relative disparity of target and distractors, the size of the stimulus display, and the global surface context are the factors which determine whether parallel processing is feasible.

Since its effects on visual search are not completely known, stereoscopic depth is an attractive feature to be investigated. The presentation of items in different depth planes seems to be perfectly transferable to our scenario of comparative visual search and promises interesting results. However, before designing the experiment – which we will call Experiment G – we should clarify its technical realization: What is the appropriate method of three-dimensional stimulus display?

The technique of stereo image-pairs has the advantage not to require special glasses, but it involves unnatural vergence of the eyes and is thus inadequate for the measurement of vergence eye-movements. Auto-stereograms are not suitable either, because they do not allow for the presentation of discrete items. From a psychological point of view, the use of shutter glasses would be ideal, but their technical integration into the eye-tracker system would be too complicated. The infra-red interface for shutter-monitor synchronization interferes with the infra-red alignment system of the eye tracker and it is difficult to combine the special glasses with the eye cameras. At least with regard to the intended exploration of three-dimensional situations, the efforts a practical realization would take would not be justified.

In face of this, we decided to use red-green glasses. Figure 7.1 shows our combination of the EyeLink head-set with adjustable red-green glasses. Fortunately, the infra-red eye cameras and the pupil-tracking hardware can find the subject's pupils through the colored glasses without any problems.

As mentioned above, one drawback of the chosen method is the restriction to monochrome stimuli. This means that an exact replication of the basic experiments A to C which employs three-dimensional distributions of the same "old" items is not feasible. We could use three different levels of brightness instead of three different colors, but due to the heterogeneous color filters, the subjects' judgement of brightness could be impaired and could vary individually. Another type of analogy to the previous experiments is more reasonable: We replace the three levels of color with three levels of *depth*. Consequently, each stimulus presents ten items in a back plane, ten items in a middle plane, and ten items in a front plane. The mismatch then consists in either the form or in the depth of a single item.

An example stimulus is shown in Figure 7.2 on page 133. The items in each hemifield consist of a red and a green component with black areas of overlap. The yellow background color has the same intensity of red or green respectively as these components. Through the red glass, subjects see the green components as dark, because their red value is zero. The red components are invisible, because they present the same red value as the background and are thus perceived in the same brightness. Vice versa, only the information on red components passes the green glass. The black areas are seen by both eyes. As a consequence, subjects perceive three virtual planes of dark items on a light background. *Light* items on *dark* background would be more similar to the situation in Experiments A to F. Due to the characteristics of the red-green glasses and the monitor, however, their presentation cannot be realized as perfectly as the chosen variant.

The virtual planes are equidistant with the distance between back and front plane being the same as the distance between the top and the bottom of the stimuli. There are no effects of perspective, hence the size of the items – as seen without red-green glasses – does not vary with their depth plane. This arrangement is important, because otherwise we could not decide whether depth mismatches are detected by their difference in depth or by their difference in size. In three-dimensional perception, the items in the back seem to be slightly larger than those in the front, but this effect is small and does not impair the overall stability of perception.

What results can be expected? It could be possible that subjects tend to scan the display using a plane-by-plane strategy in order to decrease the number of necessary vergence eye-movements. We can take a look at the three-dimensional gaze trajectories to figure out the strategies that are applied. Another indicator could be speed of processing (SP): If planewise scanning is performed, SP measures the mean gaze speed for processing *one plane*, i.e. ten items instead of thirty. Hence, extraordinarily high SP values would signify that plane-by-plane trajectories play an important part in Experiment G.

In Experiment G, the detection of depth mismatches should require shorter reaction time (RT) and lead to higher values of saccade length (SL), area coverage (AC), and speed of processing (SP) than the detection of any mismatches in the "flat" scenario. This assumption is based on the studies mentioned above which proposed special properties of binocular disparity to enable parallel search. In the case of comparative visual search, parallel search can only be *locally* parallel, but nevertheless it should result in higher search efficiency. In order to detect further differences in eye-movement parameters between the scanning of two- and three-dimensional stimuli, the other eye-movement variables, namely FD, SL, and FW, will be analyzed as well.

Another expectation is that the detection of mismatches in the back plane should be more difficult than in the front plane, because the perception of the back plane is impaired by the two other planes being located in front of it.

Finally, vergence movements are assumed to be weaker during the search for a depth difference than during the search for a form difference, since the perception of depth is not restricted to Panum's fusional area (see above), whereas form recognition is because it requires precise inspection. To investigate the differences between depth and form search in Experiment G, subjects should be informed about the type of mismatch in advance.

7.2 Method

7.2.1 Subjects

Sixteen students from different faculties at the University of Bielefeld were recruited as subjects for this experiment. They were paid 7 DM for their participation. All of them had normal or corrected-to-normal vision; no pupil anomaly or color blindness was found.

7.2.2 Materials

The stimuli were shown on a computer screen with a resolution of 1024×768 pixels. They consisted of distributions of simple geometrical items (triangles, squares, and circles). Each item was represented by a red and a green component on the same vertical position but with a horizontal shift between them. The background color was yellow, while the areas of overlapping red and green components were black-colored. The subjects wore red-green glasses and were placed in front of the monitor with an eye-screen distance of 50 cm. They perceived the items as being located on one of three different virtual planes parallel to the surface of the screen: One plane 11 cm *behind* the screen, another *on* the screen, and a third one 11 cm *in front of* it.

The size of the items was about 24 pixels in diameter. Their locations were pseudo-randomly chosen under the condition of not touching or overlapping each other. The random distribution slightly tended to create regions of similar item forms and depths in order to allow the investigation of eye movements in homogeneous versus heterogeneous areas (see Section 2.3).

Each stimulus picture presented two hemifields separated by a vertical black line and containing 30 items each (see Figure 7.2). These rectangular hemifields were 22 cm high and 16 cm wide with a horizontal gap of 4 cm in between. The items in each hemifield were equally balanced for form and depth. Except for their translation, the hemifields were identical with respect to the form and the spatial distribution of the 30 items – with exactly one exception: One mismatch between corresponding items was integrated, consisting in either their form or their depth.

7.2.3 Apparatus

While the subjects solved their task, their eye movements were measured with the SMI EyeLink system, which is a video-based binocular eye tracker (cf. Section 2.1.3). They wore a head-set with two miniature infrared eye cameras fastened to it, yielding eye movement data without imposing severe restrictions on the subjects' field of view or their head movements. The stimuli were presented on a 20" Sony Trinitron monitor.

For each eye, the time of occurrence, the duration, and the screen coordinates of the subjects' fixations were recorded. The absolute spatial precision of the system lay within 0.7 to 0.9 degrees of visual angle corresponding to about 0.6 to 0.8 cm on the screen. Due to the high frame rate of the infrared cameras, the temporal resolution of the measured data was 4 ms.

Preceding each experiment, a calibration procedure had to be carried out, demanding the subject to fixate a target that appeared in different positions on the screen. Before the presentation of each stimulus, a drift correction using a single target in the center of the screen was performed.

7.2.4 Procedure

All subjects were tested individually. Their task consisted in detecting the mismatch between the two hemifields that were simultaneously presented on the computer screen. They had to report target detection by pressing any button on a computer mouse being placed in front of them.

Each subject viewed 60 pictures which were newly generated for every subject and thus differing between trials as well as between subjects. Either the first 30 or the last 30 pictures being shown (counterbalanced between subjects) contained a depth mismatch, whereas the remaining 30 presented a form mismatch. The subjects always knew the type of mismatch they had to search for in advance.

7.3 Results and Discussion

7.3.1 Characteristics of Three-Dimensional Scanpaths

As in the previous experiments, the gaze position is measured as the point where the visual axis intersects the monitor screen. The only innovation is that there are *two* gaze positions at a time now, one for each eye. In general, the vertical coordinates of these two positions do not differ from each other, while the shift between their horizontal locations is variable and indicates the subject's *vergence angle*. This is the angle between the two visual axes, which is usually adjusted by means of *vergence eye-movements* if attention is shifted between points of different distances from the observer. Vergence movements and *conjugate movements*, i.e. binocular saccades within a depth plane, are controlled in completely different ways. Vergence movements are slow (4–20^o/s) and can be altered or stopped in response to stimulus changes, while saccades are fast (200–700^o/s) and unchangeably "programmed" before their realization (for detailed information about vergence eye-movements, see Schor & Ciuffreda, 1983).

Furthermore, the horizontal shift between the measured gaze positions allows to geometrically calculate the depth of the actually inspected point in space. If the left eye's horizontal gaze position is measured to be right of the right eye's gaze position, this point is located in front of the screen, in the opposite case it is behind the screen. Figure 7.3 attempts to visualize a pair of scanpaths for the example stimulus. This is obviously more problematic than in the monocular case. Hence, no durations or numbers are displayed, while the saccades are marked with thick lines. The green lines signify saccades of the left eye, the red ones those of the right eye. These shades correspond to the components of items: The left eye perceives only the green components, the right eye only the red ones. Thus, viewing Figure 7.3 through red-green glasses, the reader will see a *three-dimensional* scanpath within a *three-dimensional* stimulus, which is impressive and illustrative at the same time.

Nevertheless, even without the use of red-green glasses the image provides some preliminary information about the gaze trajectory and its relation to the underlying stimulus. The subject starts his search for a form mismatch exactly from the center of the screen (where the target for drift correction has been shown immediately before stimulus onset). After an orientation phase he begins his comparative search at the top of the stimulus with most of the items being located in the middle plane, i.e. on the screen level. Accordingly, there is no shift between the gaze positions of his left and his right eye. The subject proceeds to the front plane, indicated by the red components being displayed to the left of the green ones.

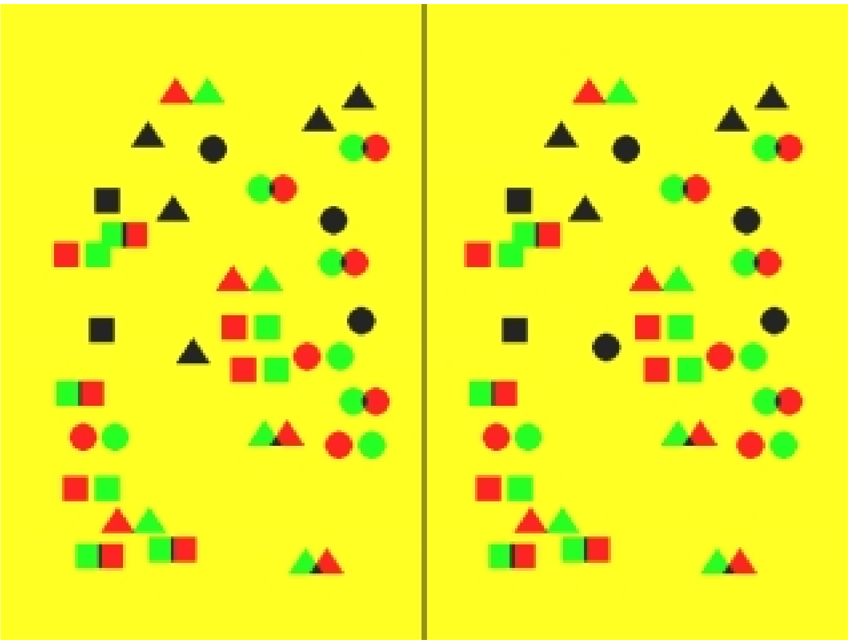


Figure 7.2: Example of a virtual three-dimensional stimulus. Viewing this image through red-green glasses leads to its spatial perception.

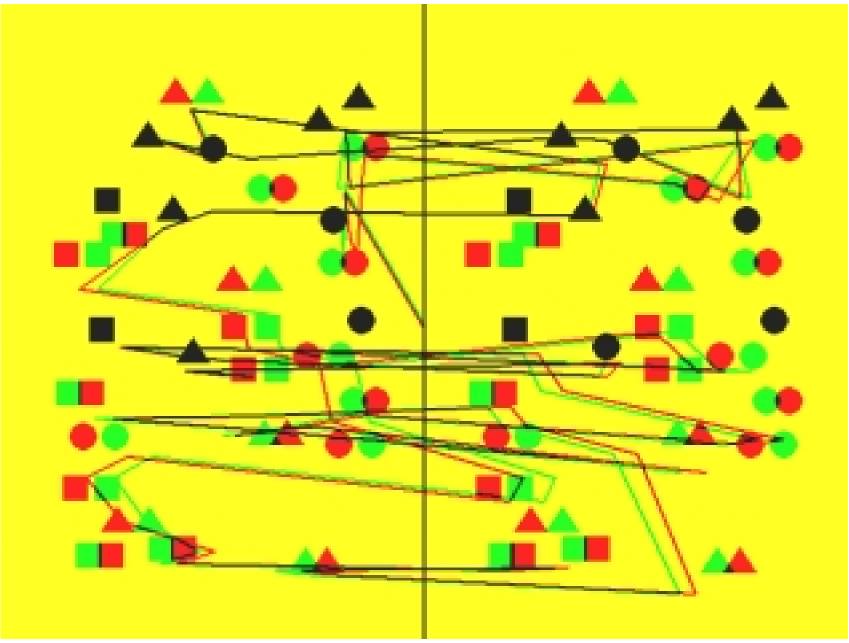


Figure 7.3: A subject's binocular scanpath in the example stimulus. The green lines signify saccades of the left eye, the red ones saccades of the right eye. They are directed at the green or red objects respectively.

In fact, we can observe the subject adjusting his vergence to the item depth, which is especially obvious in the lower left corner of the image. The horizontal shift between the two gaze positions on the screen is identical to the horizontal shift between complementary item components. At this point, the subject changes to the back plane and inverts the shift between his gaze positions with respect to the screen level. Finally, he detects the form mismatch located on the middle plane at the center of each hemifield.

This illustration demonstrates that, in fact, binocular eye tracking allows to record three-dimensional scanpaths. However, it must be stated that the explicitness of the data chosen for this illustration is clearly above average. Errors in measurement as well as incomplete or inaccurate vergence movements during comparative search lead to considerable "background noise", as we will see below.

7.3.2 Baseline Data of Vergence Eye-Movements

Before proceeding to the analysis of vergence eye-movements during comparative visual search, some of their basic features with regard to the chosen scenario have been investigated. During task completion, a subject's vergence movements are likely to be interrupted or inhibited by conjugate movements, i.e. by binocular saccades, in order to optimize search efficiency. These effects can be disentangled only if there are baseline data for unbiased vergence movements. Therefore, in an informal pre-study, the gaze trajectories of two additional subjects were recorded. They were *not* instructed to search for a mismatch, but rather to inspect items on certain depth planes. During each trial, they were told several times to direct their attention to a specified plane. One of the subjects was used to viewing red-green images and thus had a very stable perception of the three-dimensional stimuli. The other subject was unfamiliar with pictures of this kind, but after some practice trials she gained an almost stable perception as well. Only immediately after stimulus onset, it always took her about one or two seconds to stabilize her perceived image.

Figure 7.4 presents the experienced subject's data. The diagram shows the spatial depth of inspected items (dotted line) and the corresponding gaze depth (solid line) for each of his fixations recorded during one trial. Although the horizontal axis refers to the number of the fixation instead of the time, the diagram can be viewed as the time course of depth values, since no systematic temporal distortion occurs. The spatial depth of a fixation is measured as the depth of the nearest neighboring item to the fixation point, while the gaze depth is geometrically computed (intersection of visual axes). Increasing depth values indicate shorter distances to the subject; the front plane is located at 11 cm, the middle plane at 0 cm, and the back plane at -11 cm relative to the screen plane. In order to compensate for the background noise mentioned above, both item and gaze depth are processed by a median filter of window size five. The results show a good temporal correspondence between these two variables, and additionally an interesting effect: Vergence movements from a near plane to a more distant one (divergence) are clearly slower than those in the opposite direction (conver-

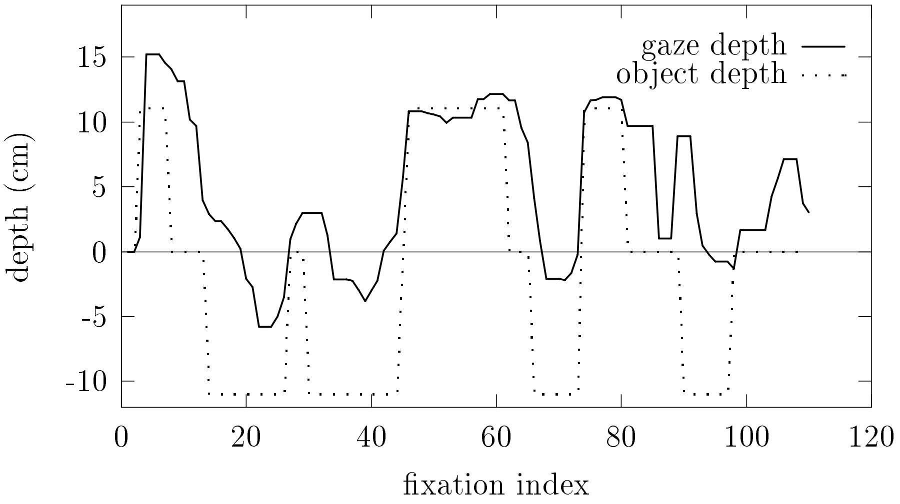


Figure 7.4: Time course of gaze depth during the inspection of different depth planes by an experienced subject

gence), preventing the subject from reaching the optimal vergence angle for the perception of the back plane. This difference in vergence velocity is well-known (e.g. Yarbus, 1967), but it still lacks an indisputable physiological explanation. In earlier times, scientists used to assume that the eye muscles were innervated only in case of convergence, and that they were just “released” for divergence, since the physiological resting position of the eyes is even more divergent than parallel view. Electrophysiological data have falsified this assumption, however. From an evolutionary point of view, the effect is plausible: Approaching items are potentially more dangerous than vanishing ones, which means that it is more important to focus attention on the former as soon as possible.

In the case of the unexperienced subject, the discrepancy between converging and diverging eye movements is even stronger (see Figure 7.5). She starts in the front plane, then moves to the middle plane, and finally inspects the back plane. The corresponding vergence movements are extremely slow and do not reach the optimal vergence angle for the back plane, although it is viewed for more than ten seconds. As shown at the beginning of this trial, the converging movements are still rather fast; they even tend to “overshoot” the respective target distance. Apart from physiological aspects, we have to consider another factor that may contribute to the extremely low velocity of divergent movements, namely the perceptual interference with those planes lying in front of the actual fixation point in space. Results from similar experiments in our group demonstrate that a clear lateral separation of depth planes enables even unexperienced subjects to produce vergence movements as shown in Figure 7.4.

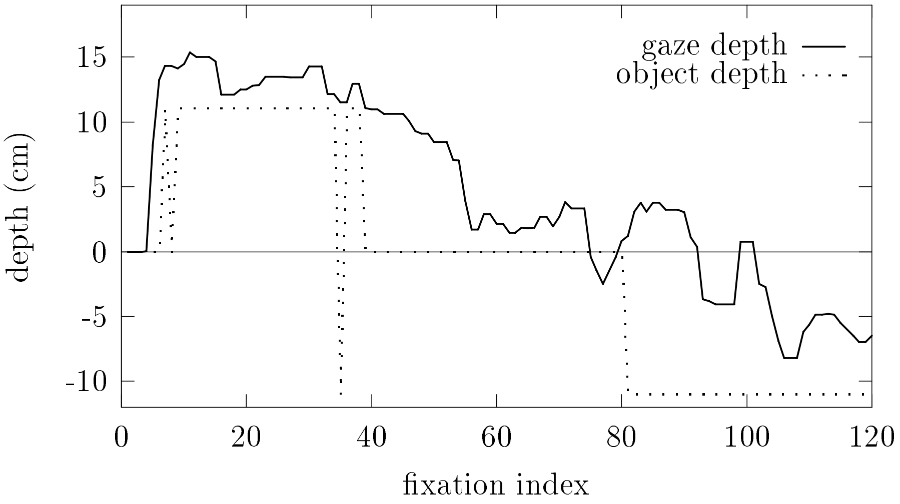


Figure 7.5: Time course of gaze depth during the successive inspection of the front, middle, and back plane by an unexperienced subject

7.3.3 3D Scanpaths in Comparative Visual Search

Now we are prepared to take a closer look at the time course of gaze depth during comparative search, as illustrated in Figure 7.3. The example presented in Figure 7.6 is compatible with our assumption: During rapid sequences of transitions between different depth planes, vergence movements are not conducted completely. Although the subject is searching for a form mismatch, task completion does not seem to require exact vergence. Another observable effect is the "hyper"-convergence while looking at items in the front plane that cannot be perfectly explained. The finding of overshooting convergent movements mentioned above may contribute to the hyper-convergence, and possibly subjects like to sustain this state, because it reduces the visibility of items in the other planes. Thus, the perceptual interference with these planes decreases, which may result in a higher search efficiency. A further reason could consist in systematic errors in gaze position measurement, because the eye tracker is still calibrated using targets on the two-dimensional monitor screen. If the computation of fixation points works correctly, however, which is very probable, the deviation between measured and applied vergence should be rather small.

The sample binocular scanpath (Figure 7.3) suggests that the third dimension increases the variability of scanpaths, since the subject does not proceed from the top of the display to its bottom or vice versa, but rather applies a plane-to-plane strategy: Starting on the middle plane, he goes on to the front plane, briefly inspects the back plane, until he detects the target in the middle plane again. Obviously, the z-axis seems to replace the y-axis as the most important structuring standard. The main advantage of a planewise strategy is that it reduces the

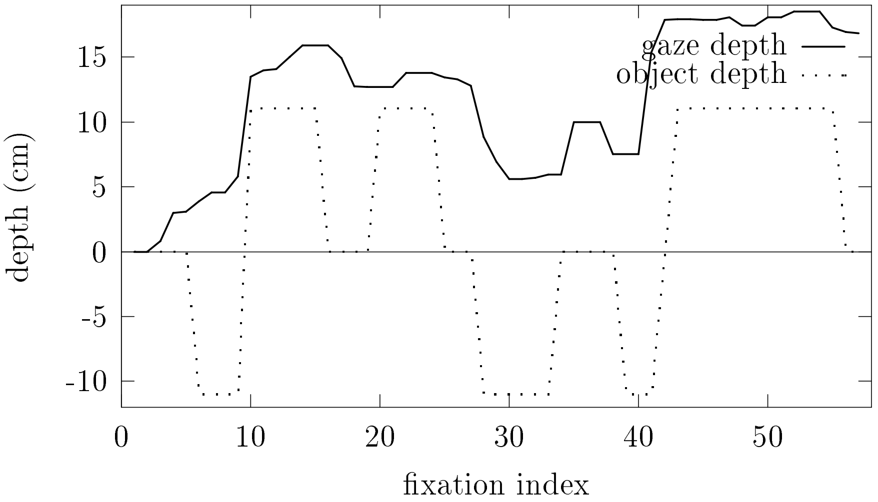


Figure 7.6: Example of one subject's time course of gaze depth during the search for a form mismatch

perceptual effort for the visual system, because fewer changes between different depth planes have to be performed. As mentioned above, gaze shifts along the z-axis are slower than those along the x- and y-axis, since vergence movements have to be employed instead of saccades. The sequential scanning of all items in the same depth plane is likely to allow a better adjustment of vergence and hence an improved discriminability of items.

How can we investigate individual preferences of certain scanpath strategies? One illustrative approach consists in accumulating all time courses of gaze and item depth recorded for the same subject. For each number of fixation, these values are represented by their arithmetic mean. Since we expect form search to induce a more precise adjustment of vergence than depth search, only the 30 trials with form mismatches are considered. Figure 7.7 shows one of the resulting diagrams with the fixation index ranging from one to 59, because fewer than five scanpaths contain more than 59 fixations and the subsequent data are thus not reliable. It can clearly be seen that the subject tends to begin her search in the front plane, to move to the middle plane afterwards, and finally – if necessary – to scan the back plane. Furthermore, the slow vergence movements from nearer to more distant planes, the interference with planes lying in front of the gaze position, as well as the assumed effect of overshooting convergence lead to a strong statistical deviation of gaze depth from item depth towards the subject.

Did all subjects prefer such front-to-back strategies, or did maybe some of them choose back-to-front schemes? In order to get a quantitative measure of "strategy alignment", we calculated two linear regressions on each subject's time course of item depth: One regression on the 30 trials of depth search and another on the 30 trials of form search. Thus, we obtained 32 straight lines of regression

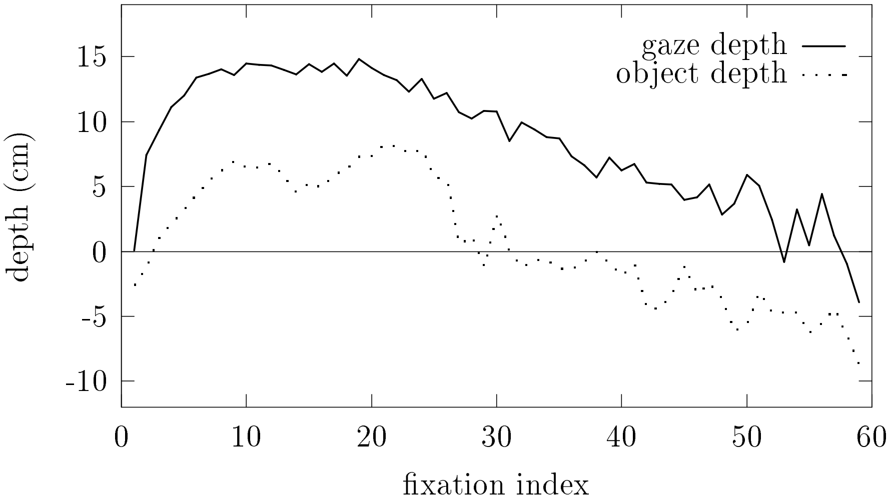


Figure 7.7: Time course of gaze depth averaged over 30 trials of one of the subjects

indicating individual alignments of scanning strategies. A straight line is determined by two parameters: The "depth offset", i.e. the depth value for the first fixation, and the "depth slope", i.e. the depth difference between successive fixations. If a line of regression presents a depth offset which is close to the front plane, for example 10 cm, together with a strong negative depth slope, e.g. -0.5 cm per fixation, we can assume the subject to use a front-to-back scheme almost every time. Vice versa, back-to-front strategies would be signified by a negative offset and a positive slope.

Figure 7.8 shows the distribution of these two-dimensional scanpath parameters for the 16 subjects, separated for depth search and form search. In this diagram, the vertical axis represents the depth offset, while the horizontal axis represents the depth slope. The global shape of the distribution reveals a negative correlation between the two dimensions which can be explained in a plausible way: Since the average depth of inspected items is likely to be close to zero, those lines of regression starting with a positive depth offset tend to have a negative slope, and those with a negative offset tend to have a positive slope.

Taking a closer look at the illustrated distribution, we find that a group of twelve individual parameters (four times depth search, eight times form search) is located near the origin of the diagram, indicating no correlation between fixation index and item depth. The respective strategies are likely to be aligned to the y-axis, just as the scanpaths in Experiments A to F. Also, it could be possible that these subjects switch between front-to-back and back-to-front strategies in such a way that the respective effects neutralize each other; however, an almost perfect neutralization is rather improbable.

Two further clusters of parameters are located in the vicinity of the "cen-

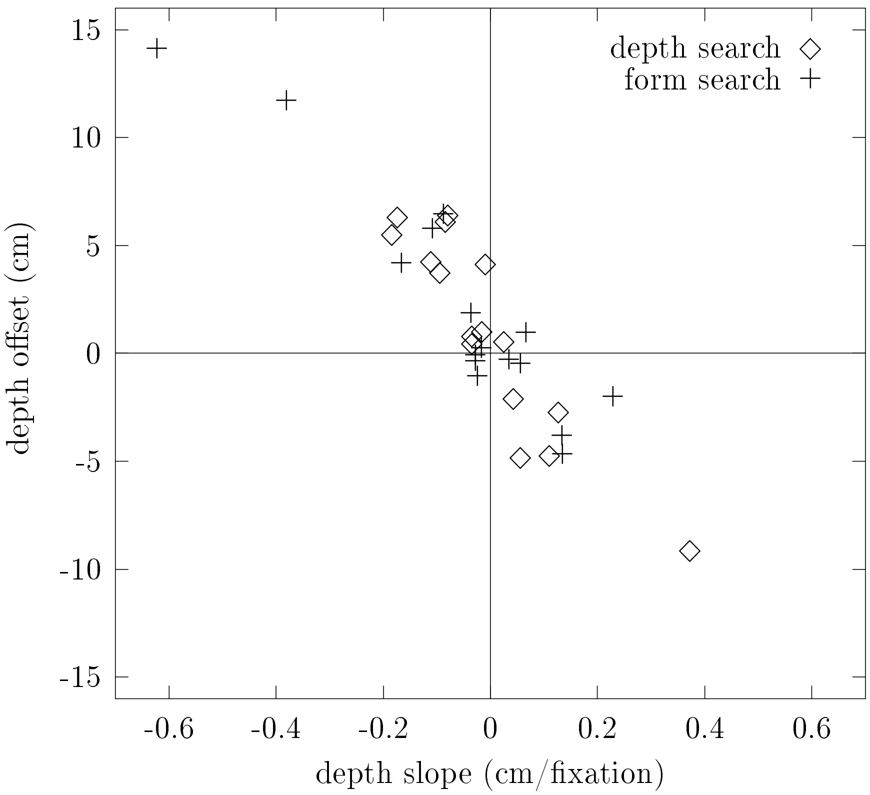


Figure 7.8: Quantitative characterization of the 16 subjects' scanning strategies, as obtained by a linear regression calculated on the time course of item depth in the subjects' scanpaths. The resulting lines of regression are represented by their depth offsets (vertical axis) and by their depth slopes (horizontal axis).

tral" group, suggesting at least tendencies towards front-to-back and back-to-front schemes respectively. Only three lines of regression clearly indicate the dominance of a specific type of strategy: Two of them (form search) signify front-to-back strategies and another one (depth search) signifies a back-to-front strategy.

These findings show that the additional third dimension causes a higher variability of cognitive strategies. Some subjects apply planewise scanning, others do not align their strategies with the depth planes. No significant difference between depth search and form search is indicated so far.

The relationship between gaze depth and item depth during comparative search, which has not been analyzed on a quantitative level yet, might reveal an effect of the type of mismatch. In Section 7.1 we assumed the condition of form search to induce a more precise adjustment of vergence than depth search. This hypothesis can be tested by separately calculating the mean gaze depth for the different viewed planes and the types of mismatch (depth versus form).

Figure 7.9 gives the data which are based on the relation between the me-

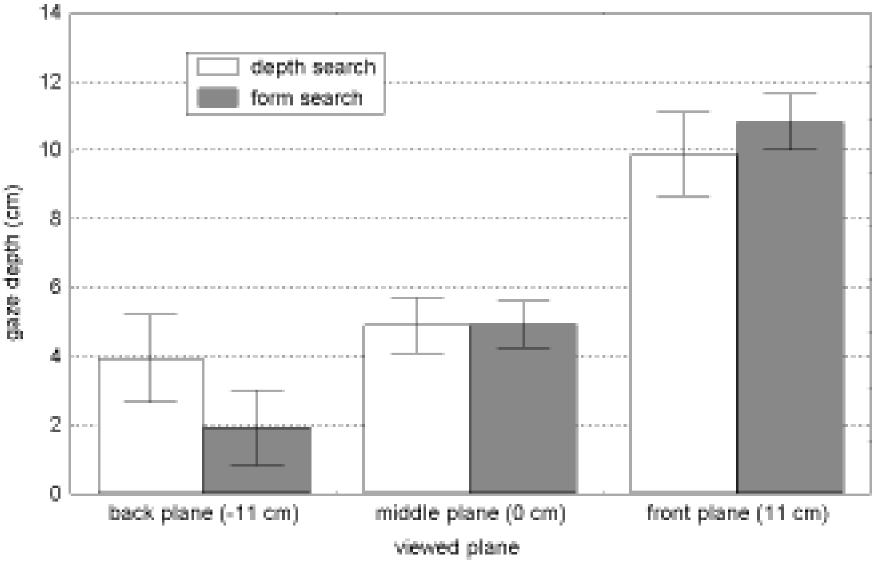


Figure 7.9: Measured depth of the subjects' gaze according to the different depth planes

dian values of gaze and item depth as illustrated above. Concerning the absolute gaze depth, the results are compatible with previous findings, demonstrating a substantial gap between gaze and item depth. Even during the inspection of the back plane, the subjects' vergence does not statistically point behind the monitor screen. Nevertheless, the two-factorial (viewed plane, type of mismatch) analysis of variance revealed a significant main effect of the viewed plane on gaze depth ($F(2;30) = 140.01; p < 0.001$). The analysis of contrasts between different depth planes showed that gaze depth is significantly higher during the examination of the middle plane (4.83 cm) than during the inspection of the back plane (2.96 cm) ($F(1;15) = 17.37$; $p < 0.001$). At the front plane, gaze depth is significantly higher (10.34 cm) than at the middle plane ($F(1;15) = 199.78; p < 0.001$) and at the back plane ($F(1;15) = 229.50; p < 0.001$). The main effect of the type of mismatch does not reach significance, which means that the average gaze depth does not differ between depth search and form search. Interestingly, there is a reliable *interaction* between the two factors ($F(2;30) = 14.25; p < 0.001$), signifying different influences of the viewed plane on the adjusted gaze depth in the conditions of depth search or form search respectively. As shown in the diagram, this difference is relatively small but obvious, and it supports our hypothesis: Form search leads to a better adjustment of vergence to the currently viewed plane than does depth search.

7.3.4 Comparison to the "Flat" Scenario

Since there were always two stimulus dimensions and since the subjects knew the dimension of the mismatch in advance, the results of Experiment G should be compared to those of Experiment B. For a start, we can take a look at reaction time (RT). As expected, RT is significantly shorter for the detection of depth mismatches (9060 ms) than for the detection of form mismatches (11458 ms) ($t(15) = -2.77; p = 0.014$). With respect to form search, RT does not significantly differ between Experiment B (10541 ms) and Experiment G, but we find a reliable difference between color search in Experiment B (7330 ms) and depth search in Experiment G ($t(34) = -2.11; p = 0.042$). This result shows that stereoscopic depth – though it leads to faster detection than form attributes – is not "special" in the sense that it facilitates higher search efficiency than two-dimensional features.

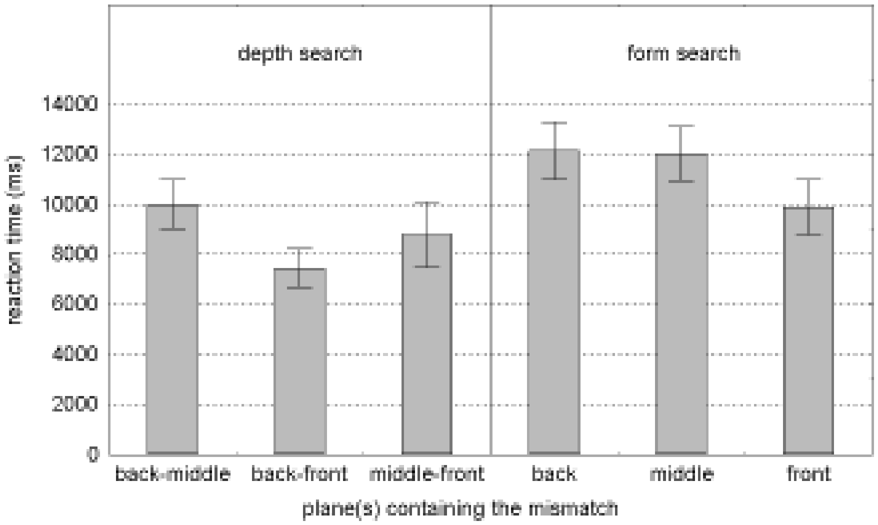


Figure 7.10: Reaction time RT for different depth locations of the mismatch

At this point, we can test our initial hypothesis that mismatches in more distant planes are more difficult to detect. Figure 7.10 presents a diagram of RT values separated for different planes containing the mismatch. The three columns on the left side refer to depth search, where every mismatch involves *two* planes at a time. An analysis of variance reveals that RT is significantly shorter for the detection of depth mismatches between the back and the front plane (7471 ms) than for those between the back and the middle plane (10021 ms) ($F(1;15) = 6.08; p = 0.026$). The RT value with respect to mismatches between the middle and the front plane (8793 ms) is located between the two others, but does not significantly differ from either. Obviously, increasing discrepancies in depth lead to faster target detection. The tendency of back-middle mismatches to induce somewhat longer RT than the equidistant middle-front mismatches – though not significant – can be attributed to the interference with the front plane which possibly "covers" back-middle mismatches.

The right hemifield of Figure 7.10 refers to the detection of form mismatches, which are always located on the same depth plane. Its most striking feature is faster RT for the detection of mismatches in the front plane (9900 ms), but the differences to the RT values for the back plane (12159 ms) and the middle plane (12048 ms) are short of statistical significance ($F(1;15) = 3.69; p = 0.074$ with respect to the middle plane). Nevertheless, it is plausible to assume that the inclusion of additional subjects to the data basis would have made this effect a reliable one, since the mismatches in the front plane are the only ones that are not obscured by layers of other items.

Which effects do the eye-movement variables reveal? For each variable, the data were entered into a four-factorial (type of mismatch, local item density, *local depth entropy*, local form entropy) analysis of variance with two levels per factor. The independent variable local depth entropy replaced the variable local color entropy used in Experiments A to F, yielding information about the depth "disorder" at a given two-dimensional location on the screen. Since Experiment G presented three levels of depth instead of three levels of color, the calculation of depth entropy was completely analogous to that of color and form entropy.

One of the basic variables, namely fixation duration (FD), shows an interesting dependence on the type of mismatch. FD is significantly higher during depth search (226.46 ms) than during form search (209.40 ms) ($F(1;15) = 29.82$; $p < 0.001$). Furthermore, the FD values obtained in Experiment B during color search (208.04 ms) as well as during form search (203.91 ms) are significantly lower than those for depth search in Experiment G ($t(34) = -3.61; p < 0.001$ and $t(34) = -3.75; p < 0.001$ respectively). In none of the previous experiments the type of mismatch exerted a significant effect on FD, which suggests basic differences between the detection of depth targets and planar targets. Conceivably, subjects are able to perceive and memorize larger areas of local depth patterns within a single fixation than they are with regard to color or form patterns. Consequently, the memorization of more information needs longer FD.

Another factor exerts a significant effect on FD, namely local item density at the fixation point. As in all previous experiments, FD is shorter in regions of low item density (211.90 ms) than in regions of high density (223.96 ms) ($F(1;15) = 29.82; p < 0.001$). Neither depth entropy nor form entropy effects on FD were found.

Saccade length (SL), the second basic variable, is significantly higher in Experiment G (73.27 pixels) than in Experiment B (60.46 pixels) ($t(34) = 4.87; p < 0.001$). No reliable difference between depth and form search was found. Conceivably, the prolonged saccades are caused by the fact that some of the subjects use a plane-by-plane search strategy instead of performing top-down oder bottom-up scanning. They minimize the necessary vergence movements at the cost of longer saccades.

Two four-factorial analyses of variance were calculated, one referring to the local parameters at the starting points of the saccades, the other referring to their end points. With regard to the starting points, SL is significantly shorter for low item density (74.63 pixels) than for high density (69.83 pixels) ($F(1;15) = 15.96; p = 0.001$), which corresponds to the results of previous experiments. Concerning the end points of the saccades, the same effect is found with SL having a value of 76.61 pixels and 66.27 pixels respectively ($F(1;15) = 76.80; p < 0.001$). Moreover, saccades landing in areas of low depth entropy are significantly shorter (69.36 pixels) than those landing in areas of high depth entropy (73.52 pixels) ($F(1;15) = 17.38; p < 0.001$). The interaction between depth entropy and type of mismatch is reliable as well ($F(1;15) = 6.40; p = 0.023$), indicating a stronger effect of depth entropy during form search (68.97 versus 75.27 pixels) than during depth search (69.76 versus 71.77 pixels). These findings suggest that subjects avoid fixations in regions of high depth entropy, i.e. in areas surrounded by items in different depth planes, because it is difficult to gather information while attending to such regions. Since it is more problematic to "predict" the form entropy at the saccade's end point, fixations preceded by *longer* saccades are accordingly more likely to be located there. The detection of form mismatches is even more obstructed by high depth entropy than the detection of depth mismatches, because form recognition requires vergence movements to the respective planes. Therefore, subjects try even harder to avoid such regions, which leads to a more pronounced effect of depth entropy on SL.

The number of successive fixations within the same hemifield (FW) does not reveal any main effects of the four factors, but shows one significant interaction, namely between type of mismatch and local form entropy. Figure 7.11 (left) illustrates the interaction. In the situation of depth search, FW decreases with increasing form entropy (2.679 versus 2.557 fixations), whereas form search leads to an increase in FW with higher form entropy (2.658 versus 2.715 fixations).

Interestingly, the analysis of area coverage per fixation (AC) exhibits an analogous type of interaction, as shown in Figure 7.11 (right). AC increases with local form entropy during depth search (3376 versus 3577 pixels), and shows an inverse effect during form search (3641 versus 3202 pixels). These results concerning FW and AC suggest that depth search is facilitated by high form entropy, because items of different forms facilitate the comparison of depth patterns. In regions of low form entropy, e.g. in areas that exclusively contain circles, it is difficult to match the memorized depth pattern with the corresponding depth structure in the other hemifield, since there are no suitable points of reference. This problem causes additional fixations and hence increases FW and decreases AC. During form search, though, patterns of form attributes have to be memorized and compared. As shown in the previous experiments, high form entropy goes along with more information to be processed, leading to higher FW and lower AC.

Additionally, AC is higher in regions of low item density (3837 pixels) than in regions of high item density (3061 pixels) ($F(1;15) = 7.40; p = 0.016$), which is compatible with earlier findings. A comparison to Experiment B reveals that form search induces lower AC in Experiment B (2874 pixels) than in Experiment E (3588 pixels) ($t(34) = -2.82; p = 0.008$) as well as depth search in Experiment E (3572 pixels) ($t(34) = -2.26; p = 0.031$). This finding indicates that the additional

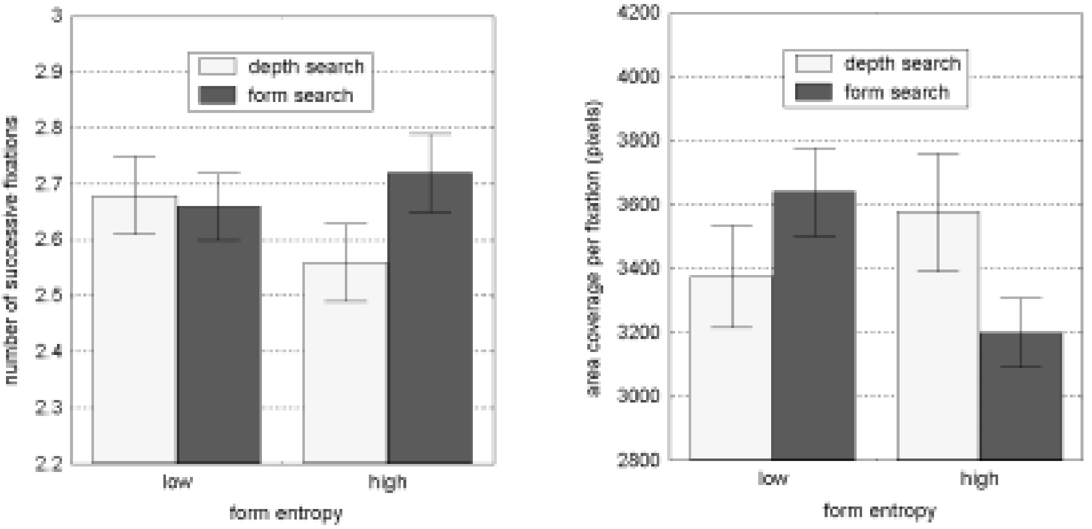


Figure 7.11: Interaction of the factors type of mismatch and local form entropy, which exerts significant effects on the variables FW (left) and AC (right)

depth information leads to larger areas being covered per fixation, at least with regard to two-dimensional form search.

The most striking difference between Experiments B and G is reflected in the search speed (SP). The mean SP value is significantly higher in Experiment G (82.62 pixels/s) than in Experiment B (52.14 pixels/s) ($t(34) = -4.09; p < 0.001$). In analogy to the results of SL, this effect can be attributed to the plane-based scanning strategies that some of the subjects appear to have applied.

7.4 Are Excursions to the Third Dimension Profitable?

The results of Experiment G show that the introduction of the third dimension leads to a broader variety of scanning strategies, suggesting that the z-axis – though represented by only three discrete levels – is perceptually “accepted” like the x- and y-axis. Some subjects use the depth structure in the stimuli as a guideline for their gaze trajectories; they start searching the front plane, then move on to the middle plane, and finally inspect the back plane; the reverse order of planes is possible as well. This behavior is caused by the facts that perceptual grouping of items in the same stereoscopic depth is especially convenient and stable and that planewise scanning needs fewer time-consuming vergence eye-movements. Vergence movements towards deeper planes are found to be extremely slow, since the perception of these planes interferes with the items lying in front of them. Moreover, physiological factors as well as most of the subjects’ little experience with artificial three-dimensional stimuli contribute to this effect.

The impaired recognition of items in the back plane seems to be responsible for the fact that form mismatches located there are – at least in a tendency – more difficult to detect. This tendency supports the view that mismatch detection is based on *binocular* perception in Experiment G, since *monocular* inspection of the stimuli should not handicap target detection in the back plane. The finding of depth mismatches between the back and the front plane to induce longer reaction times than others, however, would be expected for monocular processing as well: In the monocular situation, the depth mismatches between back and front plane correspond to a larger horizontal shift between the target items than do other depth mismatches.

Fundamental differences between the detection of depth and form mismatches, i.e. two-dimensional and three-dimensional mismatches, have been found. In order to enable target detection, form search needs more precise vergence movements than depth search. The basic eye-movement variable *fixation duration*, which has not been affected by the factors type of mismatch and experiment so far, substantially increases during depth search. This finding suggests that depth search facilitates the memorization of larger local regions for each step of comparison, causing increased fixation duration. In regions of low form entropy, depth search is less efficient, because the correspondence of information between the hemifields is less obvious.

This pattern of results is compatible with physiological data found in literature (e.g. Schor & Ciuffreda, 1983). Nevertheless, the insight gained with the help of Experiment G is of a special quality. Experiment G shows the importance and fruitfulness of studying the interactions of perceptive and cognitive factors during the completion of a complex and demanding task. For instance, the investigation of dynamically applied binocular vision, optimized for efficiency, yields detailed information about how our brain adapts the participating processes in order to facilitate fast target detection. Observing the visual system in ecologically adequate action enables us to combine the knowledge about isolated factors to achieve a more comprehensive understanding of the brain's architecture.

To answer the question posed in the headline: The excursion was definitely profitable and should by all means be extended. First, we must find out whether the effect of hyper-convergence during the inspection of the front plane is caused by a systematic error in measurement or rather by perceptive factors ("overshooting" and reduction of interference). Further research with the help of a spatial calibration procedure will be carried out in order to clarify this point. Second, as discussed in Section 7.1, the use of shutter glasses instead of red-green glasses is preferable, because it provides more natural conditions and allows us to present colored stimuli. Hence, this technical upgrade should be realized.

Within the scope of the present work, Experiment G can be considered as a suitable finale of the investigations of comparative visual search, because it provides an interesting outlook on possible research based on more realistic scenarios. Undoubtedly, our visual system has been evolutionary designed for a three-dimensional environment, and only in three-dimensional space it reveals all its capabilities.

The final part of this work is concerned with the construction and verification of computer models which try to reproduce the subjects' gaze trajectories on the basis of the empirical results. As explained in Chapter 1, only the combination of empirical research with approaches to simulations of the observed effects can lead to a broad understanding of the factors and processes involved in comparative visual search.

Chapter 8

A First Approach to Modelling: Random Walk

8.1 What can We Learn from Computer Simulations?

In many fields of research, computer simulations are the perfect complement to empirical studies. Empirical work, i.e. the conduction of experiments and the analysis of the resulting data, enables us to generate and to test hypotheses about our topic of investigation. In general, this process is aimed at the construction of a more or less formal *model* which ideally explains all of the experimental observations. Isaac Newton, for instance, studied the effects of gravity as well as the trajectories of planets. Based on empirical data, he found a common principle connecting those seemingly distinct topics. He formulated the "law of gravitation", a simple equation that can be viewed as his "model" of gravity. It does not explain anything about the *nature* of gravity, but it allows to calculate the gravitational force between any two objects with known masses and distance from each other.

This model led to a considerable advancement in theoretical physics. With the help of differential equations it enabled scientists to predict the trajectories of two objects in space attracting each other. However, its limitations soon became obvious: It was impossible to derive equations describing the trajectories of three or more objects interacting by gravitational forces. The only way of predicting the course of the objects' positions, speeds, and accelerations is a numerical computation. Numerical methods employ an iterative calculation of a system's state, usually based on some suitable discretization of the system's time evolution. The smaller the applied discretization steps, the higher the precision gets and, unfortunately, the arithmetical expense as well. Therefore, calculating the detailed behavior of a complex system by means of paper and pencil is simply not feasible. Although Newton and his contemporaries knew all the necessary "rules", they were not able to predict gravitational trajectories beyond a certain complexity.

Nowadays, the situation is different, because electronic assistants with enormous arithmetical capabilities support our work. Modern computers are able to solve problems like the one mentioned above more than a billion times faster than human beings. At present, even rather intricate systems have become accessible to *computer simulations*, for example the weather, flying airplanes, chemical reactions, flowing liquids etc. In some cases, the system's complexity and its sensitivity to small perturbations in its parameter values thwart our best efforts for long-term predictability ("chaos"). For instance, even a completely computer-based weather forecast using a gigantic database and precise measurements with high spatial and temporal resolution will in all probability never be able to predict the weather situation at a certain location for more than a few days in advance.

Nevertheless, comparing the results of computer simulations with the corresponding empirical data tells us to what extent a simulation is adequate. Serious discrepancies would indicate that we had not yet taken into account all the essential features of the system under study, hence our model would be incomplete or partially wrong. In this case we would have to perform additional experiments or measurements leading to a refined model which again would have to be empirically tested. As long as new findings or new theories are available, these steps can be repeated over and over until a satisfactory model has been established. Computer simulations can thus substantially enhance our understanding of complex processes.

Referring to the present context, what is the use of computer simulations in Cognitive Science? As already argued in Section 1.1, our brain constitutes the most complex structure we know. For obvious reasons it is impossible to observe the activity of all its neurons and their interactions in order to obtain complete information about ongoing processes. And even if we *could* do that, the interpretation of the recorded signals would still be more than problematic. However, we already possess useful knowledge about the brain with respect to at least two aspects: First, we are to some extent familiar with the various methods of communication between neurons and how information processing and learning are achieved at the neural level. Second, much is already known about the brain's functional structure, i.e. the specialization of particular brain areas for particular functions. For instance, there are the *visual cortex* for handling visual input and the *motor cortex* that controls motor actions. Some of the main connections between these areas are recognized as well, though the situation is extremely intricate and far from being well-explored.

Accordingly, there are two predominant types of models aiming at the reproduction of mental processes:

Neural models: This biologically motivated approach is based on the simulation of a limited set of neurons and their interactions, so-called *artificial neural networks.* A neuron is modelled as a simple processor calculating a mathematical, multidimensional function on its inputs from many other neurons. In turn, it sends the result of this calculation to a number of connected "colleagues". Processes of learning and adaptation are realized as changes in the strength of connections between neurons. There are various kinds of network architectures and learning algorithms, covering a wide range of possible situations.

Phenomenological models: The intention of these models is not to directly imitate physiological (i.e. neural) processes, but to reproduce perceptive and cognitive functions at a higher level. Here, functional subsystems (assemblies of neurons, whole brain areas, or distinct brain functions) are modelled by algorithms which heuristically mimic their effects. The inner structure of these areas is completely disregarded for the benefit of reduced complexity to be modelled. Hence, even high-level processes become accessible to this type of computer simulations.

In recent research, a variety of computer models have been applied to a wide range of mental processes. With regard to visual information processing in general and visual search in particular, both neural and phenomenological approaches have been attempted. Edelman and Weinshall (1991), for instance, proposed a neural model of 3D object recognition. It incorporates a two-layer neural network (input layer and representation layer) which stores information by means of unsupervised Hebbian learning. During the training process, the model is "presented" with images of a limited set of 3D objects from different viewpoints. The neurons in the input layer receive activation signals according to the patterns of the objects' salient points in the images, e.g. vertices or edge elements. This information is encoded by making neurons in the representation layer respond selectively to specific conjunctions of features, i.e. patterns of activation in the input layer. After training, the model is able to recognize the "familiar" objects in most cases even from novel viewpoints. Interestingly, the viewpoint-dependence of the model's recognition rate is qualitatively similar to empirical human data.

A computational approach to human pattern recognition was reported by Caelli & Bischof (1994). According to the proposed model, pattern recognition involves processes on three different levels:

Low level: Feature extraction, realized by applying appropriate filter masks to the 2D intensity patterns

Intermediate level: Segmentation and recognition of the extracted features with the help of local similarity measures

High level: Analysis of the geometrical relations between recognized features, resulting in a symbolic representation of the scene

On each level, "classical" algorithms standardize the data representation, realizing desirable properties of pattern recognition like rotation and size invariance. After suitable training, this model is capable of, for instance, recognizing facial expressions even in "unknown" faces.

As to the simulation of visual search, Grossberg, Mingolla & Ross (1994) combined four specialized neural networks for this purpose. Although the communication between the networks is continuous and asynchronous, the model's mode of operation can best be understood in terms of four successive steps of processing, each of them realized by a specific neural network:

Step 1: Preattentive processing of the visual input, resulting in a retinotopic representation of the stimulus features

Step 2: Boundary segmentation and surface formation; the scene is grouped into separate candidate regions

Step 3: Selection of a candidate region to be searched; this can be influenced either by bottom-up salience or by top-down priming of target features (cf. Chapter 4)

Step 4: Comparison of feature groupings with the stored target representation; if no target is detected, a new candidate region is selected in Step 3

This model is able to reproduce empirical distributions and slopes of reaction time and may explain psychophysical data of single-target, multi-target, form-color, and color-color conjunctive search. Its modular structure allows to observe its performance at least at the interfaces between the elementary networks.

A phenomenological approach was tested by Smith (1989): He modelled response time distributions for two-choice reactions like those required in picture matching and visual search. His approach was based on the so-called *Accumulator Model.* This model assumes that the magnitudes of inspected stimulus features lead to corresponding cognitive activations in the respective dimensions of sensory effect. The activations vary randomly over time; both the positive and the negative deviations of activity from a pre-defined standard value are added to an accumulator A or an accumulator B respectively. If one of these accumulators – assigned to the relevant dimension – exceeds a specific threshold value, the corresponding alternative of the two-choice reaction is initiated. After the adjustment of the standard and threshold values, this rather simple model is able to closely replicate response time distributions for a wide variety of tasks.

Motivated by the above approaches, we pose the following question: What type of simulation is adequate for comparative visual search? As we have seen in the previous chapters, task completion requires the interaction of many different mental processes and functions. Therefore, it does not seem useful to start modelling with a multi-layer neural simulation trying to explain all effects from the single-neuron activities to the joint operation of attention and working memory. It appears more reasonable to begin with a very simple phenomenological model and to learn from its shortcomings in order to arrive at an improved and more realistic simulation.

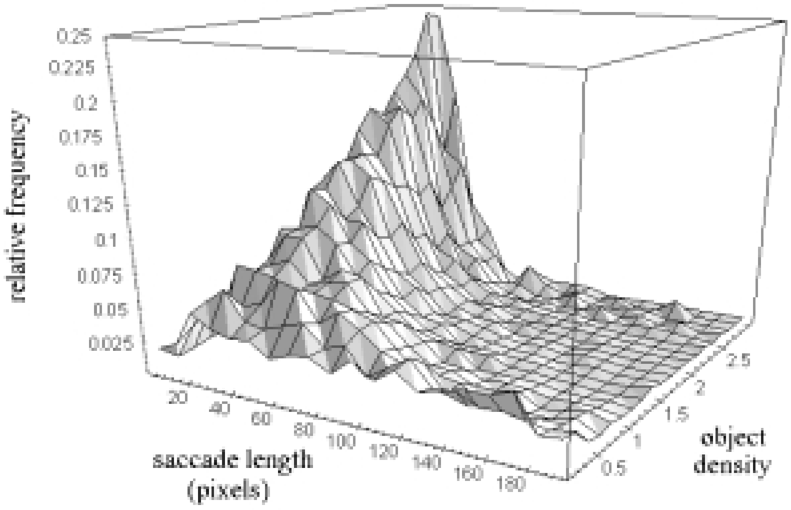


Figure 8.1: Histograms of empirical saccade length for different local item densities at the saccade's center (Experiment A)

8.2 A Random-Walk Model of Comparative Visual Search

Our idea is to start with a so-called "Random-Walk Model". It produces virtual, random eye movements according to the statistical distribution of empirical eye-movement parameters measured in Experiment A (cf. Chapter 3). Moreover, the influence of local stimulus characteristics on these parameters is taken into account as well. This rather vague description of the model's mode of operation shall be illustrated in a more comprehensible way: For a start, let us take a look at the variable *saccade length*. How does the model choose the length of the next saccade starting from a certain gaze position within the stimulus picture?

Figure 8.1 is a 3D diagram of empirical saccade length distributions – obtained in Experiment A – as a function of the local item density at the saccade's center. The diagram is made up by combining the histograms of saccade length for each level of item density, so that a "frequency landscape" emerges. If we "cut" the enclosing cube in parallel to the paper plane, the surface of intersection will show the histogram of saccade length according to the "item density" coordinate of our cut. For instance, the "frequency-by-saccade length"-plane at an item density value of 2.0 represents the distribution of saccade length for those saccades with a local item density of 2.0 at their center. With increasing item density, both the mean and the maximum value of the histograms obviously move towards shorter saccade lengths. This finding is rather plausible: In regions of higher item density, items are located more closely to each other, hence saccades tend to be shorter.

When generating its saccades, the Random-Walk Model statistically reproduces these empirical saccade length distributions. According to the local item density at the actual gaze position, the model chooses the length of the subsequent saccade randomly with the corresponding plane of the histogram in Figure 8.1 serving as the random distribution. As a consequence, the Random-Walk Model generates shorter saccades with growing item density, just as its human "examples".

Strictly speaking, saccade length is determined in *two* successive computational steps: At first, a temporary length is chosen according to empirical distributions referring to the saccades' *starting points* instead of their *centers* as in Figure 8.1. Second, the item density at the center of this temporary saccade is used to calculate the final saccade length in the way described above. Why so complicated a procedure? The strongest effect on saccade length was found to be induced by the item density at its center, being considerably more intense than the one caused by item density at its starting point. In order to integrate the stronger effect into the model, the weaker one had to be implemented as well to enable the estimation of the next saccade's center.

Whereas local item density has a substantial influence on most basic eye-movement variables, the effects of color and form entropy are rather small. Hence, the Random-Walk Model does not account for entropy to the benefit of a more transparent model structure. On the whole, the model is based on six different sets of empirical distributions measured in Experiment A:

- Distributions of saccade length varying with local item density at the starting points of saccades
- Distributions of saccade length varying with local item density at the center of saccades
- Distributions of fixation duration varying with local item density at the fixation point
- Distributions of the number of successive fixations within the same hemifield varying with local item density at the first fixation point
- Distributions of absolute angles of saccades within one of the hemifields
- Distributions of angles between successive saccades within the same hemifield

Furthermore, the model reproduces the empirical duration of saccades which mainly depends on the saccade length. Figure 8.2 shows saccade duration as a function of saccade length, as measured in Experiment A, plus the line of regression. The noticeable deviation of the duration values from their linear regression in the range from 150 to 300 ms is due to the fact that only a small number of saccades in comparative visual search has these lengths (cf. Figure 3.5).

The linear regression has been taken as an approximation of the functional relationship between saccade duration and length:

$$\text{saccade duration} = 29.1\ \text{ms} + 0.149\ \frac{\text{ms}}{\text{pixels}} \cdot \text{saccade length} \tag{8.1}$$

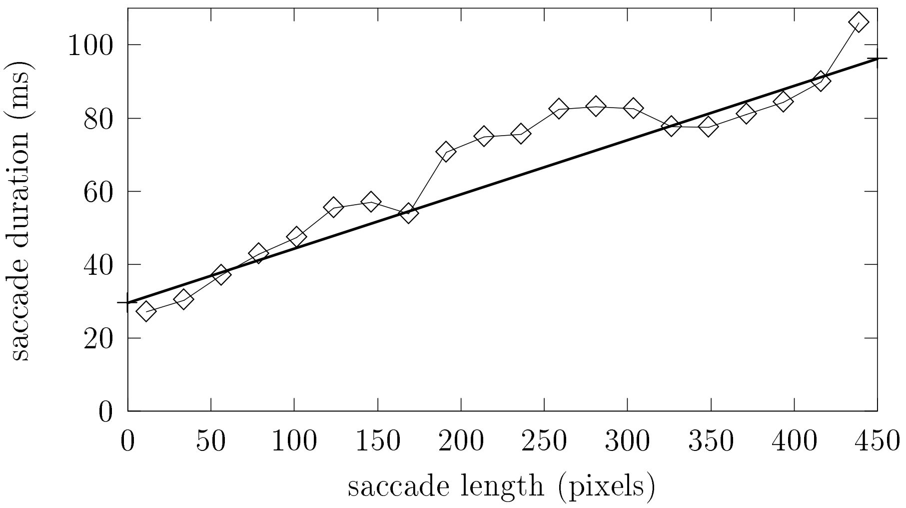


Figure 8.2: Mean saccade duration as a function of saccade length plus the corresponding line of regression

In addition, some simplifying assumptions have been introduced in order to reduce the complexity of our simulation:

- The only valid fixation points are the centers of all 60 items that constitutes a stimulus picture, i.e. the simulated fixations are "rounded" to the nearest neighboring item.
- The artificial gaze trajectory starts at the center of a randomly chosen item.
- When changing between hemifields, the gaze position always jumps to the item corresponding to the one that was fixated right before.
- The search and comparison phase terminates as soon as the gaze position reaches one of the two differing target items. That is, there are no instances of missing the target.
- The detection and verification phase consists in (randomly chosen) two, three, or four saccades changing hemifields with single, prolonged fixations in between.

Altogether, the Random-Walk Model is a strongly simplified approach. Nevertheless, it is capable of generating search trajectories and reaction times that could be compared with those of human subjects. Since it incorporates the basic eye-movement "regularities" that have been found empirically, it is possible to study the following question: To what extent can human gaze trajectories be modelled as a statistical process and when do we have to go beyond such description and take into account additional factors and processes that contribute to eye movements in comparative search?

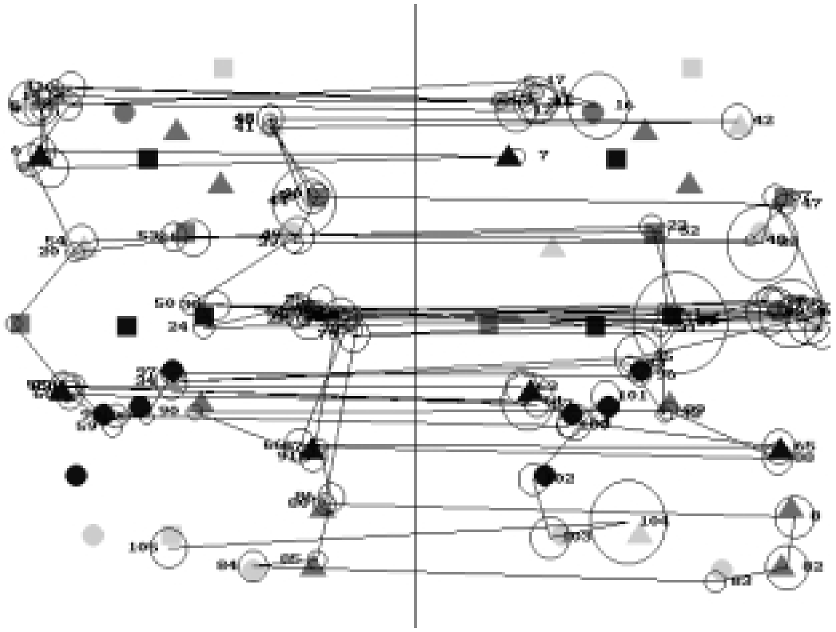


Figure 8.3: Example of a model-generated scanpath. Fixations are numbered; circle size indicates fixation duration.

8.3 Evaluation of the Model

Figure 8.3 presents an example scanpath generated by the simulation. The fixation points do not always hit the centers of the items, because the spatial error in measurement is incorporated in the Random-Walk Model as well. At first sight, the gaze trajectories produced by the Random-Walk Model seem to resemble the empirical ones. A closer inspection, however, reveals that the scanpath employed by the model just wanders about and does not resemble any human search strategy. This deficit is reflected in the resulting reaction times measured in 10 000 runs of the model. The model needs a mean duration of 30.0 seconds for target detection, whereas the subjects needed only 11.0 seconds on average in Experiment A.

The *distribution* of reaction times gives even more evidence. Figure 8.4 shows a histogram of reaction time yielded by the model in comparison to the data obtained from the subjects (as previously mentioned in Section 3.3). Here, a fundamental difference between empirical and simulated search is made transparent: While the distribution of reaction times produced by the model follows an exponential decay law, the human data exhibit a plateau-shaped maximum for reaction times in the range of approximately three to ten seconds. What is the reason for this striking discrepancy?

In fact, exponential decay of reaction time is a characteristic feature of completely *unstructured search*. Here, “unstructured” means that neither any kind of strategy nor a memory to remember the hitherto existing scanpath comes to bear. Why do such conditions lead to exponential decay?

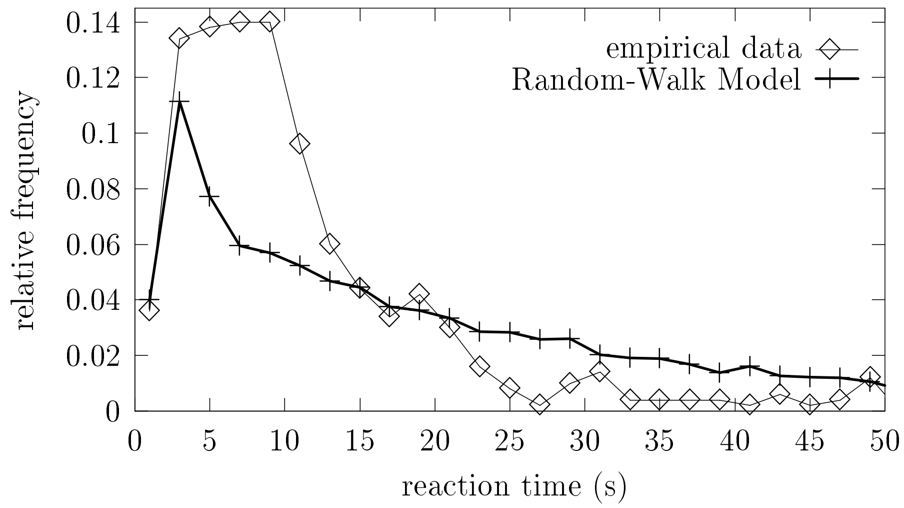


Figure 8.4: Histogram of empirical versus simulated reaction times

Let us assume that we want to conduct a perfectly unstructured comparative search. We randomly choose one of the items and compare it with its counterpart in the opposite hemifield. If these items are identical, we choose another item by random – it may be the same again – and perform a further comparison. We repeat this procedure until the detection of the mismatch. With n items on each hemifield, the probability p_1 of solving the task by the first step already resolves to:

$$p_1 = \frac{1}{n} \tag{8.2}$$

Due to the lack of strategy and memory, the probability of detection stays the same with regard to the second step. However, we have to take into account that this second step is performed only if the first step has not been successful, thus with a probability decreased by the factor $(1 - p_1)$. Generally speaking, we derive the following equation for the probability p_s of task completion in step s:

$$p_s = \frac{1}{n}\left(\frac{n-1}{n}\right)^{s-1} \tag{8.3}$$

This is obviously a function which declines exponentially with the number of comparison steps. Taking a look at Figure 8.4 again, we find a substantial similarity between this type of function and the distribution of reaction times yielded by the Random-Walk Model. In order to properly compare those distributions with Equation 8.3, we have to transform the "steps" into "times". The average time subjects need to compare corresponding regions in a stimulus picture is found to be approximately 600 ms, hence we use this value for the intended transformation. Moreover, we have to take into consideration a delay caused by the final detection and verification phase, which is set to 1200 ms.

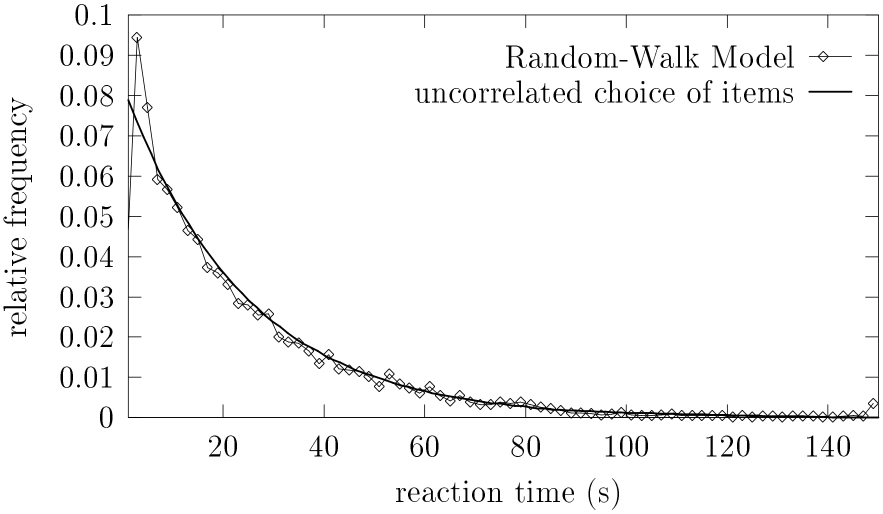


Figure 8.5: Histogram of reaction times yielded by the Random-Walk Model and by the random choice of items respectively

Figure 8.5 presents a comparison of reaction time histograms for the Random-Walk Model and unstructured search with the time constants defined above. The similarity between the two functions is obvious. It is corroborated by Figure 8.6 showing the same data in logarithmic scaling, demonstrating the exponential characteristic of the model's reaction time distribution.

Why does the distribution of *empirical* reaction times exhibit a clearly different shape (see Figure 8.4)? As we have seen in Chapter 3, subjects generally search the whole picture using a global scanpath. It is only by mistake that the same pair of corresponding items is compared more than once during task completion. If the target has not been detected after checking all items, the subject must start another search "cycle". In this case subjects are likely to apply a completely different scanpath, probably because they hope to detect the mismatch more easily when approaching it from a different direction. However, some subjects lose their patience at this stage and give up the structure of their search behavior. They start comparing items in different areas of the picture and try to solve the task without making any effort to arrange another global strategy.

These considerations can explain the shape of the empirical reaction time histogram. A complete search cycle takes about seven to twelve seconds, varying strongly between subjects. Let us assume that, during this cycle, subjects are scanning a sequence of item pairs without any pair being visited more than once. Since each pair will equally likely turn out as the target, each step of scanning within this sequence, from the first to the last one, has the same probability of leading to target detection. The constant probability of detection in the course

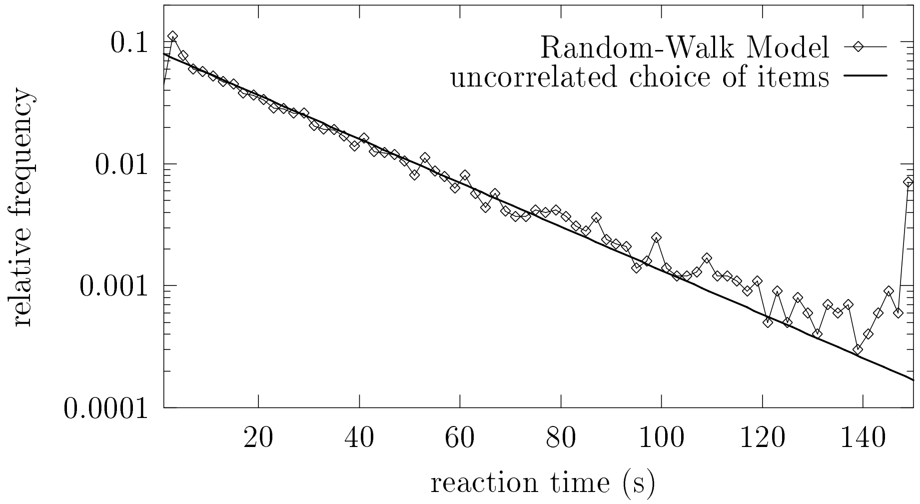


Figure 8.6: Histogram of reaction times using logarithmic scaling of frequencies

of such a search cycle will cause a plateau in the reaction time histogram.

As mentioned above, subjects may switch to unstructured search if they do not detect the target during the first search cycle. In analogy to the Random-Walk Model, such a behavior should lead to exponential decay in the reaction time histogram. In fact, the empirical reaction time distribution shows such a characteristic for reaction times longer than ten seconds. What impact on this distribution can be expected from those subjects that start a *second* search cycle instead of employing unstructured search? Strong between-subjects differences in search speed seem to prevent the emergence of a second (or third) plateau in the reaction time histogram.

Which conclusions can be drawn from the above findings? By way of re-analyzing the course of empirical gaze trajectories, we have arrived at the following answer: The global ("high-level") search strategies clearly differ between the Random-Walk Model and the human subjects. Obviously, the implementation of empirical parameters and dependences did not substantially enhance the model in comparison to pure random search. Although the trajectories generated by the Random-Walk Model seem to resemble those of the subjects, they are lacking that kind of global strategy that distinguishes the empirical scanpaths. The Random-Walk Model is adequate with respect to low-level features of scanpaths, e.g. fixation duration and saccade length, but human performance that incorporates cognitive processes on higher levels is not reflected at all.

The results reported in this chapter clearly indicate that our extensive analysis of comparative visual search has not accounted for quantitative measures of global scanpath characteristics so far. We have to acknowledge that not only cognitive processes on the level of local memorization and comparison control the subjects' behavior, but also those processes on a higher level, which enable the planning and tracking of global, self-avoiding scanpaths. This *vertical* structure of at least two different levels of information processing should be incorporated into further approaches to computer models. As a consequence, the first step towards an improved model consists in a systematic investigation of basic characteristics of global scanning strategies. In the following Chapter 9, an appropriate study is described in combination with the development and evaluation of various algorithms for a more accurate simulation of global scanpaths.

Chapter 9

Geometrical Aspects of Global Scanning Strategies

9.1 Introduction

The questions at the core of investigations in visual search are: Which are the strategies we apply when exploring our visual environment? What factors determine the sequence in which we inspect a given number of items?

These questions are far from being novel, and there are numerous approaches that have tried to provide at least partial answers. Gaze trajectories in realistic scenes have been the subject of several studies. Many of these investigations were conducted by computer scientists intending to "teach" vision systems to behave like the human visual system. To date, however, human vision is still much more powerful than even the best approaches to computer vision. It seems thus useful to find out more about the biological system and to imitate parts of it technically (for a review of efforts along these lines see e.g. Ballard, 1991).

Most models of eye movements in realistic scenes use spatial filters in order to detect an image's most salient points that are 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, whereas the locations of the following fixations are based on the analysis of *increasing* spatial frequencies. Another approach (Rimey & Brown, 1991) uses a Hidden Markov Model that is capable of learning gaze trajectories. It optimizes its scanpaths iteratively towards highest efficiency of gathering information in a given scene.

However, the scenarios of the studies mentioned above are far too complex with regard to our interest in basic properties of scanpaths. In realistic scenes, the subjects' attention is guided by high-level factors, e.g. by the functional or conceptual relationships between items or the individual relevance of items to the subjects. As comparative visual search aims at discovering fundamental properties of human scanpaths (see Chapter 3), such factors have deliberately been excluded. Hence, our stimuli were of a plain, abstract, and clearly parametrized nature.

Since gaze trajectories in comparative visual search yield only coarse information about global scanpaths, i.e. the sequence of items being compared (cf. Chapter 6), we used a simplified scenario and a simplified task: Subjects viewed only *one* display instead of two hemifields. Their task was to look exactly once at each dot in the presented random distribution of dots, starting at a specified starting dot.

Undoubtedly, the chosen scenario is a strong simplification of those utilized in the studies mentioned before. Moreover, the task of visiting each item exactly once 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. Although there is evidence for low-level processes of *inhibition of return* facilitating serial visual search (Posner & Cohen, 1984; Klein, 1988; Tipper, Weaver, Jerreat & Burak, 1994), these processes are not sufficient for the construction and tracking of efficient, i.e. self-avoiding and complete scanpaths in comparative visual search or in the task described above. Therefore, the subjects' scanpaths are likely to be influenced by higher cognitive processes than are usually involved in natural situations. In particular, *path planning processes* are expected to take place, because subjects have to keep in memory which dots they have already visited during task completion.

In spite of these ecological shortcomings, our setting enables us to investigate scanpaths 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) bias the observed strategies. The demand of attending exactly once to each item brings about an enhanced comparability of scanpaths. If we restrict the analysis to *acceptable* paths (i.e. those paths that meet this demand), it is easy to define a measure of similarity between two paths: The degree of similarity of a path A to another path B can be calculated as the number of “jumps” or edges between dots that appear in path A as well as in path B. Since there are 29 edges in every acceptable path, the resulting similarity values range from 0 to 29 (the direction of jumps is disregarded such that two paths visiting all items in nearly reverse order are still regarded as very similar, which appears to be a reasonable property in the present context).

The study reported in this chapter investigates geometrical regularities of scanpaths and is aimed at obtaining baseline data on global human scanpaths. Based on these data, we want to develop models of human scanpath selection and compare their results to the empirical data with a view to a better understanding of global scanpath characteristics. Although the chosen scenario seems to differ substantially from comparative visual search, there are two important requirements on efficient scanpaths that are identical for both paradigms: Self-avoidance and completeness. Therefore, we may hope that the results of the present study can be transferred to comparative visual search.

9.2 Measurement of Global Scanpaths

In this first approach to global characteristics of human scanpaths, we used a rather simple setting: Subjects were presented with a set of 30 dots (diameter of 0.5 degrees of visual angle) on a black background being randomly distributed within a square area (lateral length of 18.0 degrees) on a black background (see Figure 9.1 left). The dots were of the same color (blue), with the sole exception of the *starting dot* which was significantly brighter than the others. Starting at this dot, the subjects' task was to look once at each dot in the display. The subjects were told not to miss any dots nor to look at any of them more than once. Furthermore, they had to attend to each dot for at least half a second to make sure that the eye tracker registrated a fixation.

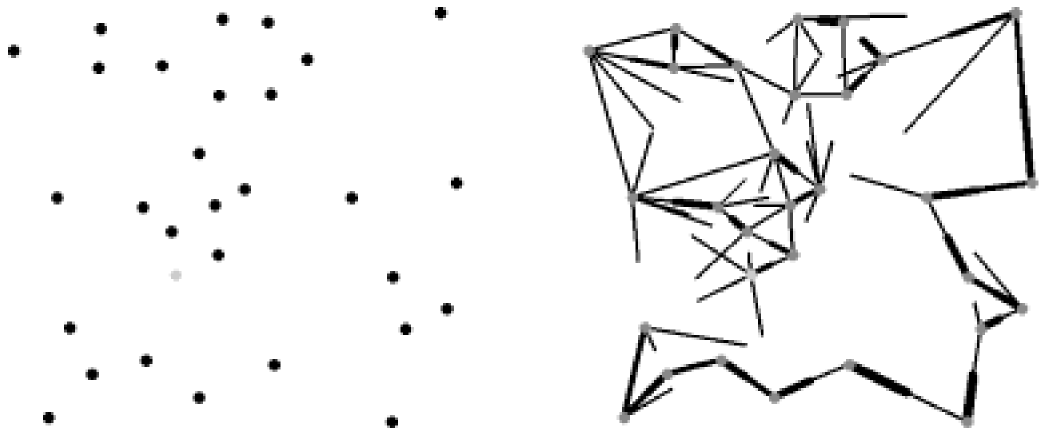


Figure 9.1: Example stimulus (left) and corresponding visualized results (right). The thicker the connecting line between two dots in the right image, the more subjects scanned these dots successively.

Five different dot configurations (stimuli) were randomly generated to prepare a set of 20 images that were used in this study. In order to investigate directional effects on the scanpaths, for instance top-down or left-right strategies according to the subjects' direction of reading, each stimulus was shown in four different orientations (rotated by 0^o, 90^o, 180^o, and 270^o). The resulting 20 stimuli were presented to twelve subjects in individually changing random order.

9.3 Data Pre-Processing

The recorded gaze trajectories were converted to item-based scanpaths. 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 procedure could not be done automatically without a severe impairment

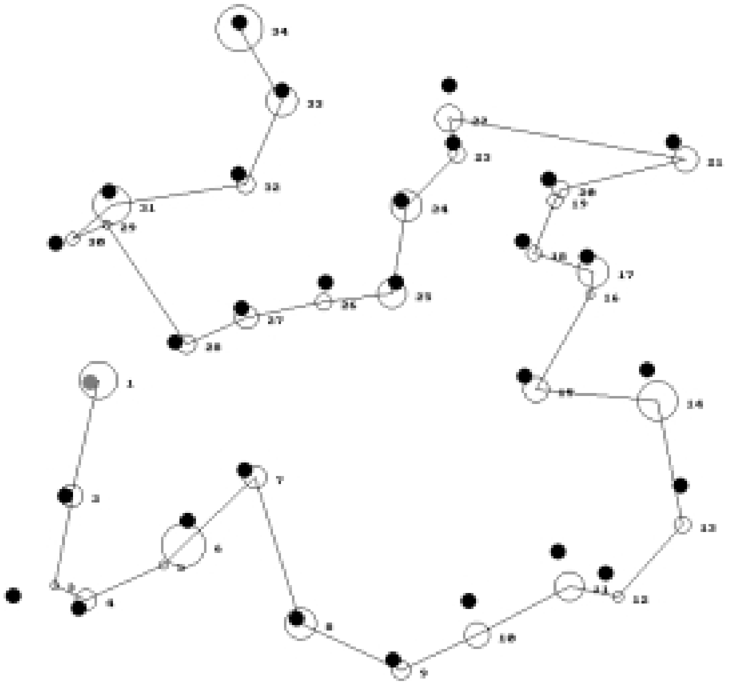


Figure 9.2: Another example stimulus with a subject's gaze trajectory. Displays of this kind were used to convert eye-movement data to item-based scanpaths.

of accuracy. The occurrence of orientational fixations, imprecise saccades as well as errors in measurement required human intelligence for the reconstruction to be correct. Consequently, an assistant – who was naive as to the purpose of the study – accomplished this task. He was successively shown all of the subjects' scanpaths on the background of the corresponding stimuli. Each fixation was plotted as a circle with a radius signifying its duration, numbers indicating the sequence of fixations, and saccades being represented by straight lines. From this information, the assistant had to decide in which order the items had been visited. He used a mouse pointer to specify this sequence of items. Figure 9.2 shows one of these displays and illustrates some of the difficulties which rendered an algorithmic solution impossible. Fixation number three is an example of an imprecise saccade: Conceivably, the subject shifted her *attention* to the dot in the lower left corner, but since the next item to be visited was rather close and required the eye movement to take a sharp turn, she executed an incomplete saccade. Inaccurate measurement can be found at fixation number 22 which evidently does not belong to its nearest neighboring dot, as the other fixations indicate. Fixation number 29 can be considered as orientational. The subject had to conduct a relatively long saccade after fixation number 28 and needed foveal information about the next group of items before choosing one of them as her following gaze target.

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 above definition. The further analyses were restricted to these acceptable paths. Figure 9.1 presents a visualization of accumulated data (right) for an example stimulus (left). Wider lines between dots indicate transitions used by a larger number of subjects. The lines are bisected due to the two possible directions to move along these edges. Each section refers to those transitions that started at the dot next to it. Sections representing fewer than three transitions were not displayed for the benefit of a clearer diagram. Within clusters of dots we find high variability of chosen edges, whereas the linear item structures on the right and at the bottom of the stimulus are scanned in what might be termed a "travelling salesman" fashion with only a few exceptions.

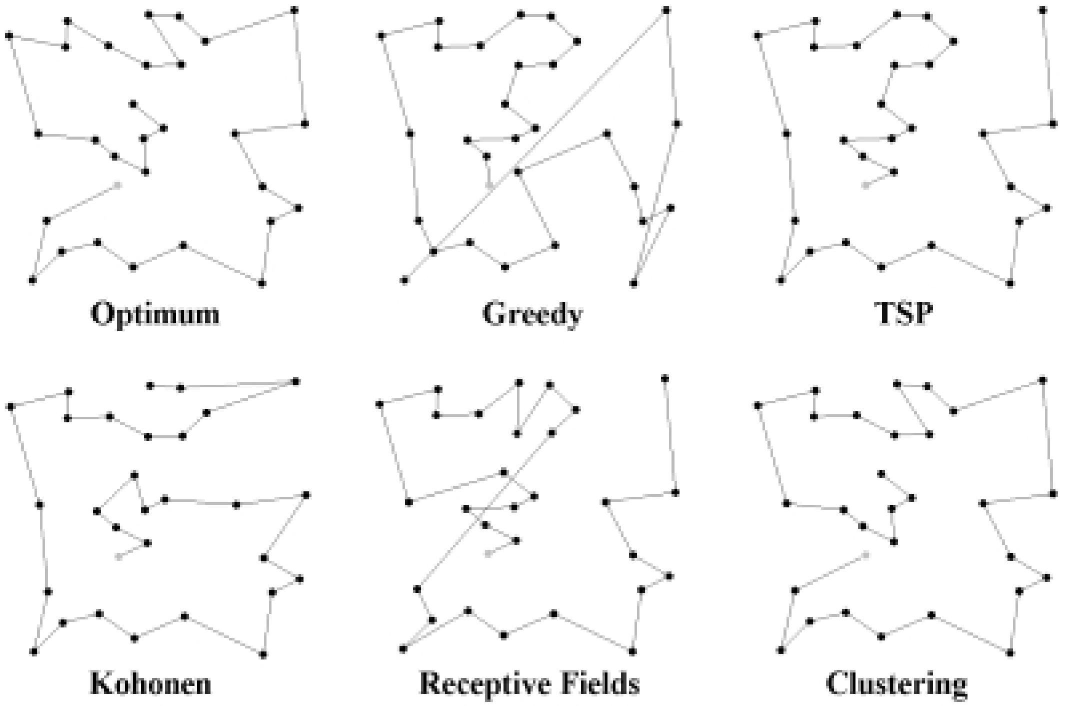
9.4 The Influence of Stimulus Rotation

Proceeding to the quantitative analysis of data, we can investigate the effect of rotating the stimuli: Are there directional influences on the scanpaths, for example caused by the subjects' reading direction? This question can be answered by comparing similarities of scanpaths (as defined above). If the scanpaths 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 was 19.43 edges per path, while the value between different orientations was 19.42. As a matter of course, this difference is not significant. Consequently, we may assume that there is no directional influence and can collapse the data for each of the five original stimuli in order to improve our statistical basis. This basis is needed for the evaluation of various *computational models* of scanpath selection which we are going to develop below. Since the empirical data show no significant dependence on the rotational orientation of the five stimuli, all models developed below do by construction not depend on this factor either. It thus suffices to evaluate them with the five dot configurations in their original orientation.

9.5 Evaluation of Scanpath Models

In order to have a baseline data for the evaluation of the models, we calculated a path with maximal similarity to the subjects' path for every stimulus. This average maximal similarity turned out to be 21.89, which means that these "optimal" paths are more similar to the empirical ones than the empirical ones are to each other (similarity value of 19.4, cf. above). Moreover, this result shows that, due to the intrinsic variability of scanpaths, no simulation can reach higher similarity values than 21.89. Additionally, as a second baseline, the similarity of completely randomly generated scanpaths to the subjects' paths was computed, yielding a value of as low as 1.75.

What are promising scanpath strategies to be modelled? We developed and tested five different models, which are described below. An example of the optimal paths as well as the scanpaths computed by the models is given in Figure 9.3 referring to the example stimulus in Figure 9.1.



Optimum Greedy TSP

Kohonen Receptive Fields Clustering

Figure 9.3: Scanpaths generated by the different models. The "Optimum" path is the one with optimal similarity to the empirical ones.

9.5.1 The "Greedy" Heuristic

As a first approach, we can test a rather plain method called "Greedy" heuristic. The "Greedy" algorithm always jumps to the dot which is geometrically nearest to the actual "gaze" position and which it has not visited yet. Although it produces plausible, locally optimized sections of scanpaths, the Greedy strategy has one drawback: On its way through the stimulus, it "forgets" to scan items of high eccentricity. As a consequence, these items have to be "collected" later, which causes unnaturally long saccades at the end of the trial. This lack of memory constitutes a fundamental difference to empirically observed strategies. Nevertheless, even this simple model reaches a similarity value of already 17.36, indicating that its strategy of always choosing the nearest item is already tremendously better than the pure random strategy.

9.5.2 The "Travelling Salesman" Algorithm

The shortcoming of the Greedy heuristic motivates the implementation of a TSP ("travelling salesman problem", see Section 3.1.1) which minimizes the *global* length of its scanpaths rather than just the length of the next jump. However, unlike standard TSP, the paths of this algorithm do *not* return to the start dot, because this would lead to inacceptable paths. Instead, the TSP Model chooses that dot to be the last one which allows to generate the shortest scanpath that is possible. The results show that this simulation gets much closer to the actual human strategies than the Greedy heuristic: The similarity value is now 20.87, which is already fairly close to the optimum of 21.89. This finding suggests that not only the local optimization of scanpaths – as incorporated in the Greedy algorithm – plays an important part in human scanpath selection, but also their global optimization.

9.5.3 Using a Self-Organizing Map

When simulating cognitive processes we should not leave aside *neural network* approaches, because they provide a biologically motivated explanation of their performance. 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 formal neurons, usually a one-dimensional chain or a 2D 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 in the network. Thus we create a map – the neural network – in which each mapped point – each neuron – represents a region or interval 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 very 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 of v; this is called the *winner*.
3. Change all neuron weights w_i towards the input vector v, with an adaptative step size h_{ij} that is a decay's function of the network distance between neuron i and the winner j. Here, ϵ is an additional global adaptative step size parameter.

$$w_i^{\text{new}} = w_i^{\text{old}} + \epsilon \cdot h_{ij} \cdot (v - w_i)\,,\ \epsilon \in [0, 1] \tag{9.1}$$

The change of neuron weights gives us an appropriate representation vector and the distribution of change around the winner produces the desired topology. 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 dots 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 a number of neurons will be mapped to the same positions. This must be resolved by a post-processing step to extract the simulated scanpath 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) in comparison to the results obtained by the plain TSP Model.

9.5.4 A Scanpath 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, artificial or natural, the term *receptive field* stands for the region of input space that affects a particular neuron. More specifically, influence from this region is not necessarily homogeneous, but dependent, for example, on the distance of the input vector from the center of the region. There may also be excitatory and inhibitory subregions, suitably represented by positive values for the interaction between input and neuron, or negative values respectively (see e.g. Hubel & Wiesel, 1962).

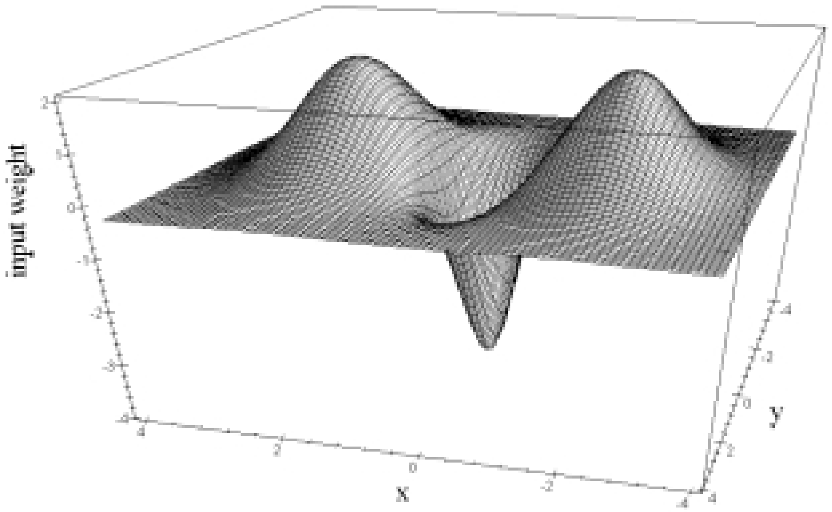


Figure 9.4: Illustration of the simulated receptive fields. The planar input space is represented by the dimensions x and y; positive values of input weight signify excitatory connections, negative values signify inhibitory connections.

With regard to our model, the receptive fields consist of an inhibitory axis and two laterally located, excitatory areas of circular shape (see Figure 9.4). We use 100 000 receptive fields that are randomly distributed over the input space.

Their size varies randomly between 80% and 120% of the relevant input space, i.e. the area of dots to be processed. 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.

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 a perceptually plausible bisection of the stimulus.

This first bisection is taken as the "level one" bisection, separating the set of dots into two subsets, which are separately processed in the following step. Each subset serves as the input to a new group of neurons with receptive fields that cover the region of the respective subset of dots. Since the input space for these neurons is smaller than it was in the first step, the receptive fields are smaller as well (see above). The resulting two "level two" bisections lead to four even smaller sets of dots which are to be bisected in the third step. This recursive procedure is repeated until none of the sections contains more than three dots, since the number three is a plausible estimate of the number of dots that can be perceived at the same time. Figure 9.5 (left) shows the bisections generated by the model with respect to the example stimulus previously shown in Figures 9.1 and 9.3. The bisections are visualized by straight lines with numbers indicating their level.

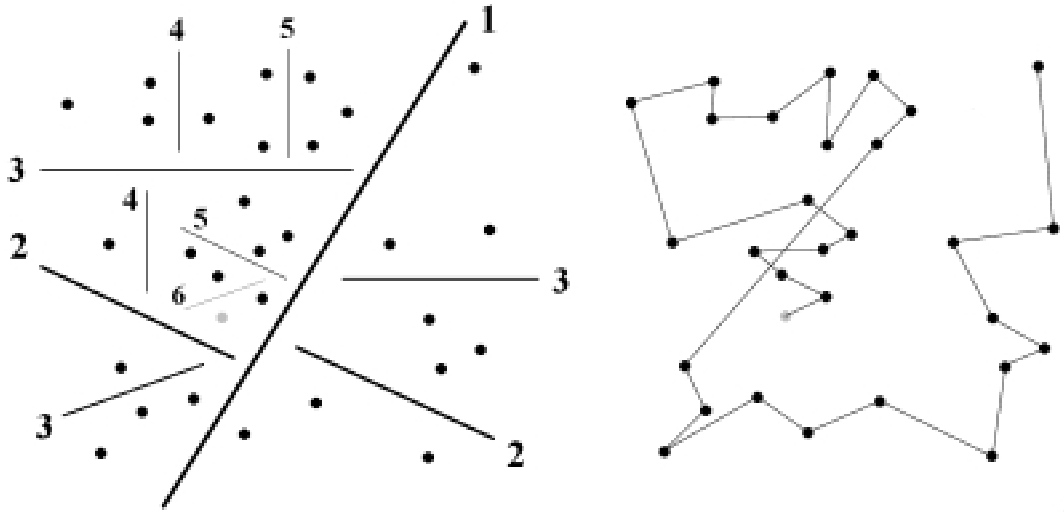


Figure 9.5: The model's hierarchical bisections of an example stimulus (left) and the resulting scanpath (right)

As a result, we get a hierarchical tree structure of bisections with its leaves corresponding to the elementary groups of dots. The derivation of this structure is our attempt to simulate a subject's perceptual processing of the visual scene.

Finally, the scanpath is derived by a TSP algorithm calculating the shortest scanpath that begins at the starting dot. In the present context, however, the algorithm does not minimize the geometrical distance to be covered, 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 to get from the group (leaf) that contains dot A to the one that contains dot B. If we choose the coefficients of the linear combination in such a way that the tree distance is substantially more relevant than the geometrical distance, the model may generate the scanpath shown in Figure 9.5 (right): The scanpath strictly follows the hierarchical tree structure, which may lead to geometrical deviations.

What are the results of the model's statistical evaluation? If the linear coefficients of the TSP component and the neural component are chosen such that the tree distance exerts a significant effect, neither the appearance of the simulated scanpaths nor their calculated similarity to the empirical paths (18.73) is convincing. This approach, at least in this rather simple form, does not yield more plausible scanpaths than does the TSP Model. The bisections introduced by the model's neural component do not seem to correspond to human strategies.

9.5.5 The Clustering Model

The fact that the TSP Model has yielded the best result so far motivates the investigation of a refined variant of it. This variant is based on the assumption that human scanpaths are based on clusterwise processing (see Section 3.3) and that linear structures of items are likely to be scanned successively (see Figure 9.1).

In a first step, this so-called *Clustering Model* divides the distribution of items into clusters. The clustering algorithm maximizes the between-cluster distances and minimizes the nearest-neighbor distances between items within the same cluster with the help of a cost function. This iterative procedure corresponds to the one used for the creation of color and form clusters in stimuli of comparative visual search (see Section 2.3), except that the clusters generated by the Clustering Model may also have linear shape. As a result, five to eight perceptually plausible clusters are calculated.

The second step consists in a TSP algorithm for scanpath generation with the restriction, however, that the "salesman" has to visit *all* dots of each cluster before he can proceed to the next one. In fact, the Clustering Model selects paths of slightly higher similarity to the empirical ones (21.12) than does the TSP Model.

9.6 Conclusions

Figure 9.6 displays a summary of the similarities in ascending order. It shows that the TSP-based models (TSP and Clustering) outperform the neural approaches

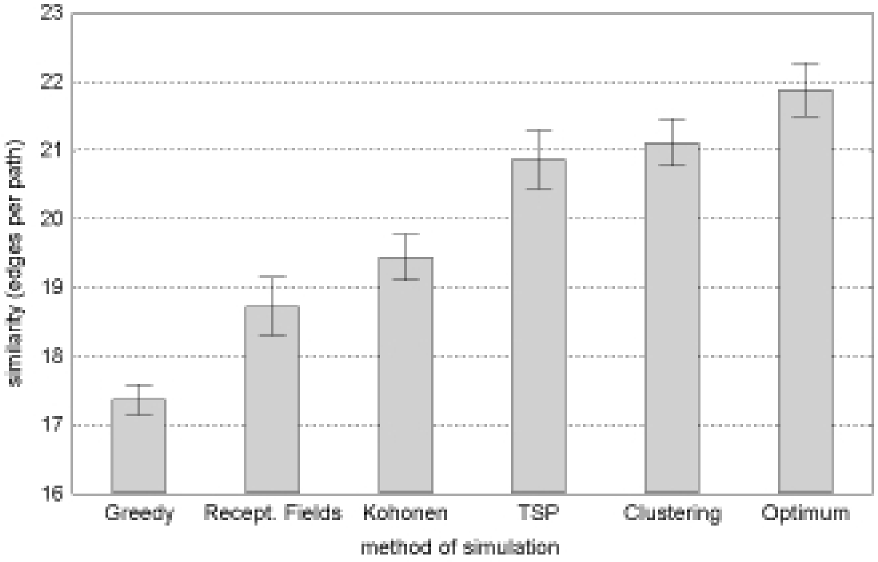


Figure 9.6: Similarity between the paths generated by the different models and the empirical scanpaths

(Receptive Fields and Kohonen), and that even the “primitive” Greedy algorithm is not far behind.

However, discretion is advisable in the interpretation of these data, since all of them are based on only *five* different stimuli. An appropriate statistical analysis requires the conduction of a more comprehensive follow-up experiment. Serving this purpose, we conducted Experiment H, which is described in the following Chapter 10.

Chapter 10

Experiment H: Scanning Strategies for Color and Form Distributions

10.1 Further Aspects of Scanpaths to be Investigated

What did we learn from the scanpath study of Chapter 9? One of the results is of particular importance for the design of Experiment H: Rotations of the stimulus plane were not found to significantly affect the subjects' scanpaths. Therefore, these rotations should be eliminated in Experiment H, making it possible to present a larger number of different stimuli to the subjects and thus to improve the validity of the experiment.

This simplification of experimental design enables us to increase the comparability between the scanpath and the comparative search scenario: In Experiment H, the items shown in the display have *color and form attributes*, just as the geometrical items used in Experiments A to F had. We can expect these features to influence the structure of chosen scanpaths, because subjects are likely to take advantage of this additional structural information. As mentioned above, their main concern during task completion 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 "scanpath units" which need less effort for memorization than single items. Thus, as in the experiments of comparative search, the stimuli should be designed with varying tendencies towards the creation of color or form clusters.

If we find that subjects make use of the color and/or form information, these effects should be integrated into the models. Furthermore, if the attributes lead to a reduction of scanpath variability, we can expect these models to yield better results than in the first study described in Chapter 9. Since the paths generated by the TSP and the Clustering Model have revealed preferable similarity to the empirical data, we should concentrate on the adaptation of these two approaches to the scenario of Experiment H.

In order to provide comparability, the experimental set-up and procedure of Experiment H should correspond to the first study as closely as possible. The introduction of color and form attributes brought about an important question: How could the starting item be indicated without changing the local color or form context in the display? We decided to use a *dynamic* cue, namely a flashing red circle around the start item which appeared two seconds after stimulus onset and vanished 1.3 seconds later. In order to increase its attractivity to the subjects' attention, the circle always appeared simultaneously with a beeping sound. As a matter of course, the adjustment of the time interval between stimulus onset and indication of the starting item is absolutely arbitrary. We can assume that, on the one hand, a shorter interval gives the subjects less time to inspect the stimulus before starting task completion and hence causes the grouping effects of color and form attributes to decrease. On the other hand, a longer interval is supposed to increase the cognitive effort subjects spend on planning their scanpaths, thus leading to "unnaturally" optimized paths. The choice of a two-second interval seems to allow a coarse inspection of the presented items and their attributes without providing any possibility of planning the scanpath step by step. Apart from this inevitable alteration, the subjects' task was the same as in the first scanpath study.

10.2 Method

10.2.1 Subjects

Twenty subjects from different faculties of the University of Bielefeld took part in Experiment H. All of them had normal or corrected-to-normal vision, none of them was color-blind or had pupil anomalies. They were paid 7 DM for their participation.

10.2.2 Apparatus

As in Experiment G (see Section 7), the subjects' binocular eye movements were measured with the SMI EyeLink system. This time, however, only fixations were registrated. The gaze position samples at 250 Hz were omitted due to their redundancy for the reconstruction of scanpaths.

10.2.3 Stimuli

The stimuli consisted of 30 simple geometrical items of three different colors (blue, green, and yellow) and three different forms (triangle, square, and circle) on a black background (see Figure 10.1). The items were identical to those used in Experiments A to F. Their spatial distribution was randomly generated with a minimum distance between neighboring items in order to avoid item overlap

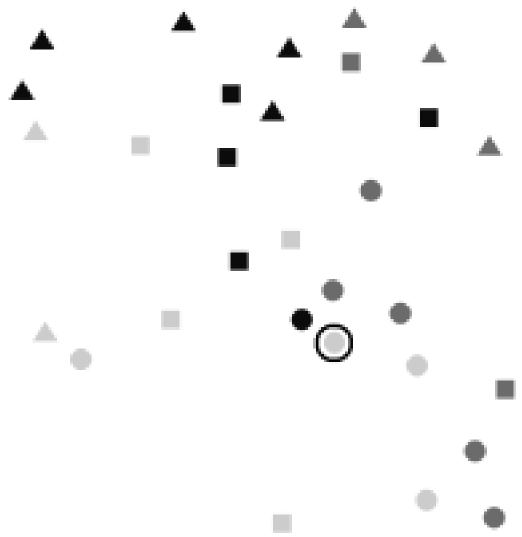


Figure 10.1: Example stimulus of Experiment H with weak color and form clustering. The starting item is indicated by a circle.

or contiguity. In each stimulus, the numbers of the occurrence of different colors and forms were balanced.

The distribution of colors and forms was not always homogeneously random, but with a tendency to create color and/or form clusters. Three different levels of clustering were used, namely “no clustering” (cluster measure 1.0), “weak clustering” (cluster measure 1.3), and “strong clustering” (cluster measure 1.7) (for detailed information about the applied cluster measure see Section 2.3). The nine possible combinations of different levels of color and form clustering constituted the categories of our stimuli. In the experiment, five stimuli of each category were presented, which were the same for each subject, but the order of presentation of all 45 stimuli was individually permutated.

Two seconds after stimulus onset, a red circle was shown around one of the items, together with a short sound (50 ms). The circle disappeared 300 ms later. This process was repeated twice with delays of 200 ms in between, signifying the location of the starting item which was always the same for the same stimulus. Figure 10.1 presents an example stimulus of the type “weak clustering” for both dimensions.

10.2.4 Procedure

A written instruction informed the subjects about their task. They had to look at the starting item as soon as it was indicated. Starting with this item, they had to fixate every shown item *once*, which meant that they should neither miss any of them nor visit any of them twice. After task completion they were to press a mouse key. The experiment started with two practice trials followed by the eye tracker calibration procedure and the 45 recording trials. As in Experiment G, each trial was preceded by a short calibration for drift correction using a single target at the center of the screen.

10.3 Results

As in the first study, an assistant converted the recorded fixations into scanpaths connecting the displayed items. In order to limit the influence of the assistant's subjective interpretation of the resulting data, he was only shown the *locations* of the items, but not their color or form *attributes*. Instead of items, he viewed distributions of dots with a brighter dot indicating the starting item, just as in the first study (see Figure 9.2). The superimposed visualization of the subject's fixations and their temporal order allowed the assistant to mark the item-based scanpath using a mouse pointer.

The proportion of acceptable paths turned out to be substantially higher than in the first study (93.3% versus 57.9%). Obviously, the additional color and form information helped the subjects not to “get lost” during task completion, apparently because the individual features of the items facilitated a more reliable memorization and recognition than their locations alone. The incorrect paths were approximately equally distributed among the nine categories of stimuli, hence they were excluded from the analysis for the benefit of a better comparability of different paths.

For a start, we can take a look at the calculated scanpaths of maximal similarity to the empirical ones with respect to varied classes of stimuli. The upper row of Figure 10.2 presents these paths for (from the left to the right column) an unclustered, a weakly color-clustered, a strongly color-clustered, a weakly form-clustered, and a strongly form-clustered stimulus. We do not find any striking evidence for the influence of color or form attributes on the subjects' action. Of course, there *are* some longer sections of scanpaths exclusively visiting items of the same color or form, but these items are always located closely together. A close analysis of the data yields strong evidence that the *location* of items remains the most important factor to determine the structure of scanpaths.

How can we find out whether there are any effects of the items' *attributes* as well? First of all, we need a quantitative measure of color and form clustering within scanpaths. An appropriate choice is the mean *runlength* with regard to these dimensions. In the present context, a “run” means a sequence of items of the same color or form respectively, being found within a scanpath. The runlength is the number of items that constitute a run, ranging from one to ten in the chosen scenario. If we want to calculate the mean runlength of multiple paths, it is useful to employ a *weighted* mean, because items in longer runs would be less meaningful otherwise. For instance, a scanpath consisting of 30 color runs (no transitions between items of the same color, runlength one) affects an unweighted mean *ten* times as strongly as a scanpath of three color runs (all colors scanned successively, runlength ten). This problem is solved by weighting each run with

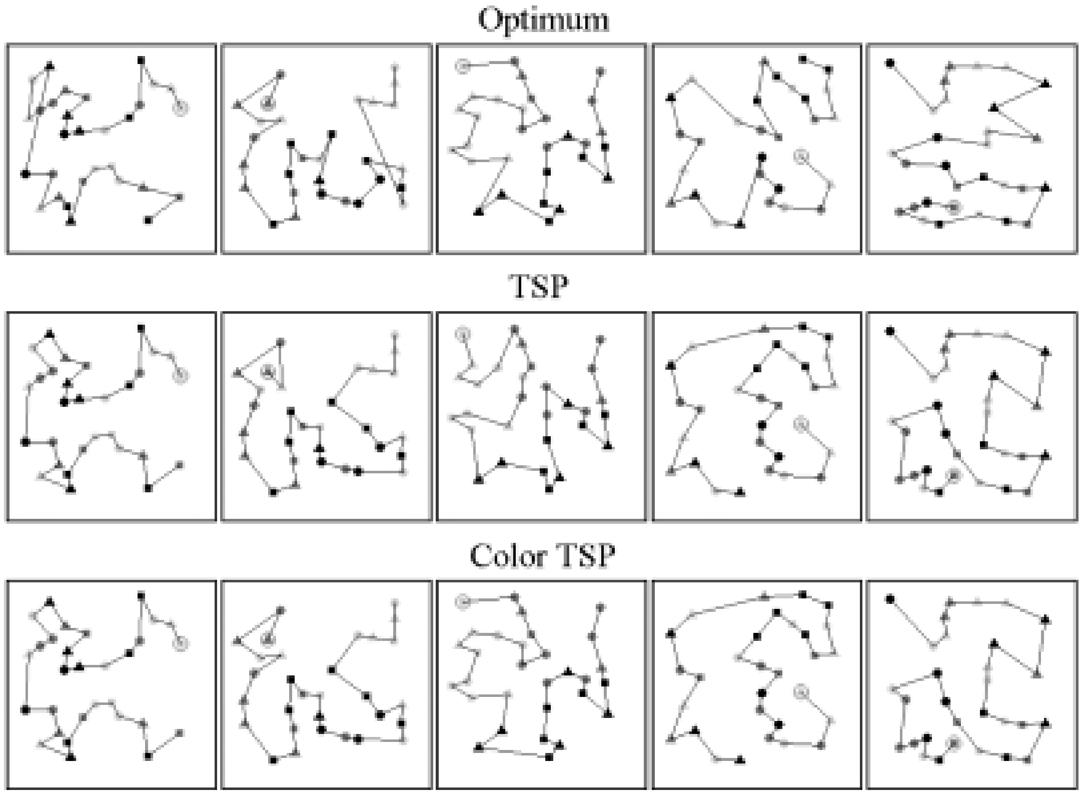


Figure 10.2: Scanpaths generated by subjects (most similar paths), TSP, and Color TSP. Circles indicate the starting items.

its runlength.

Now that we have a measure of the subjects' color and form clustering, how can we test whether its results indeed reflect the influence of item attributes or rather the geometrical structure of the stimulus? Even a subject who completely ignores colors and forms is supposed to generate longer runs with increasing strength of spatial clustering in the stimulus. This is due to the fact that, in this situation, items with the same features move closer together. Since subjects exhibit a tendency towards generating scanpaths of minimized length ("TSP"), neighboring items are likely to be scanned successively.

Fortunately, there is an absolutely "color and form blind" model which can provide us with reference data, namely the TSP Model. As in the first study, this model yields paths of good similarity to the empirical ones (see second row in Figure 10.2). If the subjects produced significantly longer color or form runs than the "travelling salesman", this would indicate grouping processes caused by the respective item attributes.

Figure 10.3 shows the subjects' mean color and form runlengths respectively for the different levels of clustering in the stimuli. The color runlengths refer to the three levels of color clustering, the form runlengths to those of form clustering. The baseline values of the TSP Model include the color *and* form runs,

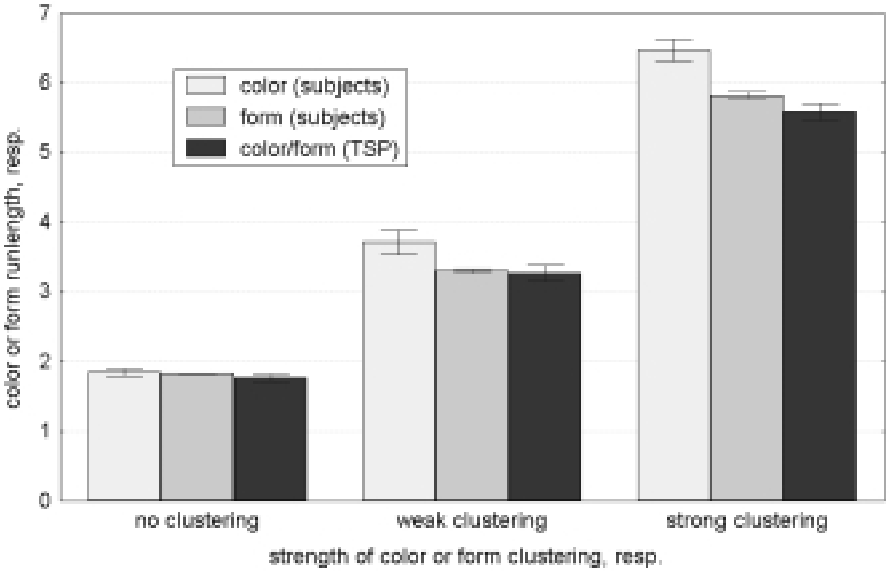


Figure 10.3: Mean color and form runlength depending on the strength of clustering with regard to the relevant dimension. TSP paths are not separated for color and form, because the TSP Model does not account for these attributes.

because there is no significant difference between them. In the “no clustering” condition, color and form runlength do not show any significant deviation from the corresponding TSP value. The “weak clustering” condition, however, leads to longer color runs, whereas the form runs remain on the TSP level. T-tests reveal that the difference between color runlength (3.716) and form runlength (3.307) closely misses significance ($t(38) = 1.92; p = 0.063$) as does the difference between color runlength and TSP runlength (3.269) ($t(48) = 2.01; p = 0.050$). The deviations seem to increase in the “strong clustering” condition, going along with a slight effect on form runs as well. T-tests show that color runlength (6.466) is significantly higher than both form runlength (5.822) ($t(38) = 3.22; p = 0.002$) and TSP runlength (5.589) ($t(48) = 4.00; p < 0.001$) now. The difference between form runlength and TSP runlength, however, remains below significance. These results demonstrate that only the items' *color* influences the subjects' scanpath strategies, whereas their form does not seem to play a part. The effect of color increases with the strength of color clustering in the stimuli.

Additionally, the two-factorial design (color clustering, form clustering) of the experiment makes it possible to test whether there is any interaction between these factors. The two-dimensional analysis of variance reveals significant main effects of color clustering ($F(2; 38) = 362.96; p < 0.001$) and form clustering ($F(2; 38) = 1269.63; p < 0.001$), as indicated by the diagram in Figure 10.3. Their interaction, however, does not reach significance. All in all, our findings yield evidence for the effect of color clustering on color runlength to be the only consequence of introducing item attributes.

Figure 10.3 provokes an important question: What might be the reason for the standard error to be substantially higher for color runs than for form runs? Are there maybe two or more distinct groups of subjects using different strategies of color clustering, demanding separate analyses? We can investigate this with the help of an appropriate visualization of the subjects' individual scanpath parameters. Each subject can be characterized by a two-dimensional vector consisting of his/her mean color and form runlengths. These vectors are divided by the average TSP runlength for easy interpretation: A vector component of 1.0 indicates that the subject's scanpaths are *not* likely to be guided by the corresponding item dimension, while increasing values signify growing effects of the mentioned kind.

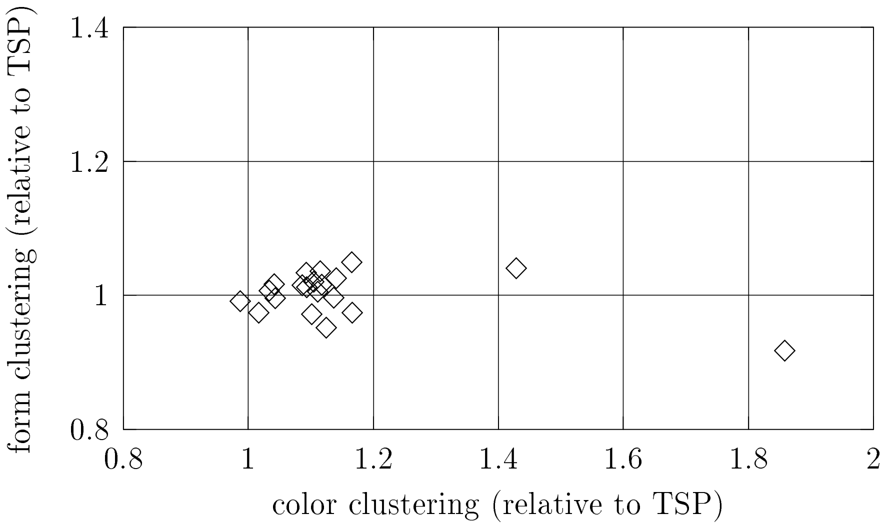


Figure 10.4: Distribution of the 20 subjects with regard to the mean color and form runlengths found in their scanpaths (relative to the TSP runlengths)

Figure 10.4 presents a cumulative diagram of the 20 subject vectors displayed as "diamonds". As can easily be seen, the distribution extends more widely on the x-axis (color runlength) than on the y-axis (form runlength). Whereas the form runlengths keep close to the value 1.0, the color runlengths range from 1.0 to 1.2 and reveal two additional "runaway" values at 1.43 and 1.86. Although these two cases show striking differences to the majority of subjects as well as to each other, it would not be justified to exclude them from the analysis. The two exceptions do not constitute a group of their own that should be studied separately. They rather indicate a wide range of applied strategies which should be incorporated into the computer models we are going to develop.

10.4 Modelling Human Scanpaths

The results of the first study motivate the adaptation of both the TSP 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 TSP Model, we should start with the adjustment of the TSP Model.

How can we "teach" the TSP algorithm to react to the color attributes in the same way as the average subject does? Basically, it should still calculate scanpaths of minimal length, but this length may *not be of purely geometrical* nature, but it must also *account for the "color distance"* between the colors of the items. An appropriate idea is to "tell" the "travelling salesman" that the distance between two items of different colors is their geometrical distance multiplied by a constant factor which we could call *distance factor*. The distance between items of the same color, however, remains identical to their geometrical distance.

Understandably, the algorithm's behavior will strongly depend on the choice of the distance factor. It is obvious that a distance factor of 1.0 leads to a standard TSP algorithm which is not influenced by color information at all. A distance factor of 1000, for instance, makes the algorithm use a minimum of transitions between different colors. At first, it visits all items of the starting item's color A, then it inspects all items of color B, and finally those of color C. Within the color groups it behaves as a conventional "travelling salesman", and the passages between them are the shortest that can be found. By adjusting the distance factor we are able to control the influence of colors and hence the average color runlength produced by the TSP algorithm. Since our goal is to adapt the TSP Model to the empirical data, i.e. to produce runlengths as generated by the subjects, the distance factor needs to be adjusted for the best match.

What is the response of our algorithm to the new definition of distance? It reveals a tendency towards the avoidance of transitions between items of different colors, because these transitions increase the overall length of the scanpath above proportion. Figure 10.5 presents a diagram of color runlength as a function of the distance factor ranging from 1.0 to 1.5. Higher distance factors cause an implausibly strong clustering strategy of the TSP algorithm and are thus not considered. The mean runlengths were measured separately for the 15 stimuli belonging to each of the three levels of color clustering (no clustering, weak clustering, or strong clustering respectively). Additionally, the empirical runlengths for these levels are shown as lines of reference.

We find the TSP runlengths to increase approximately linearly with the distance factor with higher levels of clustering leading to steeper runlength slopes. Surprisingly, we cannot select a certain value of the distance factor to yield the "best" runlengths. For each of the clustering levels, the intersection between the runlength of the TSP Model and the subjects' runlength occurs at a different distance factor. 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 distance factors with increasing color clustering in the display.

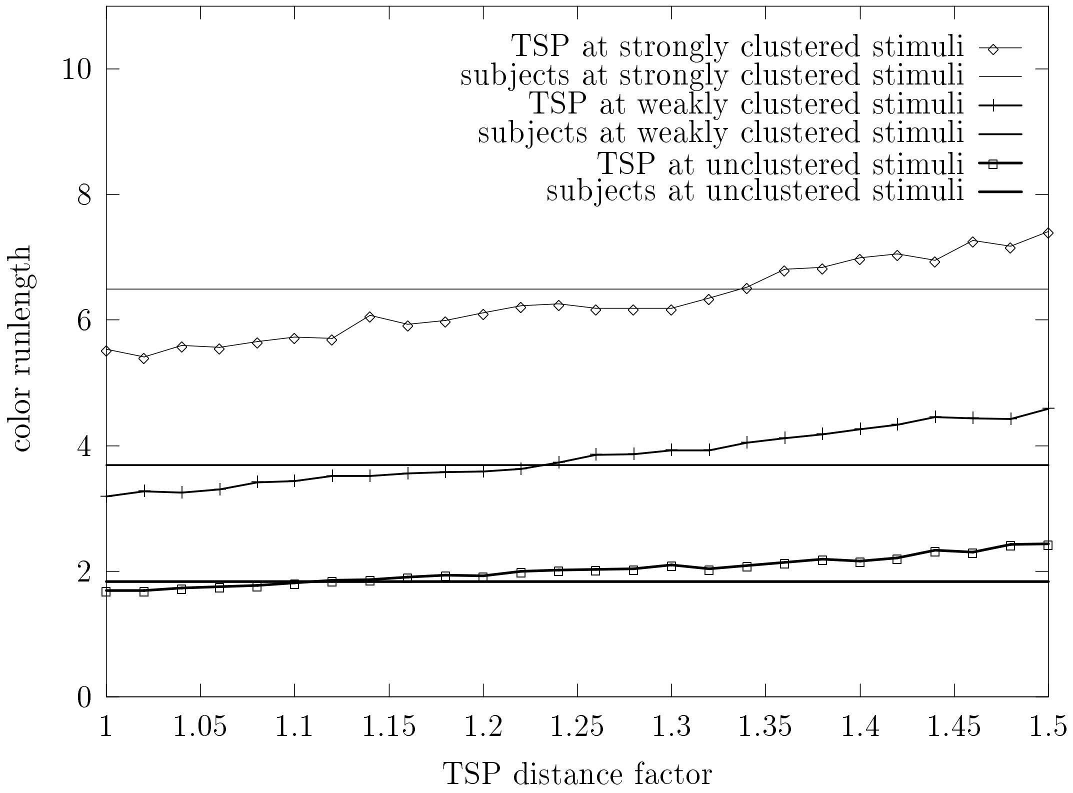


Figure 10.5: Color runlength generated by the TSP Model as a function of the type of stimulus and the introduced distance factor between items of different colors

It is both important and interesting to investigate whether the adjustment of color runlengths goes along with increased *similarity* of the TSP paths to the empirical scanpaths. Figure 10.6 shows a diagram of similarity which corresponds to Figure 10.5. Here, the lines of reference indicate the *optimal* similarity that is possible for the respective level of clustering. With an increasing level of clustering, these values diminish and hence signify a higher variability of scanpaths. This finding contradicts one of our initial hypotheses: We assumed the occurence of feature clusters to decrease the variability of scanpaths. However, the wide range of individual strategies (compare Figure 10.4) leads to substantial differences between scanpaths, especially with respect to large color clusters.

With the TSP Model, we find distinct maxima of similarity to the empirical scanpaths for the different levels of clustering. The “no clustering” level has its maximum at a distance factor of 1.09, for “weak clustering” it is 1.16, and “strong clustering” leads to a maximum plateau between 1.24 and 1.36. These values (except for the “strong clustering” condition) are slightly lower than those for the runlength intersections shown in Figure 10.5, but their relationship is clearly

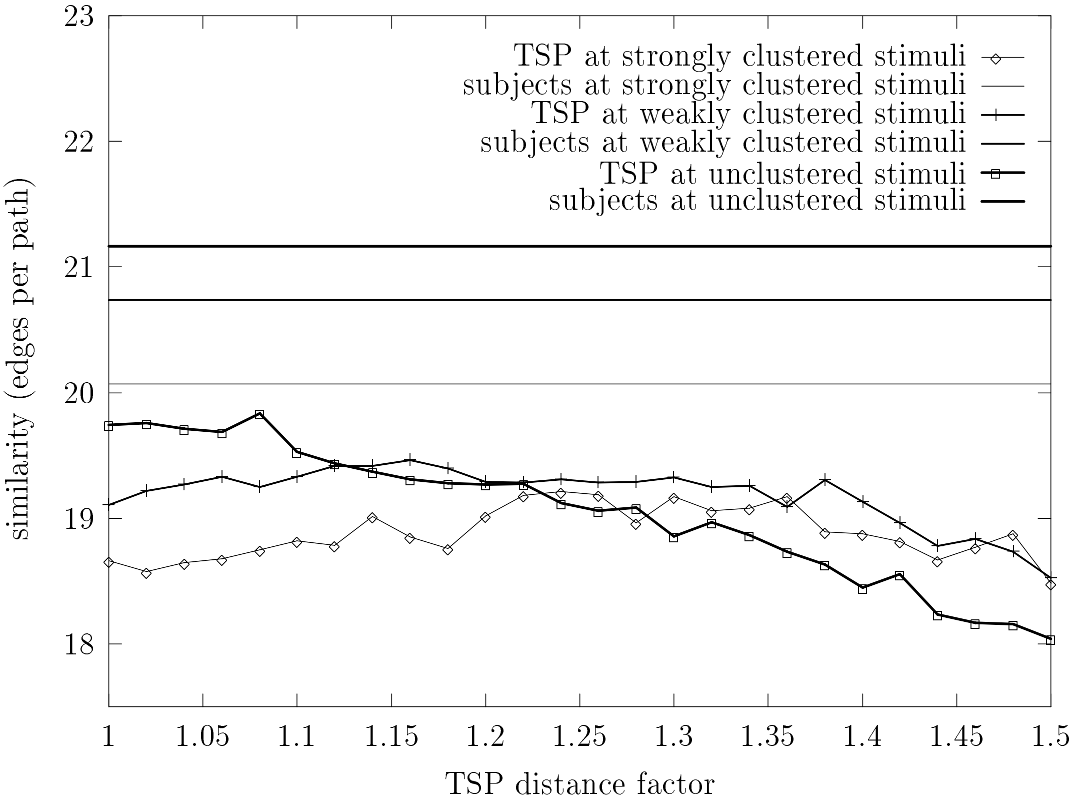


Figure 10.6: Similarity between empirical and TSP paths as a function of the type of stimulus and the distance factor

compatible with the runlength results.

Looking at the data obtained so far, we must consider if the introduction of a distance factor as described above is an adequate method of modelling the color effects. Since it needs different distance factors depending on the type of stimulus, we have to pose the question whether this approach is really plausible. An alternative idea consists in using a distance factor for *sequences* of transitions rather than for single transitions. Starting with the value 1.0, the distance factor for a whole group of successive transitions within the same color decreases linearly with the number of items in that group. This arrangement makes the choice of longer color runs increasingly attractive to the TSP algorithm. However, the results turn out to be in some way inverse to the previous ones: For increasing levels of color clustering, the alternative method needs *decreasing* “rewards” for long color runs in order to produce scanpaths of good similarity to the empirical ones.

In order 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 modelling is, however, to use as few freely adjustable parameters

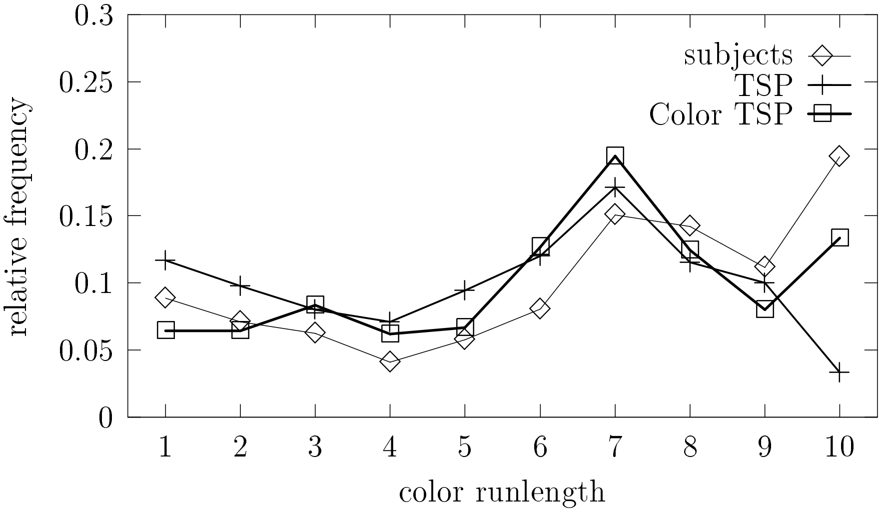


Figure 10.7: Histogram of color runlength in paths of subjects, TSP, and Color TSP in strongly clustered stimuli

as possible, because the more of these parameters are integrated into a model, the more likely it is to yield reasonable results even without being adequate. Therefore, we should keep our desired model, which we could call *Color TSP Model*, as simple as possible. An acceptable idea is to extend our initial approach to derive this model. Figures 10.5 and 10.6 suggest a linear dependence of the required distance factor on the strength of color clustering. We remember that the three levels of color clustering correspond to the values 1.0, 1.3, and 1.7 respectively, of the cluster measure. An optimization algorithm discovers the parameters of this linear relationship to yield paths of best similarity to the subjects' scanpaths:

$$\text{TSP distance factor} = 0.264 \times \text{color clustering} + 0.799 \tag{10.1}$$

Some example paths generated by the resulting Color TSP Model are shown in the lower row of Figure 10.2. In fact, some subtle differences to the TSP paths (middle row) can be found that indicate a better correspondence to the empirical paths (upper row). This impression is reflected in the similarity values: While the TSP Model yields paths of the value 19.18, the Color TSP Model reaches 19.51.

At this point, it is interesting to take a look at the *distribution* of color runlengths. In the conditions of no or weak color clustering, the differences between the subjects, the TSP Model, and the Color TSP Model are rather small. Figure 10.7 presents the respective histograms for the condition of strong color clustering. We find that the runlengths one to six are overemphasized by the TSP Model, whereas the Color TSP Model gets closer to the empirical frequencies. The most obvious difference between the two models can be found at runlength ten, which is almost neglected by the TSP Model, but strongly represented in the paths of

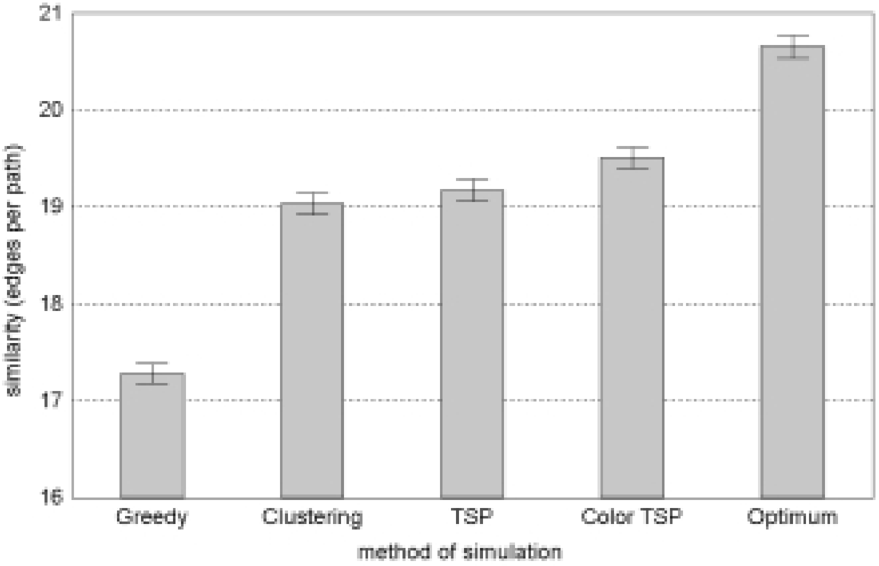


Figure 10.8: Similarity between the empirical scanpaths of Experiment H and those yielded by the different models

the subjects as well as in the Color TSP Model.

Finally, we should adapt the *Clustering Model* which was described in the context of the first study. This can easily be achieved by implementing the stimulus-dependent distance factor for both the first phase (calculation of clusters) and the second phase (cluster-based TSP) of the Clustering Model. The functional relationship between distance factor and color clustering in the stimulus which has been applied in the Color TSP Model (Equation 10.1) leads to optimal results for the Clustering Model as well. However, the Clustering Model does not yield better similarity to the empirical paths (19.03) than the Color TSP Model or even the TSP Model.

Figure 10.8 shows a survey of similarity values with regard to the empirical data, based on the similarity of 840 generated paths of each model to the 840 measured empirical scanpaths. Additionally, an analysis of variance was calculated. The Clustering Model turns out to yield a significantly higher value (19.03) than the Greedy heuristic (17.28) ($F(1;839) = 327.49; p < 0.001$), whereas its difference to the TSP Model (19.18) does not reach significance. The Color TSP Model, though, reveals a significantly higher value (19.51) than the TSP Model ($F(1;839) = 28.09; p < 0.001$). The optimal value (20.65), however, is still significantly better than the Color TSP value ($F(1;839) = 197.96; p < 0.001$). All other contrasts are significant as well.

As a conclusion, we can state that our Color TSP Model is a clear improvement over the standard TSP Model. Moreover, it constitutes the best scanpath model for items with color and form attributes that has resulted from our research; nevertheless, it does not yield optimal paths. It uses just two parameters that have been chosen in accordance with empirical data. The same is true for the Clustering Model, which does not yield better paths than the “color-blind” TSP Model. The difficulties encountered in surpassing the plain TSP Model indicate that the geometrical optimization of scanpaths is the main common principle of human strategies, even if additional color and form information is provided.

Chapter 11

The Three-Level Model of Comparative Visual Search

11.1 Requirements for an Adequate Model

As shown in Chapter 8, the direct reproduction of low-level eye-movement parameters is insufficient for a complete simulation of comparative visual search. Although immediate replication of statistical properties of eye-movement variables 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 would clearly be more comprehensible to use a model that incorporates only these interpretations instead of the raw empirical data. This model should generate fixations and saccades on the basis of assumed mental processes and their parameters (strategies, visual acuity, attention, working memory etc.) derived from the results of Experiments A to H. If the model is able to produce scanpaths that are similar to the empirical ones, it supports our interpretations and hypotheses, otherwise these have to be revised.

Which of the empirical results obtained in Experiments A to F should be replicated by our model? The desirable set of simulated effects includes all findings that have contributed to the main conclusions about perceptive and cognitive processes during comparative visual search. The following list shows all effects to be modelled as well as the corresponding interpretations to be tested:

Reaction time (RT): The empirical distribution of RT, in particular the characteristic plateau between approximately three and ten seconds, should be reproduced. Supported by the results of the Random-Walk Model, this plateau can be considered an indicator for *structured search strategies*. Moreover, the relative changes of mean RT between experiments are important to be replicated as well. Although in many cases mean RT values are not significant (e.g. see Chapters 4 and 6), they can indicate *differences in search efficiency* between experiments as well as *between color and form search*. Except for Experiment D, RT should be shorter for color search than for form search.

Fixation duration (FD): FD should depend linearly on the local item density at the fixation point, while it should not show any dependence on the local color or form entropy. These important findings have been viewed as evidence for *two distinct visual subsystems* (localization versus identification, see Chapter 3), contributing to the search process in different ways. FD should *not* vary between experiments or between color and form search, because no such effects have been found. Obviously, a constant mean duration of about 200 ms leads to optimal search efficiency, regardless of the variant of comparative visual search. The situation is different for three-dimensional stimuli (Experiment G). However, since the current model is restricted to two dimensions, we can consider mean FD a fundamental, invariant parameter of comparative visual search. Finally, FD should be considerably longer in the final phase of detection and verification than in the phase of search and comparison.

Saccade length (SL): Increasing local item density at the saccade's starting point should decrease SL, because higher local concentration of information requires shorter saccades for its inspection. If subjects are informed about the relevant dimension for target detection (Experiments B to D), SL is slightly longer during color search than during form search. Specific color search enables the subjects to perceive and process the information of more items within a single fixation. Furthermore, SL should be longer in the simulation of Experiment E than of Experiments A to D because the mirror symmetry between the hemifields goes along with partially longer distances between corresponding items. These longer distances, in turn, cause imprecise saccades and long within-hemifield corrective saccades which increase SL. Additionally, the distribution of SL with regard to stimuli of translational symmetry (cf. Figure 3.5) and mirror symmetry (cf. Figure 5.6) should be reproduced. During the verification phase, SL should be shorter than during the search phase.

Number of successive fixations (FW): FW should decrease with increasing local item density at the first fixation point after switching between the hemifields. This effect is complementary to the findings concerning FD: In regions of higher item density, more information can be perceived within single fixations, hence fixations become longer. Assuming a constant capacity of working memory, fewer fixations are necessary to "fill up" working memory, i.e. FW declines. During specific search (Experiments B to D), FW is significantly higher for form search than for color search. Conceivably, less information per fixation can be processed during form search, demanding more fixations to reach optimal memory usage. Since FD remains constant, the processing of form information seems to take longer than the processing of color information, which is compatible with the results obtained in the pre-studies (see Chapter 2). Finally, the FW values should be lower in the verification phase than in the search phase.

Area coverage per fixation (AC): Corresponding to SL, AC should decrease with increasing local item density and local color and form entropy. These effects are stronger for AC than for SL, because SL is only an indirect measure of search efficiency. Saccades of the same length can cover small or large areas, depending on the *angles* between them. AC, however, directly corresponds to the size of the area being processed within a fixation. In Experiments B to D, where subjects know the relevant target dimension, AC should be smaller for form search than for color search, which is compatible with SL as well. Moreover, the "double stairway" shape of AC with regard to cluster sizes one to six in Experiment F should be replicated. This finding was interpreted as evidence for *three items* being the optimal cluster size applied in comparative visual search.

Speed of processing (SP): In Experiment A, SP should not present a difference between color and form search, since subjects are not informed about the type of mismatch. The SP value for form search does not change from Experiment A to Experiments B and C, while SP considerably increases with respect to color search. This finding suggests that color processing is faster than form processing and that form processing may imply color processing to some extent (see Chapter 4). Moreover, SP should be lower in Experiments D and E in comparison to the corresponding Experiments B and A respectively. The increased task complexity in Experiments D and E forces the subjects to memorize fewer items at a time, resulting in reduced SP. Compatible with the requirements for AC, SP should present a "double stairway" appearance in Experiment F.

Probability of missing the target (PM): PM should decrease with increasing local item density at the target items. This effect, which is observed in all experiments, can be attributed to the increased memory load in regions of high item density. The fewer items are memorized at the same time, the easier it is for subjects to detect a possible target. In Experiment D, color targets induce higher PM than do form targets, in all other experiments the situation is reversed. Generally, the feasibility of holistic comparison of local information leads to high search efficiency and low PM for specific color search (Experiments B and C). Experiment D allows specific search as well, however, the detection of matches is substantially more difficult than the detection of mismatches. Under these conditions, subjects seem to overestimate their own capacities, because they still proceed faster during color search which causes increased PM.

To some extent, this list does not only include the main aspects of results we expect our model to reproduce, but it also contains information about important perceptive and cognitive processes to be modelled. In the following section, we will combine the list with further considerations in order to derive an appropriate structure for our improved model.

11.2 The New Concept of Modelling

First of all, a fundamental decision with regard to our model has to be made. In Section 8.1 we discussed the distinction between *neural* and *phenomenological* models. Which of these types is favorable with respect to our improved model?

In order to test the assumed interactions of different mental processes during comparative visual search, a phenomenological model seems to be sufficient. If we want to test the physiological, i.e. neural, plausibility of our assumptions, however, a neural model is inevitable. Since only a few speculations about processes on the neural level have been made in this work, it is reasonable to start with a phenomenological model. The investigation and modelling of neural processes will be achieved in future research.

As argued in Chapter 8, 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. This means that an adequate model should incorporate a *vertical organization* of mental processes, i.e. a hierarchical scheme of functional modules, corresponding to the human "brain architecture" as far as we know it (see Velichkovsky, 1990).

The main shortcoming of the Random-Walk Model is the exclusion of high levels, leading to unstructured search behavior. In contrast, the subjects use a self-avoiding global scanpath which helps them optimize their search efficiency.

Another inadequacy of the Random-Walk Model is its restriction to the centers of items as possible fixation points. Only a small proportion of empirical fixations is located close to the center of an item. Nevertheless, is the assumption of item-to-item scanpaths really implausible? The discussion of strategies applied in Experiment A (Chapter 3) suggests a possible *dissociation* between attentional scanpaths and gaze trajectories. As mentioned in the introduction (Section 1.2) and tested in one of our pre-studies (Chapter 2), shifts of attention can be performed without moving the eyes. This effect was studied in many different experiments (Posner, 1980; Wright & Ward, 1994), indicating that, during a fixation, attention can be moved with a certain velocity. One of the authors (Tsal, 1983) proposed a constant velocity of one degree per eight milliseconds, however, his experiments were based on rather coarse measurement and were not perfectly consistent, so this result should be viewed with discretion.

Consequently, it is reasonable to assume that *attention* is shifted between single items. The *eye movements*, however, do not exactly correspond to these attentional movements. In general, the fovea, i.e. the measured gaze position, follows the focus of attention in order to provide sufficient visual acuity for item recognition. If two or more items are located close to each other, it is not necessary to successively fixate each of them. Rather, a single fixation near the center of the item group enables subjects to perceive all relevant information. While fixating this point, subjects shift their attention between all items in the group in order to memorize them or to compare them to the memorized ones. This method has the advantages of avoiding a time-consuming execution of multiple, short saccades

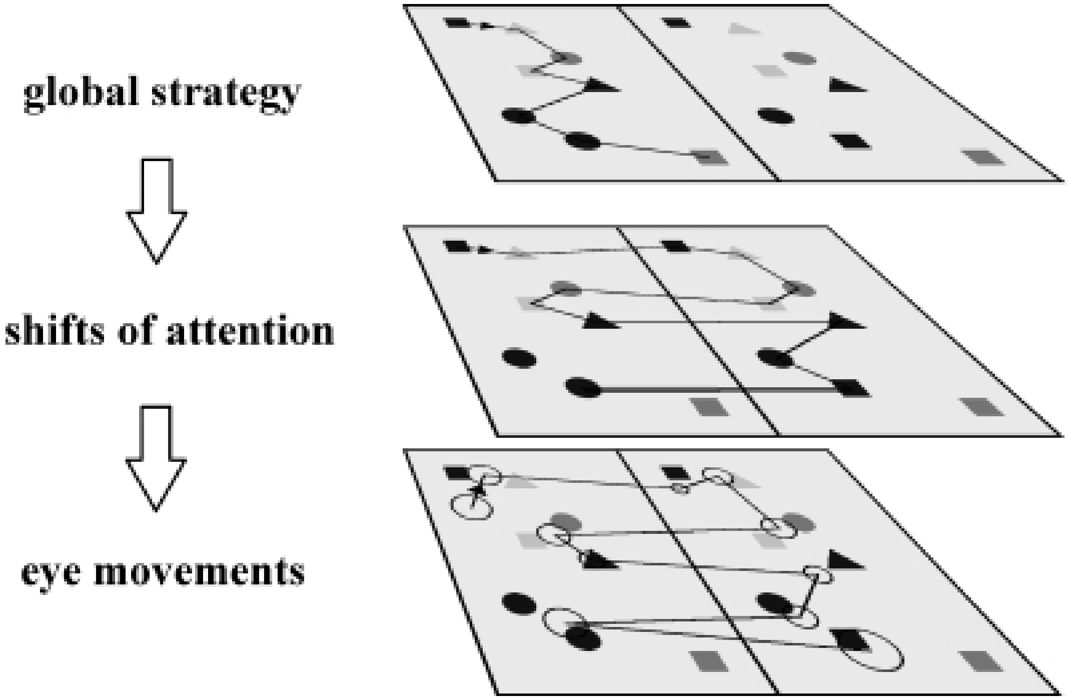


Figure 11.1: Scheme of the Three-Level Model. The example stimulus contains only eight items per hemifield for the benefit of a clearly arranged illustration.

and providing a stable perceived image during the processing of information. Hence, in most cases single fixations do not correspond to single items, although the attentional scanpath visits all items in a serial, regular fashion.

These considerations motivate a further refinement of the aforementioned vertical organization of mental processes involved in comparative visual search. The proposed lower level should be split up into two distinct levels, one containing the attentional processes including the use of working memory, the other consisting of those processes that realize the eye movements.

Figure 11.1 presents the structure of our resulting *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 scanpath 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 on this intermediate level. Finally, the lower level is responsible for executing eye movements. The gaze position always follows the attentional focus in order 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.

It is difficult, however, to model the final *verification phase* in the context of these levels of processing. Since there is only little empirical information available, it does not seem reasonable to speculate about the processes underlying this verification phase in the context of the Three-Level Model. Thus, the simulation of this phase is adopted from the Random-Walk Model. Since this model is an exact statistical reproduction of empirical low-level data distributions (saccade length, fixation duration etc., see Chapter 8), the parameters during the verification phase are disregarded in the analysis of scanpaths provided below. Further investigations on the relation between the search phase and the verification phase will have to be carried out in order to refine the Three-Level Model in this respect.

The integration of the three individual levels into a single model is described in the following sections.

11.2.1 The Upper Level: Global Strategy

The model's global scanning strategy is based on the Color TSP Model developed in Chapter 10. Thus, the value of color clustering for the left hemifield – serving as a reference – is calculated in order to determine the appropriate distance factor for the Color TSP with the help of Equation (10.1).

Which item should be chosen as the starting item? Since most subjects tend to start their search at the top of the display, it seems adequate to consider the five items with the uppermost positions in the left hemifield as relevant. Another cognitively plausible requirement for the starting item is to allow the construction of a short scanpath, because the results of Experiment H suggest that *minimal length* is the most important criterion for the choice of empirical scanpaths. Consequently, the Three-Level Model calculates a Color TSP path within the left hemifield for each of the five relevant starting items. The path of minimal length is chosen as the model's global scanning strategy for comparative search.

As a matter of course, the assumption of a complete item-by-item scanpath to be developed immediately after stimulus onset is implausible. In contrast, subjects are likely to start searching with a very coarse strategy in mind and to refine it "on-line" to a locally item-based scanpath. The resulting scanpath, however, might be the same in both cases. We do not understand the dynamic *development* of global scanpaths so far, but we know some features of the static *results*. This knowledge constitutes the basis for the Color TSP Model and we assume it to be transferable to comparative search. Hence, it is legitimate to anticipate the generation of global scanpaths in the Three-Level Model.

11.2.2 The Intermediate Level: Shifts of Attention

Attention is modelled 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, a new global scanpath is calculated (cf. Section 11.2.1) with the item in focus serving as the starting item. The search process, as described in this section, continues, guided by the new scanpath. 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 the results of Experiment F, the maximum number of objects to be memorized at a time is set to three for most experiments (see below). This maximum number can only be reached as long as the group of simultaneously memorized items can be perceived at the same time. The empirical data show that subjects generally memorize spatially small groups of items at a time, conceivably in order to execute *holistic* comparisons (see Chapter 5). Thus, the Three-Level Model assumes a specific *radius of attention.* All items to be memorized must fit into a circle of this radius, which is the same for both unspecific search and form search. According to an estimation based on the empirical area coverage per fixation measured in Experiment A, the radius was set to 30 pixels. The radius for specific color search was defined as 37 pixels, as suggested by the differences in efficiency measures between color and form search in Experiment B.

In the case of item *clusters* (Experiment F), another constraint is imposed on the model's maximum number of items to be memorized: It is not possible to memorize items of different clusters at the same time. This restriction leads to a clusterwise scanning strategy.

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 F indicate that the number of required between-hemifield saccades strongly depends on the *number of memorized features.* The number of memorized features is the number of different colors plus the number of different forms contained in the set of items that are currently stored in memory. Memorizing three green triangles at the same time, for instance, means *two* memorized features, while memorizing a green triangle and a yellow square at the same time corresponds to *four* memorized features. Moreover, the asymmetry between color and form processing indicated by Experiments B and C suggests form information to be less relevant for color search than is color information for form search. Simplifying this asymmetry, form features are not added to the number of memorized features in the case of specific color search, whereas color features are added in any case. Approximating the data obtained in Experiment F, the number of between-hemifield saccades is determined by the following linear equation:

$$\text{number of saccades} = 0.225 + 0.3875 \cdot \text{number of memorized features} \tag{11.1}$$

The resulting values are rounded by a weighted random function. A result of 1.775, for example, is rounded to 2 with a probability of 77.5% and rounded to 1 with a probability of 22.5%. This number of between-hemifield saccades is integrated into the process of comparison at random points in time.

In accordance with Tsal's (1983) results, the speed of attentional shifts was chosen: It is constant at one degree of visual angle per eight milliseconds. The pre-tests (see Sections 2.2.2 and 2.2.3) indicated 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 unspecific search. Since it is impossible to derive precise quantitative information about this span from the empirical data, we adjusted it in such a way that the resulting fixation duration is similar to the empirical one. 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.

In order to model the effects of the factors *experiment* and *type of target*, we introduced two parameters that characterize the influence of these factors:

Memory load: This is a measure of the difficulty to compare local groups of items between the hemifields. The memory load is set to 1 for Experiments A, B, C, and F, because all of them allow for *holistic* comparison of local information, i.e. no mental transformation is required. For Experiments D and E, the information differs between the hemifields in these situations. Hence, we assumed the memory load to be twice as high as in the other experiments and accordingly set the memory load to 2.

Target detectability: Basically, the detectability of a *mismatch* depends on its dimension. According to the empirical probability of missing the target, the target detectability is set to 3.3 for color mismatches and to 2.7 for form mismatches. Mirror symmetry between the hemifields (Experiment E) impaires the target detectability, as indicated by the empirical data. As a consequence, the values of target detectability are reduced by 30% in Experiment E (2.31 and 1.89 respectively). *Match* detectability, however, does not seem to depend on the relevant dimension. Target detectability in Experiment D is thus set to 1.89 both for color and form matches.

What effects do these parameters have on the modelled shifts of attention? Target detectability exerts only one effect: It determines the probability of target detection. Additionally, this probability depends on the *number of memorized features* (see above) at the moment of comparing the target items to each other, i.e. scanning one of the target items while the other is kept in memory. As indicated by the results of Experiment F, the number of memorized features is closely related to the usage of working memory. Of course, the more features are memorized at the same time, the lower should be the probability of target detection. In the model, the probability of target detection is simply defined as target detectability divided by the number of memorized objects.

The other experiment-dependent parameter, memory load, determines the amount of information to be processed at a time. The maximum number of items to be memorized at a time is divided by the memory load which means that, in the case of memory load 2, this maximum number varies randomly between 1 and 2. Furthermore, the radius of attention is divided by the memory load. These two definitions make the model memorize smaller bits of information in situations of higher memory load, reflecting our interpretation of the respective empirical findings discussed in Chapter 5. In turn, the span of attention spent on a single item is multiplied by the memory load, because we assume processes of comparison to require more time if the memory load increases. Additionally, this adjustment is important for the invariance of fixation duration to the factors experiment and type of target which is one of the main empirical "landmarks" in comparative visual search.

11.2.3 The Lower Level: Eye Movements

As explained above, the eye movements 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 but one item cannot be processed within the same fixation due to a long distance separating the two items; this distance might be estimated by extrafoveal vision. 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.

Since the situation of mirror symmetry in Experiment E leads to longer within-hemifield saccades than does translational symmetry in Experiment A, the probability of a saccade to be imprecise is assumed to increase with the distance to be covered by the saccade. Accordingly, a statistical standard deviation between the intended target of the saccade and the actual end point of the saccade ("saccadic noise") is defined in such a way that an increasing distance to be covered causes higher saccadic noise. The following simple equation was found to yield values of saccadic noise that resemble the empirical data:

$$\text{saccadic noise} = \frac{\text{distance}^3}{1.2 \cdot 10^6 \;\; \text{pixels}^2} \tag{11.2}$$

Interestingly, neither the assumption of a linear nor a quadratic dependence of the saccadic noise on the distance is able to explain the empirically hypothesized effects of saccadic noise.

Equation (11.2) is individually applied to both the horizontal and the vertical distance to be covered by a saccade. Subsequently, a random value is calculated on the basis of a Gaussian random distribution with the obtained value of saccadic noise taken as its standard deviation. This random value is *subtracted* from the ideal saccade length, i.e. the length that would move the gaze position exactly onto the intended item. This implementation of imprecise saccades is, of course, not plausibe in that the lengths of saccades tend to be too short and can never be too long. Nevertheless, this model of saccadic noise reflects to some extent the properties of empirical saccades: They are likely to be shorter than the distances between corresponding items.

In the situation of item clusters (Experiment F), saccades can be assumed to be more precise than in the other experiments due to the additional structural elements in the display. Accordingly, in Experiment F the saccadic noise is set to 50% of its value defined by Equation (11.2).

If a saccade does not enter the radius of attention of the intended item or any of the items to be compared, another, *corrective* saccade is executed aiming at the same target. This procedure may be repeated until a valid gaze position is reached. Between the saccades of this kind, *orientational* fixations are executed which have a random duration between 90 and 110 ms.

The saccadic noise allows us to easily integrate another empirical observation into the Three-Level Model, namely the fact that sometimes subjects seem to compare the memorized items in reverse order for the benefit of a shorter saccade between the hemifields. In the model, the order of items to be compared is reversed if the first fixation point in the relevant hemifield – due to saccadic noise – is closer to the last item than to the first item in the intended sequence of comparisons. Thus, we model the inversion of the steps of comparison as a *result* of saccadic noise. The empirical data do not tell us whether this view is correct.

The duration of saccades is modelled as a linear function of the respective saccade length (see Equation (8.1) in Chapter 8).

Finally, the empirical error in the spatial eye-movement measurement is simulated as well. The eye-movement data are randomly shifted in accordance with the average error of OMNITRACK1 (see Section 2.1.1). This feature of the Three-Level Model improves the comparability between empirical and simulated gaze trajectories.

11.3 Artificial Versus Empirical Scanpaths

Does the Three-Level Model meet the requirements listed in Section 11.1? In order to answer this question, the model was "presented" with 10 000 randomly generated stimuli for each of the relevant Experiments A to F. This amount of simulation data is about ten times larger than the empirical database. Accordingly, the standard error in the measurement of dependent variables is considerably smaller, making it possible to rate the model's performance without calculating analyses of variance. The eye-movement data-files generated by the computer implementation of the model have exactly the same format as those yielded by the OMNITRACK1 system. Hence, the same programs for data extraction as in Experiments A to F could be used.

First of all, we can take a look at some example scanpaths of the Three-

Level Model. Figure 11.2 presents such an example with regard to Experiment A. As can clearly be seen, this scanpath resembles the empirical ones more closely than do the paths generated by the Random-Walk Model: Both a structured search strategy (top-down scanning) and grouping processes (e.g. within fixations number 14 to 16) can be observed. Neither of these important features could be achieved with the Random-Walk Model.

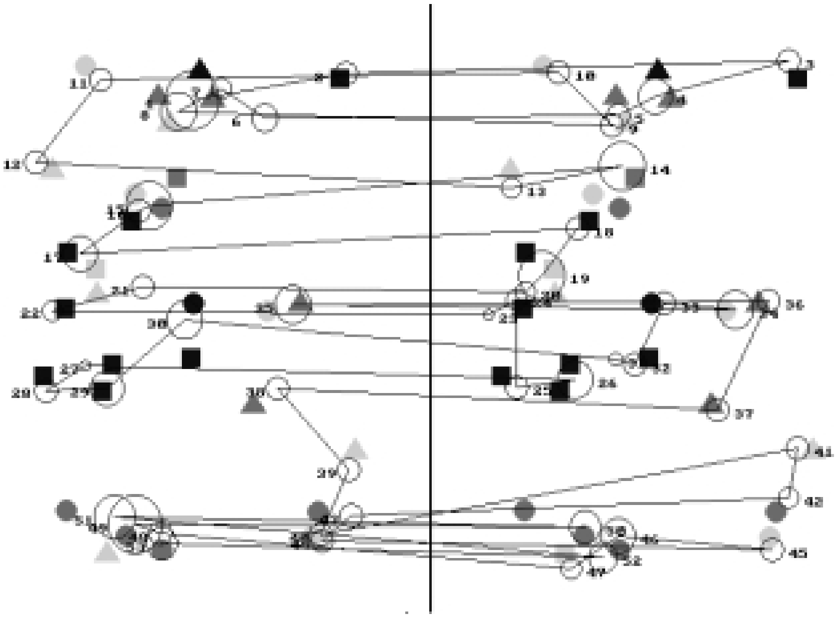


Figure 11.2: Example scanpath generated by the Three-Level Model in the situation of Experiment A. Fixations are numbered; circle size corresponds to fixation duration.

A different stimulus geometry is provided by Experiment E, namely mirror symmetry between the hemifields. Figure 11.3 shows a respective example scanpath generated by the Three-Level Model. Again, the display is searched in a top-down fashion. As a matter of fact, the lengths of saccades between hemifields are adapted to the varying distances between corresponding items. Fixation number 12 demonstrates the occurrence of an imprecise saccade caused by a long distance to be covered. Here, a corrective saccade has to be executed in order to focus on the intended item. The lower part of this stimulus is not covered by the plotted gaze trajectory, because the target is already detected after scanning about half of the items.

Figure 11.4 illustrates an example scanpath on clustered items, i.e. in the situation of Experiment F. The chosen stimulus contains six clusters of size four and – since 30 cannot be divided by four – two clusters of size three. The model starts scanning at the top of the display and wanders down cluster by cluster. This strict clusterwise scanning, however, is not a result of modelled perceptive influences of the connecting edges. The strategy is rather "forced" by making all clusters "invisible" to the model, except for the currently processed cluster.

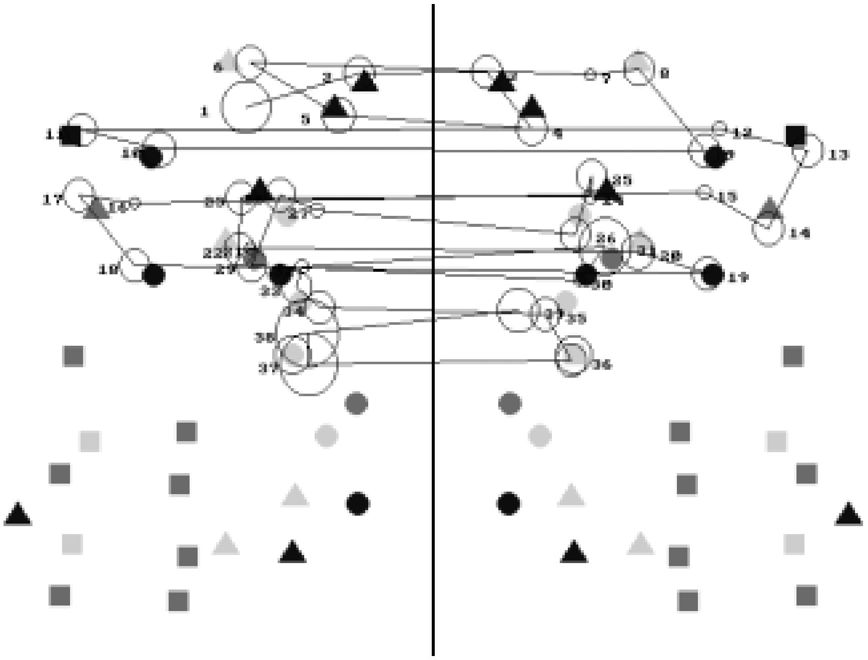


Figure 11.3: Example scanpath generated by the Three-Level Model in the situation of Experiment E (mirror symmetry)

Nevertheless, the generated scanpaths are undoubtedly similar to the empirical ones.

More important than this qualitative, superficial inspection of scanpaths yielded by the Three-Level Model, an extensive quantitative analysis is necessary in order to rate the model's performance. In the following paragraphs, the results of this analysis are shown in comparison to the corresponding empirical data. It should be noted that the mean values of the same variable may slightly vary between different statistics due to different designs of the respective analyses.

11.3.1 Reaction Time (RT)

Figure 11.5 shows the distribution of RT in Experiment A for both the subjects and the Three-Level Model. As can clearly be seen, the plateau in the empirical data between three and ten seconds is qualitatively replicated by the model, indicating a structured search strategy that may be similiar to the subjects' strategies. The decrease in the relative frequency after this plateau, however, seems to present a steeper slope for the model data than for the empirical data. This finding suggests that the model does not sufficiently reproduce the considerable variability of scanpaths between subjects, especially with regard to search efficiency. Consequently, the "noise" in the empirical data is higher, leading to slighter slopes in the respective histograms.

As illustrated in Figure 11.6, the RT values for Experiments A to E are almost identical for the subjects and the model. The only difference might be that the

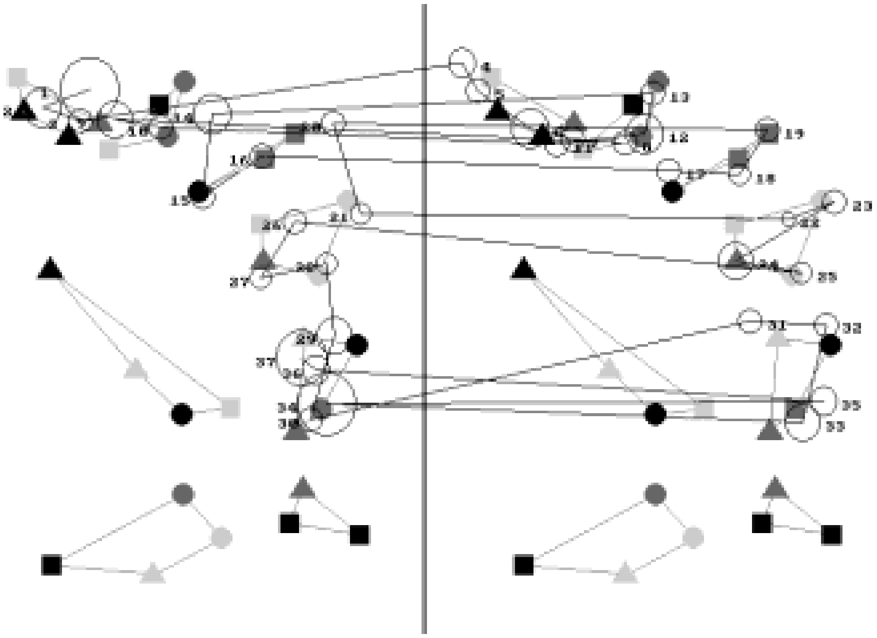


Figure 11.4: Example scanpath generated by the Three-Level Model in the situation of Experiment F (clustered items)

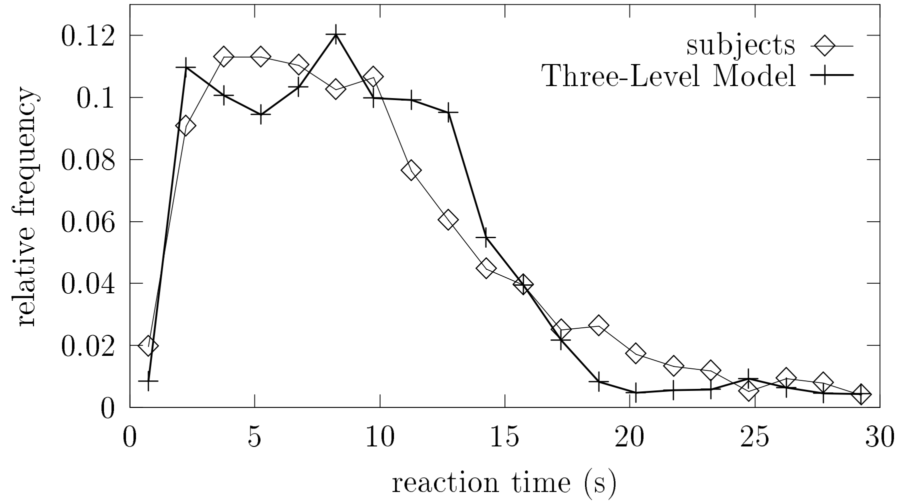


Figure 11.5: Histograms of reaction time in Experiment A

model does not yield shorter RT for form search in Experiment B than in Experiment A. In other words, the model does not take advantage of being informed about the relevant dimension. Since the corresponding difference in the empirical data does not reach significance, however, it is not clear whether such an effect should be modelled. All in all, the RT results are in line with the specified requirements.

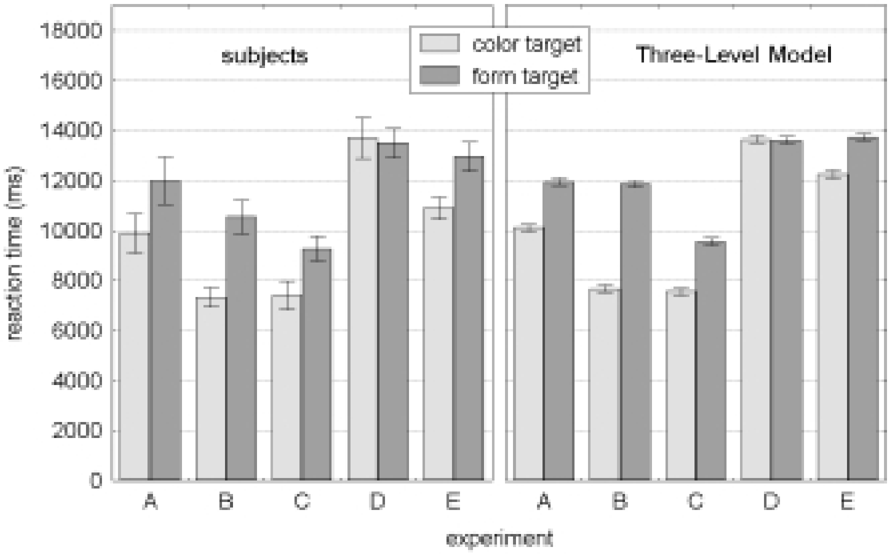


Figure 11.6: Reaction time in Experiments A to E

11.3.2 Fixation Duration (FD)

The diagram in Figure 11.7 shows FD as a function of local item density at the fixation point in Experiment A. In fact, the simulated fixation duration increases approximately linearly with item density. While this increase corresponds to the empirical data for density values above 1.7, it is slightly steeper for lower density. It might be possible that there is a constant "basis duration" of empirical fixations in addition to the duration caused by information processing. 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 11.8 presents FD in dependence on color and form entropy at the fixation point in Experiment A. Obviously, simulated FD is neither significantly influenced by color entropy nor by form entropy; the same is true for empirical FD. Hence, the structure of the Three-Level Model is compatible with the fundamental parameters of information processing indicated by empirical FD.

As shown in Figure 11.9, FD does neither vary substantially between experiments, nor between color and form search. This result is not surprising, since the constancy of FD – as a "landmark" in comparative visual search – was taken as a basis for the adjustment of some of the model's parameters.

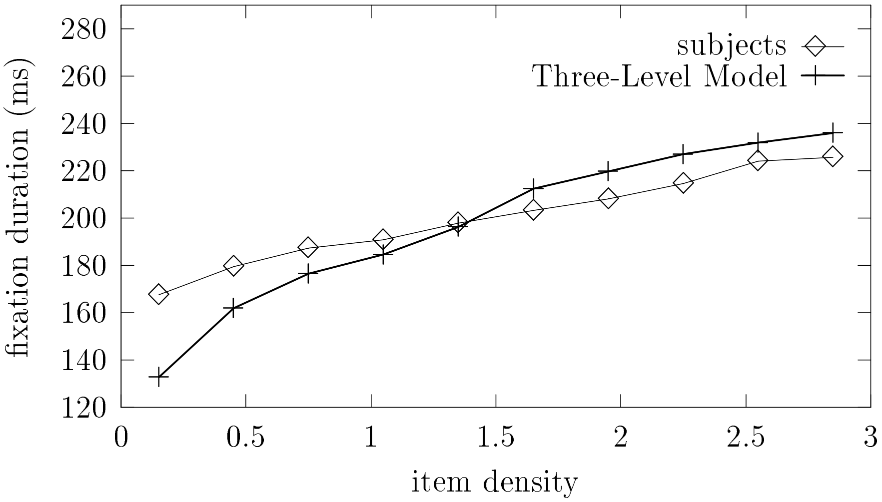


Figure 11.7: Fixation duration in Experiment A as a function of local item density

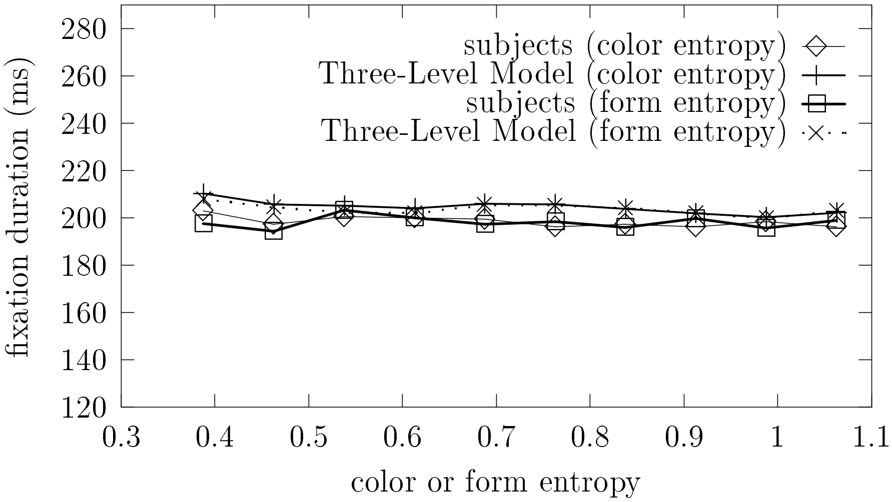


Figure 11.8: Fixation duration in Experiment A as a function of local color and form entropy respectively

11.3.3 Saccade Length (SL)

Figure 11.10 presents SL as a function of the local item density at the saccade's starting point. The model data exhibit a close similarity to the empirical ones, however, at very low density (< 0.5), the Three-Level Model generates longer saccades than do the subjects. To some extent, this finding corresponds to the

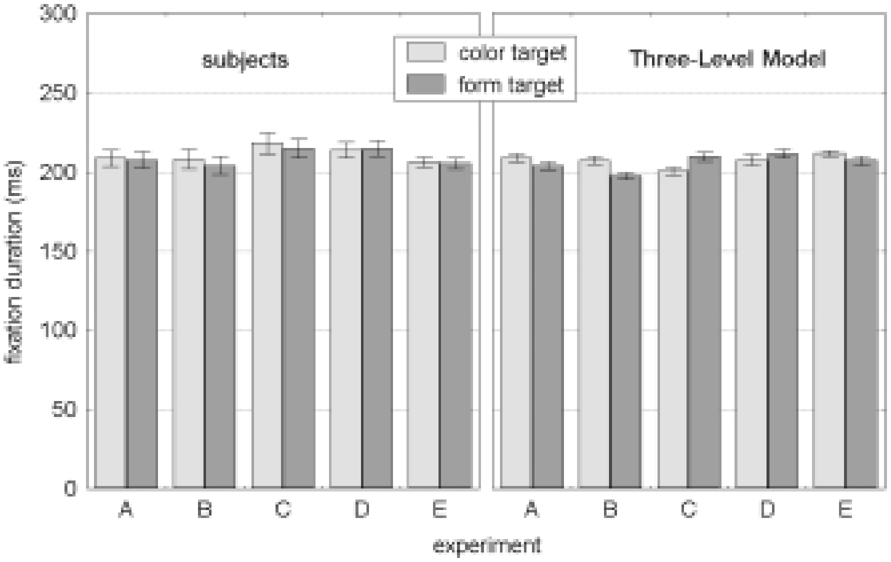


Figure 11.9: Fixation duration in Experiments A to E

results of FD (see Figure 11.7): In both cases, the model replicates the empirical data rather closely, but at low item density it seems to "exaggerate" the empirical effects. With regard to SL, this suggests that there could be inhibitory cognitive mechanisms that confine the within-hemifield saccades to a certain maximum length even if only little information is provided by the currently inspected region of the stimulus. Maybe such a confinement is necessary in order to preserve a structured scanning strategy.

Between the experiments, SL presents a pattern of results that resembles the empirical data (see Figure 11.11): If the relevant dimension is known (Experiments B to D), SL is longer for color search than for form search. Furthermore, SL is longer in Experiment E than in Experiment A and the shortest SL is found in Experiment D. However, loosely speaking, the model seems to overemphasize some of these effects: The differences between color and form search in Experiments B and C as well as the difference between Experiments A to C in comparison to Experiment D are greater than the corresponding empirical ones. This finding indicates the cognitive processes that determine SL to be influenced less substantially by the experimental setting than are the simulated processes in the Three-Level Model.

When analyzing saccade length, we should not disregard the *between-hemifield* saccades. There are two different experimental situations leading to distinct distributions of between-hemifield saccade length: Translational symmetry (e.g. Experiment A) and mirror symmetry (Experiment E). Figure 11.12 shows the respective histograms for both the subjects and the simulation. Concerning Experiment A, the model replicates the distribution of empirical saccade length nearly perfectly; only the frequency "peak" at 350 pixels seems to be higher with the model than

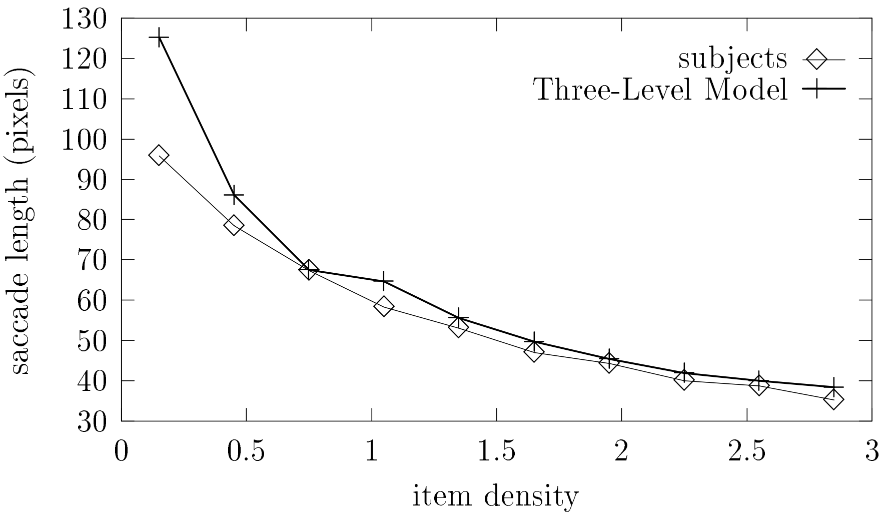


Figure 11.10: Saccade length in Experiment A as a function of local item density

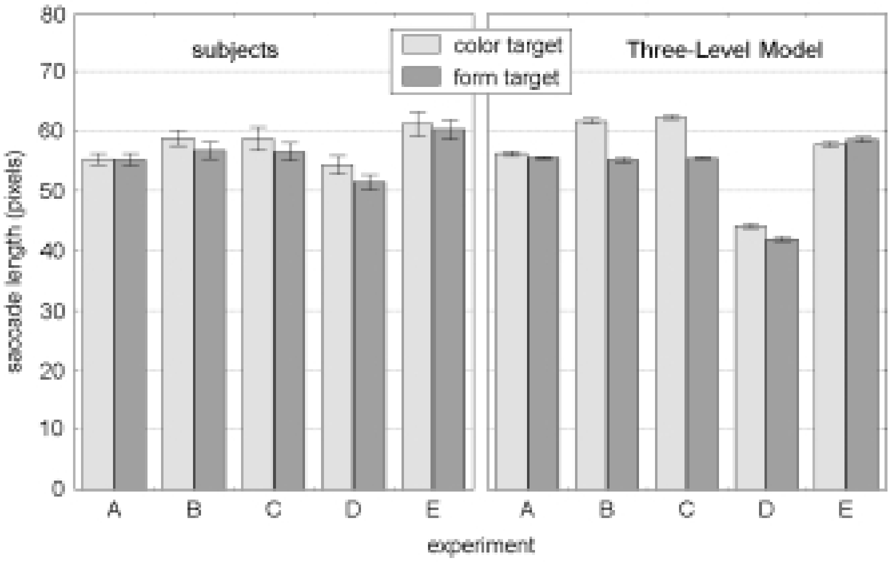


Figure 11.11: Saccade length in Experiments A to E

with the subjects. This effect can be considered analogous to the one obtained in the analysis of RT (Figure 11.5): The "noise" in the empirical data is higher than in the model data, hence the empirical histograms are affected by stronger blur.

As to Experiment E, the almost homogeneous distribution of empirical saccade length is closely reproduced by the Three-Level Model, although there seems to be a higher proportion of short saccades (< 300 pixels) in the simulation data than in the subjects' data. In particular, the model reproduces the frequency decay

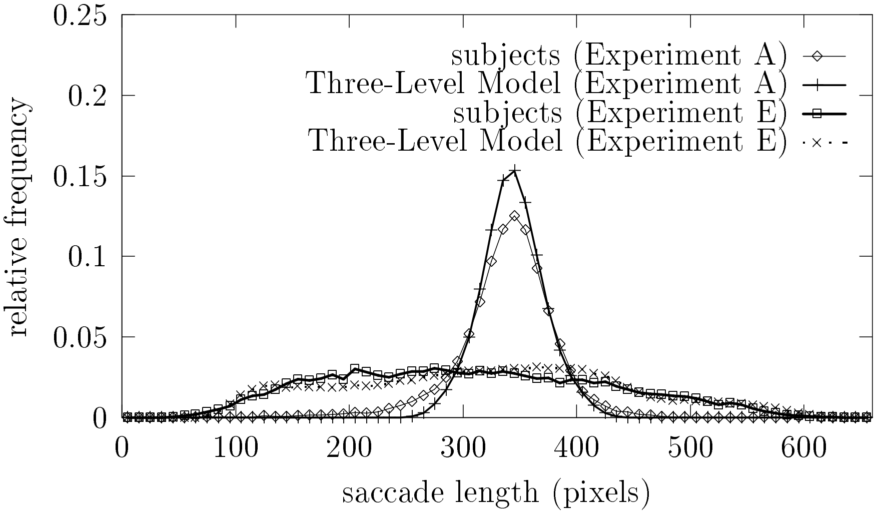


Figure 11.12: Histograms of between-hemifield saccade length in Experiment A and in Experiment E respectively

for long saccades rather precisely. This can be taken as evidence for both the occurrence of imprecise saccades and the subjects' tendency to execute saccades between the innermost points of corresponding stimulus regions to be responsible for the lack of saccades longer than 580 pixels in Experiment E. Altogether, the Three-Level Model seems to simulate the basic eye-movement variable saccade length in an adequate way.

11.3.4 Number of Successive Fixations within the Same Hemifield (FW)

Figure 11.13 shows FW for different item densities at the first fixation point in Experiment A. Generally, the effect of item density on the model's FW is slightly weaker than on the empirical FW, but the situation is reversed for low density values: Here, the effect on the model data is stronger. As observed in the analyses of FD and SL, the empirical effect of low item density is weaker than it could be expected on the basis of the Three-Level Model. For most values of item density, simulated FW is smaller than empirical FW.

This finding is reflected in Figure 11.14 which presents FW as a function of the factors experiment and type of target. Although the range of simulated FW values corresponds roughly to empirical FW, some shortcomings of the model can be found: In Experiment B, FW does not meet the requirement of being lower for color search than for form search. Moreover, in contrast to the empirical data, simulated FW is higher in Experiment E than in Experiment D. All in all, the results indicate that some basic factors that influence FW have been understood

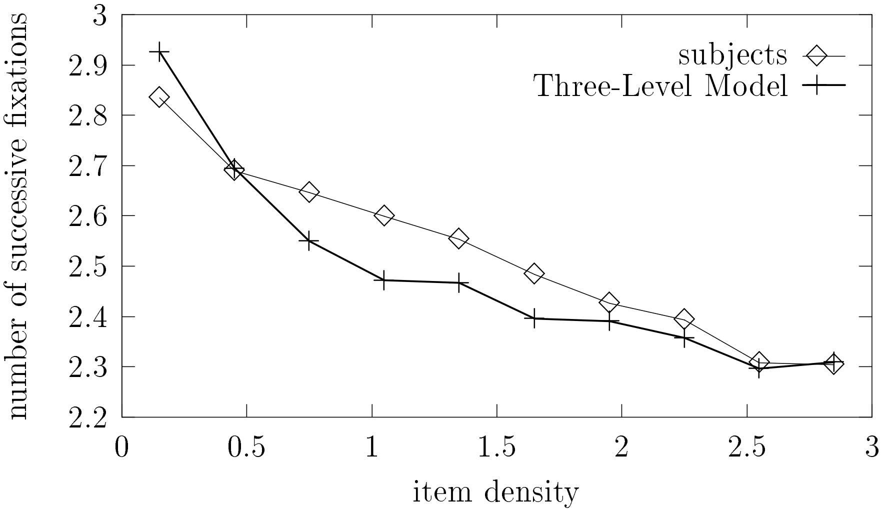


Figure 11.13: Number of successive fixations in Experiment A as a function of local item density

– as suggested by the dependence of modelled FW on item density –, while the role of other possible factors has to be reviewed.

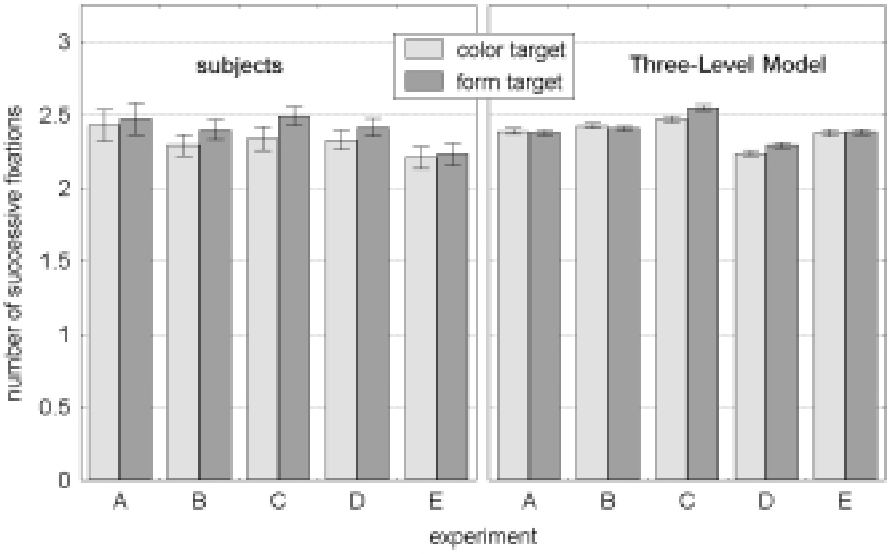


Figure 11.14: Number of successive fixations in Experiments A to E

11.3.5 Area Coverage per Fixation (AC)

The analysis of AC is of particular importance, because empirical AC is affected both by local item density and by local color and form entropy. Figure 11.15 presents AC as a function of local item density at the fixation point. The density effect seems to be stronger for the data of the Three-Level Model than for the empirical data. Again, low density values exert a stronger influence on the simulated than on the empirical data.

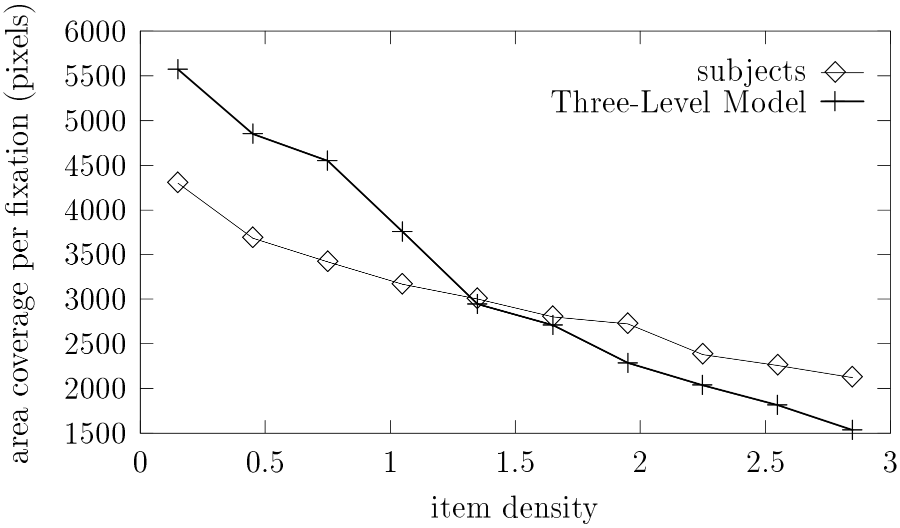


Figure 11.15: Area coverage in Experiment A as a function of local item density

As to color and form entropy, the effects on AC may be somewhat weaker for the model than for the subjects. Figure 11.16 shows that, in contrast to low item density, neither low color entropy nor low form entropy exerts particularly strong effects on simulated AC. Hence, we can not assume “inhibitory” cognitive processes to occur in low entropy regions as we do with regard to low density regions. Although the entropy effects in modelled search are possibly weaker than in empirical search, the factor entropy seems to be incorporated in the Three-Level Model in a plausible way, as the results of FD and AC indicate.

The AC values for different experiments and types of target look similar for the subjects and the model, as shown in Figure 11.17. However, there is one important discrepancy: The AC differences between color and form search in Experiments B and C, i.e. the feature-specific effects of bottom-up and top-down control of attention (see Chapter 4) do not match. Especially the empirical effect of top-down control (Experiment C) seems to be inverted by the Three-Level Model: In the simulation data, *form search* rather than *color search* benefits from top-down control, as indicated by the empirical data. Hence, the role of AC in the investigation of control of attention has to be examined more closely.

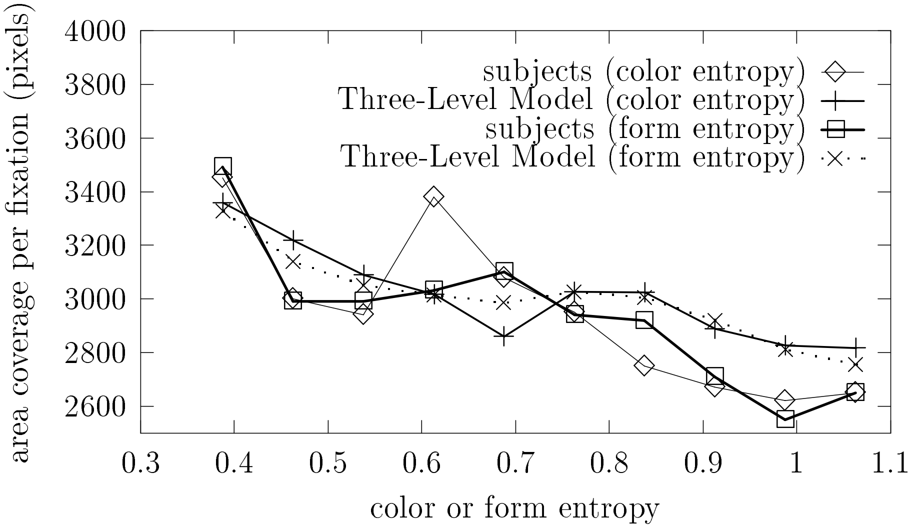


Figure 11.16: Area coverage in Experiment A as a function of local color and form entropy respectively

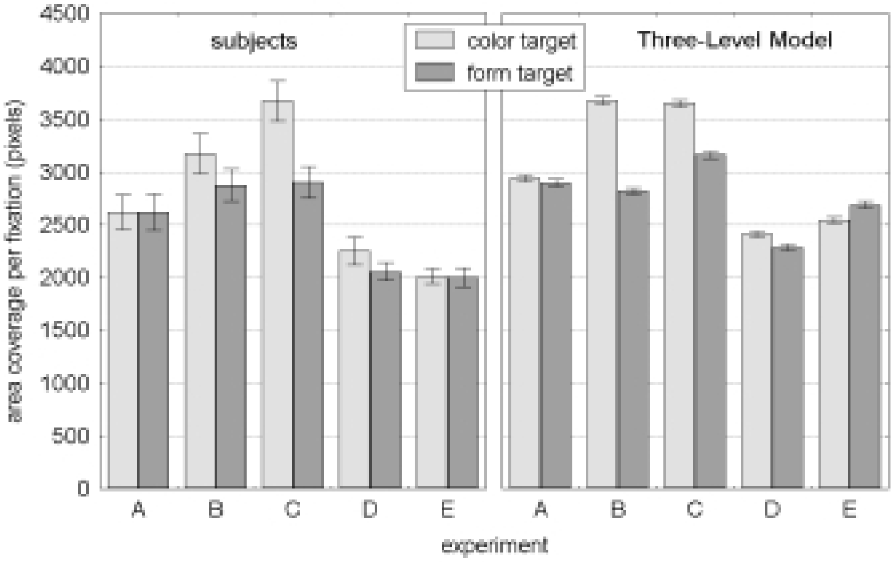


Figure 11.17: Area coverage in Experiments A to E

Do we find the “double stairway” pattern in the model’s AC results for Experiment F? Figure 11.18 allows us to compare the respective empirical and simulated data. In fact, the Three-Level Model replicates the expected pattern rather closely. The only difference to the empirical data is that the model’s absolute values are shifted towards higher AC. Nevertheless, this result supports our interpretation of the “stairway effect”.

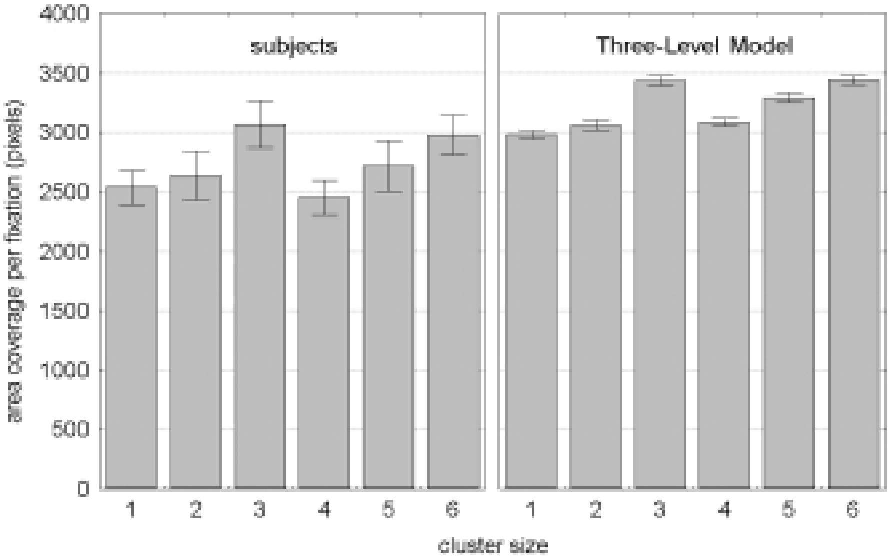


Figure 11.18: Area coverage as a function of cluster size in Experiment F

11.3.6 Speed of Processing (SP)

The comparative diagram in Figure 11.19 shows that the empirical SP values are almost perfectly reproduced by the Three-Level Model. Obviously, the only divergence can be found in the fact that – as indicated by the difference between Experiment B and Experiment C – top-down control of attention leads to an increase in SP in the model data, while it does not in the empirical data. This situation corresponds to the AC results. It seems that the lower entropy values in Experiment C, caused by one stimulus dimension being held constant, "motivates" the Three-Level Model to generate higher AC and SP values during form search. Undoubtedly, the question why these effects do not appear in the empirical data deserves a closer investigation in future research.

Figure 11.20 presents the SP results for Experiment F. Again, the "double stairway" characteristic is reproduced by the Three-Level Model. However, the influence of cluster size seems to be weaker, i.e. the SP contrasts between neighboring cluster sizes are smaller than they are in the empirical data. This suggests that the cluster structure in the stimuli exerts additional perceptual influences on the subjects, facilitating comparative search.

11.3.7 Probability of Missing the Target (PM)

As shown in Figure 11.21, there is a close correspondence between empirical and modelled PM values. Interestingly, the simple formula for the calculation of PM seems to be a plausible model for the human capability of target detection as a function of memory load and task difficulty.

This view is supported by the analysis of PM in dependence on the local item

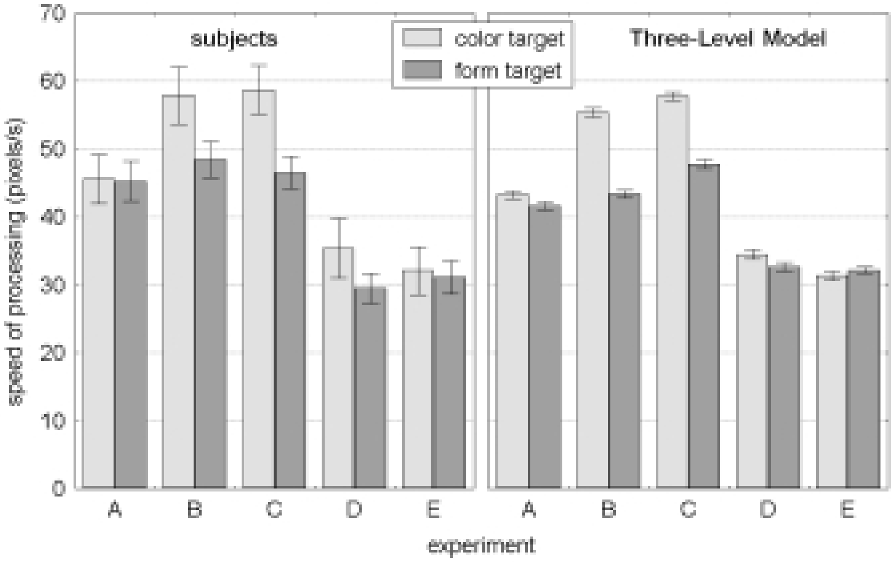


Figure 11.19: Speed of processing in Experiments A to E

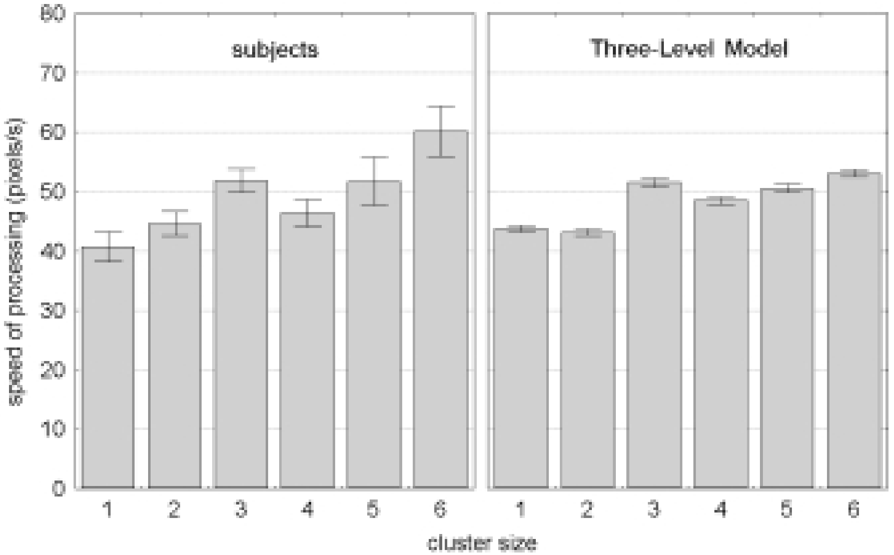


Figure 11.20: Speed of processing as a function of cluster size in Experiment F

density at the targets in Experiment A: In the simulation, PM is 11.4% for low density and 17.2% for high density. This effect is similar to the empirical one (10.5% versus 15.3%). All in all, the simulation data suggest that the factors determining PM are understood to a reasonable extent.

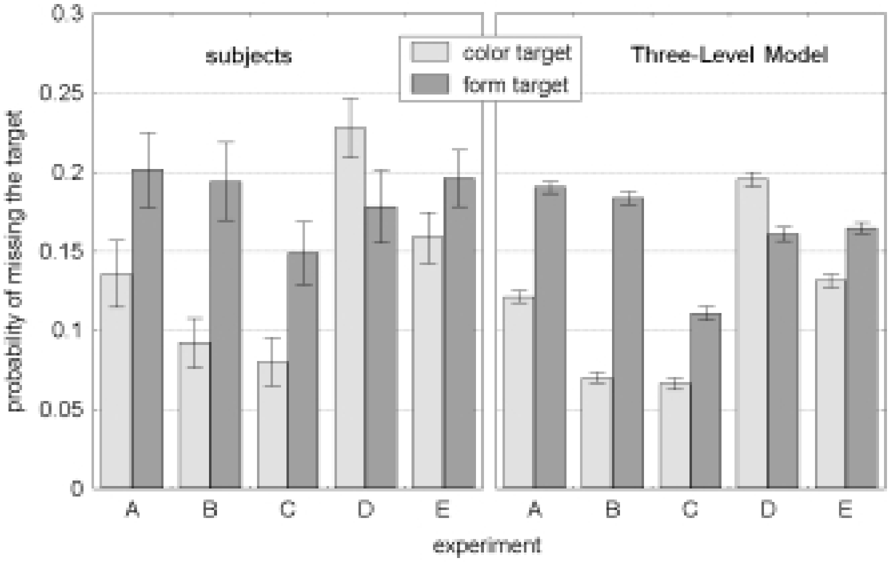


Figure 11.21: Probability of missing the target in Experiments A to E

11.4 Rating 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 basic as well as the derived eye-movement variables and their dependence on local stimulus parameters are replicated on the basis of simulated perceptive and cognitive processes. This can be considered as a substantial improvement over the Random-Walk Model which reproduced the low-level features of scanpaths *directly* with the help of experimentally obtained distributions of the respective low-level variables.

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 “getting lost”, i.e. giving up structured search during task completion. On the other hand, this discrepancy could be caused by the higher variability found in empirical eye-movement data, e.g. in the analysis of reaction time. Higher variability, in turn, may reduce the significance of measured effects. This point should be clarifield by future research.

Summarizing, the results are well in line with the assumed vertical organization of mental processes involved in the completion of comparative visual search. Furthermore, the conclusions drawn from Experiments A to F 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. This similarity is found with respect to each eye-movement variable in each of the six experiments, presenting only few cases of deviations. 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.

Since the stimuli contained no semantic information so far, the algorithmic simulation of empirical findings is suitable and its results can relatively easily be interpreted. Advanced studies, however, will go along with the introduction of semantic content, leading to difficulties with regard to modelling. Nevertheless, the studies of computational models have proven to be especially insightful and should thus be continued, even if the effort for their implementation increases.

Chapter 12

Conclusions and Outlook

In the preceding chapters, we have demonstrated how the measurement of eye movements for the task of comparative visual search can help to gain insight about a considerable number of issues involved in the perception of 2D visual displays. The results obtained with regard to the basis scenario (Experiment A) have already demonstrated one of the essential advantages over conventional visual search and same-different experiments, namely the possibility of investigating perceptive and cognitive processes during a complex task. The rather simple stimuli and resulting short reaction times in standard visual search experiments as well as the tachistoscopical presentation of items in same-different tasks lead to a limited set of results, i.e. reaction times and error rates depending on stimulus features. The isolated investigation of these rather "atomic" tasks and the shallow information content of the measured variables severely restrict insight into the structure of underlying mental processes.

In comparative visual search, the brain can be "observed" while solving a more structured task which demands a global search process as well as local steps of memorization and comparison. The recorded eye-movement data have revealed interesting aspects, for instance the linear dependence of fixation duration on local item density and its complete independence from local color and form entropy. These findings have enabled us to draw conclusions about the factors which control the involved perceptive and cognitive activities (see Chapter 3).

Slight variations of the task (Experiment B) and the stimuli (Experiment C), followed by the comparison of eye-movement variables between these experiments, have unveiled differences in the processing of color and form: While irrelevant *form* information can be disregarded quite effectively, this seems to be much harder for irrelevant *color* information. The results suggest that form processing might *include* color processing to some extent. Again, the conclusions are not based on artificial reaction time measurements, but rather on eye-movement data characterizing the "brain at work".

The same holds for Experiments D and E, which have investigated further experimental modifications. In Experiment D we have studied the task of match detection instead of mismatch detection, leading to increased difficulties of task completion and consequently to more thorough search strategies. Experiment E has examined the effects of applying mirror symmetry rather than translational symmetry between the hemifields, presenting similar results as Experiment D. In contrast to the situation in Experiments A to C, holistic comparison of local information is impossible, hence “ad hoc” processing is ineffective. Increasing demands on working memory lead to a smaller amount of information being memorized at the same time.

The role of working memory has been more closely investigated in Experiment F. Additional line elements have been successfully used to “prescribe” certain perceptual groups of items for the subjects, suggesting that the memorization of three items at a time is the most efficient strategy. Moreover, the effects of color and form entropy on working memory have been established and even quantified.

Experiment G has successfully explored the use of three-dimensional stimuli. Spatial scanpaths demonstrate a higher variability than “flat” ones, because item depth can be used for the alignment of scanpaths in addition to the two lateral dimensions. The detection of depth mismatches induces increased fixation duration, which has been constant for color and form search within and between Experiments A to F, suggesting the occurrence of fundamentally different perceptual processes. Apart from these findings, the interaction of vergence and conjugate eye movements during search processes has been investigated, yielding even physiologically relevant data.

The results of the first simulation of eye movements in comparative visual search by the Random-Walk Model has exhibited that the data obtained so far have not been sufficient to characterize human scanpath *strategies*. Accordingly, we studied scanpath regularities being applied during the inspection of randomly distributed dots. Based on the results of this study, Experiment H was conducted, employing dots of different colors and forms. While the form information did not significantly affect the subjects’ scanpaths, the color attributes were partially used to structure the gaze trajectories. The best scanpath model has been found to be a simple “travelling salesman” algorithm with a tendency towards successive scanning of equally colored items.

This knowledge has been used as the basis for an improved model, called the “Three-Level Model”. It accounts for three different levels of information processing: the *strategy* level (referring in particular to Experiment H), the *attentional* level (with parameters taken especially from Experiment F), and the *eye-movement* level (mainly based on Experiments A to E). Shifts of attention are assumed to be controlled by a global strategy, and eye movements are considered as the result of moving attention. With only a few exceptions, the Three-Level Model is capable of explaining the characteristics of gaze trajectories observed in Experiments A to F, confirming the overall consistency of empirical results and their interpretation.

Which starting points for further research on this paradigm seem to be most interesting and most important? One essential point will be the transition from abstract items to more realistic stimuli involving perception on a higher level. Since the basic characteristics of comparative visual search have been extensively

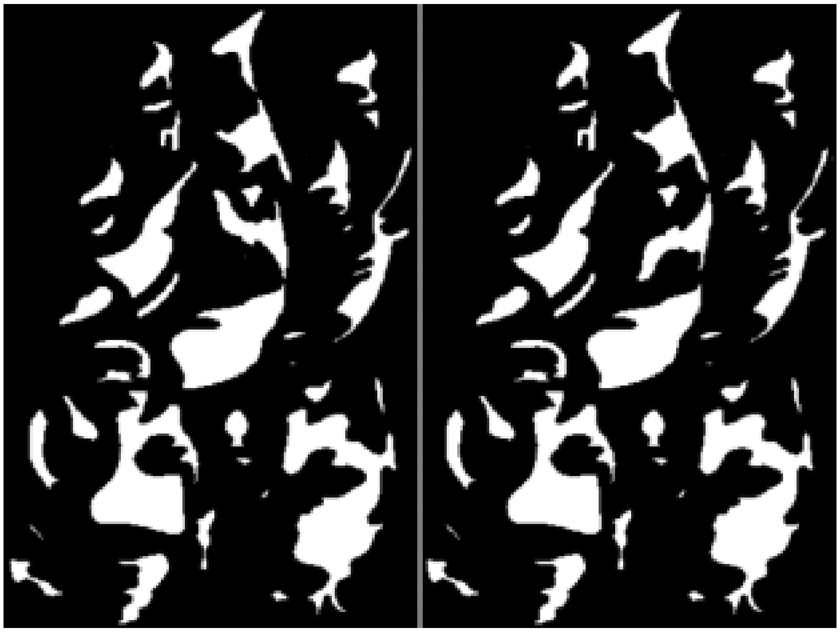


Figure 12.1: Possible stimulus for the investigation of low-level versus high-level perception during comparative visual search. Rotating the picture by 180^o leads to a switch between these two perceptual levels.

studied in this work, it seems reasonable to increase the complexity and realism of the search pictures. The clustered items used in Experiment F can be considered as a first step in this direction. Single items were combined to item groups, changing the subjects' perception of the stimuli. However, no modifications towards ecologically more adequate scenes were introduced.

An immediate use of realistic stimuli as shown in Figure 1.7 would be too drastic. Their high complexity and the influence of subjective, not quantifiable parameters would make the interpretation of findings very difficult if not impossible. Although there have been studies on visual search proving that naturalistic scenes can be reasonable stimuli (e.g. Henderson, 1992; Humphreys, Keulers & Donnelly, 1994; Wolfe, 1994), a direct step from completely abstract to realistic stimuli would lead to gaps in the theoretical background. Instead, we should *gradually* increase the complexity of stimuli between experiments in order to always maintain an adequate framework for the incorporation of new results. A suitable pilot experiment could investigate the differences in eye movements between the presentation of *abstract* and *meaningful* stimuli. Figure 12.1 shows an appropriate kind of stimulus for this purpose. The geometrical, colored items have been replaced with white items of irregular shape, but still the subjects' task is to find the mismatching pair. In our example, the mismatch consists in one of the items being turned around its center by 180 degrees, which is a possible standardization of mismatches in this scenario.

What does the example stimulus have to do with meaningful information?

Rotating Figure 12.1 to its upside-down position leads to the answer: Now, the items are perceptually grouped to meaningful units, namely to a number of different faces. Strictly speaking, no faces are actually being shown, but the patterns of black and white induce the perception of light and shadow as caused by facial structures. This brilliant perceptual feat cannot be achieved in the case of upside-down reversal, because human beings are almost exclusively trained to recognize upright faces.

The idea behind this sort of stimuli – which could be called "Mooney" stimuli due to the psychologist Mooney who "invented" them – is that we can use the same material for the investigation of low-level and of high-level perceptual processes. We just have to turn the stimuli around in order to switch between these two levels. Conceivably, the eye-movement variables strongly depend on the orientation of these stimuli and hence yield basic data on the differences which occur when meaningful information is involved. Starting with this pilot experiment, the way to the investigation of realistic stimuli could be paved.

As a matter of fact, the ecological adequacy of comparative visual search is not only determined by the semantics of the utilized stimuli, but also by the way of their presentation. Indicated by the results of Experiment G, the third dimension plays an important role, strongly influencing the subjects' scanpath strategies. Furthermore, scenarios of this kind allow us to study the interaction between *natural* physiological processes (vergence movements and saccades) and stimulus features. Although Experiment G can be viewed as a first approach to three-dimensional stimuli, it still involves several "unnatural" points that should be eliminated in future research:

- Due to the applied method of red-green images and glasses, the stimuli are monochrome. With respect to further series of experiments, the technique of shutter glasses should be used to include color information and to enable unbiased vision.

- The third dimension is represented by only three distinct depth planes. In order to provide an ecologically plausible situation and comparability among the three dimensions, these planes have to be replaced by a depth continuum.

- The items themselves should be three-dimensional as well. As a first approach, we could use various types of regular and irregular polyhedrons. To achieve the natural appearance of these items, the location of a light source has to be specified and the resulting shades of item colors have to be displayed.

- A realistic type of perspective should be integrated into the stimuli, which basically means that the size of items decreases with their virtual distance from the subject's eyes.

The semantic as well as the presentational steps towards higher ecological adequacy should go along with corresponding "updates" of the Three-Level Model of comparative visual search. The major problem is assumed to be the integration of perceptual processes on a higher level, for instance the recognition of faces. The application of *artificial neural networks* seems to be a promising way of solving this problem. The artificial neural networks are able to learn the components of faces and their geometrical relationships from a database of example faces. If they are combined with an appropriate simulation of attentional mechanisms, they should be capable of imitating human perceptual processes on the corresponding stage. As demonstrated in the present work, the future experiments and models should go hand in hand in order to provide a valid and detailed understanding of the interacting processes that participate in comparative search performance.

Undoubtedly, the work in hand has successfully established first landmarks in the wide and merely untouched field of comparative visual search. The results of the described experiments and models have revealed fruitful aspects for empirical research as well as for computer simulations of mental processes. A huge amount of interesting work is still waiting to be done, and maybe the present findings will motivate scientists of related fields to participate in the exploration of this promising paradigm.

References

Bacon, W.F. & Egeth, H.E. (1994). Overriding stimulus-driven attentional capture. *Perception & Psychophysics*, 55, 485–496.

Ballard, D.H. (1991). Animate Vision. *Artificial Intelligence*, 48, 57–86.

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.

Bruce, V.G. & Morgan, M.J. (1975). Violations of symmetry and repetition in visual patterns. *Perception*, 4, 239–249.

Caelli, T. & Bischof, W.F. (1994). Computational approaches to human pattern recognition. *Spatial Vision*, 8 (1), 57–76.

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.

Dixon, P. & Just, M.A. (1978). Normalization of irrelevant dimensions in stimulus comparisons. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 36–46.

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.

Edelman, S. & Weinshall, D. (1991). A self-organizing muliple-view representation of 3D objects. *Biological Cybernetics*, 64, 209–219.

Egeth, H.E. (1966). Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1, 245–252.

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

Enns, J.T. & Rensink, R.A. (1990). Influence of scene-based properties on visual search. *Science*, 247, 721–723.

Friedman-Hill, S. & Wolfe, J.M. (1995). Second-order parallel processing: Visual search for the odd item in a subset. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 531–551.

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

Grossberg, S., Mingolla, E. & Ross, W.D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, 101, 470–489.

He, Z.J. & Nakayama, K. (1992). Surfaces versus features in visual search. *Nature*, 359, 231–233.

Henderson, J.M. (1992). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology*, 46, 319–341.

Howard, J.H. & Kerst, S.M. (1978). Directional effects of size change on the comparison of visual shapes. *American Journal of Psychology*, 91, 491-499.

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.

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

Humphreys, G.W., Keulers, N. & Donnelly, N. (1994). Parallel visual coding in three dimensions. *Perception*, 23, 453–470.

Humphreys, G.W. & Müller, H.J. (1993). Search via recursive rejection (SERR): A connectionist model of visual search. *Cognitive Psychology*, 25, 43–110.

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

Hung, G.K., Wiler, J., Weiss, F.L. & Curry R.K. (1993). Random and direct path eye movements during target search. *Medical Science Research*, 21, 389–391.

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*, 397–410. Amsterdam: Elsevier North-Holland.

Julesz, B. (1969). Binocular depth perception. In W. Reichardt (Ed.), *Processing of optical data by organisms and by machines.* New York: Academic Press.

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

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.

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.

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

Koesling, H., Pomplun, M. & Ritter, H. (submitted 1997). The effects of structural information on perceived numerosity in two-dimensional object distributions. *Perception and Psychophysics.*

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

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.

Mackworth, N.H. (1976). Stimulus density limits the useful field of view. In R.A. Monty & J.W. Senders (Eds.), *Eye Movements and Psychological Processes.* New York: John Wiley.

Menkhaus, G. (1997). Präattentive Fokussierungspunktgenerierung zur Objekterkennung mit künstlichen neuronalen Netzen. *Diplomarbeit*, Technische Fakultät der Universität Bielefeld.

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.

Nakayama, K. & Silverman, G.H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, 320, 264–265.

Nattkemper, T.W. (1997). Untersuchung und Erweiterung eines Ansatzes zur modellfreien Aufmerksamkeitssteuerung durch lokale Symmetrien in einem Computer Vision System. *Diplomarbeit*, Technische Fakultät der Universität Bielefeld.

Nodine, C.F., Kungel, H.L., Toto, L.C. & Krupinsky, E.A. (1992). Recording and analysis of eye-position data using a microcomputer workstation. *Behavioral 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*, 261–292. London: Academic Press.

O'Toole, A.J. & Walker, C.L. (1997). On the preattentive accessibility of stereoscopic disparity: Evidence from visual search. *Perception & Psychophysics*, 59 (2), 202–218.

Palmer, S., Rosch, E. & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance*, 9, 135–151. Hillsdale: Erlbaum.

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

Pashler, H. (1988). Cross-dimensional interaction and texture segregation. *Perception and Psychophysics*, 43, 307–318.

Pomplun, M. (1994). Aufmerksamkeitsverteilungen in ambigen Bildern und ein vielseitiges Programmpaket für Eyetracking-Experimente. *Diplomarbeit*, Technische Fakultät der Universität Bielefeld.

Pomplun, M., Ritter, H. & Velichkovsky, B.M. (1996). Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception*, 25, 931–948.

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.

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.

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

Saccuzzo, D.P. (1987). *Psychology: From research to applications.* Boston: Allyn and Bacon.

Santee, J.L. & Egeth, H.E. (1980). Selective attention in the speeded classification and comparison of multidimensional stimuli. *Perception & Psychophysics*, 28, 191–204.

Schmidt, R.F. (1985). *Grundriß der Sinnesphysiologie.* Berlin: Springer.

Schor, C.M. & Ciuffreda, K.J. (1983). *Vergence eye movements: Basic and clinical aspects.* Boston: Butterworths.

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

Smith, P.L. (1989). A deconvolutional approach to modelling response time distributions. In D. Vickers & P.L. Smith (Eds.), *Human Information Processing: Measures, Mechanisms, and Models.* Amsterdam: Elsevier.

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

Steinman, S.B. (1987). Serial or parallel search in pattern vision. *Perception*, 16, 389–398.

Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics*, 50, 184–193.

Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics*, 51, 599–606.

Theeuwes, J. (1993). Visual selective attention: A theoretical analysis. *Acta Psychologica*, 83, 93–154.

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. (1992). Spreading suppression or feature integration? A reply to Duncan and Humphreys (1992). *Journal of Experimental Psychology: Human Perception and Performance*, 18, 589–593.

Treisman, A. & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97–136.

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.

Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance*, 9 (4), 523–530.

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. (1990). The vertical dimension of mental functioning. *Psychological Research*, 52, 282–289.

Velichkovsky, B.M. (1995). Communicating attention: Gaze position transfer in cooperative problem solving. *Pragmatics and Cognition*, 3 (2), 199–222.

Velichkovsky, B.M., Sprenger, A. & Pomplun, M. (1997). Auf dem Weg zur Blickmaus: Die Beeinflussung der Fixationsdauer durch kognitive und kommunikative Aufgaben. In R. Liskowsky, B.M. Velichkovsky & W. Wünschmann (Eds.), *Usability Engineering: Software-Ergonomie '97.* Stuttgart: Teubner.

Watanabe, T. (1988). Effects of irrelevant differences as a function of the relations between relevant and irrelevant dimensions in the same-different task. Journal of Experimental Psychology: *Human Perception and Performance*, 14, 132–142.

Wieners, K.H. (1995). Implementierung und Evaluation eines Modells zur ontogenetischen Entwicklung von Merkmalskarten im Visuellen Kortex bei Primaten. *Diplomarbeit*, Technische Fakultät der Universität Bielefeld.

Wolfe, J.M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research*, 34, 1187–1195.

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.

Yarbus, A.L. (1967). *Eye movements and vision.* New York: Plenum Press.