Visual Representations in a Natural Visuo-motor Task

by

Jeff Bennett Pelz

Submitted in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

Professor Mary M. Hayhoe

Department of Brain and Cognitive Sciences

The College

Arts and Sciences

University of Rochester

Rochester, New York

1995

Curriculum Vitae

The author was born in Palo Alto, California on March 25, 1956, and grew up in Ann Arbor, Michigan. He attended the Rochester Institute of Technology, earning a Bachelor of Fine Arts degree in Fine Arts Photography in 1980 and a Masters of Science in Imaging Science in 1986. He began his studies at the University of Rochester in the fall of 1990 at the Center for Visual Science. He pursued his research under the direction of Professor Hayhoe, and received his Master of Arts degree in 1993. He is currently on the faculty of the Center for Imaging Science at the Rochester Institute of Technology.

Acknowledgments

A great many people contributed, directly and indirectly, to the body of work that makes up this dissertation. The guidance and encouragement provided by Mary Hayhoe and Dana Ballard were invaluable. Discussions with Steven Shimozaki, Keith Karn, Per Möeller, and Michael Spivey also helped shape the experiments. Early work by Mary Hayhoe, Dana Ballard, Steven Whitehead and Feng Li provided the base on which this work was built. The work was only possible because Alan Russell, Andrew Forsberg, Deborah Bancroft, and Timothy Becker shared their ideas and skills so generously. Portions of the work represented in this thesis have been supported by NIH RO1 EY05729, NIH R24 RR06853, and AFOSR-91-0332. Additional support was received from the Rochester Institute of Technology's Center for Imaging Science Industrial Associates Program. Finally, the whole experience could never have taken place without the love, support and patience of Myra, Rachel, and more recently Madeline Pelz.

Abstract

This thesis investigates the role of foveating eye movements in vision. A block-copying task, in which fixations are tied closely to perceptual, cognitive, and motor events, was used to explore visual representations and visuo-motor coordination in the context of complex, ongoing behaviors. The experiments required the development of a laboratory facility capable of monitoring concurrent eye, head, and hand movements at a fine time-scale. The primary finding was that subjects made frequent eye movements to serialize the task into simpler subtasks which were executed sequentially, minimizing working memory load. This suggests that only a sparse, transient representation of task-relevant information is maintained, rather than an extensive, task-independent reconstruction of the visual world which is commonly assumed to be the goal of vision. A series of experiments in which the 'cost' of fixating a pattern to be copied was increased, or the information content of the pattern was reduced, demonstrated that the trade-off between frequent eye movements and working memory load is flexible, and that subjects are capable of dynamically adjusting that balance based on task demands.

Measurements of subjects' eye and head movements during the block-copying task led to the observation that some subjects executed independent eye and head movements, dissociating their spatial and temporal trajectories, an observation inconsistent with current models of eye and head movements that postulate a common gaze shift goal. In another series of experiments, subjects performed the block-copying task while performing one of two secondary tasks. The added cognitive load led to dramatic changes in subjects' head trajectories, and affected the strategies subjects used in constructing the duplicate pattern.

The experiments provide support for a different approach to studying visual processing, in which vision is viewed as more top-down than previously supposed. The task takes on particular importance, because behavior cannot be divorced from the immediate task(s). Fixations are shown to play crucial cognitive roles in perception, including binding task-relevant information to variables in working memory, and indexing the execution of sequential programs.

Table of Contents

List of Tables

List of Figures

1. Introduction

Visual perception is an extended process that naturally occurs in the context of complex behaviors in dynamic, feature-rich environments with literally thousands of potential fixation targets. Yet its study has been limited in large part to the perception of simple, static stimuli in reduced environments with the subject holding fixation, or perhaps making simple eye movements. This is in part because of a limitation in instrumentation and in part motivated by a belief that complex systems can be understood in terms of simpler component processes. Instrumentation now exists that allows the study of vision in more natural settings. Moving beyond the investigation of small subsystems presents several challenges, but in return we are able to ask new kinds of questions about vision in its normal behavioral context. Studying vision in this context (and in the associated complex environments) compels us to consider the nature of the internal representations used by humans in the performance of complex, real-world tasks. It is not yet clear what type of internal representation exists, nor the manner in which it is created. At one extreme, successive fixations made over a period of time could be integrated into a high-fidelity, general-purpose representation of the immediate environment. At the other extreme, transitory task-specific representations could be computed for each fixation frame, with little carry-over between fixations. Such moment-by-moment representations do not match our conscious perception, but there is evidence that subjects do not have access to the kind of high-fidelity, internal representations that introspection might lead one to postulate, and the 'internal store' doctrine has been questioned by several investigators. Gibson [1966], O'Regan [1992], and others have proposed that the same subjective impression of an internal replica of the visual environment could be achieved through active perception of the environment by treating that environment as an 'external store,' though little evidence has accrued to date to support the proposal.

The study of how internal representations are formed and utilized leads to questions about how working memory is used in performing complex, multi-step tasks. The nature of internal representations is intimately linked to the study of short-term, or working memory. Baddeley defines working memory as "... a system for the temporary holding and manipulation of information during the performance of a range of cognitive tasks such as comprehension, learning, and reasoning." [Baddeley 1986, p. 34] Rich internal representations require substantial working memory resources, a resource that we know to be limited. Complex visuo-motor tasks that we perform daily also rely on working memory. Most studies of working memory have focused on determining the upper bounds of memory, so while we know a great deal about the limits of memory systems, we know less about how memory is actually utilized in natural tasks, and we have little understanding of the computational role those limitations place on the system as a whole.

Recent advances in instrumentation now make it possible to perform experiments in which subjects perform complex tasks under natural conditions. In the past, most research on visual perception and eye movements has been limited to the examination of single events.

Experiments on the oculomotor system have focused on understanding the mechanical properties and limitations of the isolated components rather than their combined contribution to the strategy of perception. The study of visual processes has been segregated from the study of motor processes like eye movements, though they must be intimately linked. Whether it was a question about how much information can be gained in a single glimpse, or the mechanics of eye movements, the research was limited to looking at the small building blocks of visual perception. But visual perception is not just what happens in 200 msec; it operates continuously. Studying isolated eye movements made in reaction to reduced stimuli forced eye movements into the reflex mode. With the exception of reading, we know very little about how visual behaviors unfold over time, or about the sequence of eye movements observers use when they are free to choose their own strategies in the context of performing natural tasks. Making accurate measurements of eye movements is difficult, and until recently it required that the head be immobilized. Kowler [1990] described some early measurement techniques in which the experimenter "photographed a droplet of mercury placed on the limbus. Translations of the head were minimized by having subjects lie on a stone slab with their heads wedged tightly inside a rigid iron frame." While such measures may now seem extreme, there is evidence that even requiring a subject to stay on a biteboard alters oculomotor performance [Kowler, *et al*., 1992, Collewijn, *et al*., 1992]. Another result of the limitations of previous eye movement monitors is that we know very little about how the eye and head work together in natural gaze changes.

Eye movements are integral to visual perception; if we are to understand vision, we will need to understand the role that eye movements play. If we no longer consider eye movements as reflexive reactions to the environment, we can begin to investigate the processing underlying eye movements. The work of Yarbus [1967] was very important in expanding the view of the role of eye movements as externally visible reflections of cognitive events. In one of Yarbus' classic experiments on eye movements during perception of complex objects, he monitored subjects' eye movements as they viewed Repin's painting, "The Unexpected Visitor." Before viewing, the subjects were instructed to perform one of seven tasks: 1) free examination of the picture, 2) estimate the material circumstances of the family in the picture, 3) give the ages of the people, 4) surmise what the family had been doing before the arrival of the "unexpected visitor," 5) remember the clothes worn by the people, 6) remember the position of the people and objects in the room, and 7) estimate how long the "unexpected visitor" had been away from the family. The pattern of eye movements and fixations varied dramatically with different instructions to the subjects. Figure 1.1 shows the scan paths for a single observer viewing the painting under each of the seven instructions.

Yarbus concluded that "... the distribution of the points of fixation on an object, the order in which the observer's attention moves from one point of fixation to another, the duration of the fixations, the distinctive cyclic pattern of examination, and so on are determined by the nature of the object and the problem facing the observer at the moment of perception."

[Yarbus 1967, p. 196] The fact that eye movement patterns are not determined by the stimulus alone, but are dependent on the task being performed suggests that eye movements are an integral part of perception and not simply a mechanism evolved to deal with the 'foveal compromise' (the uneven distribution of photoreceptors across the retina that allows both high resolution and a wide field of view).

While Yarbus' work demonstrated the coupling of eye movements and perception, it is not clear how his results relate to more normal conditions. Yarbus' subjects were required to view the painting for three minutes after each instruction. While there were

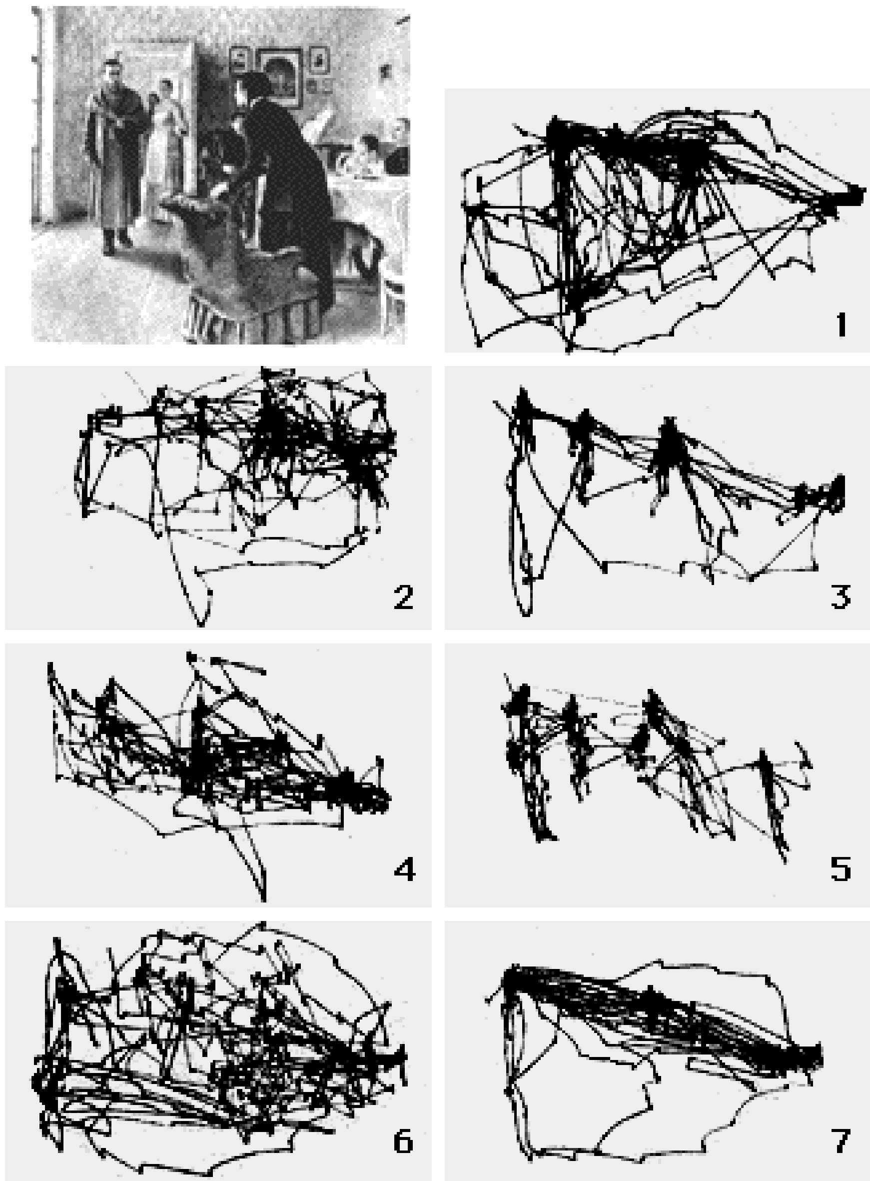


Figure 1.1 Eye movement records for one subject while viewing the painting "An Unexpected Visitor." [Yarbus 1967, Fig. 109].

clear differences in the eye movements recorded under the different instructions, this time scale is more normally associated with cognitive processes than perceptual processes. We are left to guess the underlying perceptual and cognitive processes tied to any individual fixation, and since the subject had no control over trial duration and no task to perform we cannot even know which fixation clusters occurred while the subject waited patiently for the trial to end. The "Unexpected Visitor" experiment demonstrated that high-level cognitive processes affect fixations, but such experiments do not allow any inferences to be made about the purpose of any fixation or series of fixations. Such inferences are only possible if the subject is performing a task where the computations necessary for task performance can be made explicit while maintaining natural behavior.

The issues surrounding the nature of visual representations also arise in the context of computer vision. Implicitly or explicitly, the design of computationally-based artificial vision systems have traditionally been driven by a desire to model the subjective experience of human vision. This approach dominated work in machine vision for a decade [see e.g., Marr 1982, Feldman 1985, Levine 1985]. In this strategy, the entire scene is analyzed, building models of progressively higher dimensionality, eventually reaching the "3-D Model Representation," an approach that fits with our intuition about what it feels like to 'see.' While observers may be conscious of 'paying attention' to various aspects of a scene, it is usually assumed that attention is a cognitive mechanism applied to at least a partially analyzed perceptual representation. Computational models have focused on algorithms for reaching high-level descriptions of the scene from various cues (e.g., the shape-from-X and structure-from-motion algorithms), or a semantic representation of the scene [e.g., *From Pixels to Predicates*, Pentland 1986]. In general, these problems have proven to be very difficult except in very restricted, artificial environments. The approach common to all these systems is a full-scale assault on the sensor data before specifying the immediate action to be planned or the overall task to be performed. But visual information necessary for any given task is not uniformly distributed in space and time, so a vision system that applies its resources uniformly, without regard to the immediate task, is inherently inefficient. For example, given the task of identifying a person standing behind a chain link fence, such a system would devote the same computational resources to processing the chain-link fence as to the face behind it.

In the last decade, some workers in the field of machine vision systems have found that allowing camera movement may simplify some of the representational problems. The task-dependent nature of biological vision systems served as a model for several early proponents of 'active vision' [Aloimonos *et al*. 1987, Bajcsy 1988, Ballard 1989, Brooks 1991]. The study of attention and eye movements was crucial in much of that work, and early papers on active vision paid homage to Yarbus' [1967] work which demonstrated the task-dependent character of human eye movements. The idea that optimal behaviors for collecting visual information are dependent on the particular task is central to the advantages offered by the 'active vision' systems that have been implemented in computer vision. The early proponents of active vision shifted their efforts away from attempts at general-purpose "image understanding," searching instead for behaviors of the agent that could efficiently probe the environment for information where and when it was needed.

Physical motion of the camera provides important advantages to active vision systems. Control over (and knowledge of) the camera's location constrains the images in a manner that allows conclusions to be drawn that are impossible without those constraints. For example, Ballard and Ozcandarli [1988] demonstrated a motion parallax system that computed the depth of objects based on their motion relative to the fixation point. The ability to fixate a single point while translating dramatically reduced the computational resources necessary to determine object distance. Systems using two cameras have shown computational economies by using disparity information derived from the two images.

The crucial aspect of these models is that they allow frequent access to the sensory input during the problem-solving process [Brooks 1991; Agre and Chapman 1987; Ballard 1989, 1991]. Agre and Chapman [1987], building on work by Ullman [1984], introduced the term 'deictic primitives' to refer to transient pointers used to mark aspects of a scene (e.g., color or shape). Such aspects are dynamically referred to by indicating that part of the scene with a special marker. The term 'deictic' comes from linguistics; a class of pronouns (e.g., 'this,' 'that,' and 'those') are termed 'deictic' words (from the Greek *deiktikos;* to show or point). These deictic words result in significant representational economies in language, allowing one to point and say "I want *those"* instead of "I want to purchase the pair of size 10, black, high-topped, canvas athletic shoes in the third row of boxes above the floor and 4 boxes to the left of the door."

The power of computer vision systems using deictic markers comes from their use of a relatively small number of these markers, binding them to areas of a scene only as long as they are relevant to the immediate task, then moving them to another area. Memory requirements for systems using deictic strategies are also dramatically lower than for non-deictic systems. Because only a small portion of the visual scene is represented at any time, there is no need to maintain a large internal store. If information from a particular region of the scene must be referred to again later, it is only necessary to store the location (or feature vector) of the marker, rather than all the information to be found at that location. Table 1.1 illustrates the simplification allowed by systems that do not rely on representations in which the positions and properties of all objects in the scene must be represented in viewer-centered coordinates. Quadrant IV represents the goal of 'traditional' image understanding algorithms; given a (static) scene, locate and identify all objects in the scene. This requires that all known models be applied to all areas of the image -- a task that has proven very difficult. Deictic variables can reduce the computational complexity by marking portions of a scene, simplifying the problem into a series of two kinds of tasks; "what" and "where." Quadrant II represents a "what" task; given a single image location, determine the identity of the marked object from among many possible models. Quadrant III is the "where" task;

given a single internal model, search the image space to determine the location of the marked model. This simplification leads to dramatically faster algorithms for each of the specialized tasks [Swain and Ballard 1991; Swain *et al*. 1992; Ballard and Rao 1994]. Once an item's identity and location are known, it can be manipulated (physically or symbolically). So the organization of visual computation into WHAT/WHERE modules may have a basis in complexity. Trying to match a large number of image segments to a large number of models at once may be too difficult. The problem can be made tractable by serializing the problem into sequences of 'what' and 'where' functions.

Table 1.1 Organization of visual computation into WHAT/WHERE modules.

	One Model	*Many Models*
One Image Part	I. Manipulate an object whose identity and location are known.	II. "What?" Identify an object whose location is known.
Many Image Parts	III. "Where?" Locate a known object in the scene.	IV. Locate and identify all objects in the scene.

These deictic representations, which allow localization and interaction with respect to the current fixation point rather than in an absolute, camera-centered reference frame are particularly valuable when vision is used to control actions, i.e., in visuo-motor tasks. Fixation-based, exocentric reference frames permit closed-loop 'servoing' actions, eliminating complex reference frame transformations, and making the systems less 'brittle' in the face of errors. Traditional systems that rely on visual localization with respect to egocentric reference frames are intolerant of even relatively small errors in localization and/or representation of the motor agent's position.

Since the utility of deictic systems has been demonstrated in computational systems, we are led to ask whether similar representational economies may be exploited in biological vision systems. Because humans have limited working memory and the eyes allow a natural implementation of deictic strategies which might allow complex tasks to be performed without elaborate internal representations being held in memory, the question is immediately raised whether humans in fact use eye movements in this way during natural behaviors. One of the goals of this thesis is to explore the applicability of the computational idea of deictic representations to human vision.

However, care must be taken in selecting a task for study. Because we do not have any direct access to the strategies employed by humans in most complex tasks, one can only infer those strategies based on subjects' behavior. In free viewing of an image or scene, subjects make a series of fixations, separated by saccadic eye movements. The resulting 'scanpaths' have been studied in attempts to infer the underlying cognitive goals of the viewers [e.g., Noton and Stark, 1971]. Without knowledge of the intentions and pre-conceptions of the subject, (knowledge probably not available even to the subject in many complex tasks), the value of scanpaths in discovering cognitive strategies is limited. While there is some similarity in scanpaths of observers viewing the same scene with the same instructions, there is also significant variability within and between-subjects. Viviani [1990] presents these objections to attempts to discover cognitive strategies from eye movement records, cautioning that such an attempt "presupposes, however, the possibility of controlling the input stimuli and subject's intentions to an extent that is seldom, if ever, attainable in the case of free visual exploration. One must therefore, settle for evidence derived from situations with more constraints on the sequence of movements than are actually present in real-life scanning." [Viviani, 1990 p.380] The "constraints on the sequence of movements" is meant quite literally: Viviani goes on to describe single and double-step paradigms that are considered sufficiently constrained.

But there is another alternative: Instead of arbitrarily constraining the task to reduced stimuli and providing explicit instructions on the pattern of eye movements permitted (e.g., as in a traditional double-step paradigm), it is possible to design a task which constrains fixations to 'useful' regions in a complex field rather than to a small number of permitted points. Ideally the task would also constrain the number of 'useful' strategies likely to be exhibited. Such a task would fall between the underconstrained case of "free viewing," and the overly restricted case where one or two targets are to be fixated in a prescribed manner. Tasks such as mental rotation and mental arithmetic have been used in attempts to understand internal representations of objects [e.g., Just and Carpenter 1976]. Such tasks require complex cognitive computation, but have little or no externally observable behaviors other than task duration. On the other hand, a task that requires complex eye movements tied to relatively simple cognitive processing, where fixations could be reliably related to cognitive processes based on the subject's progress in the task, could provide a useful way to examine visual behavior in complex tasks. This thesis explores such a task involving a sequence of visual, motor and memory components in copying a pattern of colored blocks.

To do this, we need to monitor subjects' eye, head, and hand movements in an unconstrained situation. Head-free eyetrackers now permit the measurement of eye and head movements at a fine time-scale, allowing us to see how humans use vision to gather visual information and guide motor movements during real-world tasks. Simultaneously monitoring the eyes, head, and hand allows the coordination of perceptual and motor behavior to be examined as well. The work by Agre and Chapman [1987] and Whitehead and Ballard [1990] demonstrated the advantages of deictic representations using computer simulations of a robot solving problems by manipulating blocks in a 2-dimensional space. These simulated block-manipulation tasks provided inspiration to investigate human performance in a related block-copying task and to study human's use of such deictic strategies in natural behavior. In order to investigate these questions, a series of experiments are presented that examine human performance in an extended, natural block-copying task. There are several key features of the task under study: It is an extended task, allowing the study of natural, ongoing behavior rather than isolated movements. Unlike meta-tasks such as mental rotation and mental arithmetic, the block-copying task is concrete - there is no ambiguity about the task to be performed or the subject's progress in the task. While the task is clearly defined, it has a loose trial-like structure corresponding to the copying of each block, and no explicit instructions need be given regarding eye, head, or hand movements. Subjects simply perform the task in whatever manner they choose, using natural eye, head, and hand movements.

The experiments presented in this dissertation demonstrate the importance of studying vision in the context of ongoing behavior. Vision can not be considered a process that operates independently of ongoing tasks. The experiments show that vision is highly task-dependent. Investigating natural behavior required the development of a new laboratory facility with the capability to monitor several aspects of complex behavior. The component systems for monitoring eye, head, and hand movements, and the subsystems' integration into the new facility are discussed in Chapter 2. The block-copying paradigm on which all the experiments are based is then described and fundamental properties of subjects' performance is discussed. The main result is that subjects use frequent eye movements to serialize an extended, complex task. Serializing a task in this manner reduces the instantaneous load on working memory by gathering information necessary to perform individual subtasks only as it is needed. This 'just-in-time' perceptual strategy apparently uses fixation as a form of deictic marker to simplify multi-step tasks, reducing memory load.

Chapter 3 explores the tradeoff between frequent fixations and working memory load. Experiments manipulating the relative cost of frequent fixations and the amount of visual information required to complete the task show that subjects modify the tradeoff between working memory load and frequent fixations depending on immediate task constraints. The new instrumentation also makes it possible to study the coordination of the eye, head, and hand during complex tasks. Chapter 4 examines the coordination of the eye and head subsystems, and their coordination with the hand. The experiments show that the performance of the subsystems, and the coordination of those subsystems show the same task-dependence found in the studies of gaze in the earlier chapters.

The results of all the experiments support the interpretation that subjects serialize complex tasks into a number of subtasks that are executed sequentially by a system with limited central resources. Chapter 5 presents a series of experiments in which subjects had to perform a secondary task in addition to the primary block-copying task to probe the limits of the central resource. Even a seemingly unrelated verbal shadowing task caused significant interference with performance of the primary, visual/spatial block-copying task. The implications of these experiments are discussed in Chapter 6.

2. The Block Copying Paradigm

2.1 Introduction

The task selected was one in which a subject is required to manually duplicate a multi-color pattern of blocks. The block-copying task is primarily sensorimotor and is naturally broken down into a series of sub-tasks, each tightly coupled to externally observable gaze and hand movements. As a result, the subject's immediate cognitive state can be meaningfully inferred by examining those movements. The task requires that the subject interact with the environment, and requires fine manual manipulation.

The task, reminiscent of the block manipulation simulations of Agre and Chapman [1987] and Whitehead and Ballard [1990], required a subject to manually duplicate a pattern made of colored Duplo® blocks. The subject was seated in front of a 110 cm x 75 cm board set at 10° from vertical that was divided into three sections: The "model" area, the "resource" area, and the "workspace" area (see Figure 2.1). The model area contained the eight-block pattern to be duplicated, the resource area contained twelve blocks from which blocks could be selected to construct the copy, and the subject was instructed to build the copy in the workspace. The selected task is a generalization of the task used by Ballard, Hayhoe, Li, & Whitehead [1992], in which subjects manipulated colored patterns on a Macintosh CRT using a mouse. The subject's head was fixed to allow eye position to be monitored using an SRI tracker. The task used in the present experiments allows the examination of natural, unconstrained movements of the eye, head, and hand. The block-copying paradigm is a natural, multi-step task requiring coordination of the eyes, head, and hand, yet the task is sufficiently explicit so that cognitive operations can be inferred from the subject's behavior.

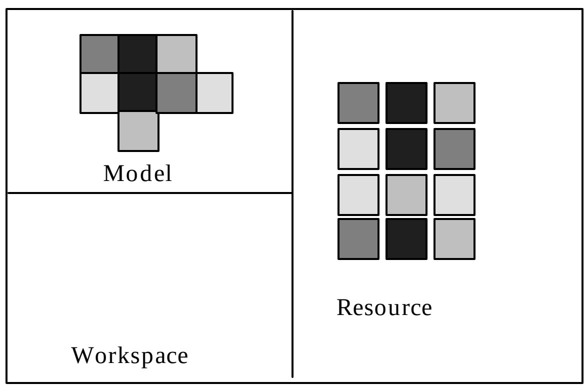


Figure 2.1 Layout of experimental working plane containing the model, resource, and workspace.

In order to complete the block-copying task, the subject must:

1) Perform a series of hand movements, moving blocks from the resource to the workspace

2) Use vision to gather information about the model configuration and blocks available in the resource area, and to guide the hand movements described in 1) above

3) Use working memory to retain information about the model pattern and to monitor progress in the task

The above list is ordered. The task cannot be completed without performing item 1), the series of hand movements. The restriction that a single hand be used to move the blocks forces the blocks to be copied serially. The order in which the blocks are selected and which blocks are moved from the resource area are left to the subject. (There are always more blocks in the resource area than are needed to complete the copy.) Items 2) and 3), use of vision and working memory, are less rigidly constrained; there are many strategies a subject could adopt. While the subject is free to adopt any strategy, gathering information about the model configuration and guiding hand movements to pick up blocks in the resource area and place them in the workspace are almost always accomplished by foveating the point of attention, so the eye movements can be used to infer the underlying cognitive operations. Therefore the block-copying task requires eye movements tied to relatively simple cognitive processes. Cognitive events are intimately tied to objects, their features, and/or movements made by the subject in completing the task. If, after placing a block in the workspace, the subject makes a fixation in the model area while moving the hand towards the resource area, we can make two inferences: First, that the subject is selecting the next block to copy and determining its color so that a block in the resource area can be targeted and second, that the subject chose to make these decisions based on the current model fixations and not solely on information from previous model fixations retained in working memory.

2.2 Methods

2.2.1 Monitoring eye position

Monocular (left) eye position was monitored with an Applied Science Laboratories ('ASL') Model E4000SU eyetracker and a 386 lab computer. The ASL is a headband mounted, video-based, IR reflection eyetracker. Figure 2.2 shows the eyetracker in use.

A collimated infrared emitting diode (IRED) illuminates the eye, resulting in a 'bright-pupil' retroreflection from the subject's retina, and a first surface reflection at the cornea (the first Purkinje image). A monochrome CCD camera (without the standard IR rejection filter) is aligned coaxially with the illuminator to image the eye. Figure 2.3 a) shows the bright-pupil and first Purkinje images as captured by the eye-camera. The eye-camera image is digitized and thresholded at two levels in real-time by the ASL control unit. The two threshold levels are adjusted manually so that pixels within the bright pupil are above threshold at one level, while only those pixels within the corneal reflection are above threshold at the second level. The centroid of the pupil and first Purkinje image are then computed by the lab computer. The ASL control unit overlays crosshairs indicating the pupil and first Purkinje centroids on the image from the eye camera. Figure 2.3 b) shows the resulting image as displayed on the 'eye monitor.'

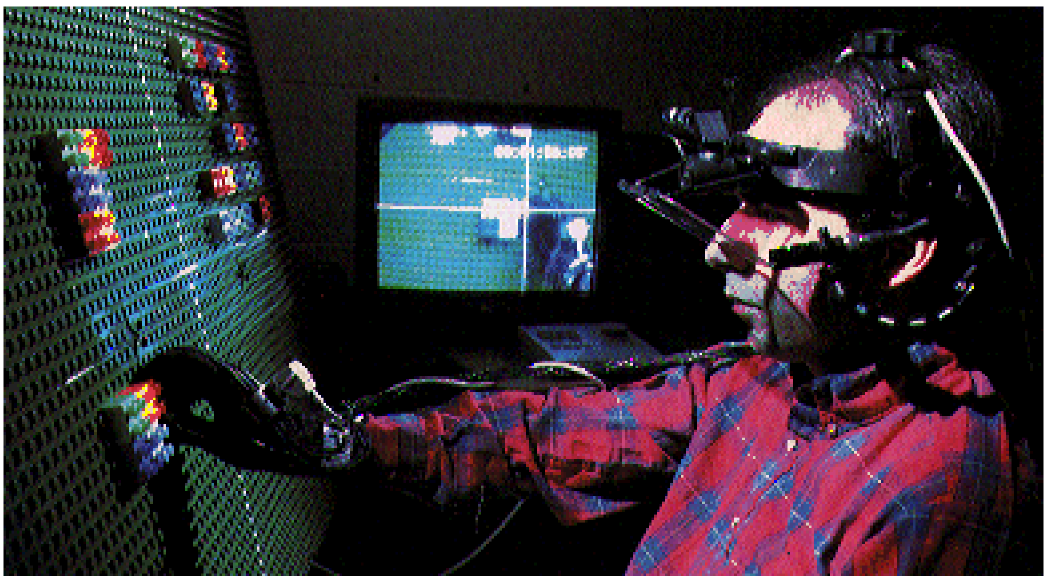
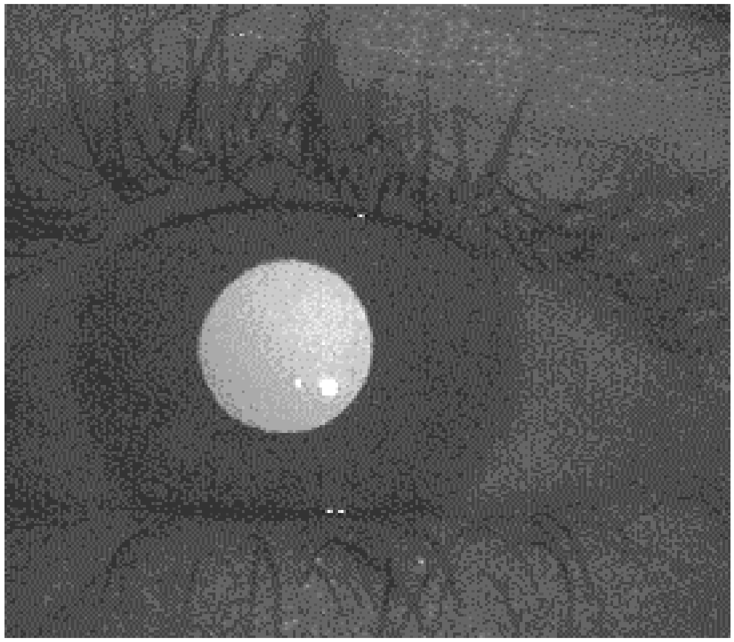
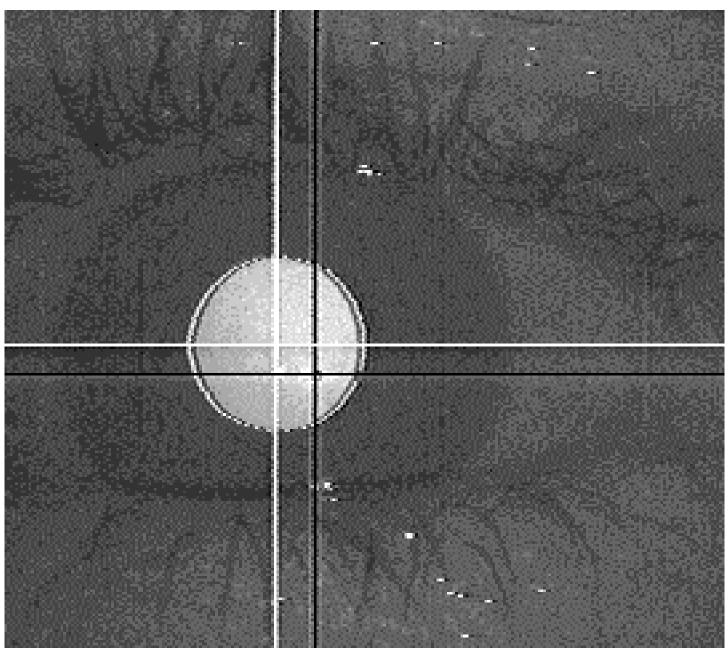


Figure 2.2 Author wearing headband-mounted eyetracker.



a)



b)

Figure 2.3 a) Raw video frame from ASL's 'eye camera' showing bright pupil and first Purkinje image. b) Crosshairs mark centroids of pupil and first Purkinje image computed by the ASL.

Tracking both pupil and first Purkinje images makes the system less sensitive to movement of the tracker with respect to the head because translation of the eye's image (caused by headband movement) causes both pupil and first Purkinje images to move together, while rotation causes differential motion of the two centroids. To reduce eye movement artifacts due to headband movement, eye-in-head position is calculated based on the relative location of the two centroids whenever both are present in the eye-camera image. If the system loses the first Purkinje image, eye position is calculated based on the pupil image alone until the first Purkinje image is re-acquired.

Because the system is video-based, eye position signals are limited to 60 Hz when a single interlace field is used for each eye position computation, or 30 Hz when a full frame (odd and even interlace fields) is used. The latency of the eye position signal is between two and three video fields (33 to 50 msec). The accuracy of the ASL's eye-in-head signal is approximately 1 degree over a central 40 degree field.

Gaze position (integrated eye-in-head and head-position and orientation) is calculated by the ASL using the eye-in-head signal described above and a head position/orientation signal from a magnetic field head-tracking system (see description of Ascension Technology 6DFOB below). The ASL reports gaze position as the X-Y intersection of the line-of-sight with the working surface, identified as the 'calibration' plane. The position and orientation of the calibration plane are defined by entering the three-dimensional coordinates of three points on the plane into the ASL. The coordinates are entered indirectly by measuring the distance and angle of each point with respect to a fixed point. The magnetic field transmitter is used as the origin; the distance to each of the three points (A,B, & C in Figure 2.4) is measured and entered

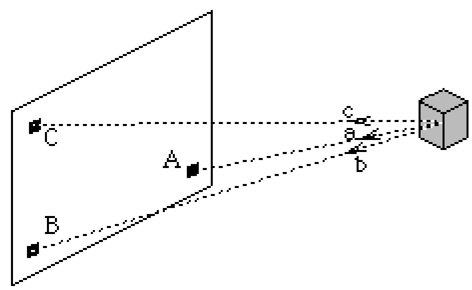


Figure 2.4 The working plane is defined by locating three points (A, B, C) on the plane with respect to the center of the magnetic tracker's transmitter.

manually. The unit-vector from the transmitter to each point is entered using a gimbal attached to the transmitter that holds a HeNe LASER and a magnetic field receiver. The gimbal is directed to each point (using the LASER to ensure an accurate angle), and the orientation of the magnetic field receiver is read by the ASL. Eye-in-head, head orientation and position, and gaze intercept are available on an RS-232C serial interface from the ASL. The digital data stream was collected on an Apple Macintosh 840AV computer for storage and analysis. In addition to this digital data stream, the ASL provides a video record of eye position. The headband holds a miniature "scene-camera" to the left of the subject's head, aimed at the scene (see Figure 2.2). The ASL creates a crosshair overlay indicating eye-in-head position that is merged with the video from the scene-camera, providing a video record of the scene from the subject's perspective on the scene-monitor, along with a crosshair indicating the intersection of the subject's gaze with the working plane (see Figure 2.5). Because the scene-camera moves with the head, the eye-in-head signal indicates the gaze point with respect to the world. Head movements appear on the record as full field image motion. The scene-camera can be fitted with a range of lenses. Figure 2.5 was made with the 3.5 mm wide-angle lens, and shows the barrel distortion typical of such lenses.

Because the scene camera is not coaxial with the line of sight, calibration of the video signal is strictly correct for only a single distance. All gaze points are in the plane of the working board, and subjects typically do not change their distance from the board substantially, so the parallax error is not significant in this task, though it can be significant in tasks not constrained to a near-vertical plane. The parallax error can be eliminated by repositioning the scene-camera below the visor so that it is collinear with the eye-camera (see Figure 2.6). While this orientation eliminates parallax error, it

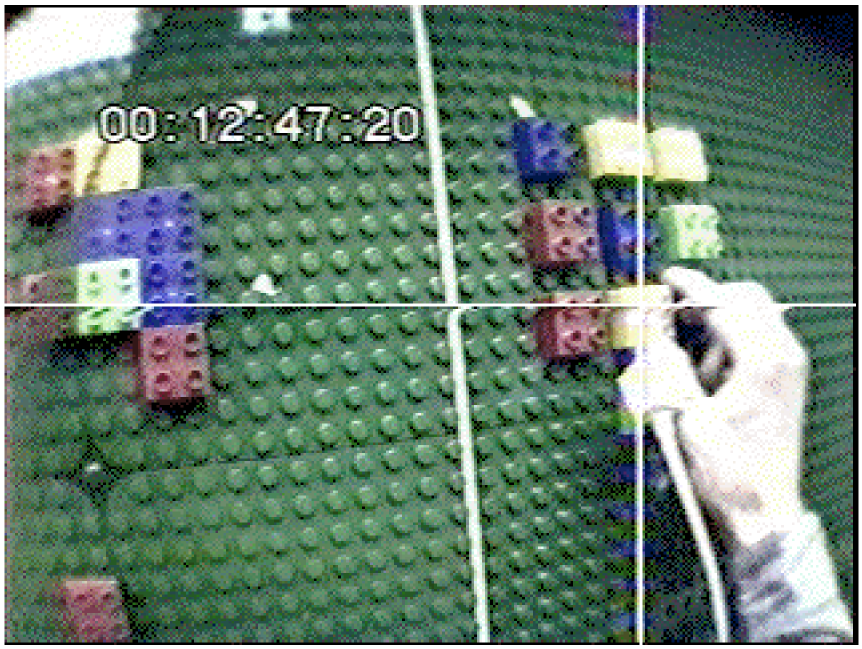


Figure 2.5 A video frame from the ASL's 'scene-monitor' shows gaze position in the scene with a white crosshair overlay on the image from the headband mounted scene-camera

severely restricts the field of view of the scene-camera. In addition, image contrast and chroma are reduced due to the poor reflectance below 800 nm and flare from the IRED illuminator.

The eye-in-space signal calculated by the ASL by integrating the eye-in-head and head position/orientation signals is not affected by parallax -- the scene camera is used only during calibration when the distance to the scene is fixed. After initial calibration, the gaze intersection is calculated by projecting the eye-in-head position onto a 'virtual calibration plane' at the same distance as the calibration plane during calibration. The vector defined by the eye center and the intersection with the 'virtual plane' is then rotated based on the head position/orientation signal, and projected onto the working plane.

The ASL was calibrated for each subject before each trial session. The subject was fitted with a biteboard and seated a comfortable distance from the work surface, typically 60 - 75 cm. Calibrating the ASL requires three steps -- 1) entering the position of the three reference points on the calibration plane (see Figure 2.4), 2) locating the calibration points (9 or 17 points; see Figure 2.7), and 3) recording the subject's pupil and first Purkinje centroids as each point in the calibration target is fixated.

The first step was described above (see page 29). In the second step, the calibration points are located on the work surface by marking them on the scene monitor and entering their absolute, three-dimensional coordinates in the ASL "environment" file. The three points used to locate the calibration plane (points A, B, & C in Figure 2.4)

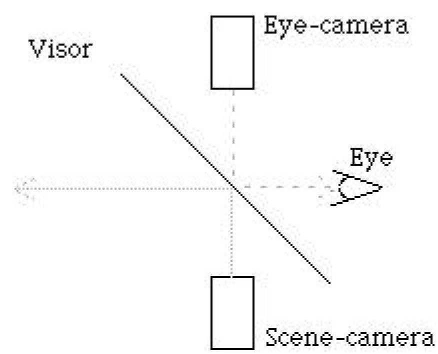


Figure 2.6 Alternate arrangement of scene-camera eliminates parallax error by aligning axes of eye- and scene-cameras.

are the bottom-right, bottom-left, and top-left points in the calibration array, respectively. In the final step, the subject is steadied by a biteboard and instructed to fixate each calibration target in turn, so that raw pupil and first Purkinje images can be grabbed at each point. The calibration function is determined by a proprietary algorithm based on the known target positions and the raw pupil and corneal reflection positions. The calibration can be performed with 9 or 17 points, as shown in Figure 2.7. The 17-point calibration target increases accuracy by allowing the target points to cover a larger area while reducing the area over which eye-position data must be interpolated. The 17-point target is especially critical when the scene-camera is fitted with a wide-angle lens that suffers from barrel distortion.

2.2.2 <u>Monitoring head and hand position</u>

The ASL relies on a magnetic head-tracker to monitor the position and orientation of the head. An Ascension Technology magnetic field tracker (Model 6DFOB, "Flock") was used to monitor the position and orientation of the head and the hand. The 6DFOB system can 'daisy-chain' multiple receivers with a single transmitter. The transmitter unit was mounted above and in front of the subject's head. The transmitter contains three orthogonal coils that are energized in turn. The receiver unit contains three orthogonal 'antennae' coils which detect the transmitters' signals. Position and orientation of the receiver are determined from the absolute and relative strengths of the transmitter/receiver coil pairs. The position of the sensor is reported as the (x, y, z) position with respect to the transmitter, and orientation as azimuth, elevation, and roll angles.

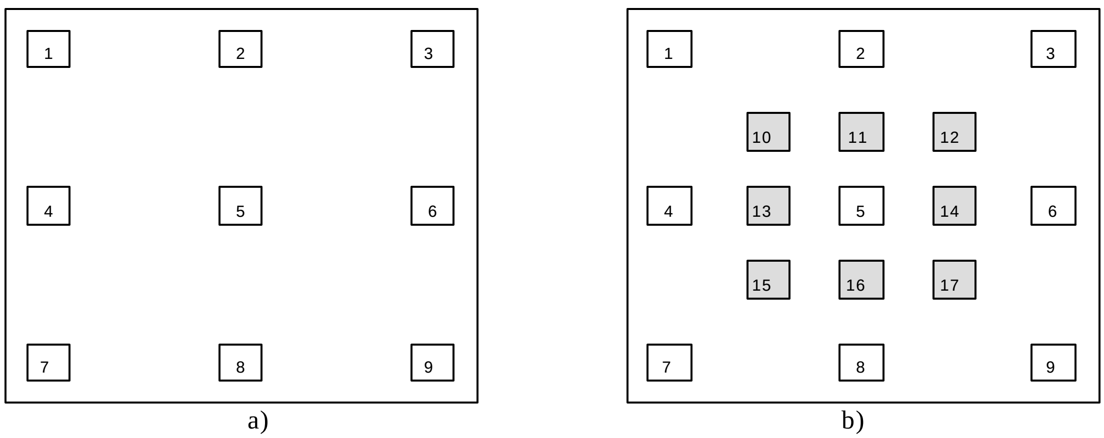


Figure 2.7 Nine and 17 point calibration targets for the ASL eyetracker.

To allow measurements over a range of transmitter to receiver distances, the Ascension 6DFOB adjusts the transmitter's field strength based on the distance of the receiver from the transmitter. The maximum distance at which a clear signal can be detected varies with the strength of the transmitted signal, but too strong a signal saturates the receiver. The maximum field strength is 1 Earth field. Position and orientation values are encoded as 16-bit integers. Distances (x, y, z) are scaled from -36" to 36", yielding a precision of 0.001" ($72''/2^{16}$), or 0.003 cm. Orientation (azimuth, elevation, and roll) are scaled from -180° to 180°, with a precision of 0.005° or 1/3 min arc.

The Ascension system has a range of temporal filter options that can be selected with software commands. There are two classes of filters: 'AC' and 'DC.' The AC filters are band-block filters designed to filter out signals caused by environmental sources operating at around 60 Hz; i.e., 120 VAC line supply, video monitors, and lighting equipment. There are three settings for the AC filters: i) 'AC filters off', ii) 'AC Narrow', and iii) 'AC Wide.' The two AC filter options (narrow and wide) differ in the width of the band of frequencies blocked by the filters. Removing frequency components in this range is not detrimental in itself because there is no appreciable component of head or hand movements beyond about 20 Hz [Rosenbaum 1991], but there is no way to implement such a filter in real-time without introducing a delay in the reported position and orientation values. Unlike the 'AC' filters, the 'DC' filter does not filter out a fixed band of frequencies. Instead, it is an adaptive filter that monitors the sensor's reported values over a period, and adjusts its time constant based on recent position/angle history. If the sensor has shown little motion for a given period, the time constant is increased in an effort to reduce steady-state, or 'DC' position/orientation reports. When sensor movement is detected in this mode, it is at first suppressed by the long time constant on the presumption that it represents noise rather than real movement of the sensor. If the movement continues it is assumed to represent real motion of the sensor, and the time constant is reduced. The user can set upper and lower bounds on the time constant to limit the degree to which the adaptive filter can adjust to varying inputs, and the speed at which the filter adapts to sensor motion. The default filter configuration ('AC wide' and 'DC' filters on) produces very low noise output at the cost of increased temporal lag between sensor movement and position reporting. Because the DC filters actively vary their time constants the actual lag introduced by the 'DC' filter is not constant.

The "Flock" sensors used for head and hand tracking were characterized to determine the accuracy and noise in the measurement system with and without the default filters. The three-dimensional position signal accuracy was dependent on the separation between transmitter and receiver. Absolute error was below 0.2 cm when the receiver was within 40 cm of the transmitter, but increased dramatically beyond that distance. Errors reached approximately 1 cm at a distance of 65 cm, and 5 cm at a distance of 85 cm. Orientation values (computed based on the relative strength of the three channels) are less sensitive to distance. Errors were below 5 minutes of arc, and unaffected by distance out to 65 cm.

See the Appendix for details of the Ascension 6DFOB calibration. Two Ascension 6DFOB units were used in the experiments. One receiver was attached to the eyetracker's headband to monitor head movements; the second receiver was taped to the subject's thumb to monitor hand movements. The hand position data was sent via a serial connection to the lab computer. The head position data was reported directly to the ASL's PC where it was integrated with the eye position signal to calculate the integrated gaze position. The raw head position signal was also sent to the Macintosh lab computer where it was logged along with gaze, eye, and hand movement signals. In some of the later experiments, the second Flock unit (used to track hand movements) was replaced with a Polhemus "Fastrak" model. This allowed a separate transmitter (dedicated to monitoring hand position) to be placed behind the board so that the distance between transmitter and receiver was minimized. Because the noise levels rise as the distance between transmitter and receiver increases, there is an advantage to keeping the distance small.

The headband mounted scene-camera was equipped with a 3.5 mm focal length wide-angle lens providing a wide field (approx. 110° horizontal) that included the board and the subject's hand. A Sony EVO-9650A Hi-8 video deck was used to record the video from the scene camera with gaze position overlay. A Hi-8 format timecode (30 frames per second) was recorded with each frame, allowing timing measurements to be made on playback. The video deck was connected to the lab computer with an RS-232C serial interface so that it could be controlled automatically by the computer. In addition to starting and ending recording sessions, the interface allowed the program to poll the deck to determine the timecode being written to tape. This timecode was written to the header of the digital data file for each trial so that the data stream and video record could be correlated.

2.3 Procedure

The model area contained the pattern to be duplicated (a pattern of eight blocks of four colors), the resource area contained the blocks to be used to construct the copy, and the copy was constructed in the workspace. The blocks were square, approximately 2.5 cm on a side. The subject was instructed to duplicate the pattern as quickly as possible without making errors using one hand (of the S's choice), but was otherwise free to choose any strategy and sequence of movements to accomplish the task. Ten subjects performed the basic block-copying task, completing from 60 to 210 block moves each (mean 160). Subjects were seated approximately 60 - 75 cm from the board (each subject selected the distance to be a comfortable reaching distance from the board). At the typical distance, the board subtended approximately 80 x 60 degrees of visual angle. Each block subtended approximately 2.25 degrees.

Subjects were fitted with the ASL eyetracker, and a calibration was performed. Early experiments were done with a 9 point calibration target; later the 17 point system was implemented. The outermost calibration points formed a rectangle approximately 25 x 20 degrees (the exact value depended on viewing distance, which was set by the subject). The eyetracker was calibrated at the beginning of each block of trials, and zeroed on a central fixation point just before each trial began.

2.3.1 Scoring of block-move strategies

The sequential nature of the block movements (enforced by requiring that blocks be moved with one hand) makes it convenient to break up the sequence of movements in each 8-block trial into smaller subtasks. Taking the movement of each block as a sub-task, it is useful to consider the sequence of eye, head, and hand movements taking place for each block move. We consider each 'block move' sub-task to begin when the previous block has been placed in the workspace and the eyes move away from that area. The sub-task is completed when the present block is "dropped" in the workspace. Each block move is made up of a set of lower-level subtasks; the subject must find (or have remembered) the color and position of the block to be moved, move the hand to "pick up" a block of the same color in the resource area, then return to the workspace to add the block to the correct position in the duplicate being constructed. It is instructive to examine the pattern of fixations used in copying each block. Those fixations serve several purposes: gathering information about the model pattern, the location of particular blocks in the resource area, monitoring progress in the workspace, and guiding hand movements picking up and dropping the blocks.

The gaze- and hand-movement 'primitives' making up each block move were used to describe the subjects' strategies. Gaze changes are labeled by the areas in which fixations occur. The primitives were labeled:

M - fixation in the ***[m]****odel* area

P - block **[p]**ickup (with fixation in the resource area)

D - block **[d]**rop (with fixation in the workspace area)

"Pickup" and "Drop" events are used to label hand movements, and it is understood that Pickup ('P') and Drop ('D') events are accompanied by fixations in the resource and workspace areas, respectively. While the gaze often left the resource area before a block was picked up, there was almost always a fixation associated with the pickup and drop events.

By reviewing the videotaped records in slow motion, the sequence of fixations and hand movements used to copy each block was used to label the strategy used for that block. For example, Figure 2.8 shows a typical 'block move' sequence schematically. The sequence begins with the first fixation after the previous block has been placed in

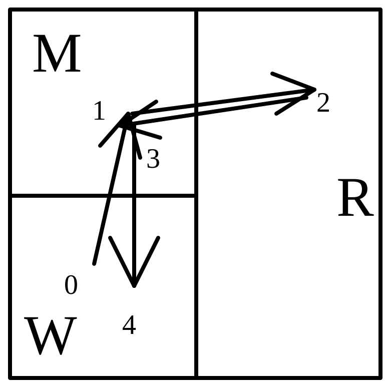


Figure 2.8 Sequence of gaze changes making up an 'MPMD' block move.

the workspace. In this case the gaze moves first to the ***[M]****odel* area ('M-...'), then to the workspace, where the hand **[P]**icks up a block ('M-P-...'), then returns to the ***[M]****odel* area ('M-P-M-...'), and finally to the workspace, where the block is '**[D]**ropped' in position ('M-P-M-D'). The block move is thus labeled as an 'MPMD' sequence.

2.4 Results

A striking aspect of subjects' performance was the highly stereotyped behavior in completing the task, marked by frequent fixations in the model area and a relatively small number of strategies. Approximately 90% of the block moves were classified into the four strategies shown in Table 2.1. Figure 2.9 a), b), and c) show the sequence of fixations for block move sequences labeled MPD, PMD, and PD, respectively.

While these four strategies were most common in the basic experiments, it was necessary to define a fifth category because in some experiments subjects made even more frequent model fixations. Block moves were categorized as ">MPMD" sequences if the subject looked more than twice into the model area during a single block move. A small number of block moves (3.3% across 10 subjects) still did not fit into any of the above categories and were labeled "other." These were typically block moves in which the subject had difficulty picking up or placing a block, or in which the subject made an error. Table 2.2 shows the six categories used to score the block move strategies.

Table 2.1 Four strategies into which ninety percent of the subjects' block moves were categorized

Model-Pickup-Model-Drop	"MPMD"
Model-Pickup-Drop	"MPD"
Pickup-Model-Drop	"PMD"
Pickup-Drop	"PD"

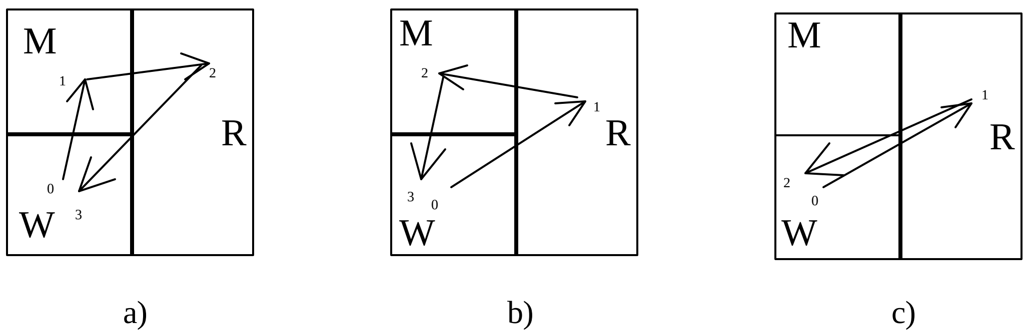


Figure 2.9 Sequences of gaze changes making up a) an 'MPD' block move, b) a 'PMD' block move, and c) a 'PD' block move.

Table 2.2 Ninety-five percent of the subjects' block moves were categorized into the first five strategies shown.

"> Model-Pickup-Model-Drop"	"> MPMD"
Model-Pickup-Model-Drop	"MPMD"
Model-Pickup-Drop	"MPD"
Pickup-Model-Drop	"PMD"
Pickup-Drop	"PD"
any other sequence	"other"

2.4.1 Basic Features of Task Performance

2.4.1.1 Relative Frequency of Strategies

A striking feature of subjects' performance in the task was the consistent reliance on frequent reference to the model. Subjects often averaged over two saccades into the model area in the course of moving each of the eight blocks. This frequent reference to the model is surprising because humans are capable of holding several items of information in short term memory; certainly enough information to remember the position and color of a few blocks. Except for one subject all subjects averaged more than one fixation in the model area for each block copied. The average number of model references per block copied was 1.51; subject jw averaged only 0.93 model references per block. Figure 2.10 shows the mean relative frequency of each of the six strategies listed in Table 2.2, averaged across 14 subjects. Figure 2.11 shows the relative frequency of strategy use for each subject. In nine of the fourteen subjects, over 50% of block moves were accomplished with two or more model fixations. In twelve of fourteen, block moves with two or more model fixations were more common than any other strategy.

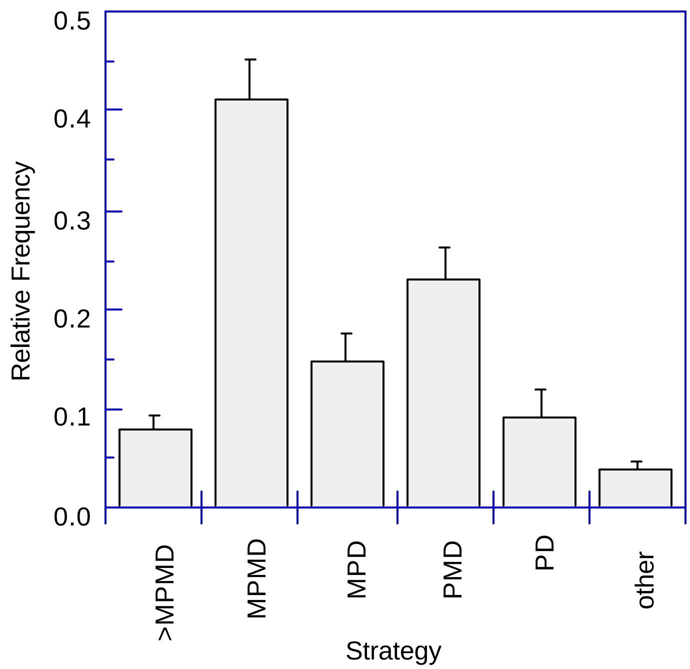
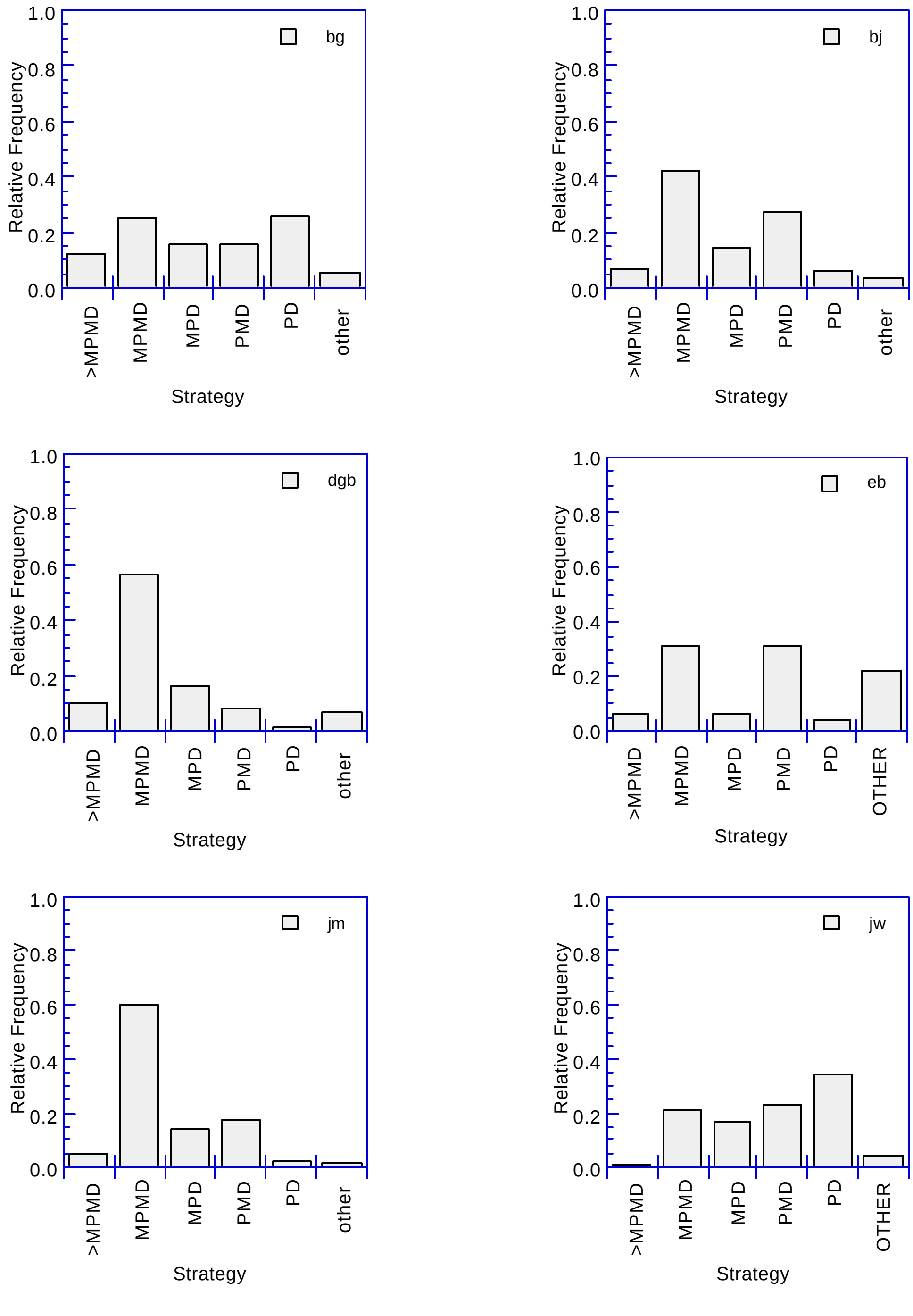
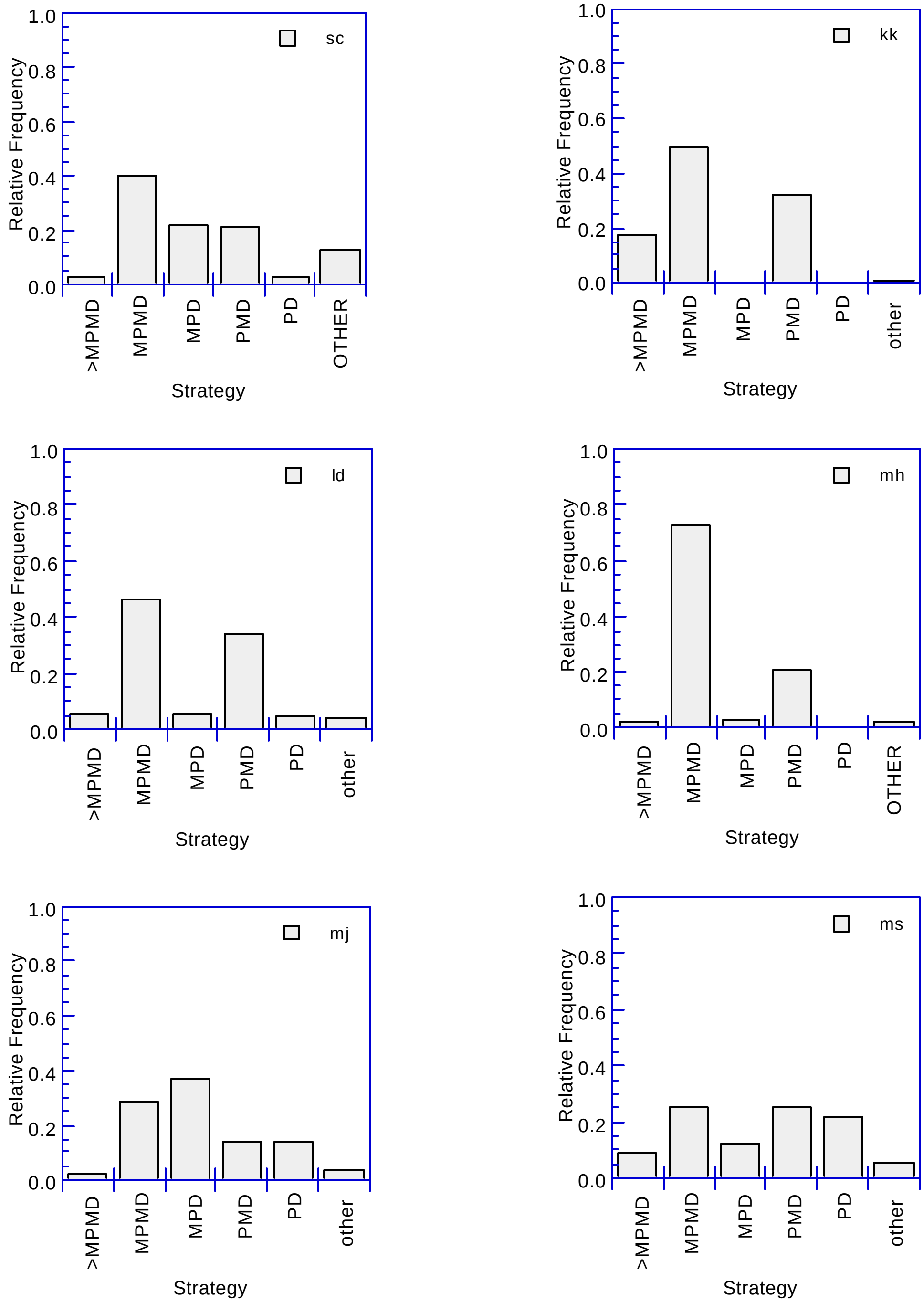


Figure 2.10 Mean relative frequency of six strategies for fourteen subjects.



(Figure continues on next page)



(Figure continues on next page)

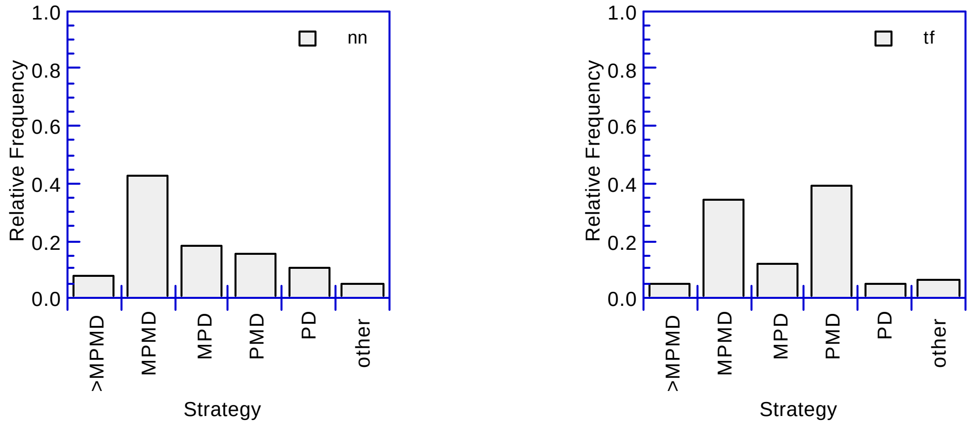


Figure 2.11 Individual relative frequency histograms of strategies used in block-copying task by fourteen subjects.

2.4.1.2 Change in Strategy Use Over the Eight-Block Trial

While subjects' performance in the task suggests that very little information about the model pattern is remembered from one block to the next, examination of the strategies used as a function of the serial block number (i.e., over the course of each trial) does suggest that some information about the model is acquired during the trial. Figure 2.12 shows that the low-memory MPMD strategy is most common for the first block, while the PD strategy increases over the trial. So, while subjects tend to rely on frequent eye movements rather than working memory, it is evident that they do use information from previous model references to some extent -- some internal representation of the model pattern is apparently built up over the many fixations used in the construction of the duplicate pattern.

2.4.1.3 Number of References to the Model per Block Copied

A useful metric of the degree to which working memory is used in task performance is the average number of model references per block copied. Figure 2.13 shows the change in the frequency of fixations in the model area as a function of serial block order. There is a significant decrease in the number of model references after the first block ($P<0.005$), and on the last block of the eight-block pattern ($P<0.05$). Blocks 2 through 7 are relatively stable at approximately 1.4 looks per block.

2.4.1.4 Trial Duration

Subjects were instructed to complete the task as quickly as possible without making errors. In addition to monitoring strategy use by each subject, the trial duration was recorded. Figure 2.14 shows the average trial duration for each subject, and the

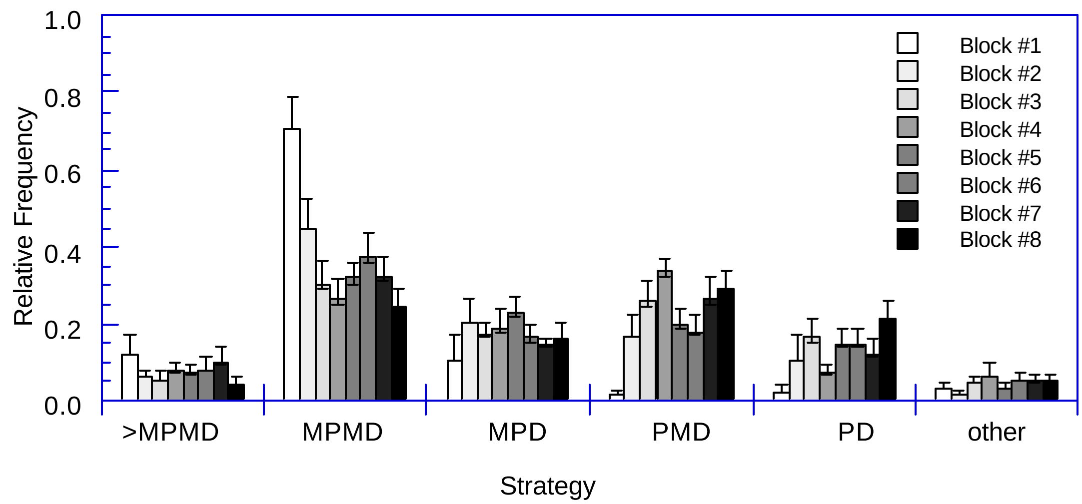


Figure 2.12 Variation in strategy use over the eight-block trial.

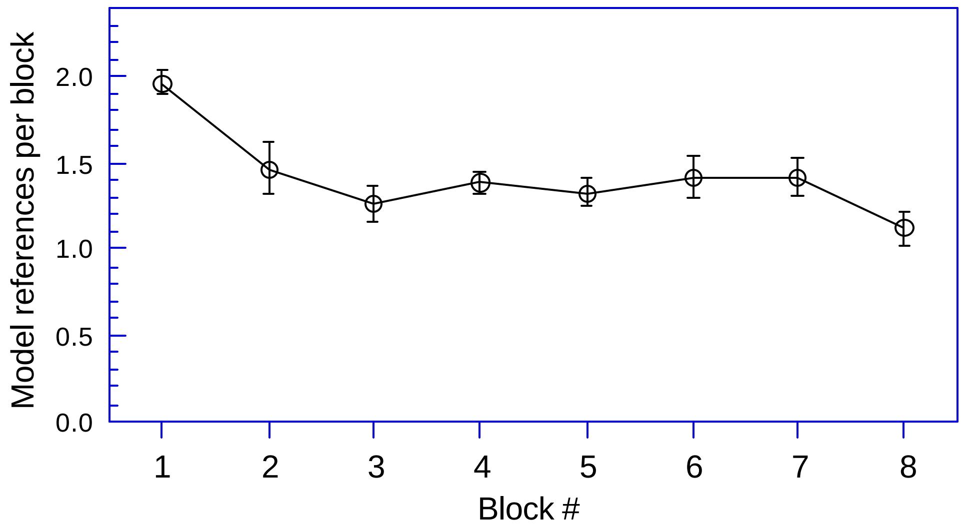


Figure 2.13 Average number of model references per block as a function of serial block order

2.4.1.5 Fixation Durations in Model Area

Because our understanding of the common occurrence of the MPMD strategy suggests that different information is extracted in the first and second model fixations, individual model fixation durations were examined to learn whether the fixation durations might indicate the complexity of the information gathered. For example, subjects might require less time to bind the color of a target block than to bind its location in the model during an MPMD sequence, or the model reference in an MPD sequence might be longer than that in a PMD sequence. Four subjects' records (eb, sc, jw, and mh) were analyzed to extract the time spent in the model area during the first model reference in an MPMD sequence (labeled MPMD1), and the second model reference (labeled MPMD2). The fixations for the model reference in MPD and PMD sequences were also measured to see whether there were significant differences between model fixation durations in those strategies. The results were idiosyncratic, as seen in Table 2.3 a). While a significant difference between the first and second model fixations (MPMD1 & MPMD2) was found in three of the four subjects, two (eb and sc) spent less time in the first model reference, and the third (jw) spent more time in the first model reference. The fourth subject (mh) spent less time during the first model reference, like eb & sc, but the difference was not significant. The inconsistency between subject jw and the other subjects can be understood, however, by noting that subject jw performed the task very differently than other subjects. While on average the MPMD strategy was used in approximately 40% of block moves, that double-look strategy accounted for only half as many of jw's block moves. The average subject moved only 12%

of the blocks using the PD strategy; jw was three times more likely to copy a block without a model reference. To determine whether the differences in model reference durations were due to different strategy use among the subjects, the number of fixations made during each model reference were analyzed. While eb, sc, & mh each averaged between 1.3 and 1.8 fixations during the first model reference (MPMD1), jw averaged 2.6 fixations. There was no such difference in the second reference (MPMD2); all four subjects averaged between 1.5 and 1.8 fixations. Table 2.3 b) shows the average duration of MPMD1 and MPMD2 model references that contained only one fixation in the model area. This presumably eliminates from analysis any block moves in which jw made several fixations in preparation for a future PD block move. This reduced the mean durations for all subjects, but the decrease was most dramatic for jw. All four subjects spent less time in the first model reference than in the second, though the differences were only significant for eb and sc. The comparison between MPD and PMD model references was less clear; all four subjects spent more time in the model in the PMD than during the MPD sequence (opposite our expectations), though the difference only reached significance for subject eb.

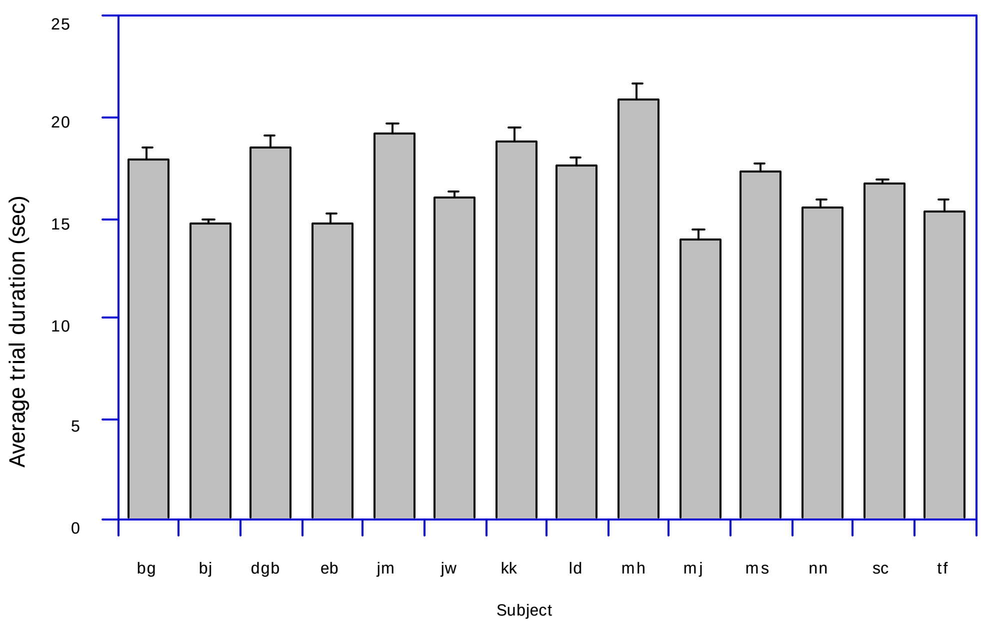


Figure 2.14 Average trial duration (time to duplicate the eight block pattern) for fourteen subjects.

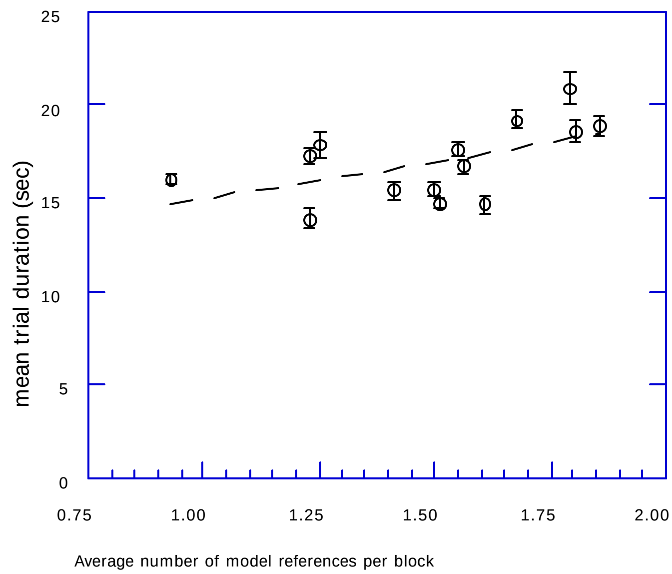


Figure 2.15 Average trial duration as a function of the number of model references per block.

Table 2.3 Mean time [msec(s.e.m.)] in model area for the first and second MPMD model references, MPD, and PMD sequences. a) Average of all model references. b) Average of MPMD1 & MPMD2 model references with only one fixation.

a) Average of all model references

Model Reference	eb	sc	jw	mh
MPMD1	385(23)	362(16)	599(37)	335(24)
MPMD2	423(15)	424(12)	472(19)	355(19)
MPD	346(45)	416(25)	469(31)	325(17)
PMD	447(15)	472(19)	470(22)	356(39)

b) Average of MPMD1 & MPMD2 model references with a single fixation.

Model Reference	eb	sc	jw	mh
MPMD1	292(14)	326(34)	389(29)	282(23)
MPMD2	418(24)	390(16)	405(24)	315(18)

2.5 Discussion

These experiments support and generalize the result of the earlier block-copying experiments performed on a CRT with a mouse [Ballard, Hayhoe, Li, & Whitehead 1992]. The fact that subjects are now free to make unconstrained movements while performing the task avoids concerns about the un-natural constraints necessary to use other kinds of eyetracking devices, and about use of a computer mouse instead of natural hand movements. It is now possible to investigate complex, natural behaviors.

In performing the block-copying task, subjects make very frequent references to the model area; the modal strategy includes two looks to the model area for every block copied. The average number of model references was greater than 1.5, and all but one subject averaged more than one look to the model for every block copied. Subjects chose strategy even though they are capable of memorizing multiple-block sub-patterns; in a preliminary experiment in which subjects manipulated colored block patterns on a Macintosh CRT using a mouse, subjects were allowed to inspect the model pattern for a variable length of time, then constructed the duplicate from memory after the model was removed [Ballard, Hayhoe, Li, & Whitehead 1992, Ballard, Hayhoe, & Pelz 1995]. Subjects could copy patterns of up to four blocks with few errors. So while subjects are capable of remembering multi-block sub-patterns, they choose not to operate at the maximum capacity of working memory when free to select their own strategy. Instead, they seek to minimize reliance on working memory by acquiring information incrementally during the task.

Subjects used frequent eye movements to serialize the complex task into a series of simpler subtasks, each placing only a small load on working memory. Rather than work from a rich internal representation built up over multiple fixations, subjects chose instead to work from moment-by-moment representations, gathered 'just-in-time' from the most recent fixation. In the extreme, individual features of objects appear to be gathered separately. The most plausible interpretation of the role of the fixations is that color is loaded in the first fixation and position in the second. The color is loaded in the first model fixation, and used to pick up a block of the appropriate color in the resource area, then the position is loaded in the second model fixation and used to place the block in the workspace. The fact that this 'extreme' case was the modal response observed in the experiment provides an important insight into the manner of representation employed by humans under natural conditions. It is important to distinguish this experiment from the class of experiments designed to find the limits of humans' visual memory. We know from those experiments that we are capable of maintaining several objects in memory simultaneously [Miller 1956, Baddeley 1986]; this experiment shows that subjects choose to operate far from that limit when performing a task that allows visual reference to the relevant information.

Performance in the task was not totally 'memoryless' however. Strategies did shift over the course of a trial. Over 90% of the first blocks were copied using ≥MPMD block moves (i.e., >MPMD + MPMD) strategies, but that number fell to only 30% for the eighth block. The average number of model references per block copied fell from 2.0 for the first block to 1.2 for the last block, so subjects did not approach each block move from 'scratch' using only a single variable to store color and position information. So while it is clear that subjects seek to minimize memory use with frequent fixations, some information is retained across fixations. Subjects' performance in the task also suggests that eye movements should be considered as more than unfortunate side-effects of the limited foveal region. Eye movements serve a cognitive role in performing the complex task, serializing the task into simple components, each with minimal memory demands. Perceptual and motor events are ordered and executed by marking elements in the scene with fixations.

The reluctance to use working memory to capacity can be explained if such memory is expensive to use with respect to the cost of the serializing strategy. The tradeoff between eye movements and working memory can be visualized with the help of schematic timelines illustrating working memory load over time. Model fixations are used to bind values from the scene into variables in working memory. Figure 2.16 is a schematic illustration of a strategy that relies heavily on working memory. At the beginning of the task, the subject inspects the model and loads the color and position of the three blocks forming the model's top row into working memory. We can think of this action as *"binding the values from the fixation point into working memory."* Several variables are needed in the strategy shown in Figure 2.16, and they must retain their contents for varying durations. In this example the position of block #3 must be held until the third block is finally placed in the workspace. Using working memory in this way to 'load up' information from the model minimizes the number of eye movements to the model area. After completing the three blocks the subject could fixate the second row of the model and bind that set of variables into working memory. This series of actions would be repeated until the entire model was duplicated. This strategy would appear as an initial block with one or more fixations (e.g., >MPMD, MPMD, MPD, or PMD) followed by two PD block moves, a sequence that was not observed in these experiments.

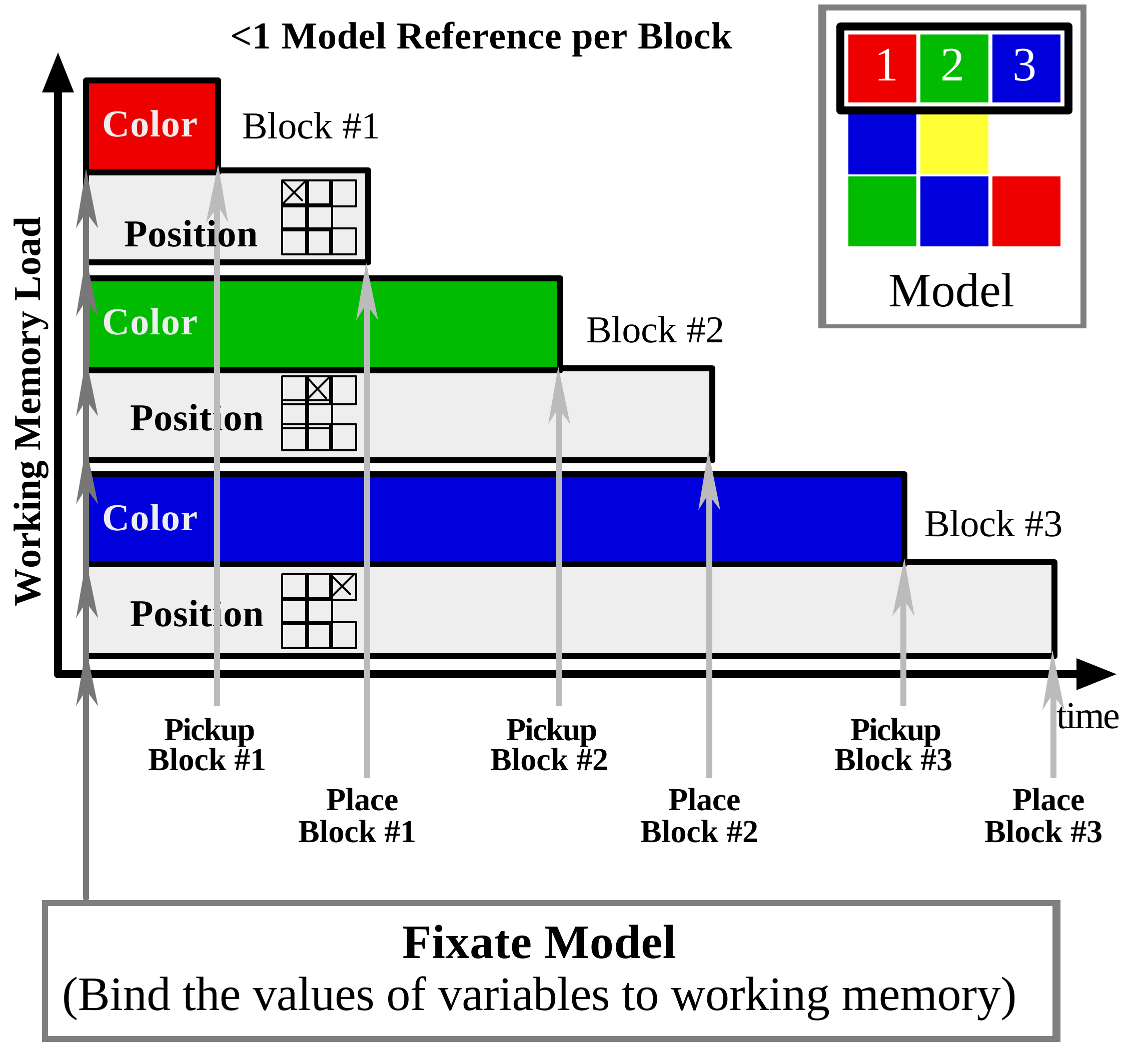


Figure 2.16 Working memory load as a function of time for a subject who executes three blocks 'from memory.'

Figure 2.17 illustrates a strategy that does not rely on working memory to maintain information about blocks other than the immediate block being copied. In this example, each block is fixated just before it is copied, requiring only that blocks' color and position to be bound to variables in working memory. Both the number of variables and the duration for which they must be held are reduced by trading off more frequent model fixations for a lower working memory load. In this case, the eyes must be moved to the model area for each block copied, but only two variables are required. Subjects are apparently using fixation as a "pointer" to elements in the environment to gain the same representational economies that machine vision systems have shown by using deictic pointers [see Agre and Chapman 1987; Ballard 1989, 1991; Brooks 1991]. This use of fixation is central to the tradeoff between eye movements and working memory load. Figure 2.17 illustrates the MPD strategy, in which each block move is preceded by a model fixation. Figure 2.18 shows the minimum memory strategy; each block is fixated twice -- once to load its color, then a second time to load its position. Such a strategy could use just one deictic variable by postponing the second model fixation until after a block of the needed color was picked up from the resource area. This MPMD strategy relies the least on working memory, requiring only a single variable, but requires twice as many model references as the MPD strategy shown in Figure 2.17. In spite of the number of eye movements required to copy the model pattern using the MPMD strategy, it was the most common strategy. Analysis of the time subjects spent in the model area during the two model references in the MPMD strategy is consistent with the interpretation that subjects are gathering different information during the two references to the model.

These results show that subjects choose to refer to the external world, acquiring (and re-acquiring) information just as it is needed, rather than relying on working memory to hold even relatively simple information relevant to the task. The frequency of eye movements used to minimize working memory suggests that working memory load is expensive relative to eye movements. In the next chapter we will explore to what degree subjects trade-off eye movements and working memory, and how the balance between eye movements and working memory can be manipulated by task constraints.

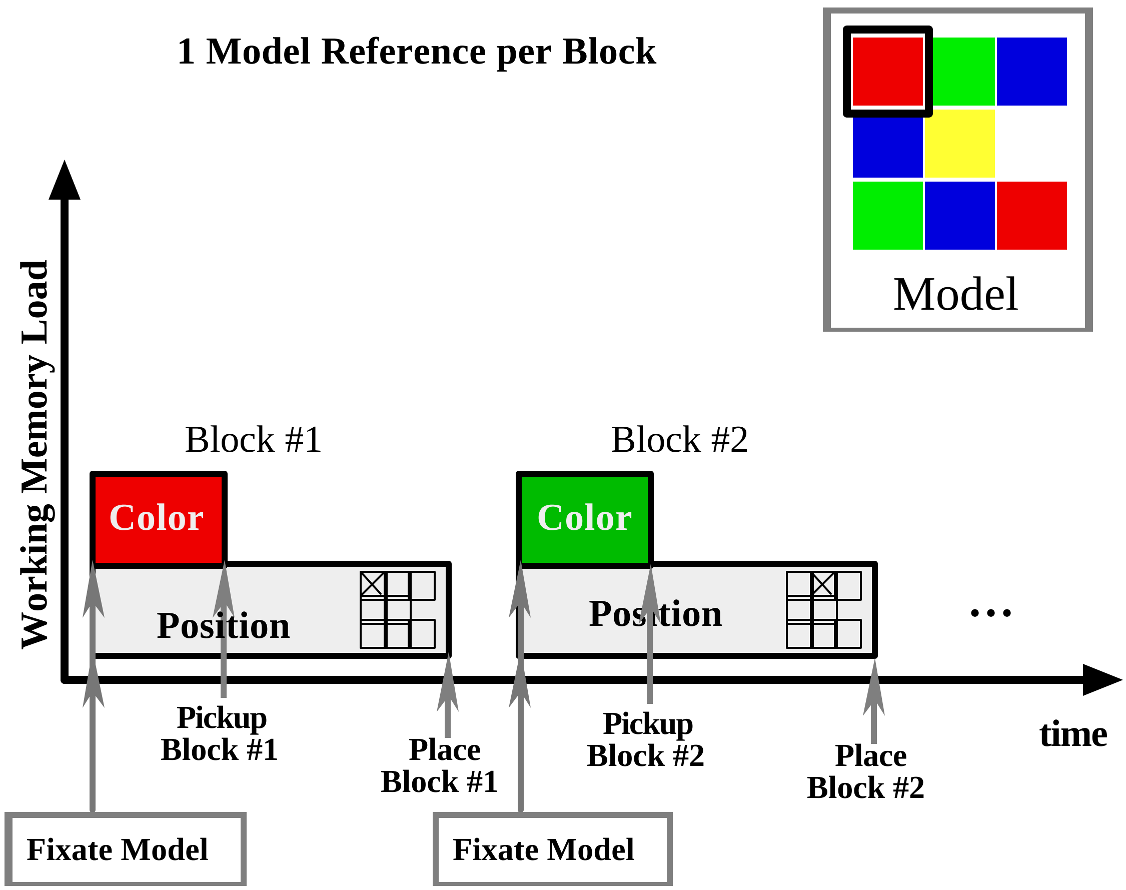


Figure 2.17 More frequent model fixations reduce the working memory load.

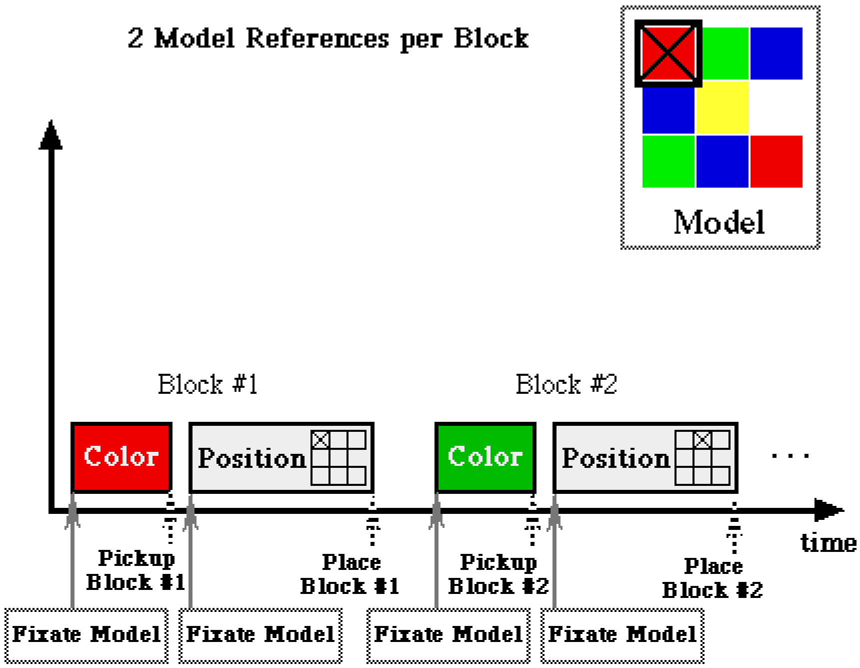


Figure 2.18 Working memory load is minimized in the MPMD strategy. Only the color or position of a single block is held at any time.

3. Tradeoff between frequent eye movements and working memory load

3.1 Introduction

The results of the block-copying task in Chapter 2 are interpreted as evidence of the high cost of working memory relative to eye movements. Subjects chose to make frequent eye movements to the model area rather than copy multi-block patterns from memory, even though memorizing multi-block sub-patterns is well within the limits of working memory. The conclusion is not that eye movements are used to the exclusion of information held in working memory, but that subjects strike a balance between the two based on their relative cost and the amount of information in the model pattern needed to complete the task. Given this analysis, we expect that the trade-off between working memory and frequent eye movements would be affected by varying those parameters. If subjects are able to dynamically adjust the balance between memory and eye movements, increasing the cost of frequent model references would shift the balance toward fewer eye movements and higher memory strategies. If the frequent model references in the control condition serve to gather information necessary to perform the task (rather than being simple artifacts of experimental design, because the eyes are so much faster than the required hand movements), then decreasing the information content of the model pattern should lead to fewer model references. Working memory load is reduced, so fewer eye movements would be needed to maintain the desired balance.

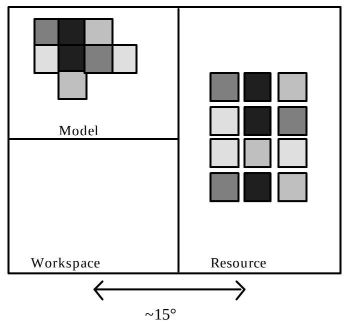
In order to investigate the significance of the frequent model references observed in the main experiment, both types of manipulation were explored: 1) increasing the cost of the model references by increasing the distance between the three areas, and 2) reducing the information content of the model by eliminating color or position information.

3.2 Manipulating Task Parameters

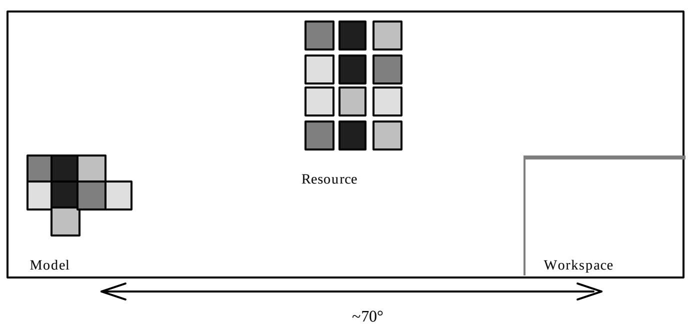
3.2.1 Increasing the Cost of Frequent Model References

To test the effect of varying the relative cost of frequent model references, the model, resource, and workspace areas were rearranged to increase the cost of the frequent gaze changes. Placing a larger distance between the three areas required much larger gaze changes than in the control configuration. Figure 3.1 shows the configuration, labeled the 'far' condition. The model and resource are separated by ~40°, and the workspace and model are separated by ~70°. This angular separation required large gaze shifts made up of eye, head, and torso movements that are more costly than are the equivalent movements in the 'near' condition.

One measure of the increased cost of eye movements in the far condition is the time it takes to perform each block move. Figure 3.2 shows the average duration of block moves as a function of strategy for five subjects in the near and far conditions (error bars show between-subject s.e.m.). Performing the task in the larger space required more time for all strategies, but the increase was not constant across strategies. As seen in Figure 3.3, the PD strategy (with no model fixations) showed the smallest average increase (285 msec), and the >MPMD strategy (with at least three model references)



a) control



b) far

Figure 3.1 Position of model, resource, and workspace areas for the control and 'far' configurations.

had the largest increase in block move duration (885 msec). An average 'temporal cost' of the model references was derived by subtracting the average increase of a PD block move from the other values, and normalizing by the number of model references. The result, 420 msec (s.e.m. = 50 msec), is an indication of the average penalty suffered for each model reference.

Nine subjects performed the block-copying task in the 'far' condition. All of the subjects had previously performed the task in the control configuration. Subjects performed from 96 to 208 block moves (mean = 160). Figure 3.4 shows the relative frequency of strategies selected for the control and 'far' conditions averaged over the nine subjects (error bars show between-subject standard error). Figure 3.5 shows the relative frequency of strategies for each of the nine subjects. All subjects showed a decrease in low-memory strategies (MPMD and >MPMD), and seven of the nine subjects increased use of the high-memory PD strategy (one subject never used the PD strategy). It is useful to examine the *shift* in strategies used by the subjects between the control and 'far' conditions. Figure 3.6 shows the average change in the relative frequency of each strategy across four subjects. Each value represents the difference between the relative frequency in the far and control conditions. The mean response is a reduction in the low memory >MPMD and MPMD strategies, balanced by increases in the higher memory MPD and PD strategies, which are performed with fewer model references.

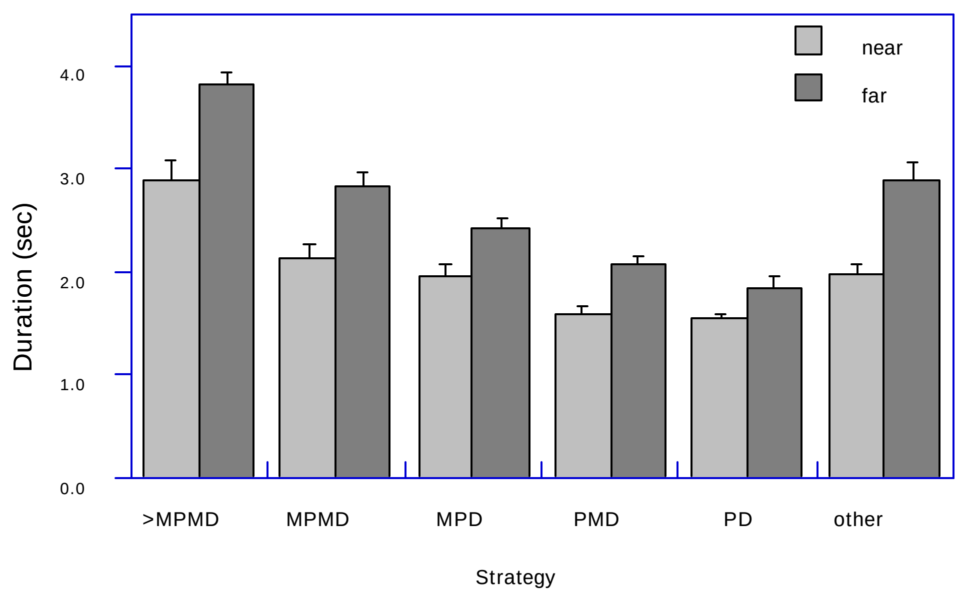


Figure 3.2 Duration of individual block moves averaged over five subjects in 'near' and 'far' conditions.

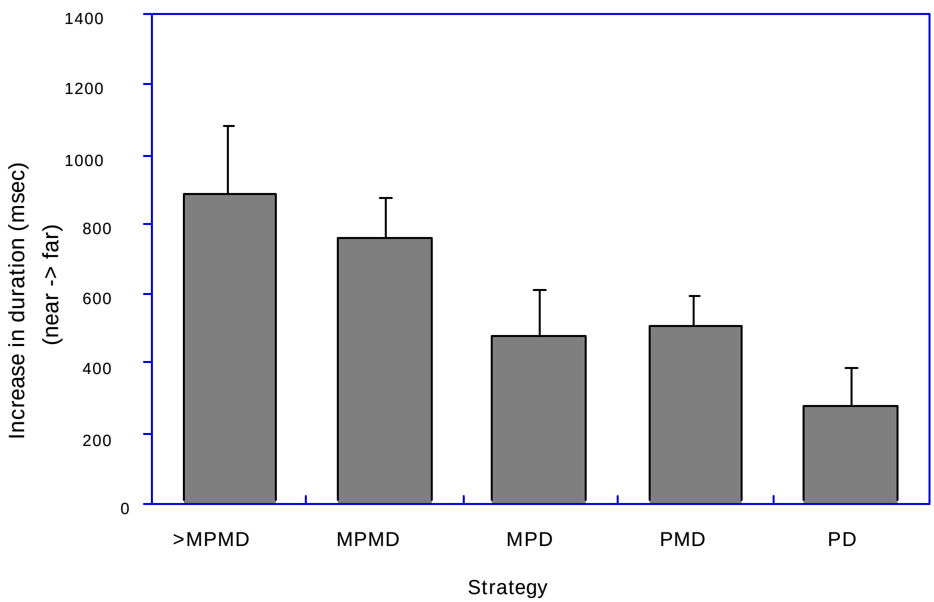


Figure 3.3 Added cost of model fixations in the 'far' condition, averaged over five subjects.

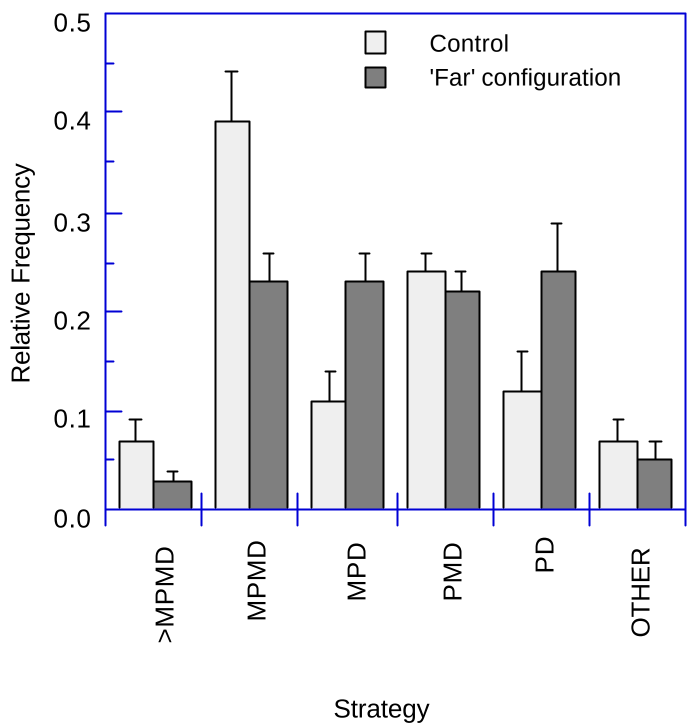
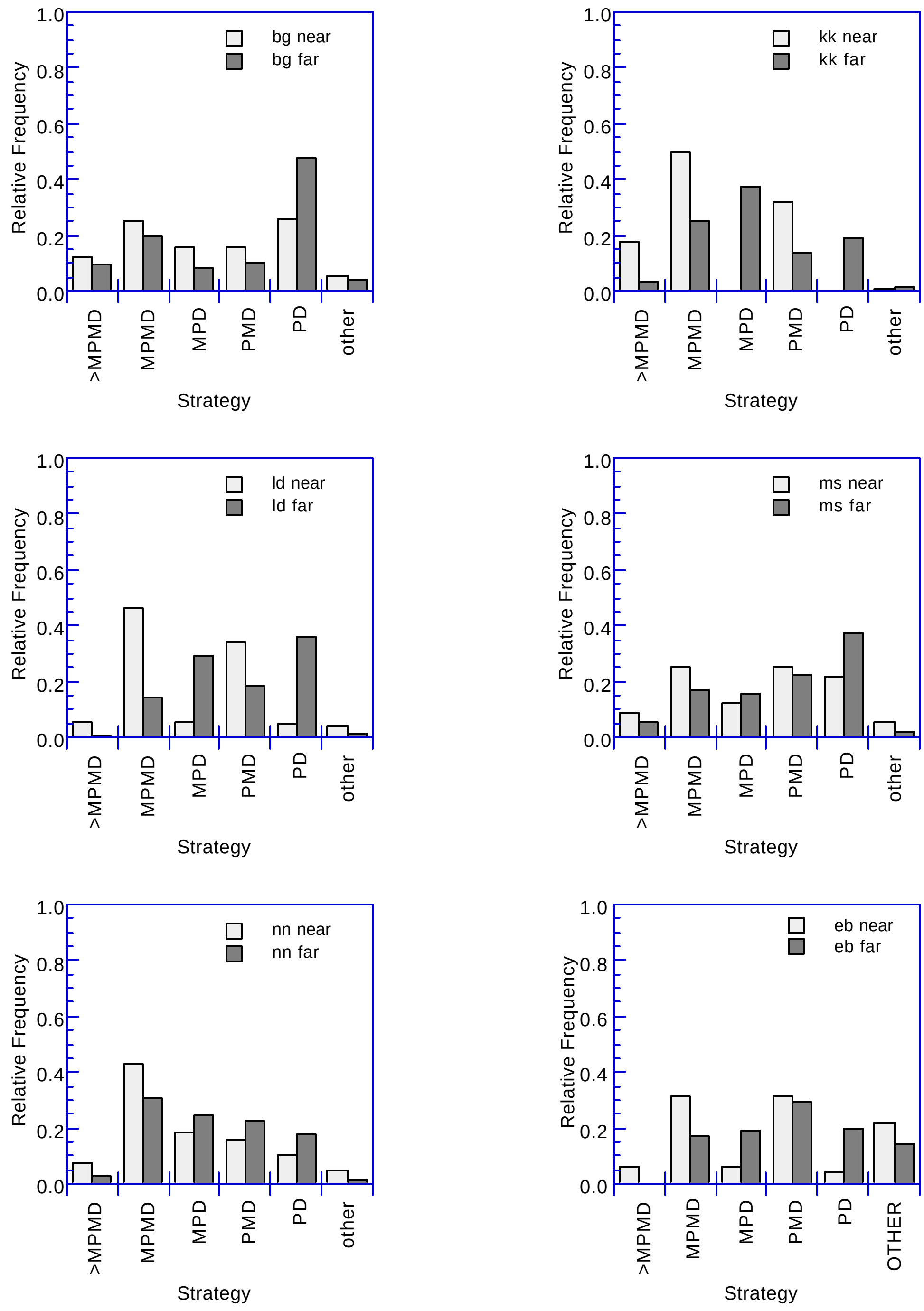


Figure 3.4 Mean relative frequency distribution for control and 'far' conditions averaged across nine subjects.



(See legend, next page)

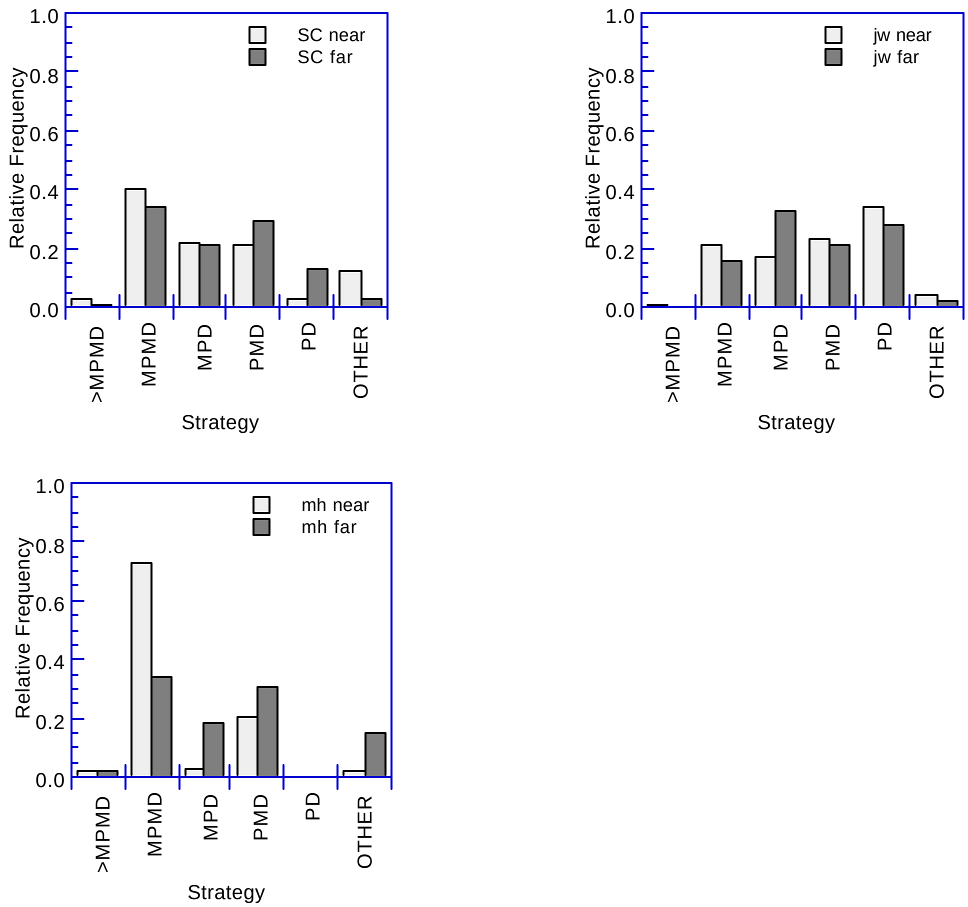


Figure 3.5 Individual relative frequency distribution for control and 'far' conditions for nine subjects.

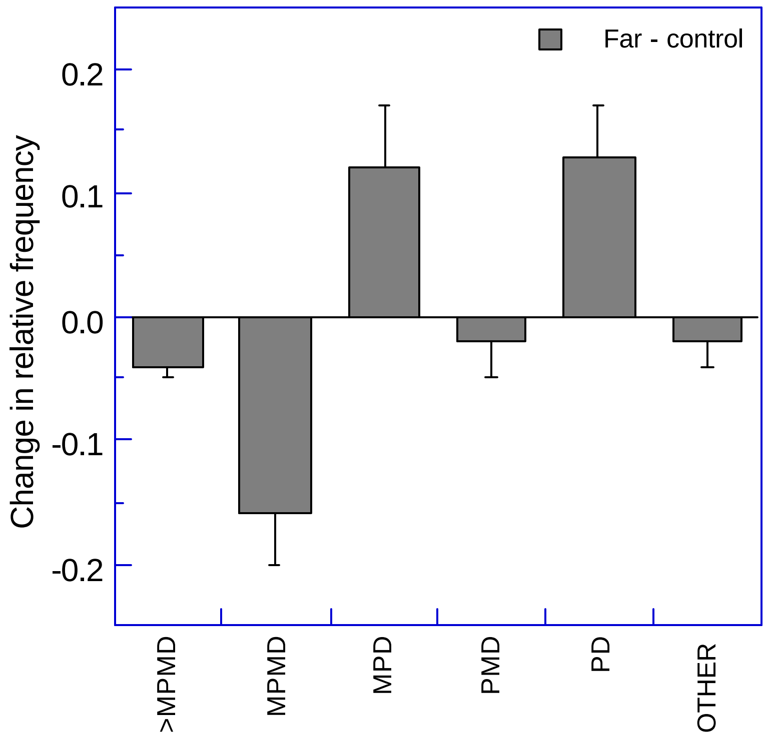


Figure 3.6 Change in relative frequency of strategies between the far and control conditions, averaged across nine subjects.

The number of multiple model fixations was reduced among eight of nine subjects. Note however that the reduction in eye movements (and the increased reliance on working memory) was not complete; none of the subjects completely eliminated MPMD strategies, nor did they copy large multi-block sub-patterns without referring back to the model. Figure 3.7 shows the average number of model references per block, averaged across subjects. The mean drops from 1.43 in the control ('near') condition to 1.17 in the 'far' condition ($P < 0.001$).

Figure 3.8 shows the individual subjects' performance in the two conditions. Eight of nine subjects made fewer model references in the 'far' condition, though even the subject who had the lowest number of looks per block in the 'far' condition (subject LD) averaged 6.4 model references per 8-block trial, providing strong evidence that the frequent model references are not artifacts of the control experiment's design.

The distribution of model references is broader in the near condition than in the far and there is large variability between subjects. When each subject's change in model references between near and far conditions is compared, the common change in task performance is more obvious.

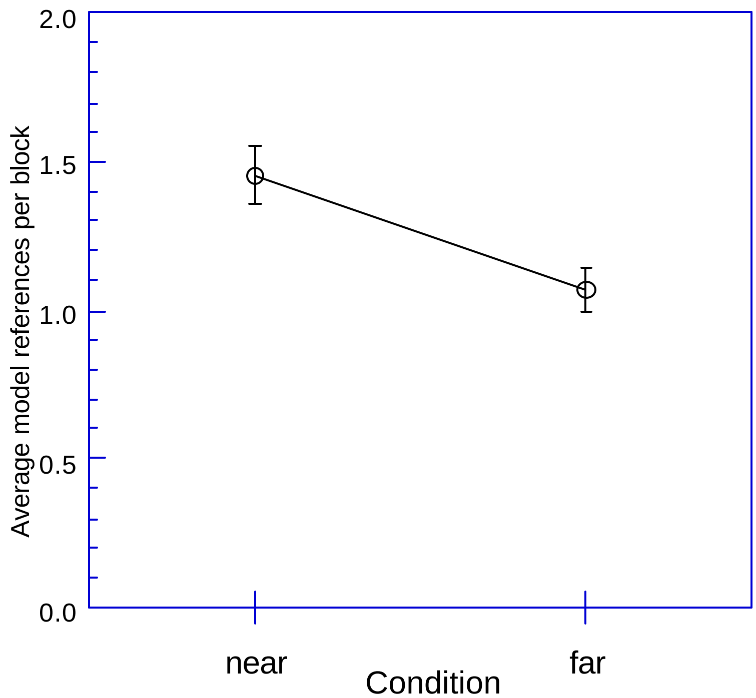


Figure 3.7 Number of model references per block moved in the near and far conditions, averaged across nine subjects.

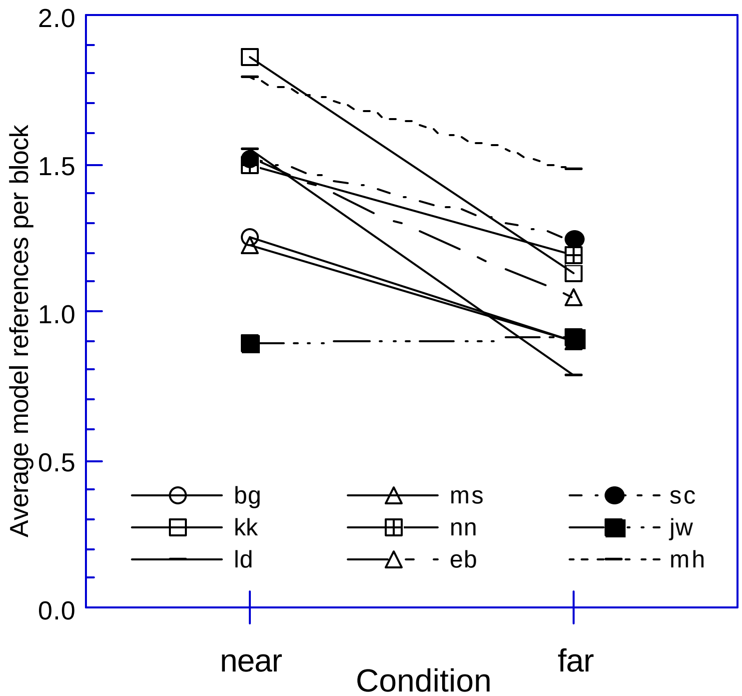


Figure 3.8 Number of model references per block moved in the near and far conditions for nine subjects.

3.2.2 Manipulating the Information Content of the Model

The experiment described above demonstrated that subjects' choice of strategy can be manipulated by adjusting the 'cost' of frequent model references. Our understanding of the model fixations as a mechanism to bind values from the fixation point into working memory suggests that reducing the amount of information needed from the model area may also reduce the frequency of model references, even in the 'near' condition. Fixations in the model area gather information about the color and position of each block. The amount of information could be reduced by 1) using a monochrome model, or 2) using a simple, predictable shape so that the relative position of each block is predetermined.

3.2.2.1 The Monochrome Condition: Eliminating Color Information

In the preliminary version of the experiment performed on the Macintosh, the model pattern and all blocks in the resource area were of a single color, eliminating the need to determine the color of each block to be copied -- the only relevant information remaining in the model pattern was the position of each block. The blocks were manipulated with a computer mouse. Blocks could be 'picked up' from the resource area by positioning the mouse cursor over a block and holding the mouse button down. The block could then be moved across the screen and 'dropped' in the workspace by releasing the mouse button. In order to reduce fine positional control requirements, blocks dropped in the workspace 'snapped' into position in a regular grid. Eye movements were recorded with an SRI Dual-Purkinje image tracker, and cursor movement was recorded throughout the task to indicate 'hand' position.

Eliminating color information from the model resulted in a decrease in model references. Figure 3.10 shows the shift in strategy averaged over four subjects in the ‘monochrome’ and control conditions, and Figure 3.10 shows the change in strategy between the two conditions. As in the ‘far’ condition (refer to Figure 3.6), low-memory strategies (MPMD+) decrease, and the frequency of the high-memory PD strategy increases. While the change in frequency of those two extreme strategies was the same in the ‘far’ and ‘monochrome’ conditions, the changes in the intermediate memory strategies (MPD and PMD) were opposite for the two manipulations. The frequency of the MPD strategy (where the subject presumably gathers both color and position in a single model reference) increased in the far condition, where the cost of model fixations is increased. The monochrome condition, on the other hand, led to more frequent use of the PMD strategy. Because the model pattern is made up of a single color, there is no need for a model fixation before a block is picked up. Figure 3.11 shows the number of model references per block in the control and monochrome conditions, averaged across four subjects. The average number of model references was reduced from 1.6 to 1.0, due largely to the decrease in MPMD strategies balanced by an increase in PD strategies, consistent with the interpretation that the frequent model references observed in the control condition are used to gather information necessary to complete the task, and not simply an artifact of the experimental design.

3.2.2.2 The Linear Condition: Reducing Position Information

The monochrome condition reduced the information content of the model by eliminating color as a variable. The informational demands can also be reduced by constraining the location of the blocks, rather than their color. In a pilot experiment,

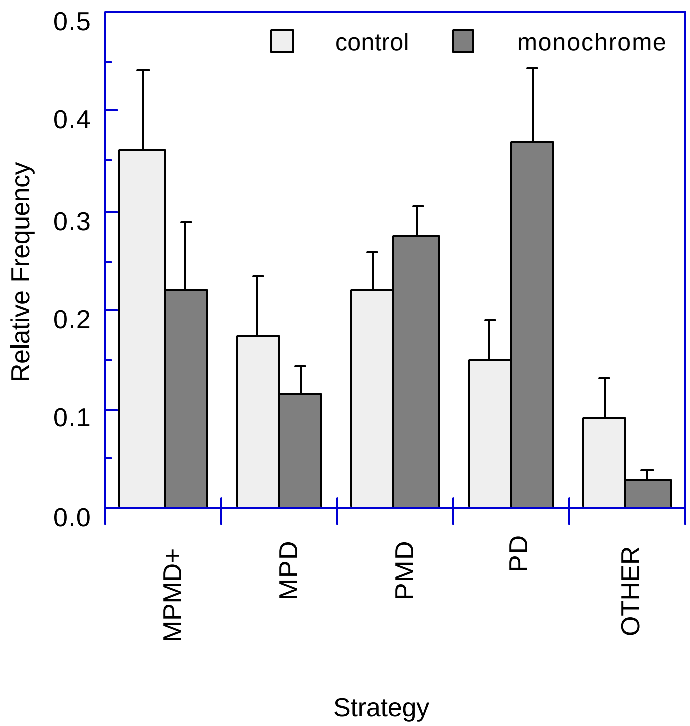


Figure 3.9 Mean relative frequency distribution for control and 'monochrome' conditions averaged across four subjects.

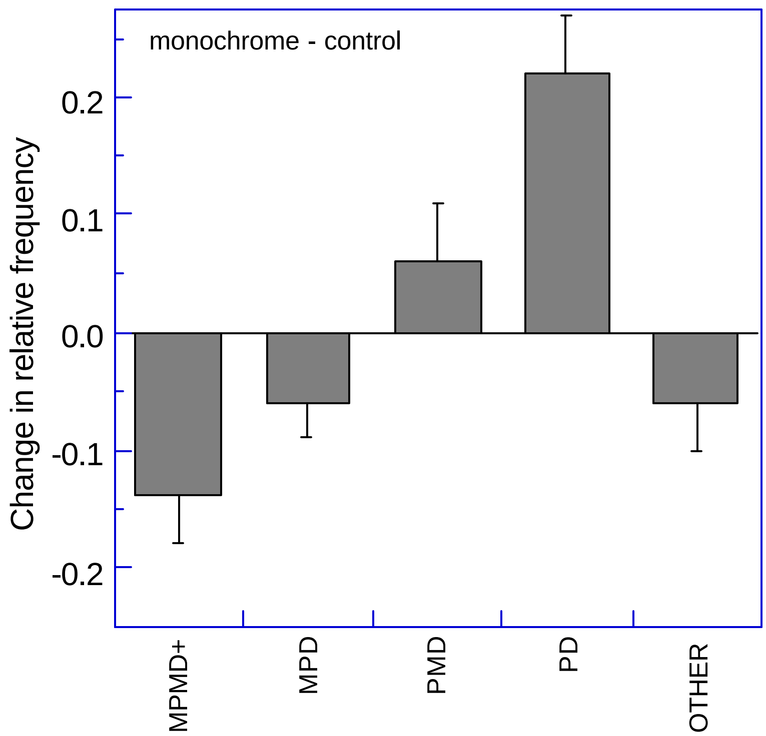


Figure 3.10 Average change in relative frequency of strategies between the monochrome and control conditions for four subjects.

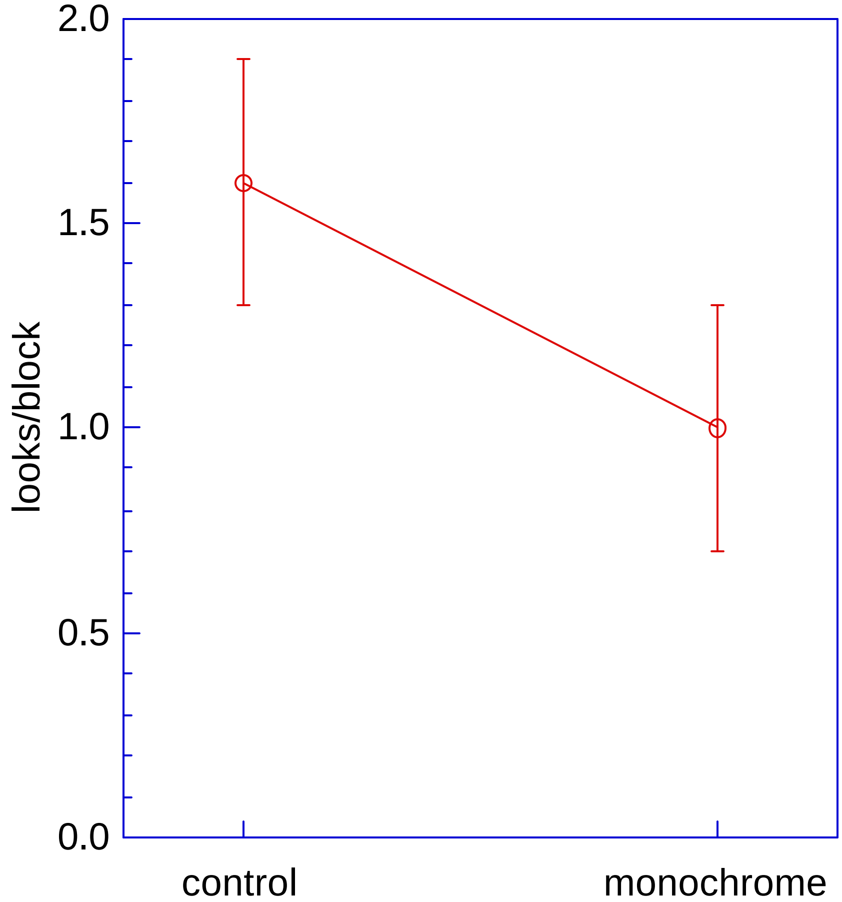


Figure 3.11 Number of model references per block in the control and monochrome conditions, averaged across four subjects

some of the colored model patterns formed a single, horizontal line, as shown in Figure 3.12. This reduced the position information needed to duplicate the pattern; after the first block was placed in the workspace, the pattern could be completed by moving consecutive blocks to the 'end' of the pattern. In the pilot experiment, the 'linear' model patterns were interspersed with the normal two-dimensional model patterns, appearing only once in each 24-trial block. Figure 3.13 shows the average relative frequency of each strategy for the linear trials and the remaining 23 trials with two-dimensional models. Figure 3.14 a) - i) shows the strategies for each of the nine subjects for the control and linear trials. The increase in the relative frequency of the PD strategy seen in Figure 3.13 approached significance ($P< 0.1$) as did the decrease in the MPD strategy ($P<0.1$).

The modest increase in the relative frequency of PD block moves in the pilot experiment (in which the linear models were interspersed infrequently with normal model configurations) suggested that subjects were changing their strategies in real-time based on the reduced spatial information content of the linear model (subjects were not told when the linear patterns would appear). Based on the results of the pilot experiment, another experiment was run in which the linear models were presented in 24-trial blocks to see whether subjects would adapt their strategies when the position information was consistently reduced. Four new subjects performed the block-copying task with control and 'linear' models, each presented in blocks of 24 trials. Figure 3.15 shows the mean across the fours subjects and the between-subjects standard error. The individual subjects' strategies are shown in Figure 3.16 a) - d). There was a larger shift toward lower memory strategies when the linear trials were presented in 24-trial blocks, but the strategy shifts are masked to some degree by the idiosyncratic performance

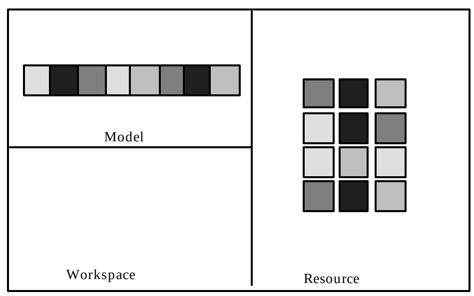


Figure 3.12 Model configuration for the 'linear' condition.

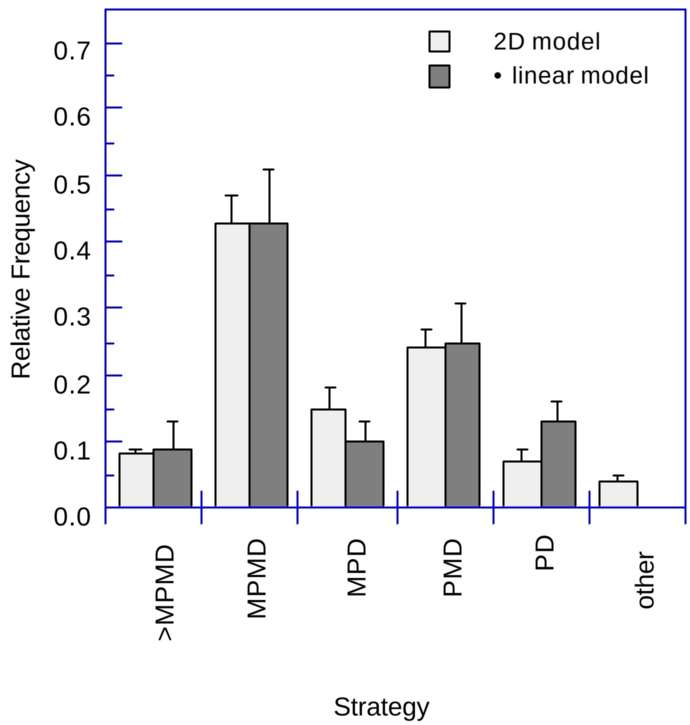
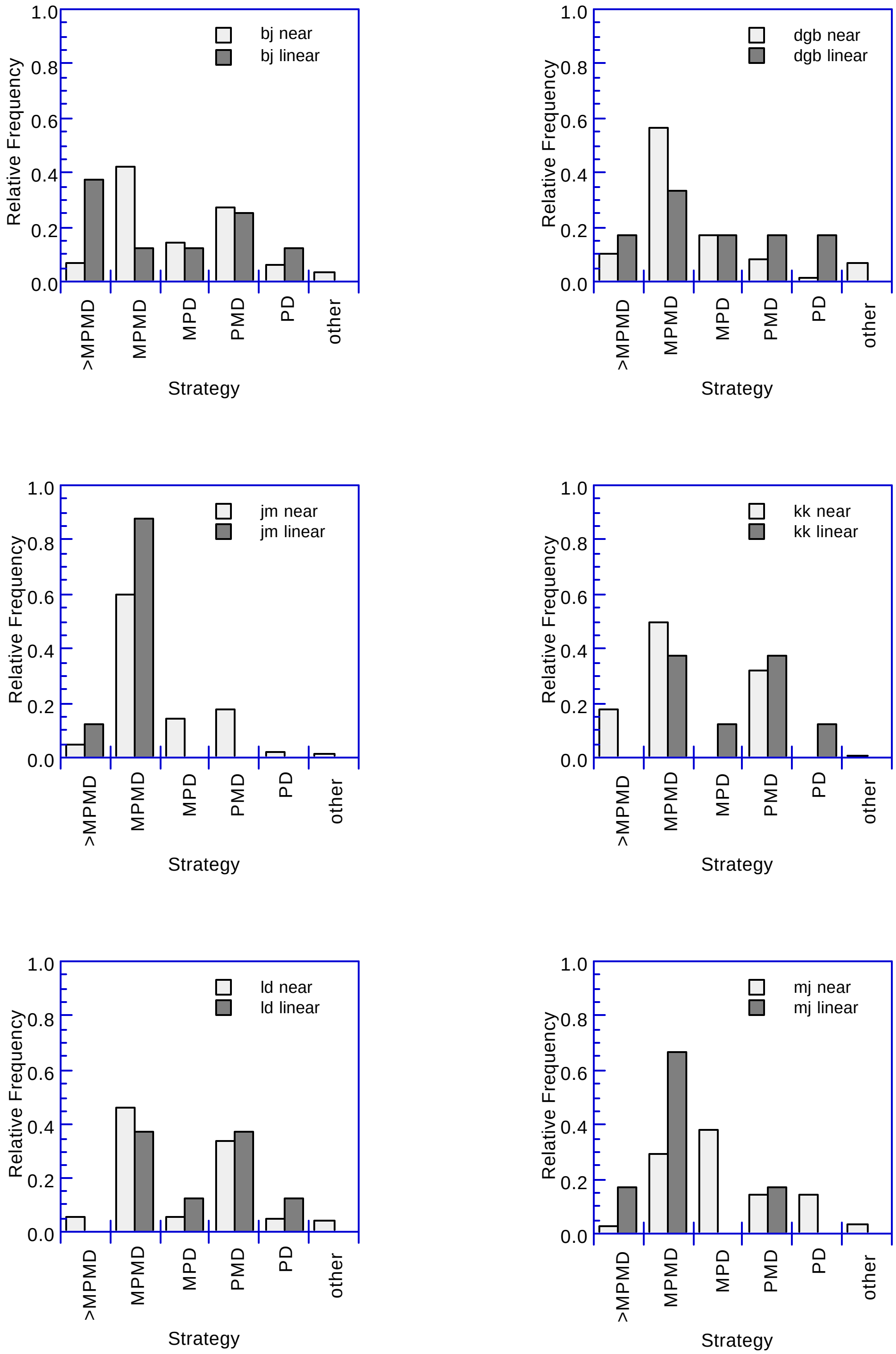


Figure 3.13 Mean relative frequency distributions for two-dimensional and linear models when linear models were interspersed with 2D patterns (one in 24 models were linear).



(see legend, next page)

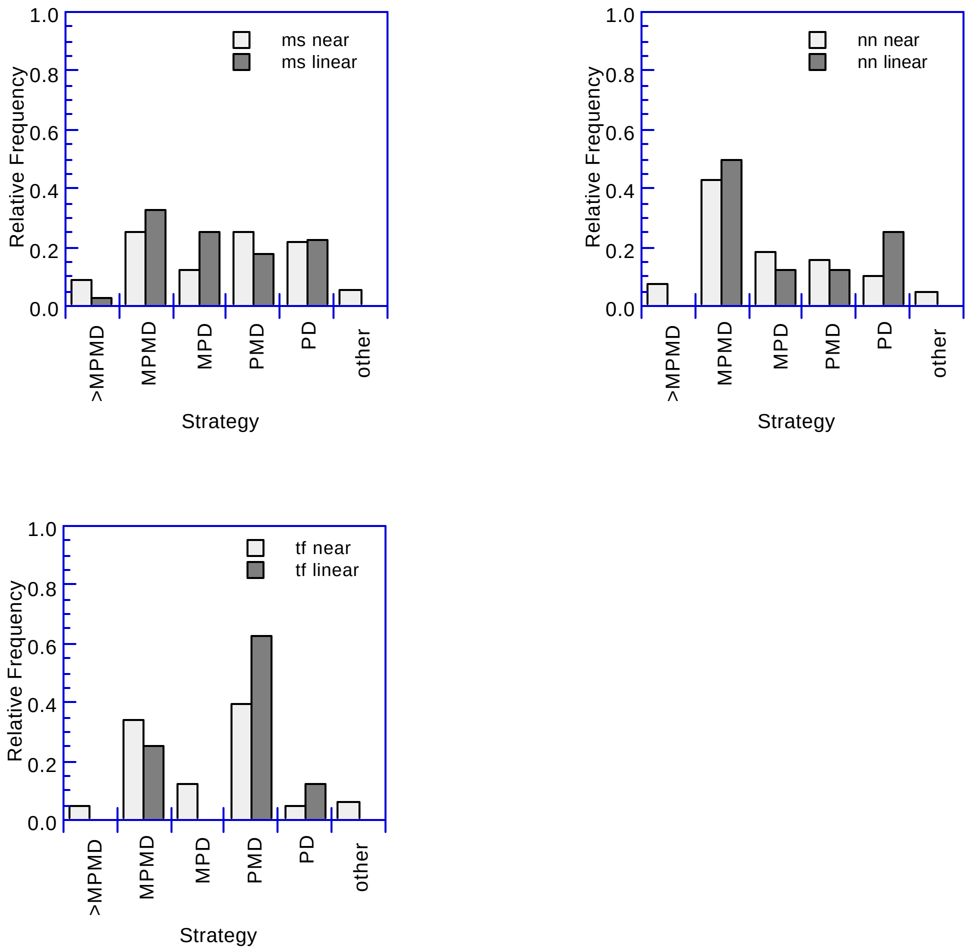


Figure 3.14 Individual relative frequency distributions for control and 'pilot-linear' conditions

of the subjects. When the results are analyzed by examining the shift in each subject's strategy between the two conditions the strategy shifts are more evident. Figure 3.17 shows the average change in strategy use (i.e., $frequency_{linear}$ - $frequency_{control}$) across the four subjects, along with the standard error between-subjects. This analysis of the linear condition showed a significant decrease in >MPMD strategies ($P<0.03$), and significant increases in PMD ($P<0.05$) and PD ($P<0.05$) strategies. The drop in the frequency of the MPMD strategy did not reach significance ($P<0.12$).

Figure 3.18 shows the mean number of model references for the four subjects in the near (control) and liner conditions. The mean number of model references per block drops from 1.43 in the control condition to 1.21 in the blocks of linear trials. Error bars show the between-subjects s.e.m.. Most of the variability between subjects is due to large mean differences between the subjects rather than variability in the change between the control and linear conditions. Figure 3.19 shows the average number of model references for each of the four subjects. A paired sample t-test shows the decrease of 0.22 model references per block to be significant ($P<0.03$).

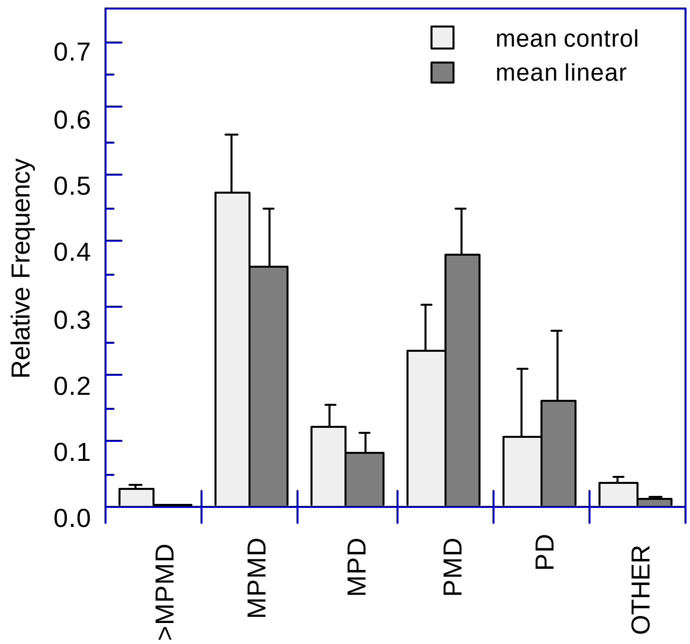


Figure 3.15 Mean relative frequency distributions in control and 'linear' conditions, averaged across four subjects.

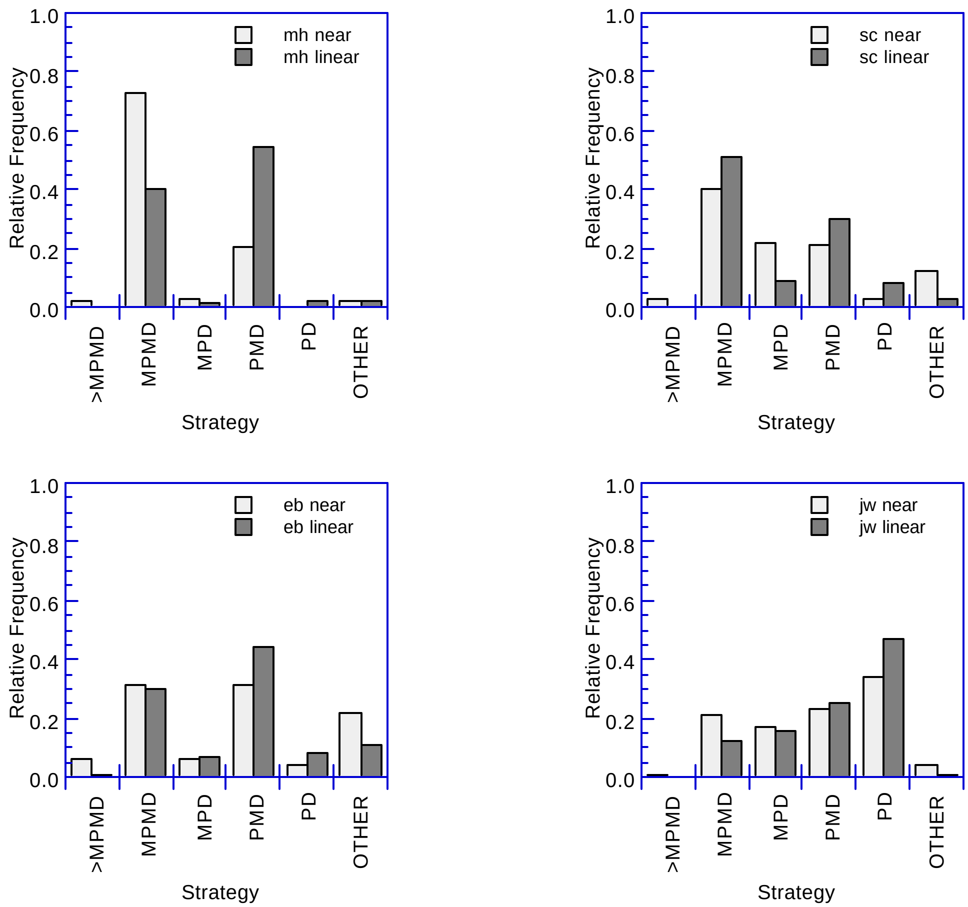


Figure 3.16 Individual relative frequency distributions for control and 'linear' conditions.

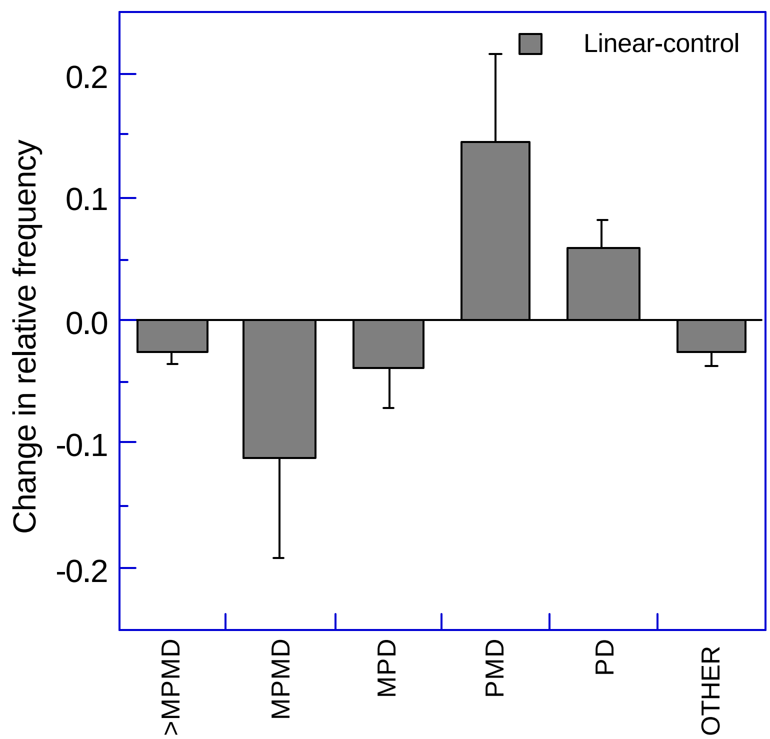


Figure 3.17 Mean change in strategies between control and linear conditions, averaged over four subjects.

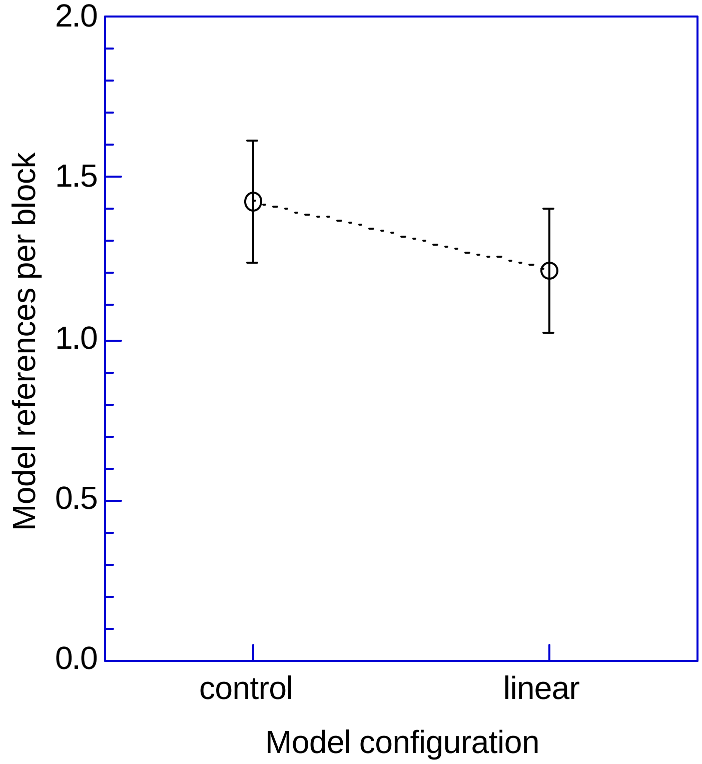


Figure 3.18 Mean number of model references in the control and liner conditions, averaged across four subjects.

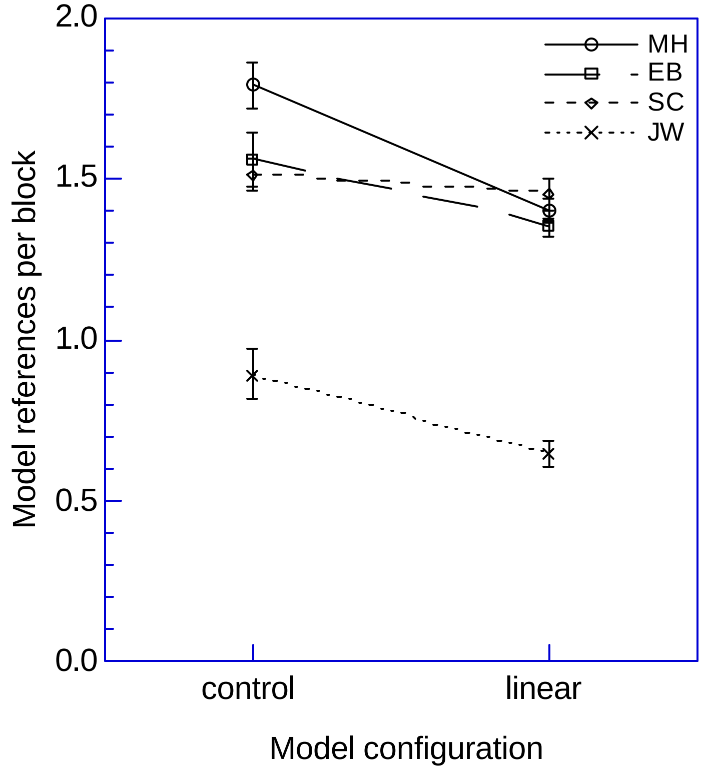


Figure 3.19 Individual number of model references in the control and liner conditions.

3.3 Discussion

When the distance between the model and workspace is increased from ~15° to ~70°, subjects had to make large eye, head, and torso movements to return gaze to the model area. Moving the model away increased the time required for all block moves, but strategies with model references required more time. The differential cost $((t_{strat} - t_{PD})/N_{model\ references})$ of eye movements to the model area was about 400 msec more than in the control condition, so there was a significant cost associated with model references in the 'far' condition. Subjects adapted their strategies to the 'far' condition, making fewer model references. The relative frequency of block moves completed with two or more model references (MPMD + >MPMD) fell significantly, balanced by increases in MPD and PD strategies, indicating that the color (and in some cases the position) of a block was held from previous model fixations. The average number of model references per block decreased by 15% in the 'far' condition. Increasing the cost of model references caused a significant reduction in eye movements to the model, but subjects still relied on repeated model references to complete the task. Although the control experiment described in Chapter 2 showed that subjects are capable of copying several blocks from memory, the mean number of model references per block did not fall below 1.0, so it is clear that when given the option of working from multi-block subpatterns held in working memory, or making frequent eye movements to the model, they choose to minimize working memory and rely instead on eye movements.

When subjects performed the block-copying task with monochrome blocks (leaving only information about a block's position relevant) subjects again shifted their strategies and made fewer model references. The largest changes were significant decreases in MPMD block moves and increases in PD moves. There was also a shift from MPD to PMD strategies. The mean number of model references per block fell from 1.6 to 1.0. The drop in frequent eye movements to the model area when color was eliminated supports the interpretation that the frequent model references in the control condition are often used to gather color and position information independently.

In the control condition, there was no significant difference between the relative frequency of MPD and PMD strategies, while in the monochrome condition the frequency of PMD sequences were more than double that of MPD sequences. In the monochrome case, where only the position of blocks in the model area is relevant, the PMD strategy is more efficient than the MPD strategy because position needs to be held for a shorter period of time before the block is placed in the workspace. This evidence that subjects are adapting their strategies based on the details of the specific task is an example of the task-dependency of visual behaviors. The eye movements are still used to serialize the task when the model blocks are all the same color, but the subtasks, and the order in which they are executed, is driven by the constraints of the task.

Linear models carry less information than do two-dimensional models because there is no positional uncertainty about the remaining blocks once the first block is placed in the workspace (presuming that the blocks are copied in order, from left to right, which was invariably the case). When linear models appeared unexpectedly and at a frequency of only 1 in 24 trials, there was a small increase in PD strategies, coming mostly at the expense of MPD block moves. When subjects knew to expect the linear model configuration, and they appeared in blocks of 24 trials, there was a more pronounced change in strategy use by the subjects. >MPMD moves vanished and the frequency of MPMD and MPD strategies fell, while the frequency of PMD and PD strategies increased. Subjects made significantly fewer fixations in the model area in the linear condition, as one would expect given the interpretation of the two model looks in the MPMD strategy as separate references to color and position. When the subjects' choice of strategies is examined in detail, however, it is clear that the interpretation is not simple. Specifically, the increase in the frequency of PMD sequences is puzzling. In the control experiment, the PMD sequence was understood to occur when the subject remembered the color of a block in the pattern from a previous fixation. After a block of that color was picked up, the block's position was determined (or confirmed) in the model reference between the pickup and the drop. But in the linear condition, once the first block is put in place in the workspace, the position of each subsequent block is determined. The relatively high frequency of the PMD strategy in the linear trials suggests that subjects are not working with the block moves as completely separate actions, but instead are 'thinking ahead' to the following block. Alternatively, it may be that it is just as easy to use visual information gathered in the model to guide the placement of the block as it is to use relative position information from the partially completed workspace.

The three experiments presented in this chapter support the interpretation of frequent model references as a strategy to reduce working memory load. They also show that the tradeoff between memory load and frequent eye movements is flexible, and that subjects can dynamically alter that tradeoff based on task demands. The difference between linear trials that occurred infrequently among 2D model patterns in the pilot experiment and those run in blocks of trials suggests that subjects adapt their strategies based on recent experience. It would be possible to investigate this interpretation of subjects' behavior further in a saccade-contingent experiment where all remaining (i.e., uncopied) blocks in the model are changed between each block move. If this is done with the linear models, it would be possible to eliminate the benefit of looking ahead, and may influence subjects' behavior. It would be instructive to see whether subjects would shift their strategies from PMD to MPD, a strategy that reduces the number of looks to the model area, but does not suffer from changes in the model area between block moves.

4. The Coordination of Eye, Head, & Hand Movements in a Natural Task

4.1 Introduction

Recent developments in eye movement monitoring systems, notably revolving magnetic field eye-coil and light, head-mounted IR reflection systems, are allowing conclusions drawn from earlier studies of eye movements to be re-examined. While many of the 'classic' results have been supported (e.g., there is no reason to suspect the pulse-step model is not valid when the head is freed), other conclusions have been called into question. Many commonly accepted metrics, such as gaze stability in the dark, peak velocity, duration, and latency of saccades appear to be different when the head is freed [Collewijn *et al*. 1992]. Five to eight week old infants' visuo-motor coordination is greatly diminished when their heads are not artificially stabilized [Jeannerod 1988], and adult subjects' ability to hold gaze stable with the head free is poorer than had been reported in earlier experiments with the head immobilized. Collewijn [1985] reported that when subjects' heads are stabilized using a biteboard, retinal slip velocities are limited to approximately 15 arc minutes per second (15'/sec). When subjects tried to minimize head movement without artificial support the slip velocities nearly doubled, increasing to 27'/sec. The retinal slip velocities rose to 97'/sec when subjects attempted to hold gaze fixed during natural head movements. Even though retinal slip velocities in this range do not reduce visual acuity significantly, these results demonstrate that oculomotor performance deteriorates by some measures when the head is freed.

Such differences are not unexpected; it seems intuitive that artificially stabilizing the head should reduce retinal slip. But not all of the changes detected when the head is freed are deficits. A striking example is the velocity of vergence eye movements; experiments performed with the head fixed have found that vergence movements are much slower than typical saccadic movements, with maximum velocities of only 20-30°/second [van der Steen 1992]. With the head free however, maximum velocities for vergence movements as high as 100°/second have been reported [Koken & Erkelens 1992]. Other types of eye movements are affected as well. In self-paced saccades between visible targets, peak velocities increased by more than 10% and durations fell, with no deficit in accuracy when the head was freed [Collewijn *et al*. 1992]. The gains were not simply the result of summing eye and head component velocities; Collewijn *et al*.'s analysis of the individual components showed that the eye-in-head component alone was greater than the peak velocity in the head-fixed condition. In fact, the peak values occurred early in the gaze change, before the head had accelerated to its maximum value. The 'profile' of the gaze changes was also different when the head was free to move. The faster acceleration at the beginning of the saccade was matched by sharper deceleration at the end of the gaze shift, resulting in a 'squarer' gaze position profile.

Experiments performed with the head fixed frequently show that subjects undershoot target location when they attempt large saccades; Carpenter [1988] reported that large saccades "almost invariably fall short of their targets," described undershoots of approximately 10% for saccades greater than 20°, and noted a 'range effect' in which the undershoots were related to the mean size of recent saccades. Becker [1991] reported a much lower frequency of overshoots for small saccades than for large ones (<50% vs. 90%). These undershoots may be an artifact of unnaturally restricting head movements. Becker [1989] reported that when the head is free to move, head movements become a regular feature of gaze shifts at approximately 20 degrees. Carpenter [1988] reported that undershoots were less pronounced in saccades to the midline than were peripheral saccades when the head is fixed. The head is of course held in a central position, and would not be expected to contribute to a gaze change toward the midline under those conditions.

Visual localization is also affected by restraining the head. In studies of pointing performance, Skavenski [1990] and Biguer *et al.* [1985] reported smaller mean errors in pointing tasks when the head was free to move naturally. Biguer *et al.* [1985] reported that when the head was fixed, errors increased with eccentricity, suggesting that the eye-in-head signal's accuracy falls off at larger angles while head position information is not degraded to the same extent. Biguer *et al*. [1982] found a correlation between head and arm movements, reporting that the head stabilized before pointing was complete, again suggesting an advantage to having the eye centered in the orbit for localization tasks.

Becker and Jürgens [1992] suggested that measurements made with the head fixed by a bite board may be ".. so unphysiological a situation as to explain in itself the lower velocities that have been observed." [p.429]. Regardless of the cause, there is ample evidence that restricting head movements affects performance, and that much of what is known about oculomotor performance may be better thought of as data regarding 'head-fixed oculomotor performance.' The simple tasks and reduced environments often used in eye movement experiments further call into question whether some of our knowledge of oculomotor performance can be meaningfully applied to 'real-world' behaviors. New instrumentation capable of studying eye and head movements without restraining the subjects' head allows the study of eye movements under natural conditions.

The coordination of the eye and head can now be studied during complex tasks as well, a capability that has shed light on the motor commands controlling eye and head movements. There is evidence suggesting that under some conditions gaze changes made up of eye and head movements result from a single command. When subjects are in the dark, and a light appears at an unpredictable position, the eye typically leads the head in an attempt to foveate the target. But this temporal dissociation of eye and head movements does not necessarily imply separate commands; EMG recordings in such cases show simultaneous innervation of the extraocular and neck muscles [Biguer *et al*. 1985]; the lag is apparently the result of the higher inertial load of the head. In the absence of further stimuli, the eye and head then perform coordinated movements that end with the eye near its primary position. There are natural behaviors where the eyes and head work together in a relatively simple, stereotyped manner. Land [1992] studied drivers' eye and head movements as they approached an intersection, looking left and right to check for traffic. Land was able to predict eye and head movements remarkably well with a very simple model, whose input consisted only of the ordered list of fixation points. Head movements were modeled with a constant duration of 400 msec, and a variable velocity (1.9°/sec per degree of gaze change). Eye movements were modeled with a constant velocity of 400°/sec and a variable duration (2.5 msec per degree of gaze change). Eye movements were modulated by a unit-gain VOR; i.e., head velocity was subtracted from eye velocity in a form of linear summation model as proposed by Robinson [1981]. Land selected the driving task because the subjects would be "too busy to exert conscious control over head or eye movements" [Land 1992, p. 318]. Another interesting finding in Land's study was the "strict synchrony" between onset of eye and head movements in his measurements. Based on this observation, Land concluded that the eye and head receive commands at the same time. But Biguer *et al*.'s [1985] EMG data suggest that Land's conclusion is inaccurate. This makes his result more interesting; because of the greater inertial load of the head, the neck muscles must receive innervation before the eyes if their movements are to begin together [Rosenbaum 1991]. In Land's driving task, subjects were planning and executing a series of gaze movements while performing a complex task, not responding to a flash of light in the dark. Rather than a central gaze command that is sent in parallel to eye and head, there appears to be a central gaze goal that, in order to be executed in "strict synchrony," requires that the head's command be initiated before the eyes'. While the tight coordination between eye and head movements reported by Land is possible, there is evidence that it is not compulsory. Kowler *et al*. [1992] studied subjects' eye and head movements while reading. They found that when the head was free to move, subjects made substantial head movements (both rotational and translational). When the subject was performing a real task with the head freed, the strict correlation between eye and head movements sometimes disappeared. Gaze changes from the end of one line to the beginning of the following line were made up of eye and head movements that were not always tightly correlated in time. One of their subjects showed several instances in which his head motion was opposite in direction from his eye movements. There were no significant differences in reading speed when the head was free. But when Kowler *et al*. had their subjects scan a page covered with characters spaced at word-length intervals (a condition which offered identical demands on the oculomotor system), freeing the head resulted in a slight increase in maximum scan rate.

In an attempt to examine the degree to which eye and head movements are correlated, Kowler and colleagues [1992] had a subject attempt to scan a square grid of targets. The subject performed the task under five different instructions; 1) scanning with the head held as still as possible, 2) scanning in a natural pattern, without regard to head movements, 3) scanning the array as fast as possible, 4) shaking the head while scanning the array of targets, and 5) moving the head and eye in opposite directions. There was a deficit in speed when the subject tried to minimize head movements, even without artificial restraints. Scan rates increased when the subject's head was free to move, and when instructed to scan as fast as possible, the amplitude of the head movements increased. When the subject attempted to scan while shaking the head, the subject tended to fall into a pattern in which eye and head movements were synchronized. Subjects reported that they found this condition very difficult. Kowler *et al*. [1992] concluded that these results ".. revealed a natural tendency to program head and eye movements concurrently in similar spatial and temporal patterns." Additionally, "We found that separate commands to the head and eyes are possible, but only with special effort and, perhaps, with some sacrifice in the precision of the visual or oculomotor performance." [p.426].

All of these issues can now be explored with new instrumentation capable of monitoring unconstrained eye, head, and hand movements. The block-copying paradigm offers an ideal environment in which to investigate these issues. These experiments have revealed a tighter linkage between fixations and actions (motor and cognitive) than had been understood until now. Every component of the task is regulated by these externally observable fixations. Gathering information about the color and position of the blocks forming the model is performed by fixating the model; hand movements to pick up blocks from the resource area and place them in the workspace are guided by fixations in the respective areas, and the gaze shifts between those are made up of coordinated eye and head movements, allowing us to study the coordination of the eye/head gaze system under natural conditions. It should be noted that not all the fixations are necessary to perform the task. Two subjects occasionally perform the pickup without fixating the resource area. This behavior is rare, occurring on only 1% to 2% of the two subjects' block moves, but in these cases the subjects had no difficulty picking up the correct block without fixating the block. So, while not all of the fixations are necessary, subjects choose to regulate the subtasks with fixations in almost every case.

4.2 Basic features of eye, head, and hand movements:

The block-copying task and the instrumentation described earlier provide a unique opportunity to measure the basic features of eye, head, and hand movements during a natural task. The task is made up of identifiable subtasks: the information gathering eye and head movements and the visually guided, coordinated actions of the eye, head, and hand. Until now, we have had very little information about such complex, natural behaviors.

A striking aspect of task performance is the regular, rhythmic pattern of eye, head, and hand movements observed while subjects perform the block-copying task. Figure 4.1 shows the horizontal components of gaze, head, and hand movements during a trial. The gaze, head, and hand intercepts with the working plane are plotted in cm. The gaze component is the horizontal point of regard on the plane reported by the ASL. The head intercept is computed by projecting a vector fixed to the subject's head to the plane. Gaze and head intercept were set to (0,0) at the beginning of each trial while the subject looked at a central fixation point. The hand intercept is defined by the horizontal and vertical components of the magnetic hand tracker's output, offset to place (0,0) at the same central fixation point. The angular position of gaze and head components varied with distance to the board, but are approximately equal to the intercept values given in cm (they are equivalent when the distance to the board is 57 cm [$\tan(1°)^{-1}$]). The rightward gaze and head movements ('up' on the plot) are made when a block in the resource area is targeted for pickup.

Figure 4.2 shows a four second section from the same trial, expanded in scale to more clearly illustrate the movements. At this scale, the task-dependent asymmetries of the movements are more evident; the velocities of the head movements are different in each direction, with higher velocities for rightward head movements towards the resource area than for the leftward movements toward the model and workspace. The hand movements also display a marked asymmetry, with longer dwell times to the left (down in the graph) for 'drops' than to the right for pickups.

While subjects almost always fixated the resource area before a pickup event and the workspace before a drop, there was a clear difference in the pattern of eye and head coordination between the pickup and drop events. Figure 4.3 a) shows a frame from the video record of a typical pickup event. As the hand nears the selected block, but before the subject grasps the block and lifts it away from the board, the gaze returns to the model area for the second model fixation or to the workspace in preparation for the drop. In contrast, subjects usually maintain fixation until the block move is completed

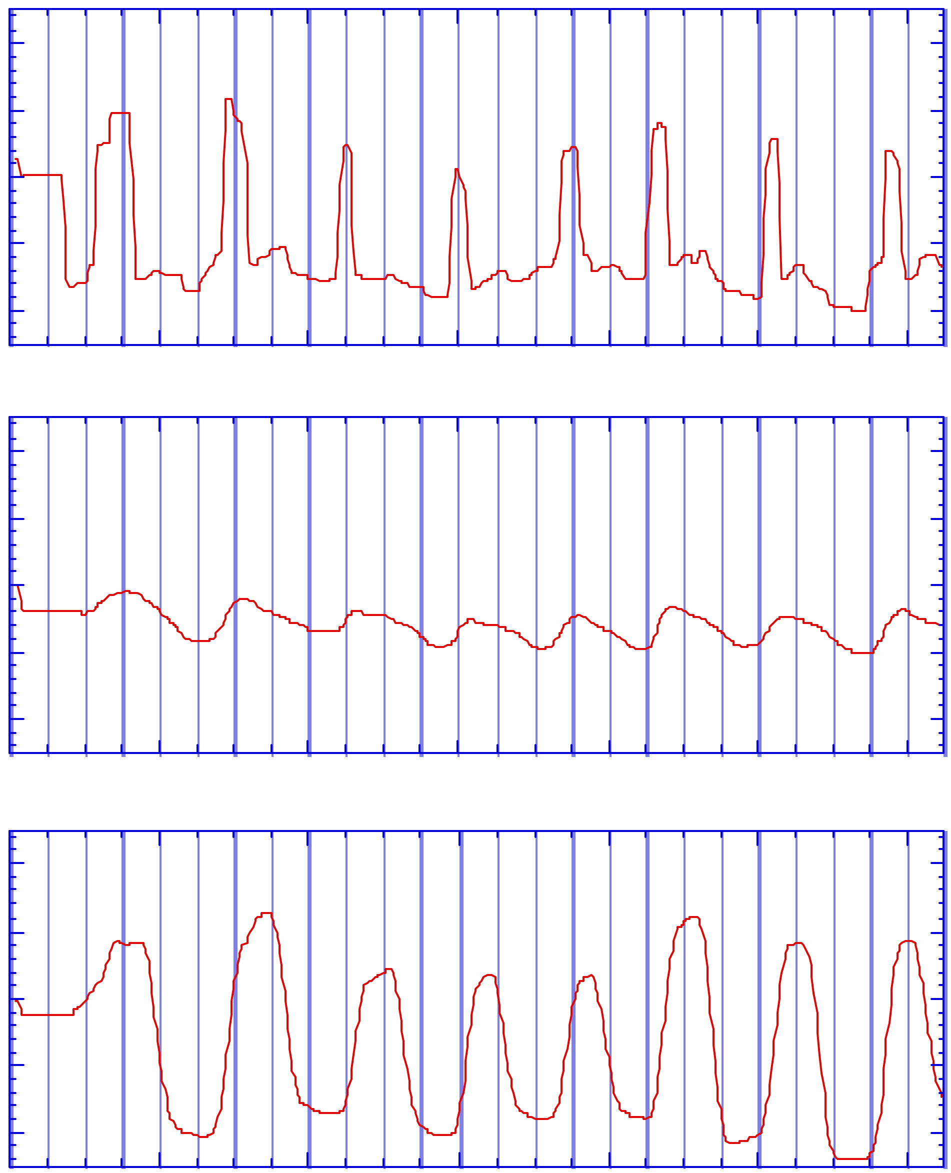


Figure 4.1 Horizontal components of gaze, head, and hand movements during the block-copying task.

when performing a drop event (see Figure 4.3 b). This task-dependent asymmetry is also evident in the longer dwell times in the workspace than in the resource area, as seen in Figure 4.2. The head intercept trace in Figure 4.2 illustrates a similar task-dependent pattern for head movements; the head is held stable longer for the putdown, and is then moved more rapidly to the right for the next pickup event (note the difference in velocities for rightward and leftward head movements).

Figure 4.4 shows the gaze and head data plotted together to make it easier to see the relative onsets of eye and head movements. It is clear from the Figure that the head movement pattern is not simply a 'low-pass' version of the gaze pattern, nor are the onset of all gaze and head movements synchronous, as Land reported in his study of drivers. In this trace, the head leads the eye by over one hundred milliseconds for rightward (up on the graph) gaze shifts to the resource area while eye and head movements leaving the resource area (towards the model or workspace) are nearly synchronous, and include instances where the eye leads the head. This task-dependent, independent programming is very different than the type of behavior reported by Land. In addition to the variation in eye/head latencies, the record shows an instance where the eye and head are moving in opposite directions (~ 6200 - 6600 msec).

4.2.1 Correlation of Eye, Head, and Hand Movements

The graphs of gaze, head, and hand movements in Figure 4.1, Figure 4.2, and Figure 4.4 demonstrate the tight coupling of these motor systems. One way to examine that coupling is to determine the degree of correlation between the different systems. A cross-correlation analysis was performed on the gaze/head, gaze/hand, and head/hand

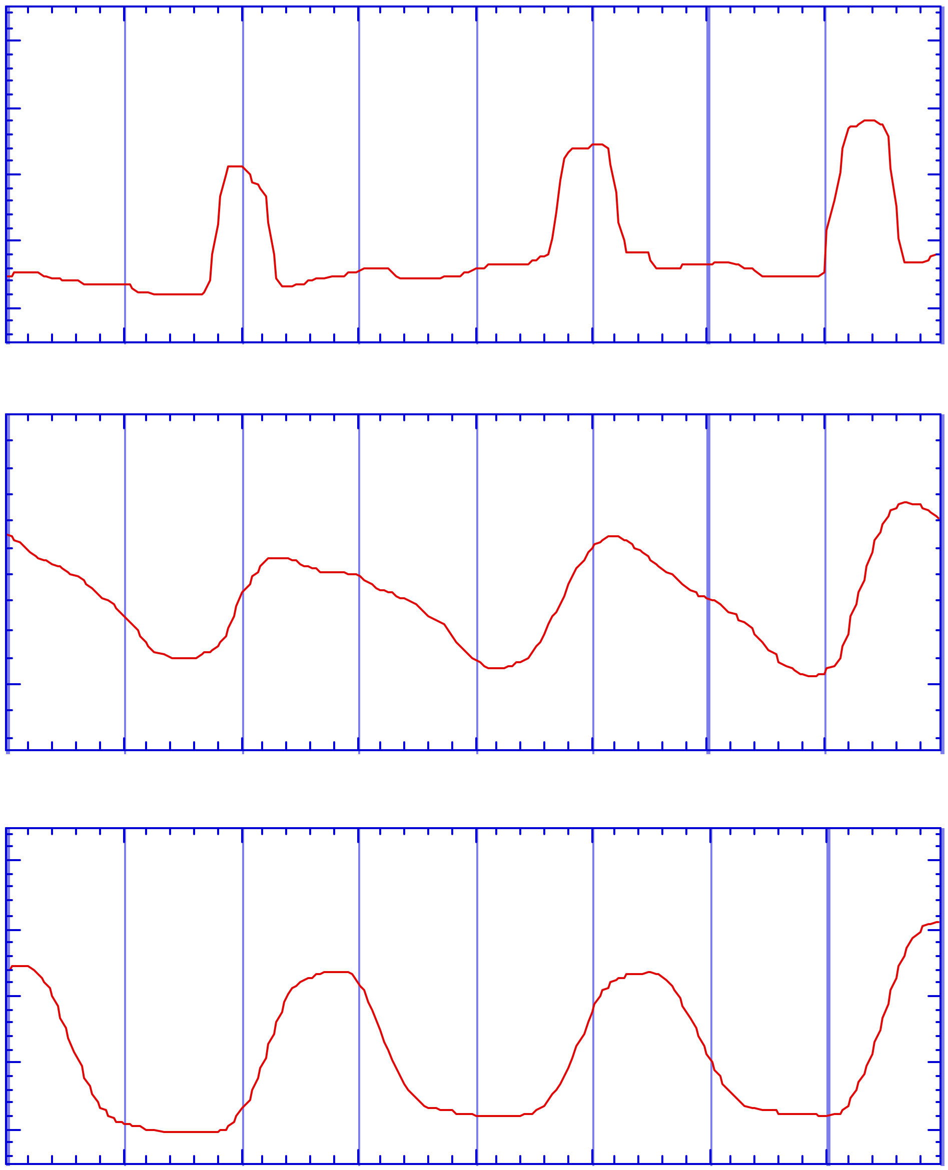
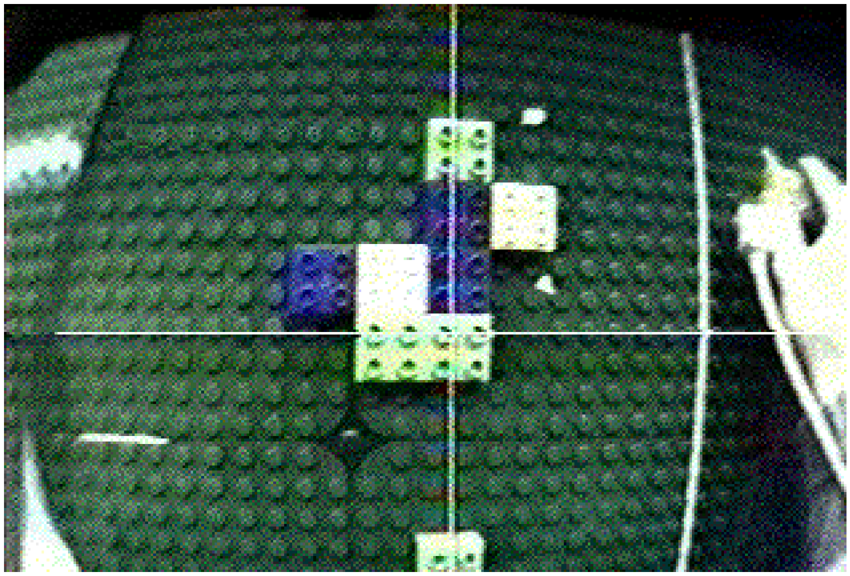
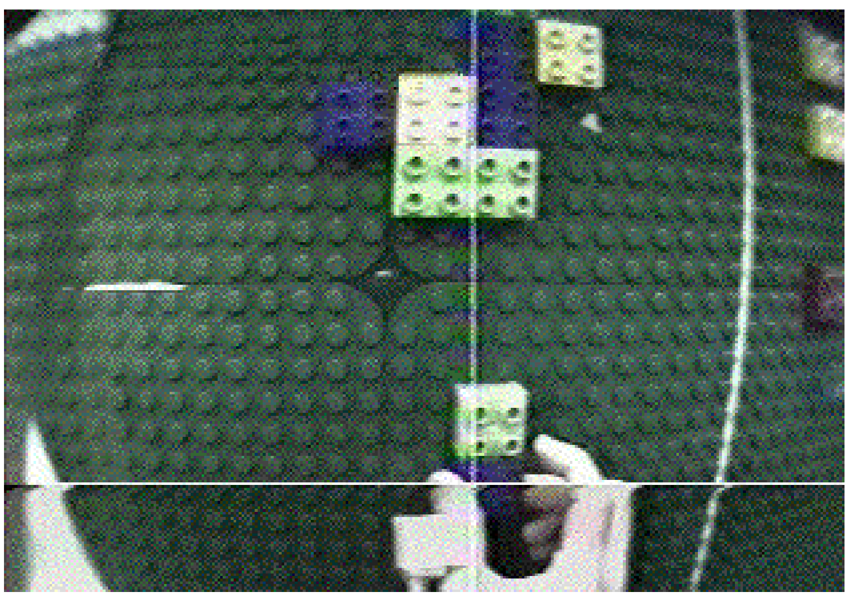


Figure 4.2 A four second section from Figure 4.1, enlarged to show more detail. (Note that the head position signal is scaled 4X.)



a) Pickup



b) Drop

Figure 4.3 Asymmetry in eye/hand coordination for block pickup a) and drop b).

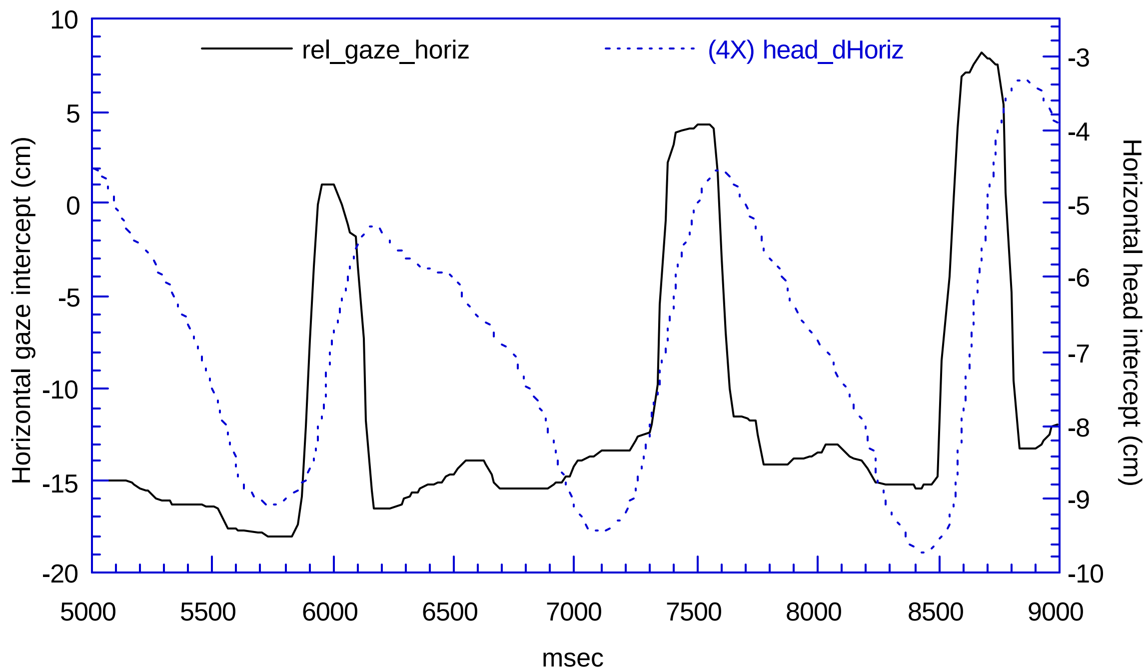


Figure 4.4 Task-dependent asymmetries in the temporal coordination of eye and head movements.

records for four subjects. Figure 4.5 shows the result of a typical analysis. The cross-correlation of the horizontal gaze and head position ("gaze/head"), gaze and hand ("gaze/hand"), and head and hand ("head/hand") signals for subject sc are shown.

In this trial, the peak correlation of horizontal gaze and head position signals was 0.69, and occurred at 324 msec, indicating that the gaze pattern leads the head pattern. It is important to note that the cross-correlation analysis provides a measure of the correlation of two waveforms, and not of individual movement onsets. As noted, the head movement toward the resource was often initiated before the associated saccade. It is evident in Figure 4.4 however, that the gaze "waveform" does indeed lead the head's; i.e., the two signals' maximum correlation occurs when the gaze is delayed by several hundred milliseconds. As would be expected from examining the gaze, head, and hand traces (e.g., Figure 4.1), the cross-correlation functions are periodic. Figure 4.6, Figure 4.7, and Figure 4.8 show representative cross-correlations for subjects jw, eb, and mh, respectively.

The peak correlation and the offset at which that peak occurred were recorded for each trial. Table 4.1 shows the mean peak correlation obtained for the four subjects (and within-subject standard error), along with the mean across the four subjects (and between-subject standard error). Table 4.2 shows the temporal offset at which the peak value for each cross-correlation occurred for the four subjects. The head and hand were most closely correlated in most trials.

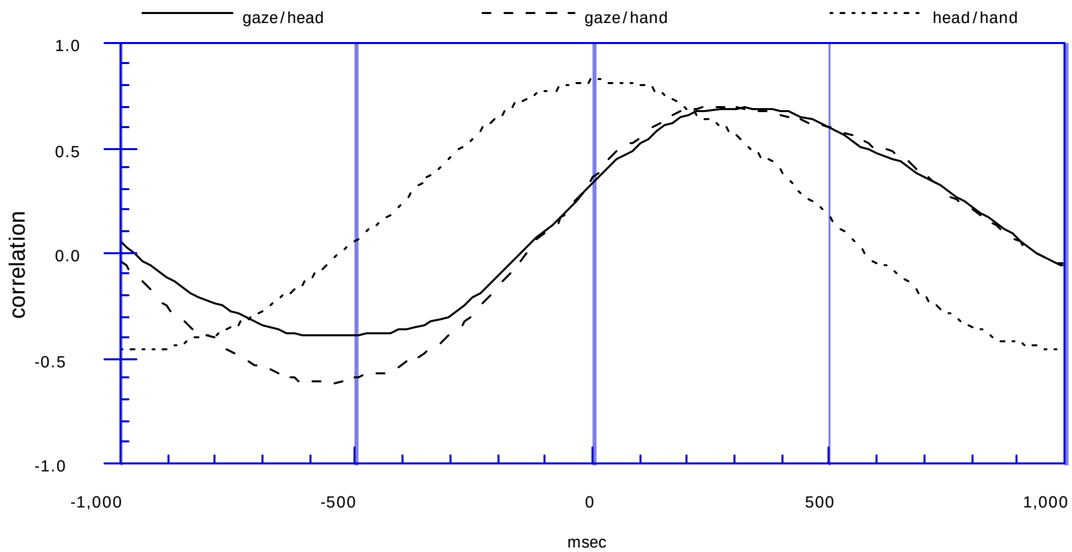


Figure 4.5 Representative cross-correlation functions for subject sc.

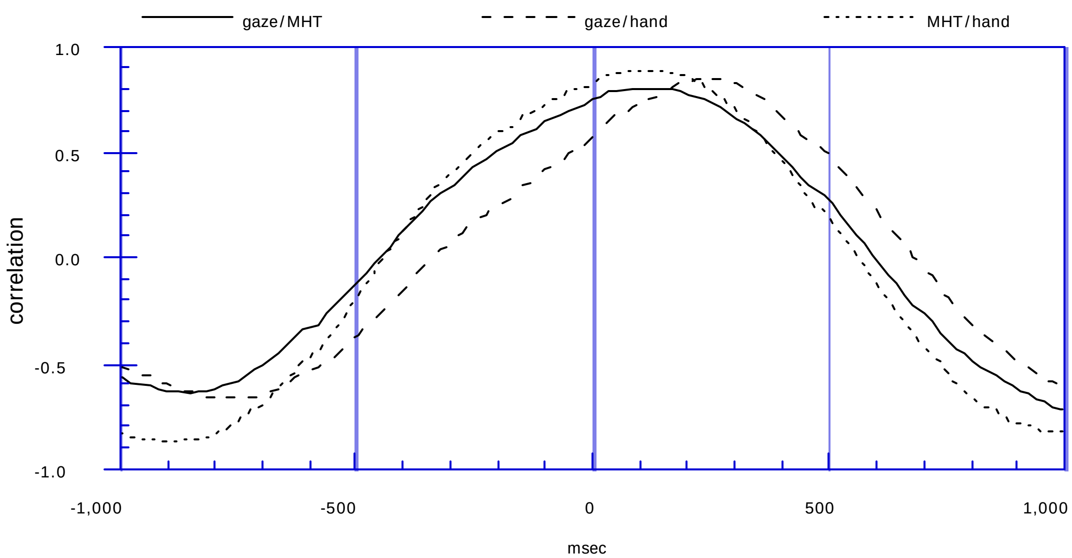


Figure 4.6 Representative cross-correlation functions for subject jw.

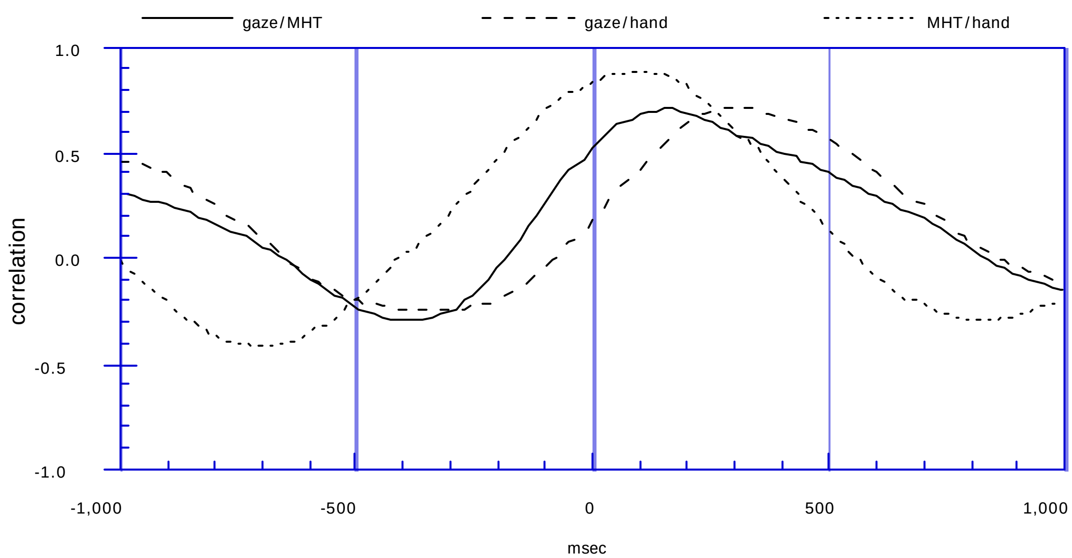


Figure 4.7 Representative cross-correlation functions for subject eb.

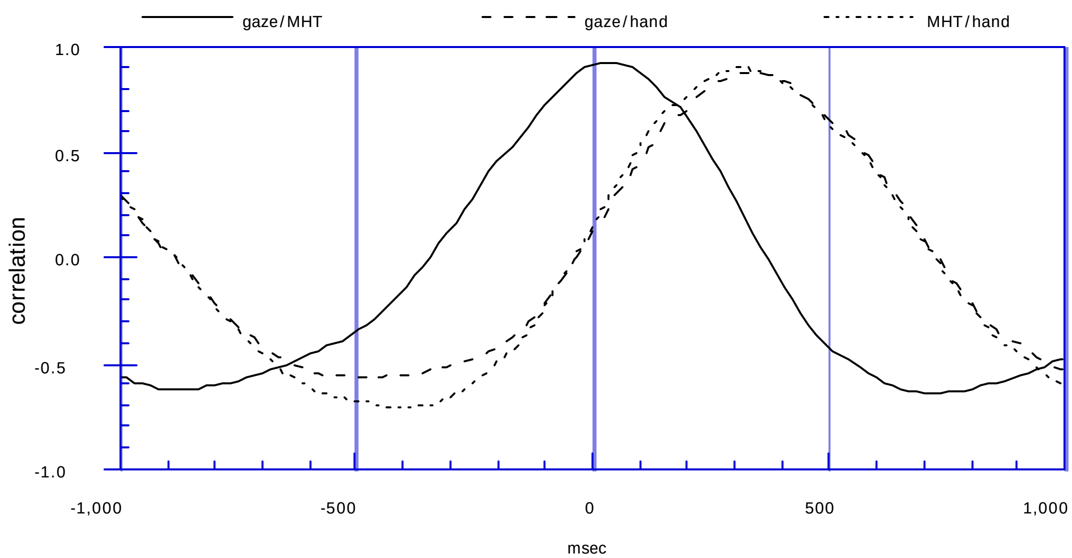


Figure 4.8 Representative cross-correlation functions for subject mh.

Table 4.1 Maximum value of the cross-correlation functions for gaze/head, gaze/hand, and head/hand.

Subject	gaze/head peak correlation	gaze/hand peak correlation	head/head peak correlation
sc	0.70 (0.01)	0.68 (0.02)	0.79 (0.02)
jw	0.69 (0.06)	0.71 (0.05)	0.73 (0.06)
eb	0.77 (0.03)	0.66 (0.01)	0.83 (0.01)
mh	0.90 (0.01)	0.80 (0.08)	0.82 (0.06)
mean	**0.76 (0.05)**	**0.71 (0.03)**	**0.79 (0.02)**

Table 4.2 Temporal offset (in msec) at which the cross-correlation reached a maximum value.

Subject	gaze/head correlation peak (msec)	gaze/hand correlation peak (msec)	head/hand correlation peak (msec)
sc	319 (16)	247 (24)	-55 (18)
jw	112 (61)	302 (81)	159 (38)
eb	136 (9)	269 (19)	111 (26)
mh	32 (5)	337 (42)	322 (54)
mean	**150 (60)**	**289 (20)**	**134 (78)**

Figure 4.9 shows the temporal positions of the peak gaze/head, gaze/hand and head/hand cross-correlations for the four subjects. The cross-correlation is sensitive to the relative timing and the shapes of the gaze, head, and hand waveforms, so the large variability between subjects may be due to timing and/or pattern differences. The mean values and between-subjects standard error are shown in Figure 4.10.

Figure 4.11 shows the peak value of the gaze/head, gaze/hand, and head/hand cross-correlation functions. Figure 4.12 shows the mean and between-subjects standard error for the peak cross-correlation values. The smaller between-subject variability is evident, with most peak correlations very close to 0.8.

4.2.2 Amplitude of Head Movements

Since the capability to monitor head-free gaze movements was developed, reports of the degree to which head movements contribute to gaze changes has varied widely, and there are few published studies reporting the dynamics of head movements measured while humans are performing complex visuo-motor tasks. There are several metrics that could be used to measure the amplitude of subjects' head movements. The simplest, the maximum range over which the head is rotated, can be misleading because a single extreme movement during a trial can increase the measure dramatically. A measure that better represents the average behavior over an extended task is the root-mean-square (RMS) variation in head orientation. Head rotation about three axes (azimuth, elevation, and roll) were recorded while subjects performed the block-copying task. The RMS measure, however, could be inflated by a change in the mean orientation of the head over the trial. This had to be taken into account because subjects typically displayed such mean orientation shifts as they progressed from the

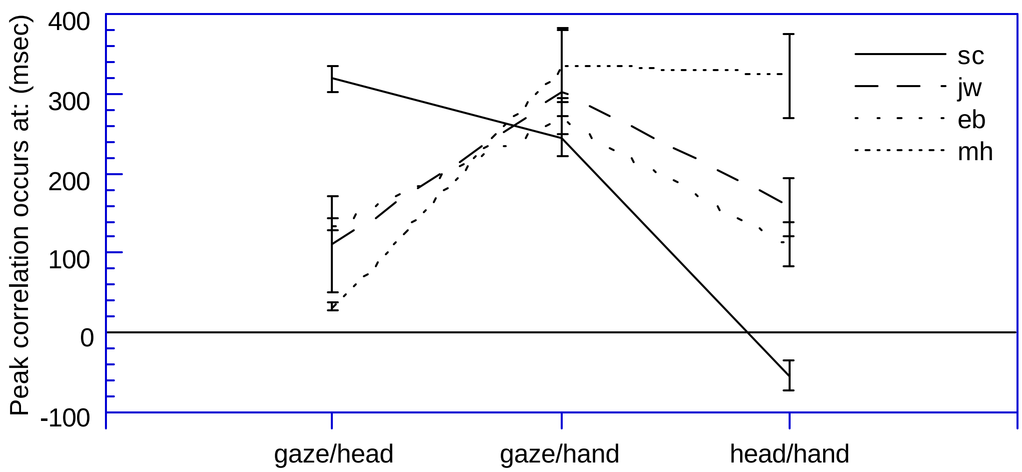


Figure 4.9 Temporal positions of the peak gaze/head, gaze/hand and head/hand cross-correlations for four subjects.

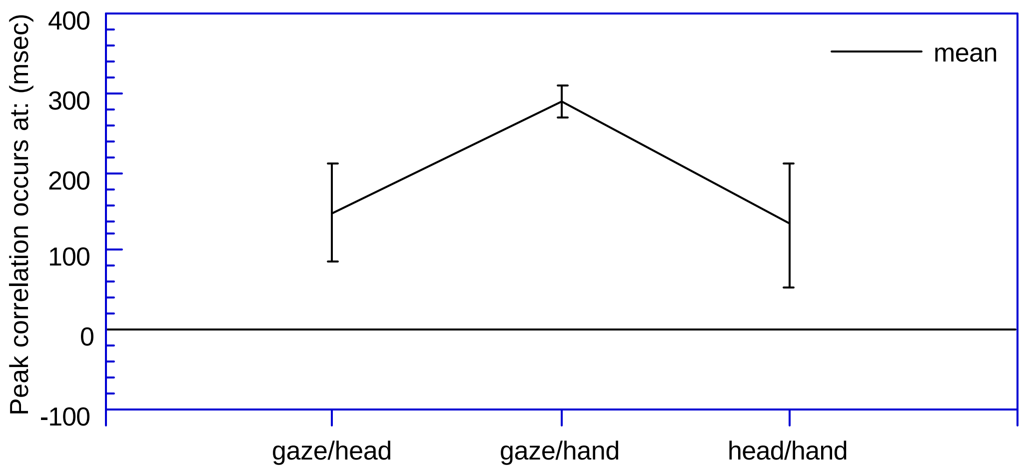


Figure 4.10 Mean of temporal positions of the peak gaze/head, gaze/hand and head/hand cross-correlations for four subjects.

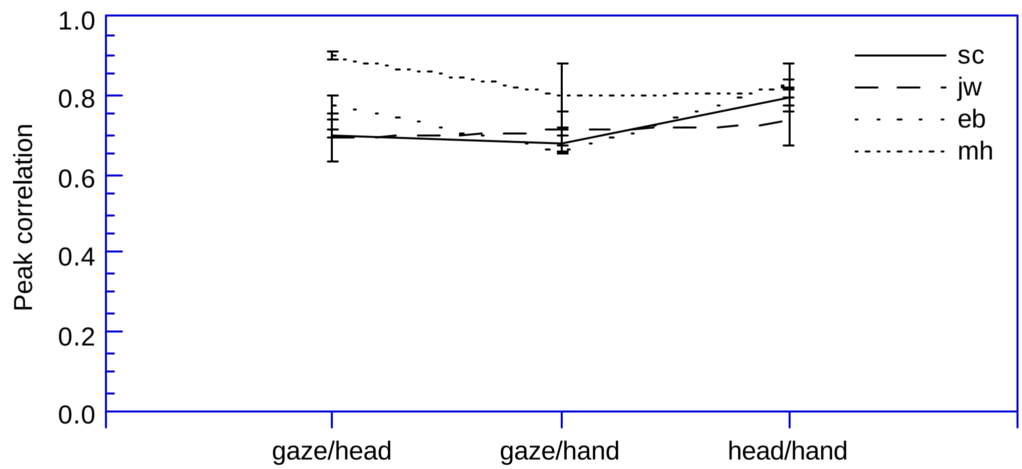


Figure 4.11 Peak values the cross-correlation functions of gaze, head, and hand for four subjects.

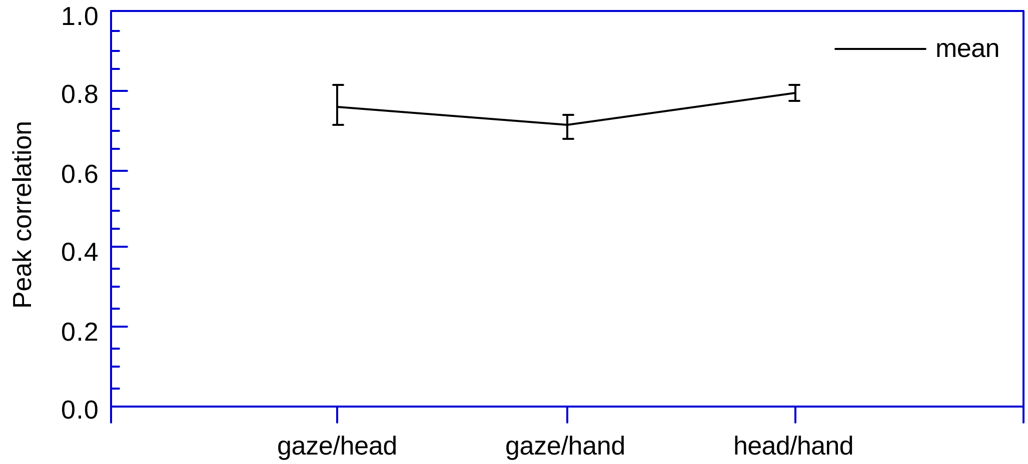
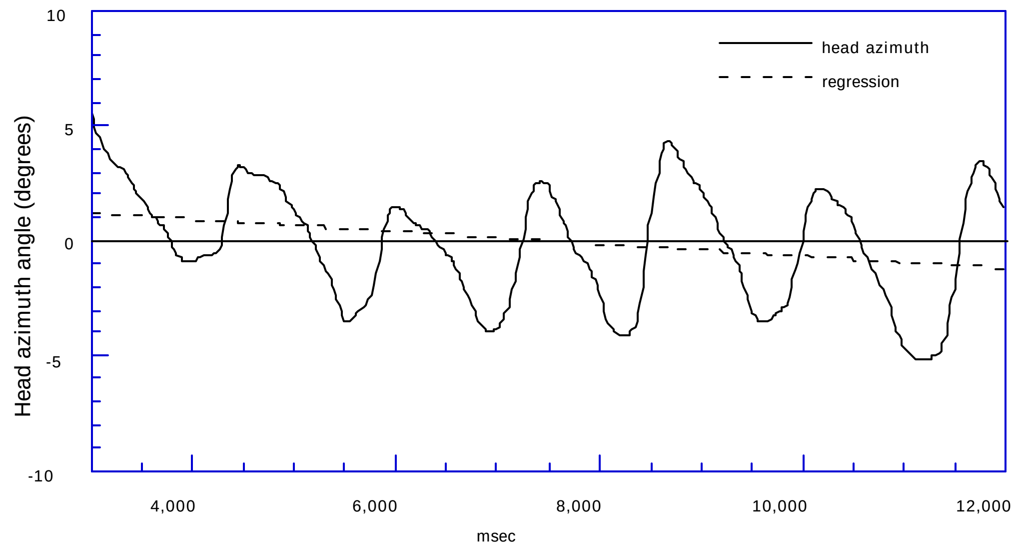


Figure 4.12 Mean peak value of the cross-correlation functions of gaze, head, and hand for four subjects

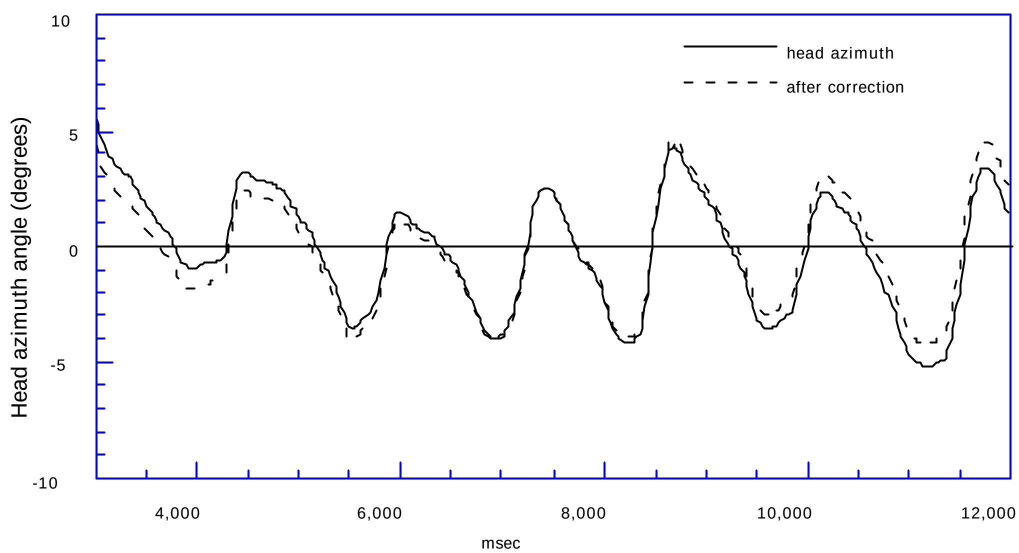
beginning of the model to the end; subjects almost always began in the upper-left corner of the model. Figure 4.13 shows the head's azimuth angle over a nine second period (the beginning and end of each trial are excluded because large head movements not representative of the ongoing task occur as the subject begins the trial and signals the end). The periodic movement is evident, as is a slow 'trend' in orientation. This low frequency component in the head's orientation signal could inflate the RMS measure, so a regression line was fit to the recorded rotation about each axis. The regression line (shown in Figure 4.13 a) was subtracted from the raw orientation record before analysis. Figure 4.13 b) shows raw and corrected azimuth records. In this case the uncorrected measure overestimated the actual RMS value by approximately 5% (2.59 vs. 2.48 degrees). Figure 4.14 shows the head's elevation over the same period, along with the regression line fit to the data. The trend (downward in this case) is of approximately the same magnitude as the trend in azimuth (<2.5° over 9 seconds), but the superimposed movement is much smaller (note that the vertical scale is different than in Figure 4.13). In this case, failing to correct for the shift in mean orientation would lead to a 25% overestimate of RMS amplitude (1.22 vs. 0.97 degrees). Figure 4.14 shows the raw and corrected elevation data.

Figure 4.15 shows mean azimuth, elevation, and roll RMS values for three subjects. As expected due to the configuration of the model, resource, and workspace areas on the working plane, the largest movements were in azimuth. The widest variation between subjects was found in the roll angle. Figure 4.16 shows the average RMS amplitude across the subjects, and the between-subjects s.e.m. Note that the RMS values are approximately

25% of the range for a normal distribution, and as seen in Figure 4.13, the range of horizontal head movements was ~10°, or about 2/3 the size of the gaze shifts.

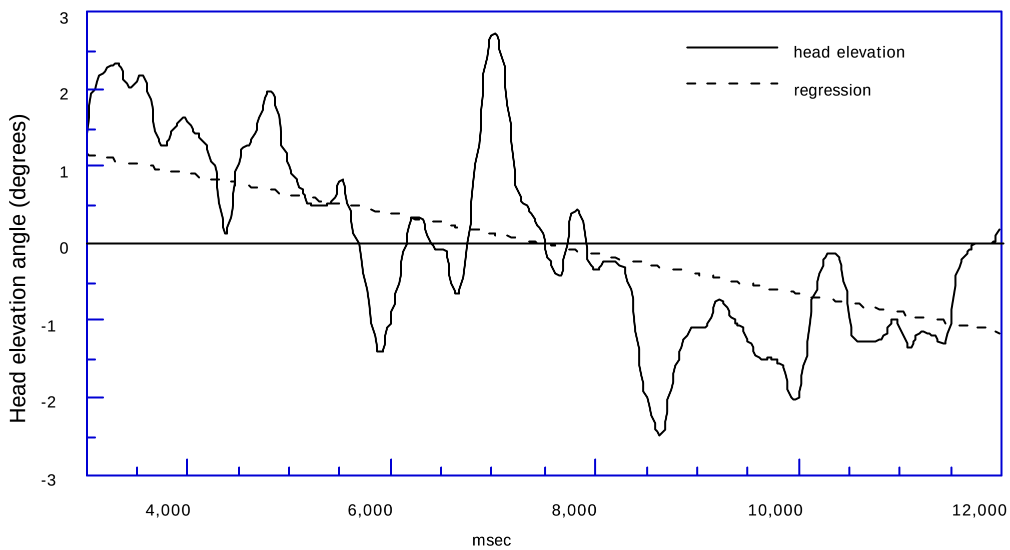


a)

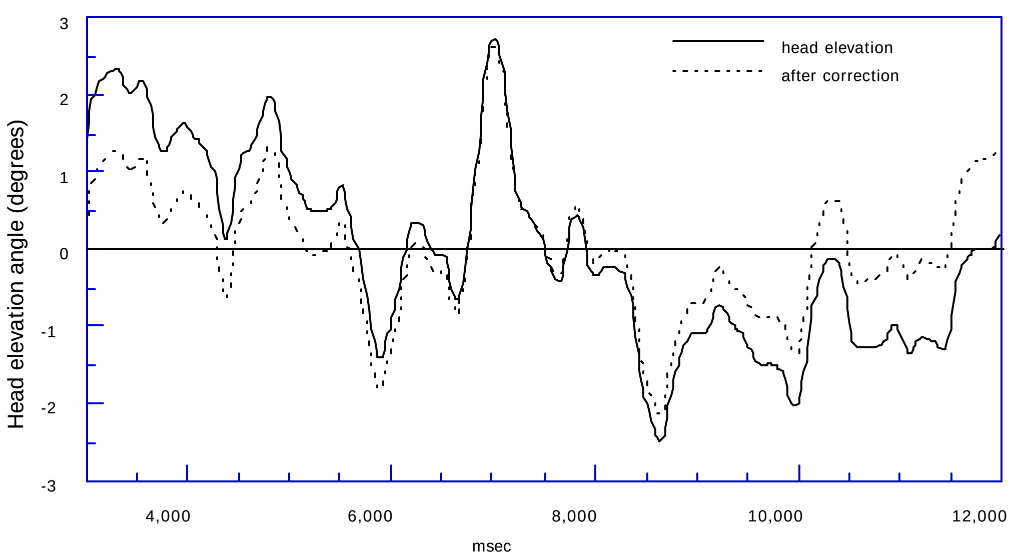


b)

Figure 4.13 a) Head orientation (azimuth) as a function of time over nine seconds with linear regression. b) Raw and corrected head azimuth records



a)



b)

Figure 4.14 a) Head elevation as a function of time over the same period shown in Figure 4.13. b) Raw and corrected head elevation records.

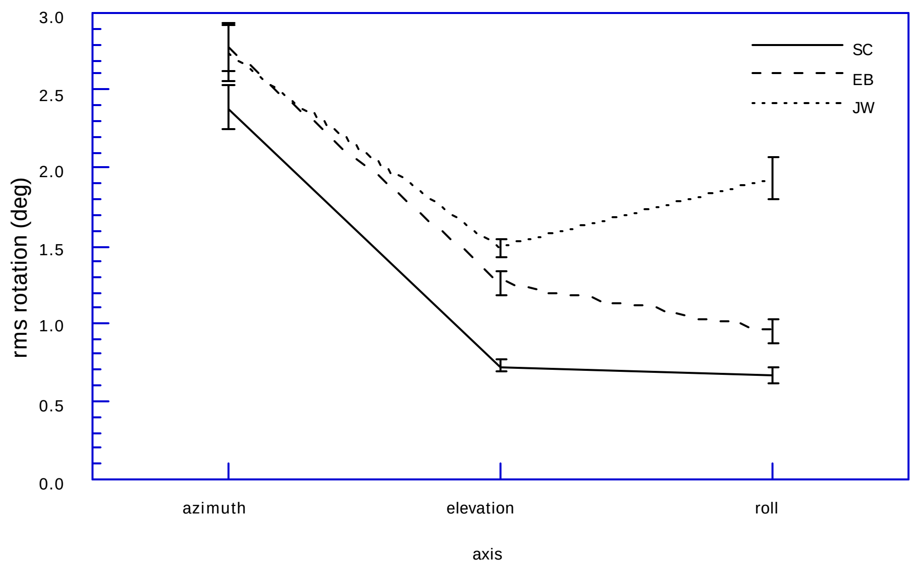


Figure 4.15 Mean azimuth, elevation, and roll RMS amplitudes for three subjects.

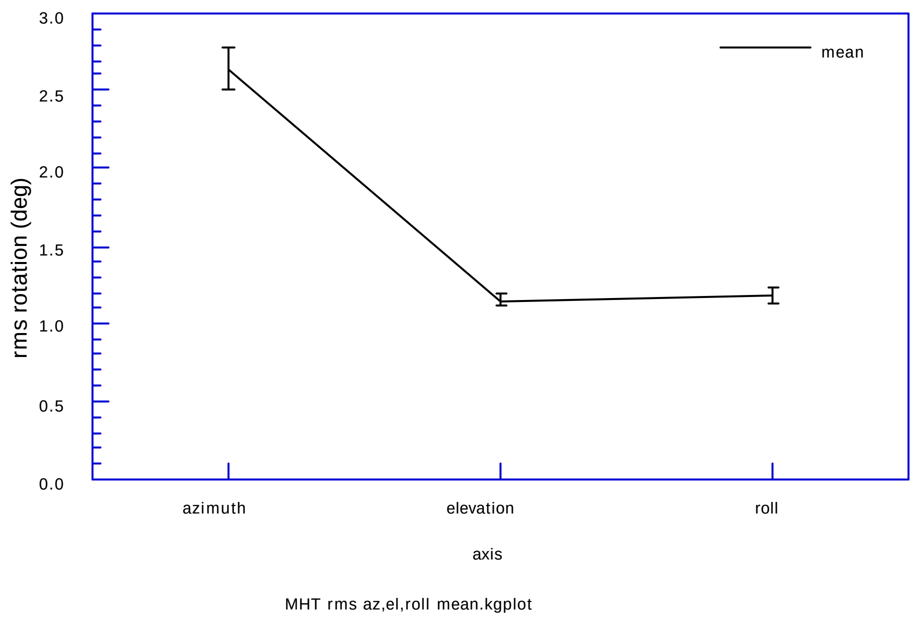


Figure 4.16 Azimuth, elevation, and roll RMS amplitudes averaged across three subjects.

4.2.3 The Coordination of Eye and Head Movements in Gaze Changes

The coordination of eye and head movements during complex visuo-motor tasks is not well understood because previous work either fixed the head or was limited to simple tasks not representative of natural behaviors. What little is known about the programming of eye and head movements suggests that a single gaze change command may be responsible for both eye and head movements under most conditions [Kowler *et al*. 1992, Land 1992]. The block-copying task provides a new paradigm in which to examine the programming of concurrent eye and head movements. In the series of fixations making up a PMD sequence (alone or as part of an MPMD sequence), the gaze moves from the resource area to the model, then on to the workspace. The resource -> model gaze change is primarily horizontal, and the model -> workspace gaze change is primarily vertical. Because of the frequency of MPMD and PMD sequences, it is useful to examine gaze and head trajectories during the PMD sequence. Four subjects' gaze and head movement records were examined. The sections of each trial with a PMD sequence were isolated, and the concurrent head movements were analyzed to determine to what degree the head movements paralleled the gaze changes as would be expected from reports of tight linkage between eye and head. Significant between-subject variability was observed in the degree to which eye and head movements were linked. Figure 4.17 a) shows a two-dimensional plot of gaze and head intersections on the working plane during a PMD sequence. Gaze and head traces are shown over an interval starting when the block is picked up in the resource area, and ending when the gaze arrives in the workspace. Figure 4.17 a) shows an example of the tight coupling of eye and head movements typical of that reported in the literature. Eye and head movements toward the model area are initiated at time t1 after a block is picked up in the resource area, and then toward the workspace at t2. Some subjects completing the extended block-copying task showed very different patterns of eye and head movements, programming eye and head movements to separate targets. For example, after picking up a block in the resource area, subjects sometimes moved their gaze to the model area while moving the head directly to the workspace preparing for the putdown. Figure 4.17 b) shows gaze and head traces during such a PMD sequence. At time t1, a leftward gaze change to the model area is initiated. At the same time, the head begins a single, diagonal movement toward the workspace in anticipation of the putdown in the workspace. At time t2, while the head is still completing its movement toward the workspace, gaze moves vertically to the workspace.

These examples represent extreme cases. Subjects showed evidence of such eye/head dissociation with varying frequency and to varying degrees. Three PMD sequences for subject eb are shown in Figure 4.18 through Figure 4.20. They are samples selected to illustrate the range of eye/head coordination observed in subjects performing the block-copying task. Figure 4.18 shows eb executing a PMD sequence in which gaze and head are apparently executing coupled eye/head movements with common spatial and temporal patterns. Figure 4.18 a) shows a two-dimensional plot of the gaze intercept as a block is picked up in the resource area, gaze returns to the model, then moves to the workspace to guide the putdown. Figure 4.18 b) shows a two-dimensional plot of head orientation over the same period, at an expanded scale to show the head movements more clearly.

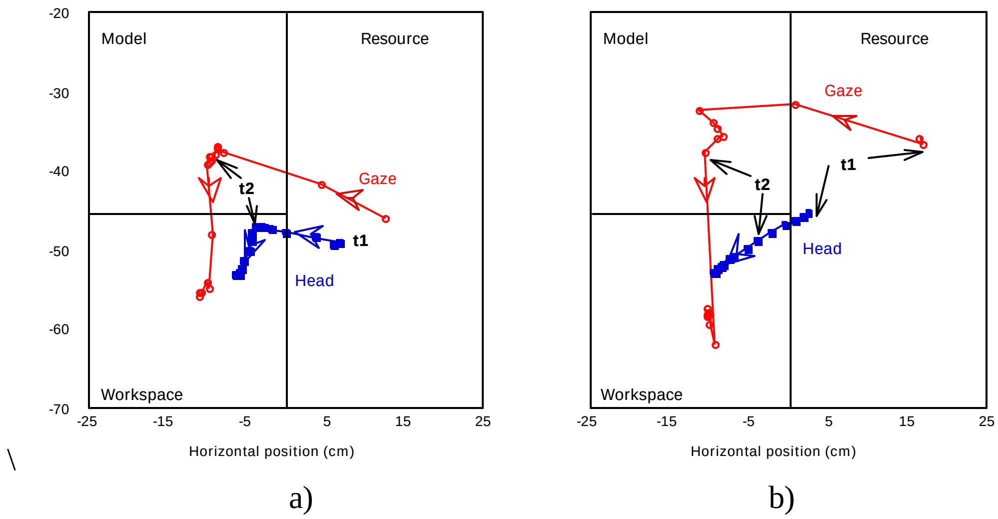


Figure 4.17 Two-dimensional gaze and head records showing a) tight coupling of eye & head, and b) dissociated eye and head movements.

The common goal of eye and head for each of the two gaze changes is evident in the plot of head position. Figure 4.18 c) shows the same head orientation data plotted as a function of time. It is evident in this plot that the two head movements are executed independently. The horizontal component of the head's motion is completed before the vertical component is initiated. Figure 4.19 shows another PMD sequence from subject eb, from a near-control trial performed on the same day as that shown in Figure 4.18. It is obvious from the two-dimensional plots of gaze (a) and head orientation (b) that the eye and head are not executing movements that were programmed with the same goals. The eyes are executing a sequential program, moving first to the model, then to the workspace, while the head is executing a single, diagonal movement toward the workspace in preparation for guiding the placement of the block, as was seen in Figure 4.17 b). Figure 4.19 c) shows the temporal coincidence of the horizontal and vertical head movements. Figure 4.20 shows another PMD sequence for subject eb. This case can be considered intermediate between the two discussed above. In this case the head follows a curved trajectory from the resource to the workspace. The horizontal head movement is initiated first, but the vertical component begins before the horizontal movement is complete, causing a curved trajectory.

Analysis of the four subjects' PMD sequences yielded the distributions shown in Table 4.3. Head movements were labeled as "Separate H & V" for block moves like that shown in Figure 4.18, "Diagonal" for cases like that shown in Figure 4.19, and "Curved" for cases like that shown in Figure 4.20. Block moves in which the vertical component of head motion was too small to meaningfully label the head movement

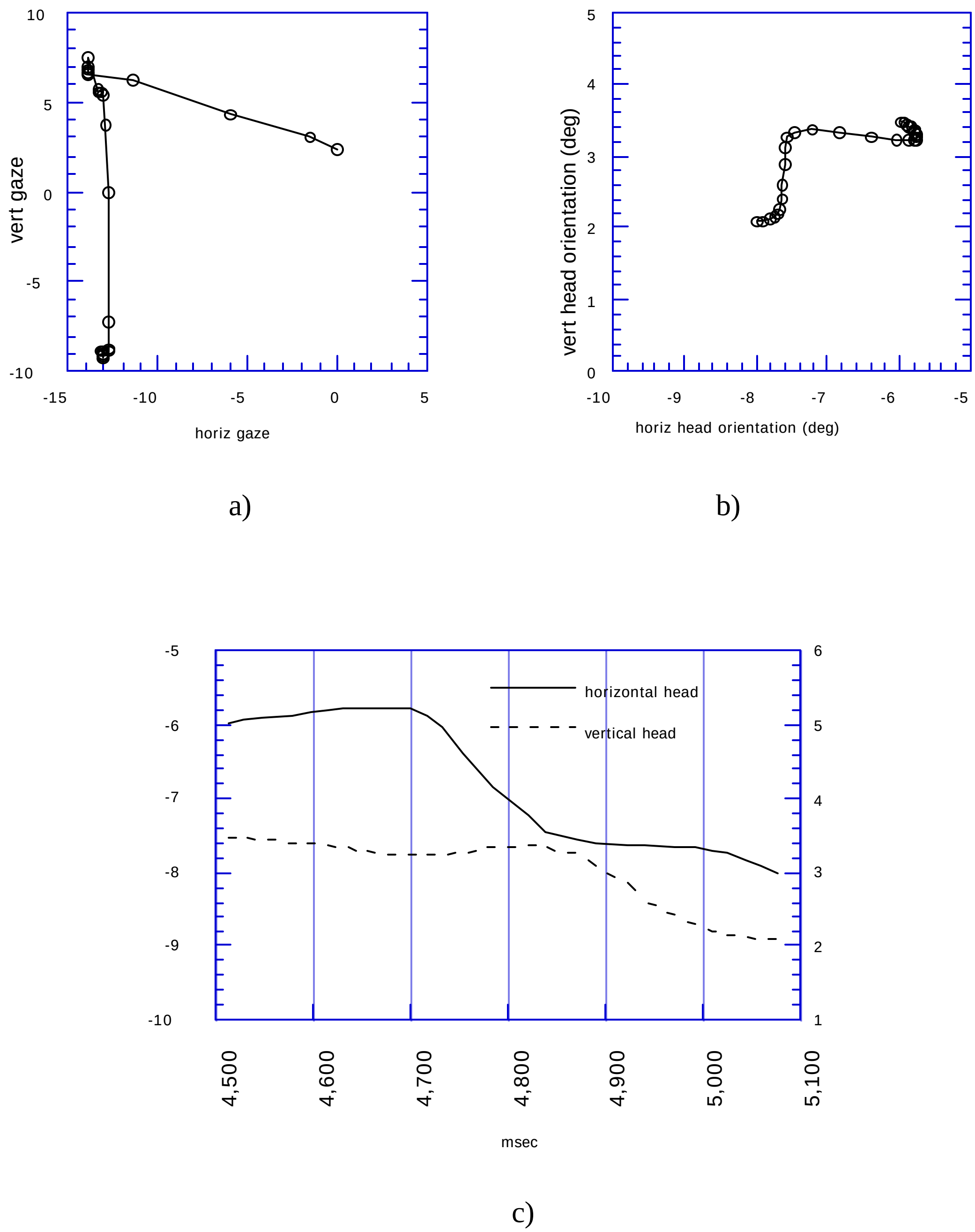
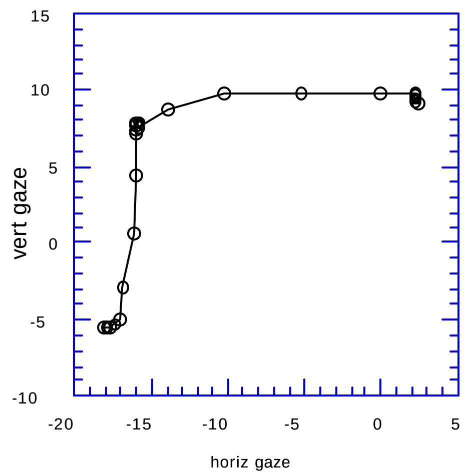
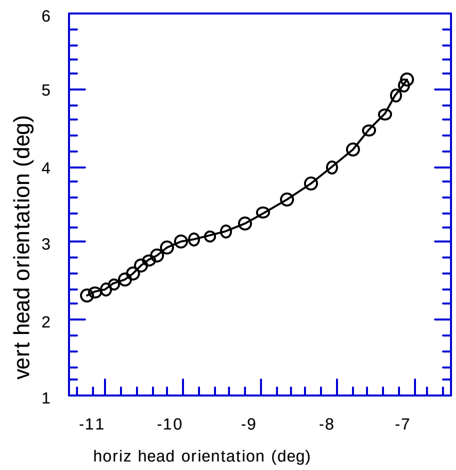


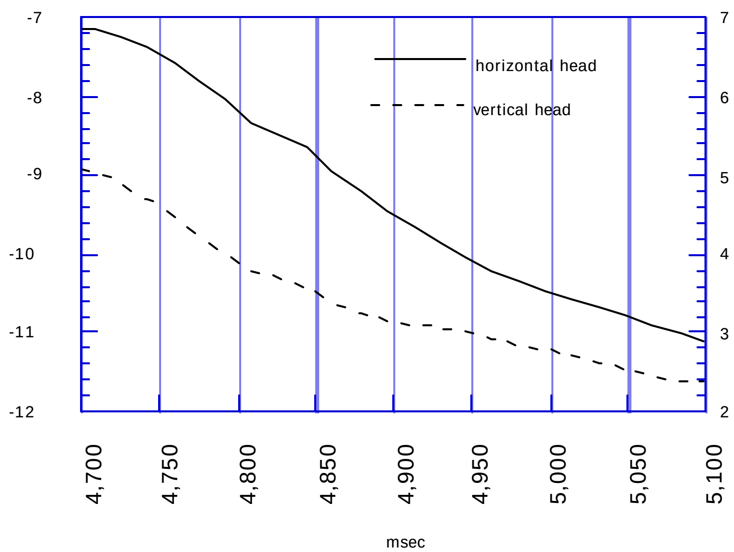
Figure 4.18 Example of common commands to eye and head. The vertical component of head motion is initiated after the horizontal component is complete. a) Gaze

intercept, b) horizontal and vertical head orientation, and c) horizontal and vertical head orientation vs. time.



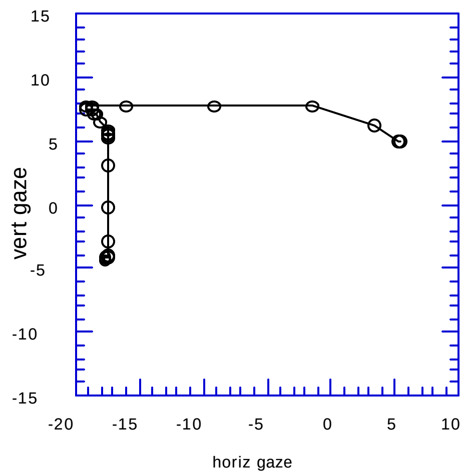


a) b)

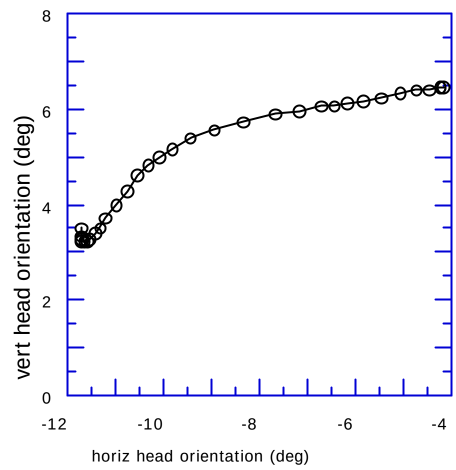


c)

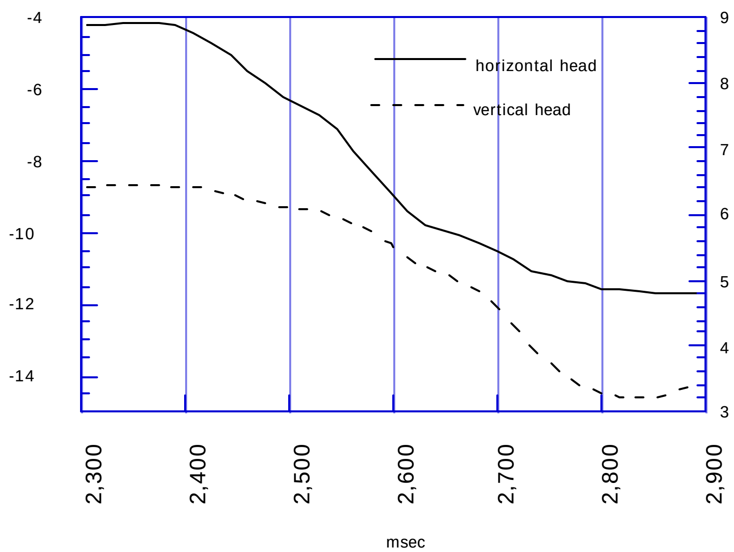
Figure 4.19 Dissociation of eye and head trajectories. Example of 'diagonal' head trajectory, where the horizontal and vertical components of head motion are executed concurrently. a) Gaze intercept, b) horizontal and vertical head orientation, and c) horizontal and vertical head orientation vs. time.



a)



b)



c)

Figure 4.20 Example of 'curved' head trajectory, where the vertical component of head motion begins before the horizontal component is completed. a) Gaze intercept, b) horizontal and vertical head orientation, and c) horizontal and vertical head orientation vs. time.

Table 4.3 Relative frequency of head movement types for PMD sequences for four subjects.

Head Movement	eb	jw	mh	sc
Separate H & V	0.45	0.22	0.67	0.35
Diagonal	0.36	.072	0.00	0.50
Curved	0.19	0.72	0.33	0.15

into one of the above categories, and cases where eye and head movements could not be reliably paired were excluded. Note the wide variation between subjects. No diagonal head movements were observed for subject mh, while jw showed 0.72. Subjects eb and sc were intermediate between those extremes.

4.3 Discussion

Subjects adopt a regular, rhythmic pattern of eye, head, and hand movements while performing the block-copying task. On first inspection the head and hand records appear almost sinusoidal, but an examination of the details of the movements shows marked asymmetries. The oscillating head and hand movements made by subjects while copying the model pattern vary depending on the specific sub-task being performed. For example, the head is moved at a higher velocity when moving towards the resource area than in the opposite direction, and the head remains stable for a shorter period for the pickup than for the drop. This asymmetry in dwell-time is more pronounced in the hand movement record, where the hand typically remains in the workspace about twice as long as in the resource area. The differences between pickup and drop actions extend to the temporal coordination of eye and hand as well. When picking up a block in the resource area, gaze was typically held on the block targeted for pickup only until the fingers were about to touch the targeted block. When the block was being placed in the workspace, however, gaze was usually maintained until the drop was complete and the hand moved away from the board. Asymmetries were also observed in subjects' eye/head coordination. The head often leads the eye in gaze changes away from the resource area, while eye and head movements were initiated at approximately the same time for gaze changes away from the resource area. These results demonstrate that movements of the eye, head, and hand, which one might expect to be controlled by low-level motor routines, are in fact dependent on the immediate task. Answers to questions like "does the eye lead the head in gaze changes?" are meaningful only when the specific task being performed is clearly understood. Refinements to the question, such as distinguishing 'predictive' saccades begin to acknowledge this relationship, but these experiments demonstrate that investigations into complex behaviors are only useful when the task is taken into account. This is crucial when considering the reports of Land [1992] and Kowler, *et al.* [1992] who both concluded that the eye and head are programmed together.

In the block-copying task, when subjects are performing a PMD sequence (alone or as part of an MPMD block move), they make two gaze changes, first from the resource to the model, then from the model to the workspace. Because subjects typically move the head toward the area where manual manipulations are performed (and hold the head still for a period), there are two goals for the eye and head motor systems. The gaze must go from resource -> model -> workspace, while the head must get from the resource to the workspace in preparation for the drop. Current models of changes in gaze postulate that both eye and head movements are driven by the desired gaze shift in body-centered or exocentric (spatial) coordinates [Guitton 1992, van der Steen 1992] and that the VOR is suppressed until the spatial goal is achieved. Based on these models, and the results of Land [1992] and Kowler *et al.* [1992], one would expect that the eye and head trajectories would be tightly linked. Some subjects did indeed move the eye and head together, first from resource -> model, then from model -> workspace. Other subjects, however, executed the movements in a very different way with gaze and head moving to different targets at the same time. The degree to which the eye and head motor programs dissociate varied between- and within-subjects.

While there is large variation between subjects, this experiment has shown clear evidence that subjects are capable of programming independent eye and head movements. Given this conclusion, one must ask why evidence of this dissociation has not been observed in the past. One cannot look at behavior divorced from the immediate task being performed by a subject and make conclusions about the capabilities of the visuo-motor system. In this case, it is not a question of whether the task was 'natural,' or whether unbiological movement restraints were forced by the test apparatus, but the nature of the primary task itself that must be considered. Land's [1992] driving task was a natural task and subjects were free to move their heads, yet his data showed no hint of eye/head dissociation. He concluded that in cases where subjects were too busy to exert conscious control over eye and head movements, they were directed to a common goal. The difference between

Land’s driving task and the block-copying task, however, is not how ‘busy’ the subjects were; it was the nature of the immediate task being performed. The driving task required large gaze changes with almost no vertical component, and there is no clear advantage to a temporal dissociation between the horizontal eye and head movements in Land's task. Kowler *et al.*’s [1992] conclusion that there was a natural tendency to program common eye and head movements was based on their study of reading and nonsense tasks requiring the same kind of eye and head movements as reading. If their data is examined from a different perspective, with an emphasis on the immediate task, a very different conclusion is possible: while they observed common eye and head trajectories in the nonsense tasks, dissociation of eye and head movements was observed in the natural multi-line reading task. Kowler *et al.* [1992] reported that subjects sometimes made eye and head movements in the opposite direction at the end of a line of text with the head moving left and down to the next line of text as the eye makes a final saccade to the right. As in the block-copying task, the reading task presents a situation in which dissociation of eye and head is an optimal strategy.

5. The Effect of a Concurrent Cognitive Load on Task Performance

5.1 Introduction

The experiments described in the previous chapters demonstrate a tradeoff between frequent eye movements and working memory in the block-copying task. In a sense, eye movements and working memory have the same logical status in terms of the value of variables currently required for the task. Requiring the subject to perform a secondary task that requires the allocation of limited central resources may affect the tradeoff between eye movements and working memory since some of the variables in working memory may be allocated to the secondary task.

The nature of working memory is not yet well understood; early concepts of working memory held that general working memory served as a unitary short-term store [e.g., Miller 1956]. This view was supported by evidence that cognitive and memory tasks in different modalities interfered with one another. But the interference was not complete, and experiments demonstrated that the performance tradeoff in dual tasks was not linear, suggesting that while all 'short-term memory' loads are somehow related, they did not simply share a unitary store. For example, Brooks [1968] performed a series of experiments probing the interaction of verbal and visual stimulus/response combinations. In one experiment the stimulus was visual; subjects were briefly shown a block capital letter, then were instructed to 'go around' the periphery of the letter (from memory) saying 'yes' for each corner that was on the top or bottom of the letter, and 'no' otherwise. They responded in one of two ways; they either spoke aloud, or pointed manually to the words 'yes' and 'no' printed on a piece of paper. Subjects were more accurate when they responded verbally to the visual stimulus held in memory. Next, Brooks presented subjects with a sentence. Subjects were required to hold the sentence in memory, 'scan' the sentence, and report 'yes' for each noun, and 'no' for all other words in sequence. In this case, the manual response was more accurate than the verbal response.

Taking such results into account, Baddeley and Hitch proposed a more complex model of working memory made up of a number of semi-independent memory subsystems controlled centrally by a limited capacity 'executive' [Baddeley and Hitch 1974, Hitch and Baddeley 1976]. Their model included separate stores for verbal and visual information, with the central executive responsible for coordinating and controlling the peripheral subsystems. In this model, verbal information is held in an 'articulatory loop' capable of holding and rehearsing up to 6 or 7 items without loading the central executive. The articulatory loop is paired with a 'visuo-spatial sketch pad' (VSSP) responsible for temporary storage of spatial information.

5.2 The Dual-Task Paradigm

Brooks' experiments provided support for a subsystem/central executive mechanism like that proposed by Baddeley and Hitch by demonstrating modality-specific stimulus/response interference in a single task. Other evidence comes from dual-task experiments in which subjects perform a primary task along with a concurrent secondary task. In classical 'psychological refractory period' (PRP) experiments, two isolated stimuli are presented with variable stimulus onset asynchronies (SOA), and the reaction times are recorded. The general result is that the reaction time to the second response is delayed as the SOA is decreased [Pashler *et al*. 1993]. The first task apparently utilizes resources that are needed to initiate the second task. There is evidence that the second response is delayed because the 'response selection' stage of the task is postponed, suggesting that the 'bottleneck' limiting performance in dual-tasks is a limited central resource that performs response selection for diverse tasks [Pashler *et al*. 1993].

Yet in the dual-task literature many tasks appear not to interfere and Pashler was unable to find interference with some saccadic eye movement tasks. The alternative conceptualization to a central bottleneck is diagrammed in Figure 5.1, after Wickens [1986, Figure 3.7]. Human processing resources are visualized in three dimensions. The left face of the solid shows stimulus space, divided into modality/code space; visual or auditory stimuli can carry spatial or verbal information. The 'length' of the solid represents stages of processing, from encoding through response. The response can be either manual or vocal. Multiple tasks utilizing resources separated within this space are thought to interfere less than tasks whose resource requirements overlap. The block-copying task would appear along the far-top edge because the task provides visual/spatial input and requires manual responses.

Pashler points out that attempts to examine complex tasks using the dual-task paradigm have suffered because they "can only be analyzed in a rather coarse way (e.g. aggregating performance errors over many seconds)" [Pashler *et al*. 1993, p. 52]. Thus a central bottleneck might not be evident in the data measured at a coarse timescale. The block-copying task, however, allows us to detect even subtle

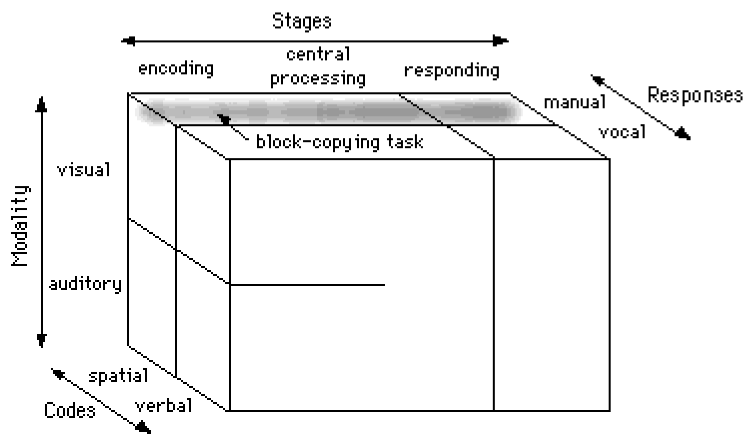


Figure 5.1 Multi-dimensional structure of cognitive processing resources proposed by Wickens [1986].

interference because movements are monitored at such a fine scale. The analyses in the previous chapters allow interference in oculomotor performance, strategy choice (and therefore memory use) and eye/head/hand coordination to be detected.

People are generally not aware of their eye movements, and many so-called 'dual-task' experiments require concurrent eye movements, even though the eye movements are rarely acknowledged as a concurrent task. Kowler [1995] showed that saccadic and pursuit eye movements invariably require the allocation of attentional resources, which suggests that the programming of eye movements should both affect, and be affected by, concurrent tasks, at least at a fine temporal scale. It may be that interference has not been detected in many dual-tasks involving eye movements because of the very brief periods involved in individual eye movements. Any interference could be hidden if the programming of eye movements can be interleaved with other tasks so that the two tasks are performed serially.

What we have found so far suggests that the block-copying task can be viewed as a sequential program of interleaved perceptual and motor actions; in order to move each block the subject must program, initiate, and monitor a large number of eye, head, and hand movements, while gathering information via frequent fixations. Thus our paradigm supports the central bottleneck idea. If this is the case we expect to see some effect of even a simple, unrelated task if we look at the individual movements at a fine timescale. The block-copying task is an interesting area for investigating to what extent all of these programs require overlapping central processing resources, and to what extent they can be carried out independently. Frens & Erkelens [1991] reported that hand movements are affected by an auditory stimulus presented near the time of a visual target for a hand movement. The primary goal of this chapter is to examine whether we would observe similar interference in a task which does not involve the same sensory or motor modalities but may require common central attentional resources.

5.3 Concurrent Cognitive Load

Two cognitive load conditions were investigated. The two differed in difficulty and in the extent to which the secondary, cognitive load task "overlapped" with the primary task.

5.3.1 The "Attend" Condition

In the 'attend' condition, the subject was instructed to perform the block-copying task, and at the same time perform a secondary task with limited working memory and attentional resource demands. The task was otherwise unrelated to the primary task. Before the trial began, the subject heard a "target character" (e.g., "F"). While performing the primary block-copying task, subjects heard a 1.1 Hz pseudo-random stream of characters (e.g., "A, K, F, V, ..."), and were instructed to repeat the target character aloud whenever they heard it in the speech stream. The target character and speech stream were created using the PlainTalk™ speech generation capabilities of a Macintosh 7100/66AV computer. The target character was selected at random from a 21 character set (A-Z, excluding C, D, P, T, and X). The characters making up the speech stream were selected from the same character set, with the probability of selecting the target character set to 0.20 per spoken character. The target occurred an average of 3.5 times per trial (based on an average rate of 1.1 Hz and an average trial time of 16 seconds). The 'attend' condition was run in two configurations; 'near-attend' trials were performed in the control configuration (see Figure 2.1). In the 'far attend' condition the model, resource, and workspace were moved farther apart on the working plane, increasing the cost of model references, as shown in Figure 3.1.

5.3.2 The "Say Other Color" Condition

The more complex cognitive load condition was labeled the 'say other color' condition. The subject was instructed to say a block color aloud as each block was copied, with the restriction that the color had to be different than the color of the block being copied. For example, while copying a blue block, the subject could say "red," "green," or "white," but not "blue." This task was expected to interfere more directly with the primary block-copying task. The 'say other color' condition was run only in the 'near' (control) configuration.

5.4 Results

5.4.1 Effects on motor coordination:

Performing the block-copying task requires the coordination of eye, head, and hand movements. Frens and Erkelens [1991] showed that gross motor (hand) movements were affected by an auditory stimulus, and Kowler [1995] has shown that saccadic and pursuit eye movements require the allocation of attentional resources, so it may be the case that programming head and hand movements also place demands on limited central resources. Referring back to the schematic timeline in Figure 2.18, the complexity of the extended block-copying task is evident. The caption "Pickup Block #1" actually requires a series of actions:

- Identify a resource block in the peripheral view
- Program eye, head, and head movements to that block
- Initiate the movements
- Monitor the movements and program corrections if needed

Placing the block in the workspace also requires a complex series of motor actions, especially in the common case where an intermediate model area fixation is programmed between the pickup and drop. If we think in terms of a sequential program in which the motor actions have to be interwoven with the information gathering model fixations, then the need to coordinate these complex actions, along with the result indicating that the addition of a cognitive load influenced the subjects' strategy, leads to the question of whether motor behavior is affected by the addition of a secondary cognitive task. Some aspects of motor behavior considered earlier can be used to study this question. The cross-correlation analysis introduced in Chapter 4 can be used to look for changes in motor behavior when an additional cognitive load is added on top of the primary task. Figure 5.2 shows the temporal position at which the peak of the horizontal gaze/head cross-correlation occurred for subject sc. There was a significant drop in the gaze/head waveform latency in the cognitive load condition for sc, who showed the greatest change in correlation under the cognitive load conditions. Not surprisingly, subjects' responses were idiosyncratic. Because the cross-correlation is a function of extended waveforms rather than the onset of individual movements, it is not possible to tell from this analysis whether the change in position of the peak is a result of a timing shift, a change in the gaze and/or head's waveform, or both. It is important to note that the primary and 'attend' tasks do not involve the same sensory or motor modalities, so any interference must occur at a higher level.

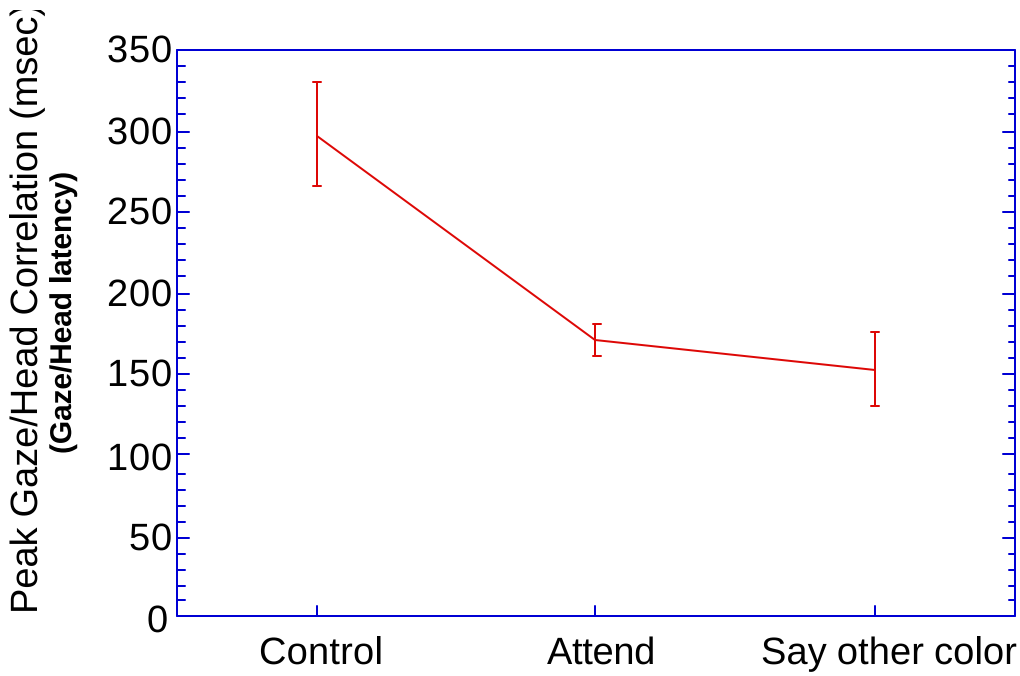
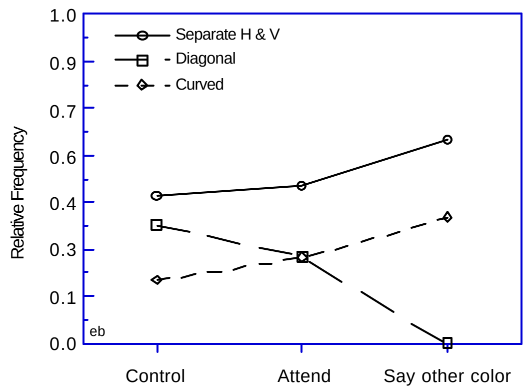


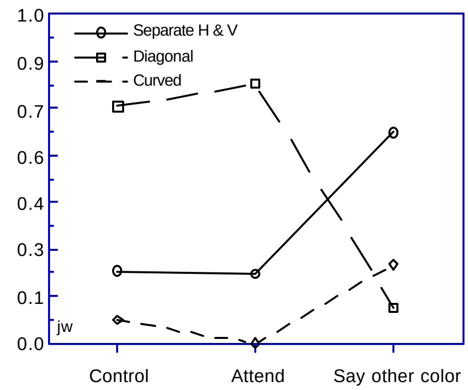
Figure 5.2 Temporal position of peak gaze/head correlation for subject sc in control and cognitive load conditions.

It was shown in section 4.2.3 that some subjects program the eyes and head to independent targets while moving blocks from the resource area to the workspace. This sophisticated behavior likely requires attentional resources. If that is the case, then the addition of a secondary cognitive load may interfere with the targeting of eye and head. Table 4.3 listed the relative frequency of targeting categories used by four subjects in the control condition. Figure 5.3 shows the effect of the two cognitive load conditions on the subjects' head targeting behavior in the 'attend' and 'say other color' conditions. As in their performance in the control condition, subjects' responses to the added cognitive load varied. Subject eb, who displayed the widest distribution of eye/head targeting strategies in the control condition, shifted towards a tighter linkage between eye and head (decreasing 'diagonal' head movements, and increasing 'separate horizontal & vertical' and 'curved' movements) under the added cognitive load. Except for subject mh (who did not make any diagonal head movements in the control condition), the frequency of diagonal (dissociated) head movements fell dramatically in the 'say other color' condition. The decrease in diagonal head movements (i.e., dissociated eye and head targets) under the 'attend' and 'say other color' conditions suggests that independently targeting eye and head movements requires the allocation of central resources. When the subject is required to perform the cognitive load along with block-copying task, those central resources are apparently too scarce to also support the independent targeting of eye and head movements.

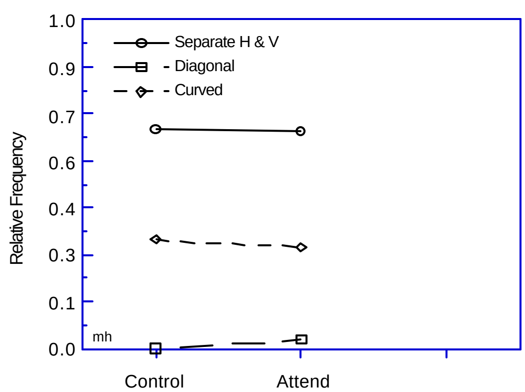
Again, subjects' response to the addition of the cognitive load was idiosyncratic, but other metrics of eye/head coordination also showed evidence that eye and head movements are more closely tied under the cognitive load conditions. The number of



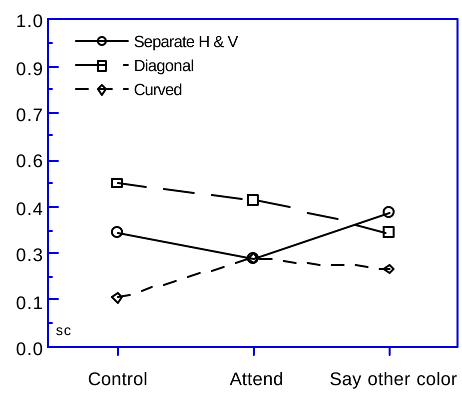
a)



b)



c)



d)

Figure 5.3 Relative frequency of eye/head targeting behavior in Control, Attend, and Say other color conditions for subjects a) eb, b) jw, c) mh, & d) sc.

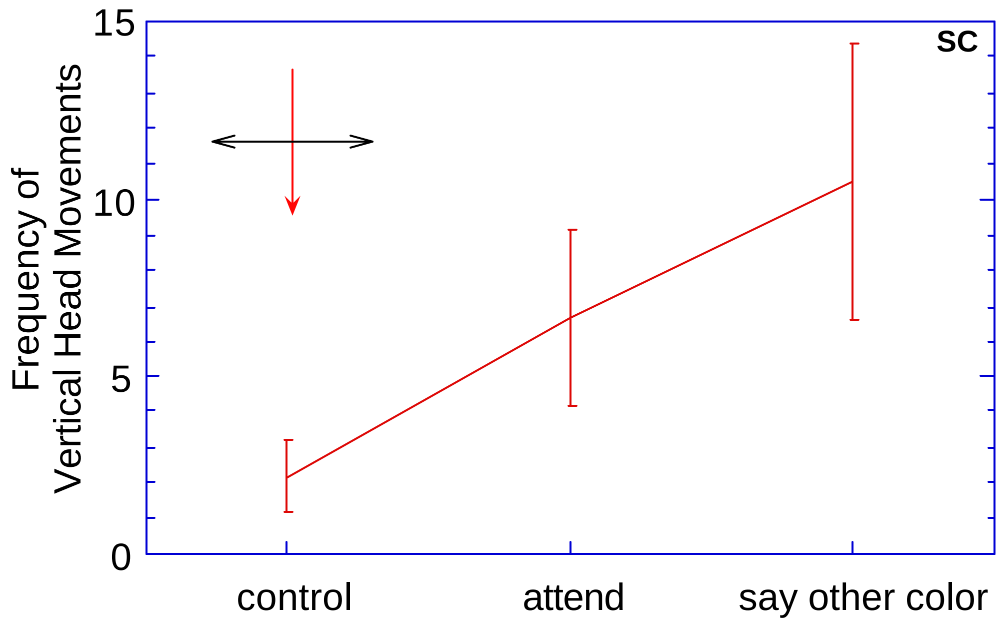


Figure 5.4 Frequency of downward, vertical head movements per trial under control and cognitive load conditions for subject sc.

vertical head movements from the model to the workspace that accompanied model to workspace gaze changes was counted. While there was considerable variability between sessions, Figure 5.4 shows that the number of vertical head movements between the model and workspace per trial increased in the 'attend' and 'say other color' conditions, again indicating a tighter linkage between eye and head under the cognitive load conditions.

So, while one might think that head movements are controlled by low-level motor routines, it appears that when an additional cognitive load is added, the eye and head become more tightly linked, temporally and spatially. Other aspects of motor behavior were affected as well. The RMS measure of head movement amplitude used in Chapter 4 can also be used to look for signs that the added load influences motor actions. Even the 'attend' condition, in which the subject was only required to remember a single target character and respond to a 1.1 Hz speech stream, affected subject sc's motor performance. Figure 5.5 shows the head's RMS amplitude along the three rotational degrees of freedom. The RMS amplitude increased along all three axes. The increases between control and 'attend' were significant in all cases (two-tailed, $P<0.005$), as were the increases from control to 'say other color' and 'attend' to 'say other color' (two-tailed, $P<0.001$). These effects suggest that many aspects of task behavior are to some degree affected by central processing limitations.

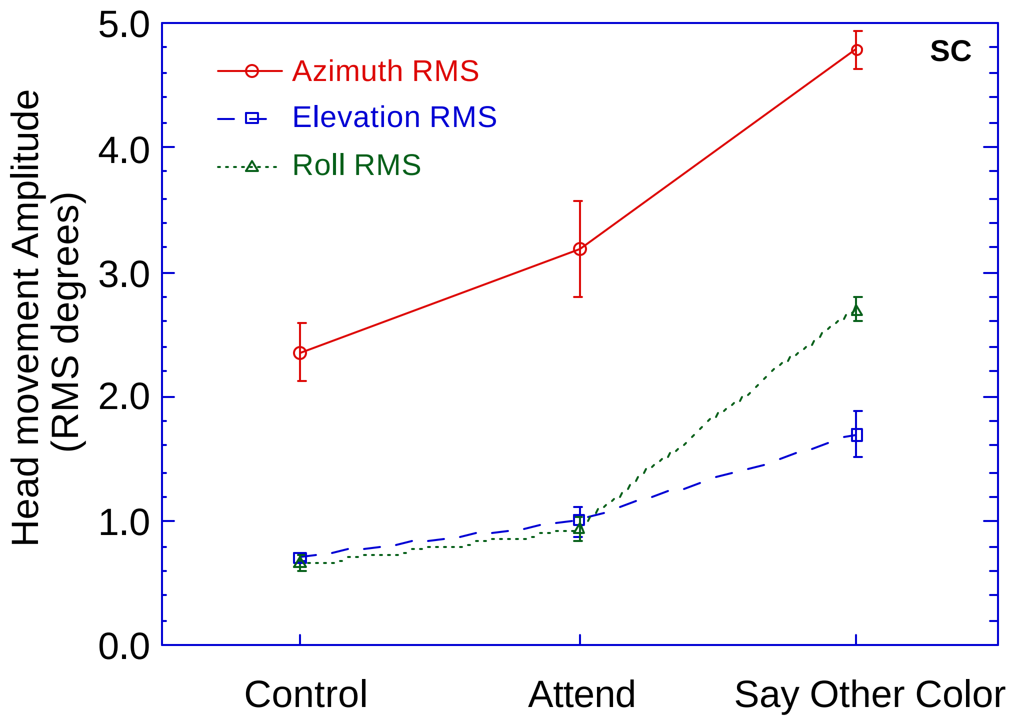


Figure 5.5 Amplitude of head movements under control and cognitive load conditions for subject sc.

5.4.2 Effects on Eye Movement Strategies

It was demonstrated in Chapter 3 that there is a trade-off between eye movements and working memory. The added cognitive load conditions presumably reduces the available working memory resources, and therefore might result in increased eye movements during the block-copying task. Figure 5.6 a) - d) show the relative frequency of each strategy for the four subjects in the control and 'attend' conditions. There was wide variation between subjects. Figure 5.7 a) - c) show the strategies for the three subjects who ran in the 'say other color' condition. As expected, there was a more dramatic shift in strategy use in this condition.

Because of the variation in strategies between subjects (in both control and cognitive load conditions) it is helpful to examine the change in observed strategies between the control and cognitive load conditions. Figure 5.8 shows the average change in relative frequency of each strategy in the two cognitive load conditions. The plotted values are the average differences between the control and cognitive load conditions (i.e., $\text{frequency}_{\text{attend}}$ - $\text{frequency}_{\text{cog. load}}$). There was a significant shift from the high-memory PD strategy to the low-memory MPMD strategy in the 'attend' condition. More dramatic shifts were observed in the relative frequency of MPMD and PD strategies, and other strategy shifts become evident in the 'say other color' condition which was expected to interfere more directly with task performance. The relative frequency of >MPMD block moves (with 3 or more model references) increased in this condition, and the frequency of the single model reference MPD and PMD strategies fell. The idiosyncratic response to the addition of the cognitive load is evident in Figure 5.9 a) - d), which show the change in strategy use for individual subjects.

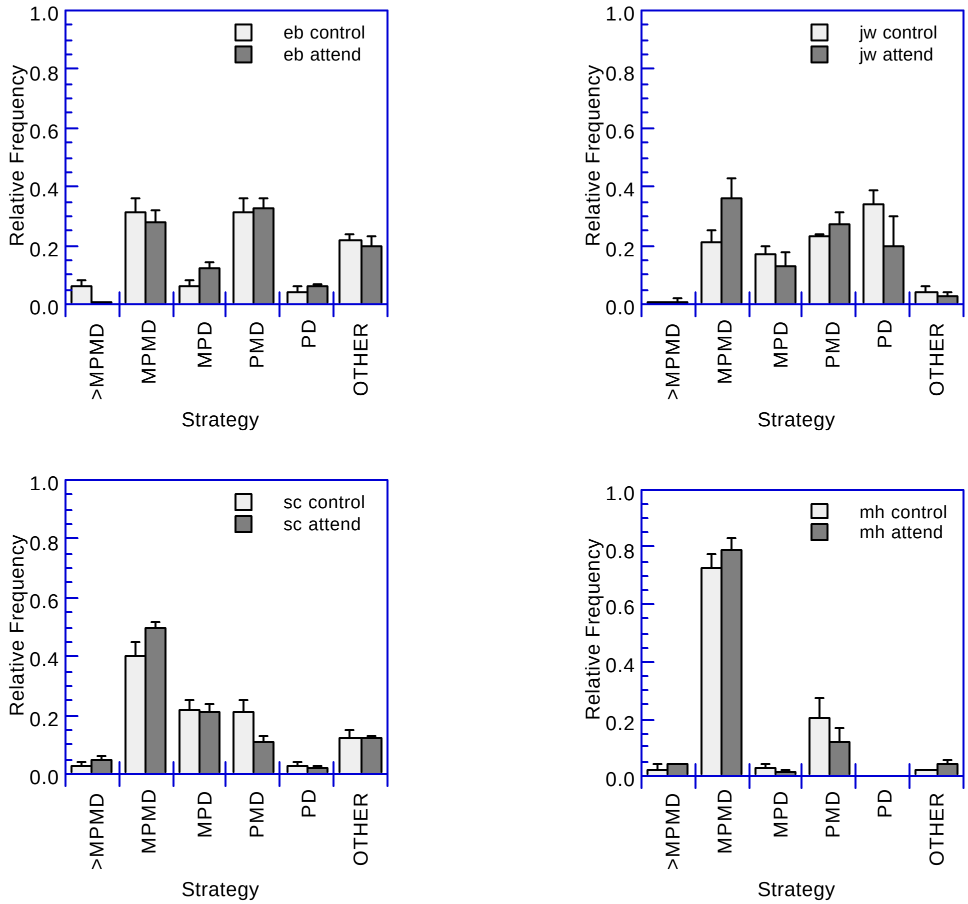


Figure 5.6 Relative frequency of strategies in control and 'attend' conditions.

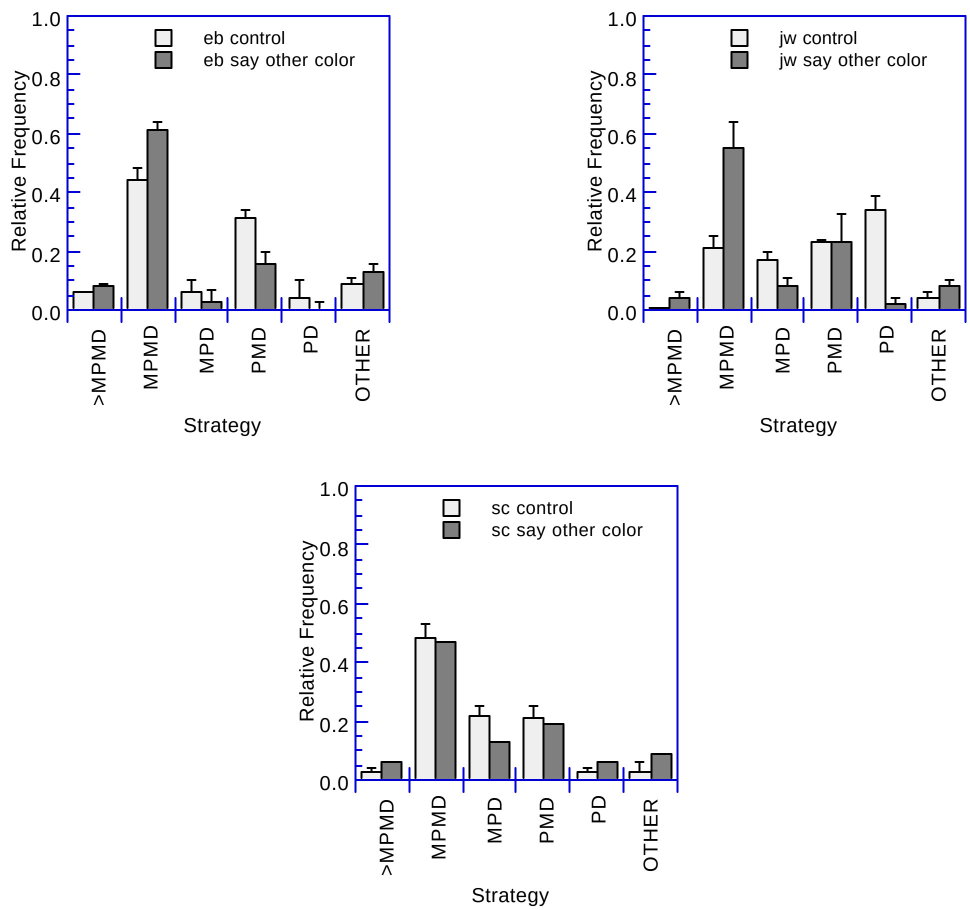


Figure 5.7 Relative frequency of strategies in control and 'say other color' conditions

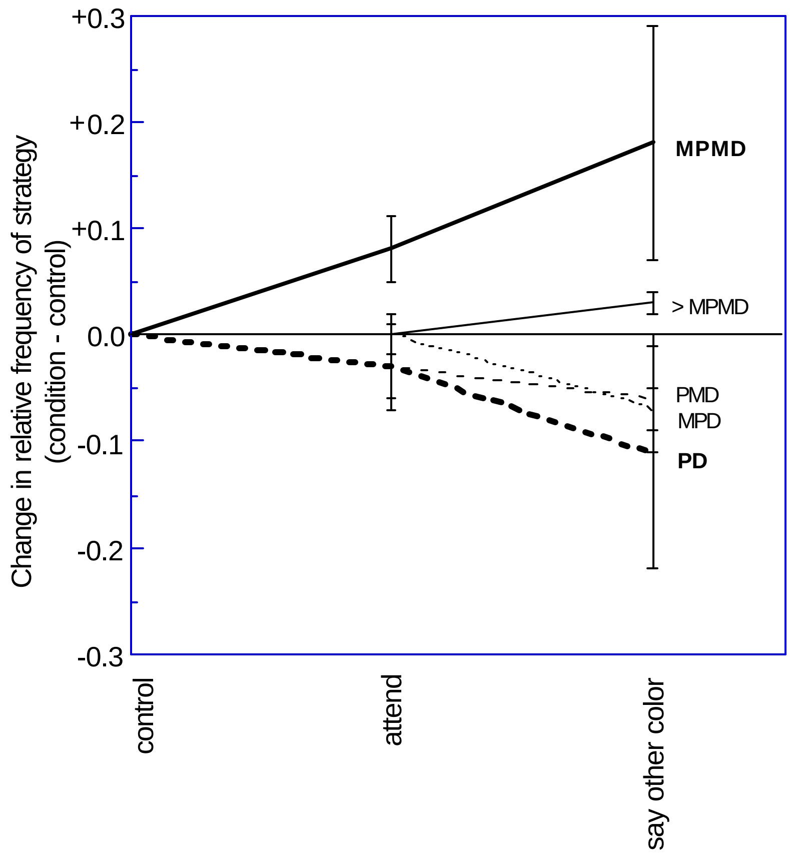


Figure 5.8 Change in relative frequency for each strategy in the control and two cognitive load conditions; "attend" and "say other color." In the "attend" condition

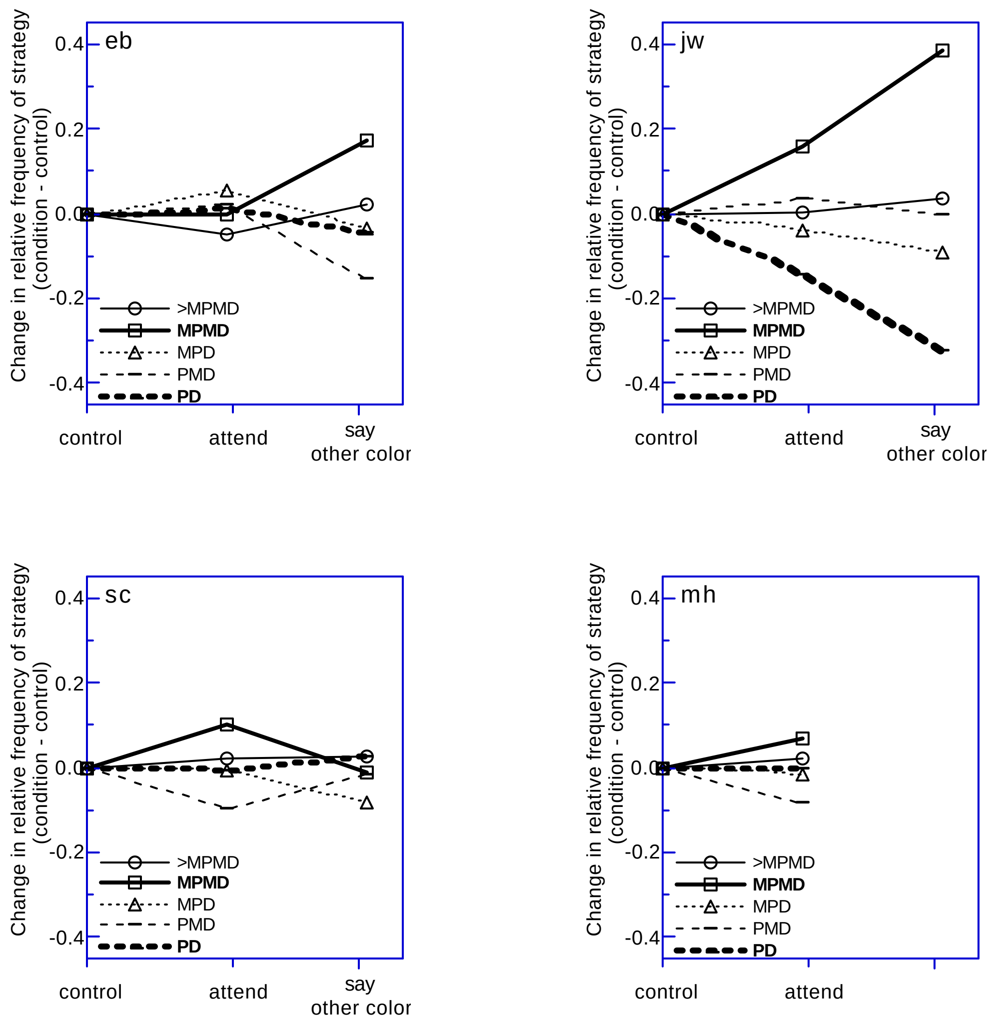


Figure 5.9 Change in relative frequency for each strategy in the control and cognitive load.

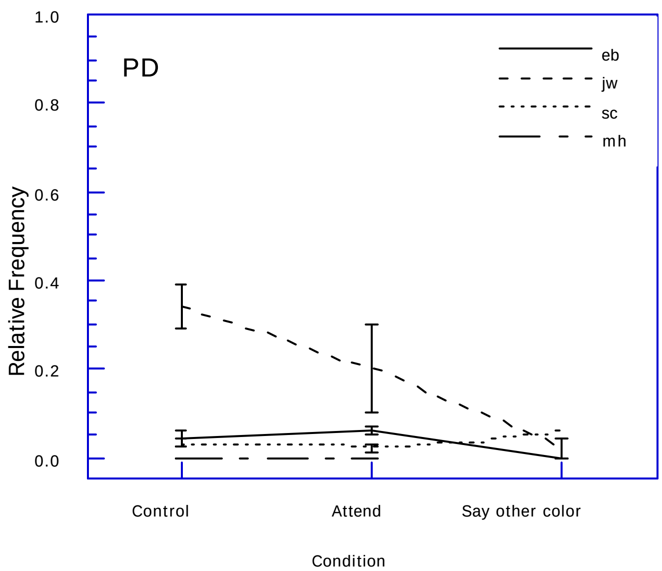


Figure 5.10 Convergence of PD strategy use under added cognitive load.

When viewed a different way, however, some aspects of subjects' performance is seen to converge toward a common reference. Figure 5.10 shows the relative frequency of the PD strategy for the three subjects. While there was a wide range of PD strategy frequencies in the control condition (subject jw had the highest fraction of PD moves, and the fewest number of model references of any subject run in the block-copying experiments), the range narrowed in the 'attend' condition, and virtually disappeared in the 'say other color' condition as jw's use of PD block moves fell to the same value as the other subjects.

5.4.3 Number of Model References Per Block

The shift caused by the addition of the cognitive load can also be analyzed in terms of average number of model references per block. Figure 5.11 shows that measure for the four subjects who performed the 'attend' condition and the three who performed the 'say other color' condition. Two of the four subjects in the 'attend' condition showed significant increases in the number of model references in the 'attend' condition, the other two did not show a significant change. The most dramatic increase was seen in the subject who had relied most heavily on working memory in the control condition: Subject jw increased from 1.0 to 1.5 model references per block. All three subjects who completed the 'say other color' condition showed significant increases in the number of model references per block compared to the control condition. As in Figure 5.10, the convergence under added cognitive load is evident.

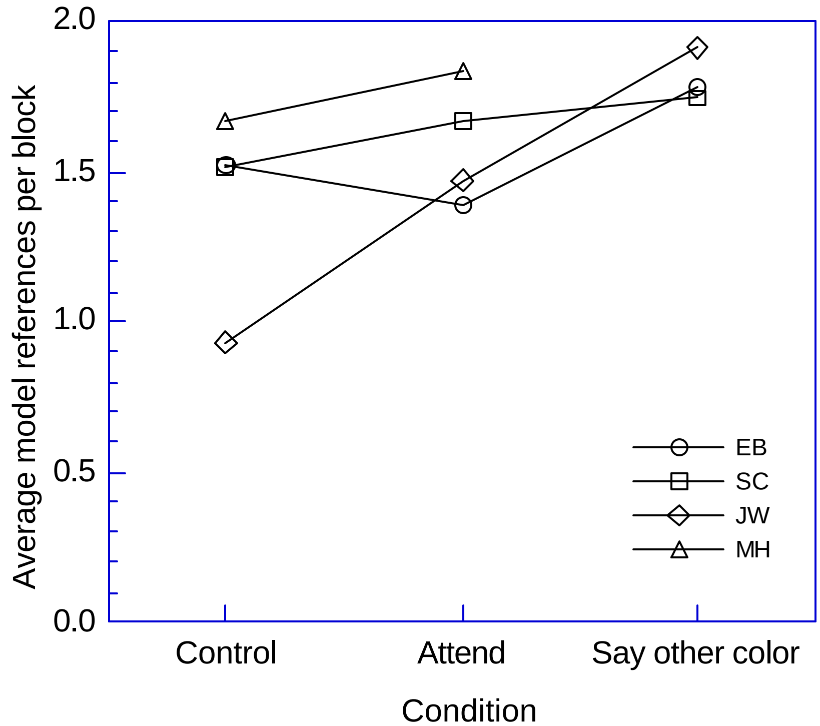


Figure 5.11 Number of model references per block for subjects in control and cognitive load conditions.

5.4.4 Concurrent Cognitive Load in the 'Far' Condition

The results of the cognitive load experiments indicate that the changes are a function of the balance between the use of frequent eye movements and working memory. The 'far' condition was shown in Chapter 3 to shift that trade-off toward heavier reliance on working memory, so it is of interest to learn how subjects react to the combination of demands in the 'far attend' condition. The combination of the 'far' configuration and the 'attend' condition is of interest because the two manipulations have been seen to shift the balance between working memory and eye movements in opposite directions; subjects reduced the number of model references in the 'far' condition, and increased them in the 'attend' condition. Figure 5.12 shows the relative frequency of strategy use for four subjects in the far control (i.e., the far condition with no added cognitive load) and the far attend conditions. Subject jw, who showed the highest fraction of PD block moves in the control and far conditions, showed very little difference in the three conditions. Subjects eb and sc, who used a significant number of PD block moves in the 'far' condition, reduced their use of the high-memory strategy dramatically when the cognitive load was added, replacing those with more MPMD and PMD block moves. Subject mh, who did not use the PD strategy in the 'attend' condition still showed an increase in MPMD block moves. The changes are reflected in the average number of model references per block copied for the same conditions, shown in Figure 5.13. The average number of model references per block fell from 1.41 in the control condition to 1.17 in the 'far' condition (see section 3.2.1). Adding the relatively simple 'attend' task to the 'far' condition resulted in a compromise, significantly increasing the average number of model references per block to 1.30 ($P < 0.05$).

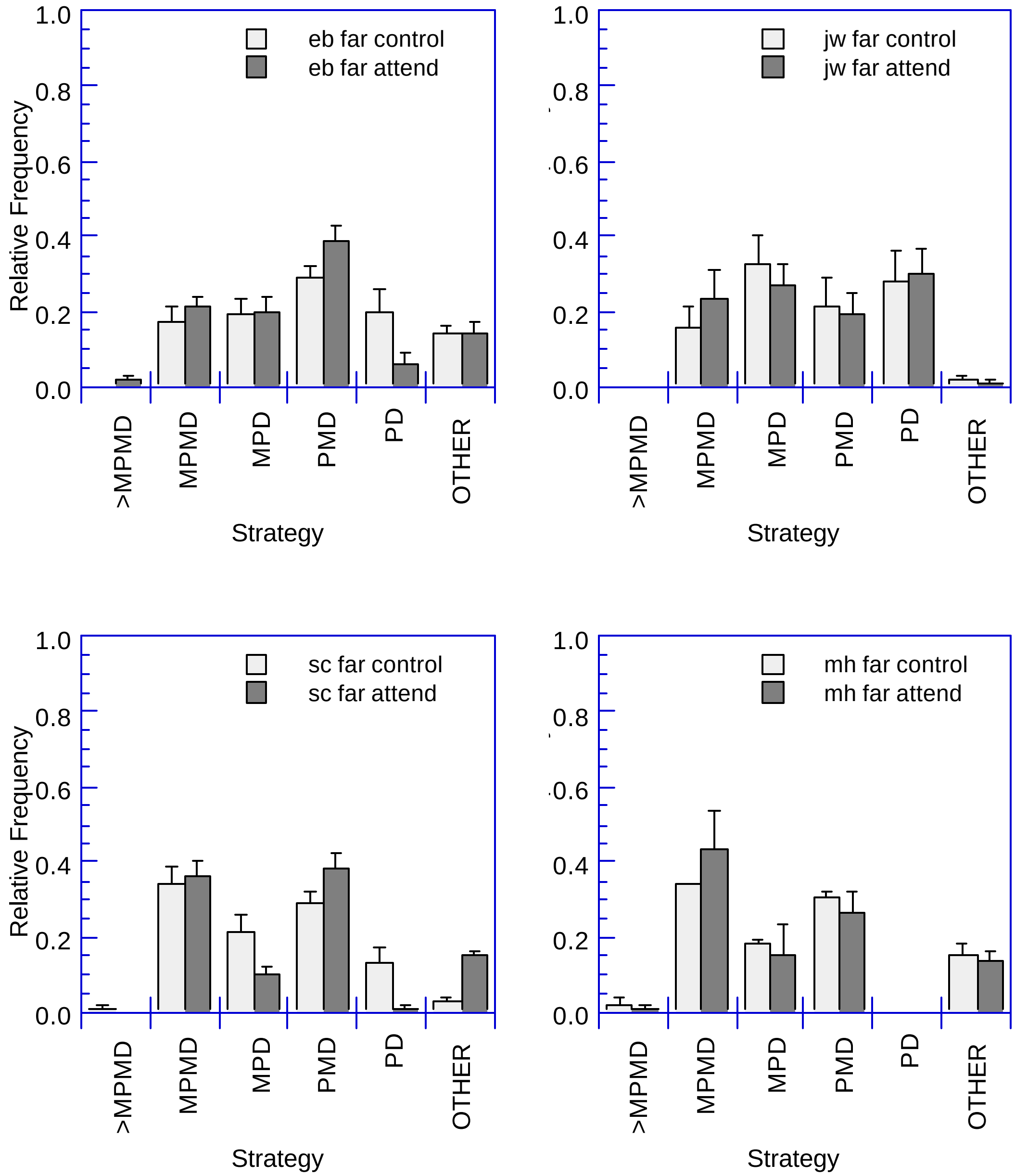


Figure 5.12 Relative frequency of strategies in 'far-control' (i.e., far with no added cognitive load) and 'far-attend' conditions.

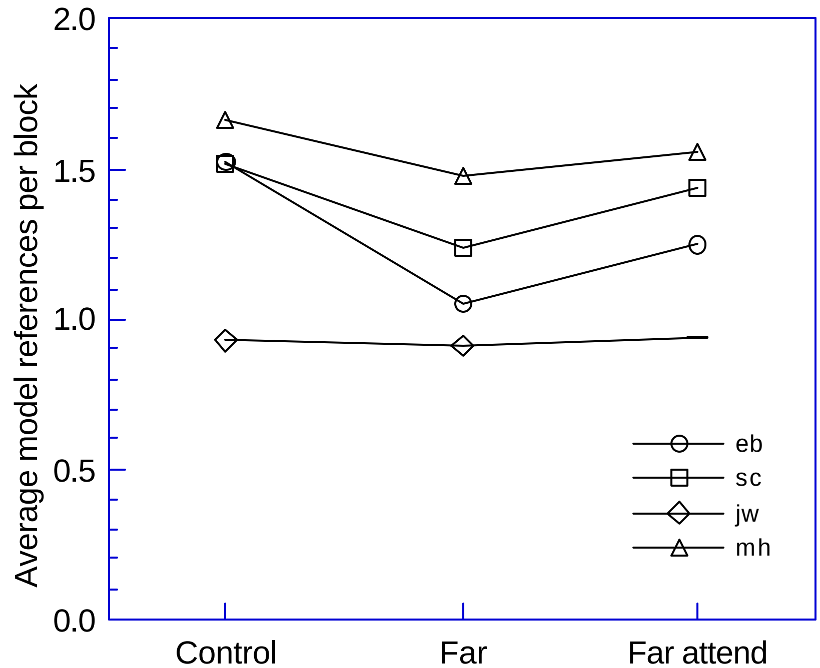


Figure 5.13 Average number of model references per block for the control, far, and far attend conditions.

5.5 Discussion

As with eye movements, taking the task into account allows issues about the cognitive control of head movements to be addressed. Subjects' responses to the added cognitive load were idiosyncratic, but the addition of a secondary task affected several measures of eye/head motor performance. The peak of the cross-correlation between gaze and head was reduced for some subjects, indicating that the eye and head were more tightly linked when the cognitive load was added, and the frequency of block-moves in which the eye and head were targeted independently was reduced dramatically. Subject eb, who showed independent targeting of the eyes and head in approximately 20% of PMD sequences in the control condition, fell to only half that number in the 'attend' condition, and to almost zero in the 'say other color' condition. The amplitude of head movements also increased for some subjects. These change in eye/head dissociation and head amplitudes are also consistent with the eye and head becoming more tightly linked in space and time.

These experiments have shown that head targeting, which might be expected to be controlled by lower-level peripheral motor routines, is also limited at a central level. Kowler [1995] has shown that programming saccadic eye movements requires attentional resources. These results suggest that head movements and the coordination of eye and head movements also require central resources. When the demands on the limited central resources are increased, the motor systems apparently 'fall back' to simpler, linked movements that require less processing 'overhead.' This suggests that the same shared central resource utilized in the primary task to maintain the color and position information about the blocks is also responsible for coordinating the interwoven perceptual and motor actions necessary to complete the task. Even though the economies of parallel computation are used to great advantage, it appears that complex behaviors are limited at a central level.

The concurrent cognitive load also affected eye movement strategies. When subjects were required to attend to a speech stream in the 'attend' condition, there was a small but significant shift in strategy use by three of four subjects, increasing the relative frequency of double-look MPMD block moves, and decreasing PMD and PD moves. The 'say other color' condition lead to more dramatic shifts in the three subjects who completed the task; MPMD block-moves were increased still further, while PD block moves dropped, especially in the subject who had the highest frequency of PD block moves in the control condition. This change in the high-memory strategy points to an interesting aspect of subjects' performance under control and cognitive load conditions. The large between-subject variation in the frequency of PD strategies virtually disappeared in the 'say other color' condition, as all three subjects converged to the same low frequency of PD block moves. The dramatic decrease in PD strategies by the subject who had previously relied on them the most heavily was not paralleled by a similar shift in strategies by the other subjects. While all subjects made more frequent model references in the 'say other color' condition, these changes also resulted in a convergence to a much smaller range of model references per block. The range of values for the three subjects who completed both cognitive load conditions was 0.6 in the control condition, dropped to 0.3 in the 'attend' condition, and to 0.15 in the 'say other color' condition. This experiment demonstrates that the addition of a secondary task, even one unrelated to the primary task, affected subjects' choice of strategy and the degree to which they relied on working memory in copying the model pattern. The convergence of different subjects' performance may suggest that there is some common minimum performance to which all subjects fall when presented with sufficient extra load. The 'far attend' condition demonstrates subjects' ability to adapt behavior based on combinations of constraints. When the model, resource, and workspace areas were separated by ~70°, subjects adapted their strategies, relying more on memory and less on frequent eye movements to the model pattern. When the relatively simple 'attend' task was added to the 'far' condition, subjects returned to an intermediate number of model references, balancing the constraints imposed by the 'far' configuration and the added cognitive load.

6. General Discussion

The goal of this dissertation was to explore visual representations and visuo-motor coordination in the context of complex, ongoing behavior. This required the development of a new laboratory facility capable of monitoring eye, head, and hand movements at a fine time-scale. The task selected was a block-copying paradigm that reflects essential perceptual, motor, and cognitive operations. The task involves important aspects of natural behavior, yet is sufficiently constrained so that the computations necessary to perform each subtask can be made explicit. While we have known since Yarbus' [1967] work that eye movements reflect cognitive events, little progress has been made in understanding this relationship [Viviani 1990]. Because of the design of the block-copying task used in these experiments, however, fixations are tied closely to the perceptual, cognitive, and motor subtasks that constitute subjects' overall behavior in the task. When subjects fixate the model area after dropping a block in the workspace, then pick up a block from the resource area, we can infer that the model fixation served to acquire the color and/or position of a block in the model pattern. Similarly, fixations in the resource and workspace areas accompanying block pickups and drops can be assumed to assist in targeting hand movements. The development of the new laboratory opened up a field of inquiry into what has until now been largely unexplored territory. The ability to monitor natural behavior at a fine time-scale, along with the advantages of the block-copying task, provide new insight into the way humans utilize visual information, working memory, and visuo-motor coordination in the performance of natural tasks.

The first observation was that subjects make very frequent eye movements, returning to inspect the model pattern again and again while copying the eight colored blocks. Eye movements were used to serialize the task into simpler subtasks, which were executed sequentially. The constraints of the task and the subjects' common, stereotyped behavior led to a relatively small number of strategies used to copy each block. Trials were analyzed, and individual block moves were categorized into one of five 'strategies,' labeled in terms of the series of fixations executed while copying the block. The modal strategy was the "Model-Pickup-Model-Drop" (MPMD) strategy, in which the subject looked first to the model, then to the resource area (to guide the block pickup), returned gaze to the model, and finally on to the workspace to guide the block drop. This strategy requires two references to the model area for each block copied. The number of model references averaged 1.5 per block, though the value was not constant over the course of the eight-block trial. The first block was higher (2.0 looks/block) and the last block was lower (1.1), indicating that some representation was built over the course of the trial. It is important to note, however, that the change was not dramatic, and the average number did not fall below 1.0, even for the last block. Because subjects were given no direction on how to perform the task, other than to complete the copy as quickly as possible without making errors, it is important that subjects chose to complete the task by referring to the model so frequently.

It is interesting to return to Table 1.1 (page 12) and note that no subjects used the alternative strategy of first locating and identifying several objects in the scene (quadrant IV in Table 1.1), then moving a number of blocks without fixating the model again. The subjects' use of temporary, task-specific visual representations suggests that vision may be much more 'top-down' than was previously thought. This thesis challenges the idea that the visual system's task is to gather information for integration into a high-fidelity, general-purpose representation of the environment without regard to the immediate task. In the classical view of visual perception, (also embraced by traditional computer vision approaches [e.g., Marr 1982]), planning and cognition was performed by referencing the internal representation. The frequent eye movements used by subjects in these experiments suggest that in real tasks, humans apparently maintain only sparse, transient representations of task-relevant information in concert with dynamic deictic markers to refer to elements in the environment. Given subjects' preference to acquire color and position information in separate model references, it appears that even representations of task-relevant information may be minimal and short-lived. Note that the retinal images during the first and second model fixations in an MPMD sequence are virtually identical, but the information held in working memory (e.g., the position or color of a block) is determined by the 'step' in the sequential program being executed. Whatever internal representation that may have been built during the first model fixation was sparse (and/or had decayed) enough to require the second model reference. The inference that different information is gathered in subsequent model references was supported by an analysis of the fixation times in the first and second model references in MPMD sequences. In moves executed with only one model fixation in each model reference, the first model fixations (color) were shorter than the second fixation (position).

Another series of experiments led to the observation that the trade-off between frequent eye movements and working memory load is flexible, and that subjects are capable of dynamically adjusting the balance based on task demands. One set of experiments manipulated task demands in two ways; by increasing the cost of model references, and by reducing the information content of the model pattern. In the first case, the cost of frequent model references was increased by placing a greater distance between the model, resource, and workspace areas. The 'far' configuration slowed performance in general, but those strategies containing multiple model references were affected the most. Subjects modified strategy use in the 'far' condition, reducing reliance on frequent model references. The average number of looks per block copied fell, but even when model references required large eye, head, and torso movements, the average number of model references per block did not fall below 1.0, indicating that subjects still chose not to copy multi-block patterns 'from memory' rather than make frequent eye movements.

Task demands were also manipulated by reducing the information content of the model pattern. In one experiment, the complexity of the model was reduced by using monochrome model patterns, leaving only information about the blocks' position. In the second experiment, a 'linear' model pattern retained color information, but the positions of all blocks were determined after the first block was placed in the workspace. In both cases the relative frequency of strategies containing two or more model references fell, leading to a reduction in the average number of looks per block, though again the average value never went below one model reference per block copied. Subjects' behavior in the 'monochrome' and 'linear' conditions supports the interpretation that the frequent model references in the control condition are used to acquire color and position information separately. If those frequent model references were merely artifacts of the block-copying task, we would not expect a reduction in model complexity to affect subjects' strategies. Task demands were also manipulated by requiring the subject to perform a secondary task while copying the block pattern. Subjects relied even more heavily on frequent eye movements to the model when a concurrent cognitive load was imposed, relying less on working memory. When the cognitive load was combined with the 'far' configuration, subjects struck a balance between the opposing constraints. The number of references subjects made to the model area under the 'far-attend' condition was midway between the observed performance in the control and 'attend' conditions. Taken together, the results of the experiments that manipulated the cost of model references or the complexity of the model pattern demonstrate that subjects use multiple model references to acquire the color and position of the model pattern, and are capable of adjusting the trade-off between frequent eye movements and working memory based on immediate task constraints.

A novel feature of this dissertation was the development of a laboratory facility to monitor eye, head, and hand movements while subjects performed natural tasks. The facility developed for these experiments also allows the study of how those movements are coordinated in natural behavior. An important observation was that some subjects programmed independent eye and head movements, dissociating their spatial and temporal trajectories. This observation is inconsistent with current models of eye and head movements that postulate a common gaze shift goal sent in parallel to eye and head motor systems [e.g., Guitton 1992, van der Steen 1992]. Recent reports by Land [1992] and Kowler *et al*. [1992] have provided experimental support for such coupling between eye and head trajectories. While the experiments reported in this dissertation resulted in wide variability between subjects, there is no doubt that humans are capable of executing independent eye and head movements to different targets. This demonstrates once again the importance of understanding the task-dependency of complex behaviors. Land [1992] probably did not observe any dissociation between eye and head movements because his task required only relatively slow, horizontal gaze changes, Kowler *et al*. [1992] studied reading and tasks requiring similar eye and head movements. While stressing the 'natural tendency' of the eye and head to move together, they noted that on some occasions the eye and head moved in opposite directions. In the present study, the two-dimensional nature of the task, along with the time pressure under which subjects worked, led to a situation where performance could be optimized by dissociating eye and head trajectories. While subjects exhibited regular, rhythmic patterns of eye, head, and hand movements while performing the block-copying task, significant asymmetries were observed, demonstrating the task-dependence of head and hand movements as well. The temporal coordination of eye and head movements, eye and hand movements, and head and hand dwell-times were different when gaze was moved into, or away from the resource area. In this case, the constraints of the subtask being performed affect subjects' motor actions.

Performing the block-copying task involves a complex set of perceptual, cognitive, and motor primitives. The planning and execution of those primitives to create coherent behavior requires the allocation of limited central resources. The secondary task used in the cognitive load conditions requires some of the same resources. When subjects performed the block-copying task with a concurrent cognitive load, several aspects of eye/head performance were affected. The amplitude of head movements and the peak cross-correlation were both affected, but of particular interest is the change in head targeting. The added cognitive load led to dramatic changes in some subjects' head trajectories. While some subjects often executed independently targeted eye and hand movements in the control condition, the dissociation of eye and head movements fell dramatically under the cognitive load condition.

One of the goals of this thesis was to explore the possible cognitive role of fixations. The experiments have demonstrated that fixations indeed play a crucial cognitive role in perception. A critical aspect of this role is their use in binding task-relevant information to variables in working memory. This conception of working memory as the currently active marked variables leads to a simple interpretation of the tradeoffs between working memory and eye movements, in which fixation can be seen as a choice of an external rather than an internal marker. These experiments also suggest that another aspect of the cognitive role of eye movements is in indexing the execution of sequential programs. Subjects' behavior in performing the block-copying task can be understood as successive application of the what, where, and manipulation primitive actions described in Table 1.1 on page 12. Figure 6.1 illustrates a sequential program executed as a subject performs an MPMD block move. What, where, and manipulation primitives are successively applied to gather information from the scene and guide movements. Each step in the program is indexed by fixations in the model, resource, and workspace areas. This demonstrates how small number of primitives used in a simple control program can be generalized to create more complex behaviors.

This interpretation of the cognitive role of eye movements may also offer insights into the classic division of visual pathways into dorsal and ventral streams [Mishkin, Ungerleider, & Macko 1983]. Positioning and binding of the dynamic markers may be performed by the dorsal stream to parietal cortex, while the ventral stream is used to acquire features (e.g., color and position) from the marked locations. Visual search would require the participation of both systems.

Interpreting eye movements and brain computations in terms of binding variables in behavioral programs blurs the distinction between perception and cognition, which have traditionally been thought of as different domains. Historically we have been accustomed to thinking of the job of perception as creating rich, task-independent descriptions of the world which are then re-accessed by cognition [e.g., Marr 1982]. These experiments suggest that the role of perception may be much simpler since it only needs to create descriptions that are relevant to the immediate task. To the extent that manipulations on a given block are largely independent of the information acquired in previous views, performance in this task suggests that it is unnecessary to construct an elaborate scene description to perform the task and that there is only minimal processing of unattended information. In addition, since color and location information appear to be acquired separately, it appears that even in the attended regions the perceptual representation may be quite minimal. These observations support the suggestion made previously that only minimal information about a scene is represented at any given time, and that the scene can be used as a kind of "external" memory [O'Regan and Levy-Schoen 1983; O'Regan 1992; Irwin 1991; Irwin 1992]. A related suggestion has also been made by Nakayama [1990].

The question of the complexity of the internal representation is often tied to the mechanisms responsible for visual stability, and the ability to integrate information across eye movements. If a high-fidelity, pictorial representation were in fact built up from information gathered over several eye movements, as suggested by McKonkie & Raynor [1976], then we would have access to a stable representation of the environment. The converse argument, that the world appears stable across eye movements, therefore we have access to a rich, internal representation, is not necessarily true. If only task-relevant information is integrated across eye movements, then even sparse representations can contribute to visual stability.

These results suggest a new interpretation of the limitations of human working memory as well. Rather than viewing the capacity limit as a fundamental limit on the brain [e.g., Just & Carpenter 1976], we can look at the limit as an inevitable consequence of an efficient system that uses deictic variables to preserve only the products of the brain's computations that are necessary for the ongoing task. In natural tasks performed in complex environments, it may be the case that retaining information from previous fixations could make tasks more difficult. When task relevant information exists in the environment, it is simpler to reference the external data with a small number of markers serving as pointers than to load all the information into working memory. Shifting a single pointer then takes the place of 'clearing' a complex data set and replacing it with a new set.

Being able to keep "14 ± 2" items in working memory instead of "7 ± 2" would enhance performance in experiments designed to determine such limits, but it would not necessarily be a benefit when performing natural behaviors. The limited number of variables are only a handicap if entire tasks have to be completed from memory; in that case, working memory may be overburdened. So while we experience the limitations of working memory when trying to remember two unfamiliar phone numbers simultaneously, we are able to copy a pattern of eight colored blocks, cook a meal, and navigate an expressway on-ramp with relative ease. Yet the state-space required to complete any one of these actions is greater than that required to remember 14 digits, and would far exceed the capacity of working memory if all the relevant information had to be maintained internally at the same time. Serializing the tasks with frequent eye movements (i.e., fixating relevant features only when that information is necessary) simplifies the tasks by keeping the instantaneous state-space to a minimum. The cost of searching alternatives and learning new behaviors scales exponentially with the number of markers [Ballard, Hayhoe, & Pook 1995], so there is great pressure to limit the number of markers and to find behavioral programs that operate with a minimum number of those markers. Having to operate with a small number of markers is seen to be less restrictive when we note that the effective capacity of working memory is limited only to the amount of information that can be 'pointed to' by the number of markers. There is also a serial vs. parallel tradeoff; instantaneous state-space requirements can be minimized by breaking tasks down into subtasks that can be executed sequentially, reusing markers freed up after the previous step.

This dissertation provides support for a different approach to studying visual processing, in which vision is viewed as more top-down than previously supposed. In such an approach, the task takes on particular importance, because the observed behavior cannot be divorced from the immediate task(s). Acknowledging the task-specific nature of vision presents a challenge: how can the results of any experiments be generalized beyond the task used in that experiment? If complex behaviors are viewed as a sequential program made up of a relatively small number of simple primitives, then experiments designed to identify those primitives are the first step toward understanding their application in a wide variety of visual behaviors.

Operation	Primitive
Select target in model	**where**
Shift gaze to model	eye & head movements (M)
Get color	**what**
Select target in resource	**where**
Shift gaze to resource	eye & head movements (P)
Move hand to fixation point	**manipulation**
Pickup	**manipulation**
Select target in model	**where**
Shift gaze to model	eye & head movements (M)
Get location	**what**
Select target in workspace	**where**
Shift gaze to workspace	eye & head movements (D)
Move hand to fixation point	**manipulation**
Drop	**manipulation**

Figure 6.1 Sequential program of 'what,' 'where,' & 'manipulation' primitives representing an MPMD block move sequence.

7. Bibliography

Agre, P.E. & Chapman, D. (1987). Pengi: An implementation of a theory of activity. *Proc. AAAI, 87,* 268-272.

Aloimonos, Y., Weiss, I., & Bandyodaphyay, A. (1987) Active Vision, *Proceedings 1st ICCV*, London. 35-54.

Baddeley, A., (1986) *Working Memory*. Clarendon Press, Oxford.

Baddeley, A. and Hitch, G.J., (1974) Working memory. In *Recent Advances in Learning and Motivation Vol. VIII,* (G. Bower, Ed.) Academic Press, New York.

Bajcsy, R. (1988) Active Perception. *Proc. IEEE 76,* 996-1005.

Ballard, D.H., (1989) Behavioral constraints on animate vision. *Image and Vision Computing, 7,* (1) 3-9.

Ballard, D.H., (1991) Animate vision. *Artificial Intelligence,* 48: 57-86.

Ballard, D.H., Hayhoe, M.M., & Pelz, J.B., (1995) Memory Representations in Natural Tasks. *Journal of Cognitive Neuroscience 7*, 1, 68-82.

Ballard, D.H., Hayhoe, M.M., & Pook, P.K., (1995) Deictic Codes for the Embodiment of Cognition. *Behavioral and Brain Sciences* (in press).

Ballard, D.H., Hayhoe, M.M., Li, F., & Whitehead, S.D., (1992) Hand-Eye Coordination During Sequential Tasks. *Philosophical Transactions of the Royal Society of London B, 337,* 331-339.

Ballard, D.H., & Ozcandarli (1988). Eye fixation and early vision: Kinetic depth. *Proceedings of the 2nd IEEE International Conference on Computer Vision*

Ballard, D.H., & Rao, R.P. (1994). Seeing behind occlusions. To appear, *Eur. Conf. on Computer Vision*, Stockholm, Sweden.

Becker, W. (1989) Metrics. In The Neurobiology of Saccadic Eye Movements. Goldburg, M.E. & Wurtz, R.H., Eds. Elsevier Science Publishers.

Becker, W., (1991) Saccades, in *Eye Movements, Vol. 8 in Vision and visual Dysfunction*, R.H.S. Carpenter Ed. CRC Press, Boca Raton.

Becker, W., & Jurgens, R., (1992) Gaze saccades to visual targets: Does Head Movement Change the Metrics? Chapter 66 in *The Head-Neck Sensory Motor System*, Berthoz, A., Graf, W., & Vidal, P.P., Eds. Oxford University Press.

Biguer, B., Jeannerod, M, & Prablanc, C. (1982) The coordination of eye, head, and arm movements during reaching at a single visual target. Experimental Brain Research, 46, 301-304.

Biguer, B., Jeannerod, M, & Prablanc, C. (1985) The role of position of gaze in movement accuracy. In M.I. Posner & O.S. Marin (Eds.), Attention and Performance XI: Mechanisms of Attention. Erlbaum, Hillsdale, NJ

Brooks, R.A. (1991) Intelligence without reason. Computers and Thought, *IJCAI-91*.

Brooks, L.R. (1968) Spatial and verbal components in the act of recall. *Canadian Journal of Psychology 22,* 349-68.

Carpenter, R.H.S. (1988) *Movements of the Eye*. Pion Press, London.

Collewijn, H. (1985) Integration of Adaptive changes of the optokinetic reflex, pursuit and the vestibulo-ocular reflex, in Adaptive mechanisms in gaze control - Facts and Theories, Eds. Berthoz & Melville Jones, 1985 pp. 51-69.

Collewijn, H., Steinman, R.M., Erkelens, C.J., Pizlo, Z., & van der Steen, J. (1992) Effect of freeing the head on eye movement characteristics during three-dimensional shifts of gaze and tracking, Chapter 64 in *The Head-Neck Sensory Motor System*, Berthoz, A., Graf, W., & Vidal, P.P., Eds. Oxford University Press.

Feldman, J.A., (1985) Four frames suffice: A provisional model of vision and space. *Behavioral and Brain Sciences 8*, 265-289.

Frens, M. A. & Erkelens, C. J. (1991) Coordination of hand movements and saccades: evidence for a common and a separate pathway. *Experimental Brain Research 85*, 682-690

Gibson, J. J. (1966) *The Senses Considered as Perceptual Systems.* Houghton Mifflin, Boston

Guitton, D. (1992) Control of eye-head coordination during orienting gaze shifts, *TINS* 15(5): 174-179.

Hitch, G.J., and Baddeley, A., (1976) Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology, 28*, 603-21.

Irwin, D.E. (1991) Information integration across saccadic eye movements. *Cognitive Psychology*, 23, 420-456.

Irwin, D.E. (1992) Visual memory within and across fixations. In *Eye Movements and Visual Cognition; Scene Perception and reading*. Raynor, K. Ed. Springer-Verlag, New York.

Jeannerod, M. (1988). Mechanisms of visuomotor coordination: A study in normal and brain-damaged subjects. *Special Issue: Methods in neuropsychology*

Just, M.A., & Carpenter, P.A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*, 441-480.

Koken, P. W. & Erkelens, C. J. (1992). Influences of hand movements on eye movements in tracking tasks in man. *Experimental Brain Research 88*, 657-664.

Kowler, E. (1990) The role of visual and cognitive processes in the control of eye movement. In *Eye movements and their Role in Visual and Cognitive Processes*. Kowler, E. Ed. Elsevier Science Publishers.

Kowler, E. (1995). Attentional Interference at Close Spatial Separations. *Investigative Ophthalmology & Visual Science,* Suppl. 35(4) S901.

Kowler, E., & Anton, S. (1987). Reading twisted text: Implications for the role of saccades. *Vision Research, 27*, 45-60.

Kowler, E., Pizlo, Z., Zhu, G., Erkelens, C.J., Steinman, R.M., & Collewijn, H. (1992) Coordination of head and eyes during the performance of natural (and unnatural) visual tasks. Chapter 65 in *The Head-Neck Sensory Motor System, Berthoz,* A., Graf, W., & Vidal, P.P., Eds. Oxford University Press.

Land, M.F. (1992) Predictable eye-head coordination during driving. *Nature,* 359: 318-320.

Levine, M.D. (1985) *Vision in Man and Machine.* McGraw-Hill, Inc. New York.

Marr, D.C. (1982). *Vision*. W.H. Freeman and Company.

McConkie, G.W. & Rayner, K. (1976) Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In H Singer & R.B. Ruddell (Eds.) *Handbook of sensory physiology 7*(4) 331-379.

Miller, G. (1956) The magic number seven plus or minus two: Some limits on your capacity for processing information. *Psychological Review, 63*, 81-96.

Mishkin, M., Ungerleider, L.G., & Macko, K.A. (1983) Object vision and spatial vision: Two cortical pathways. *Trends-in-Neurosciences 6*(10) 414-417.

Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In C. Blakemore (Ed.), *Vision: Coding and Efficiency*. Cambridge Univ. Press, 411-422.

Noton, D., & Stark, L. (1971). Eye movements and visual perception. *Scientific American*, *224,* 34-43.

O'Regan, J.K. (1992). Solving the "real" mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology, 46,* 461-488.

O'Regan, J.K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research, 23,* 765-769.

Pashler, H., Carrier, M., & Hoffman, J. (1993) Saccadic Eye Movements and Dual-task Interference. *Quarterly Journal of Experimental Psychology, 46A*(1), 51-82.

Pentland, A. (1986) *From pixels to predicates*. Ablex Publishing. Norwood, NJ.

Robinson, D.A. (1981) Control of eye movements. In Brooks, V.B. (Ed.) *Handbook of Physiology,* American Physiological Society, Bathesda, MD. 1275-1320.

Rosenbaum, D.A. (1991) *Human Motor Control.* Academic Press, San Diego.

Skavenski, A.A., (1990) Eye movement and visual localization of objects in space, in *Eye movements and their role in visual and cognitive processes*, Kowler, E., (Ed.) Elsevier.

Swain, M.J., & Ballard, D.H. (1991). Color indexing. *Int'l. J. Computer Vision, 7,* (1), 11-32.

Swain, M.J., Kahn, R.E., & Ballard, D.H. (1992). Low resolution cues for guiding saccadic eye movements. *Proc., Computer Vision and Pattern Recognition Conf*., Urbana, IL.

Ullman, S. (1984). Visual routines. *Cognition, 18*, 97-157. (Also in S. Pinker (Ed.), *Visual Cognition*. Cambridge, MA: Bradford Books, 1984.)

van der Steen, J., (1992) Timing of coordinated head and eye movements during changes in the direction of gaze. Chapter 71 in *The Head-Neck Sensory Motor System*, Berthoz, A., Graf, W., & Vidal, P.P., Eds. Oxford University Press.

Viviani, P. (1990) Eye movements in visual search: cognitive, perceptual, and motor control aspects. in Kowler, E. (ed.) *Eye movements and their role in visual and cognitive processes. Reviews of Oculomotor Research V4,* Elsevier, 353-383.

Whitehead, S.D., & Ballard, D.H. (1990). Active perception and reinforcement learning. *Neural Computation*, *2*, (4), 409-419.

Wickens, C.D. (1986) The effects of control dynamics on performance. In Boff, K.R., Kaufman, L., & Thomas, J.T. V.B. (Eds.) *Handbook of Perception and Human Performance, Vol. 2: Cognitive processes and performance.* John Wiley & Sons, New York. 1-60.

Yarbus, A.F. (1967) *Eye Movements and Vision,* New York, Plenum Press.

8. Appendix

Calibration of the Magnetic Tracking System

The ASL relies on a magnetic head-tracker to monitor the position and orientation of the head. An Ascension Technology magnetic field tracker (Model 6DFOB, "Flock") was used to monitor the position and orientation of the head and the hand. The transmitter unit was mounted above and in front of the subject's head. The transmitter contains three orthogonal coils that are energized in turn. The receiver unit contains three orthogonal 'antennae' coils which detect the transmitters' signals. Position and orientation of the receiver are determined from the absolute and relative strengths of the transmitter/receiver coil pairs. The position of the sensor is reported as (x, y, z), and its orientation as azimuth, elevation, and roll angles.

The Ascension 6DFOB uses an auto-ranging system to increase the useful range while keeping noise levels down. The maximum distance at which a clear signal can be detected varies with the strength of the transmitted signal, but too strong a signal saturates the receiver. To allow measurements over a range of transmitter to receiver distances, the Ascension 6DFOB adjusts the transmitter's field strength based on the distance of the receiver from the transmitter. The maximum field strength is 1 Earth field. Position and orientation values are encoded as 16-bit integers. Distances (x, y, z) are scaled from -36" to 36", yielding a precision of 0.001" (72"/2^{16}), or 0.003 cm. Orientation (azimuth, elevation, and roll) are scaled from -180° to 180°, with a precision of 0.005° or 1/3 min arc.

The Ascension system has a range of temporal filter options that can be set with software commands. There are two classes of filters; 'AC' and 'DC.' The AC filters are band-block filters designed to filter out signals caused by environmental sources operating at around 60 Hz, (e.g., 120 VAC line supply, video monitors, and lighting equipment). There are three settings for the AC filters; i) 'AC filters off', ii) 'AC Narrow', and iii) 'AC Wide.' The two AC filter options (narrow and wide) differ in the width of the band of frequencies blocked by the filters. Removing frequency components in this range is not detrimental in itself because there is no appreciable component of head or hand movements beyond about 20 Hz [Rosenbaum 1991], but there is no way to implement such a filter in real-time without introducing a delay in the reported position and orientation values.

Unlike the 'AC' filters, the 'DC' filter does not filter out a fixed band of frequencies. Instead, it is an 'active filter' that monitors the sensor's reported values over a period, and adapts its time constant based on recent position/angle reports. If the sensor has shown little motion for a given period, the time constant is increased in an effort to reduce steady-state, or 'DC' position/orientation reports. When sensor movement is detected in this mode, it is at first suppressed by the long time constant on the presumption that it represents noise rather than real movement of the sensor. If the movement continues it is presumed to represent real motion of the sensor, and the time constant is reduced. The user can set upper and lower bounds on the time constant to limit the degree to which the 'active filter' can adapt to varying inputs, and the speed at which the filter adapts to sensor motion. The default filter configuration ('AC wide' and 'DC' filters on) produces very low noise output at the cost of increased temporal lag between sensor movement and position reporting. In addition, the actual lag introduced by of the 'DC' filter is not constant.

The "Flock" sensors used for head and hand tracking were characterized to determine the accuracy and noise in the measurement system with and without the default filters.

Position Calibration:

The first series of measurements was performed to measure the accuracy of the position signal of the magnetic tracker over a range of transmitter-to-receiver distances. Because the goal was to measure absolute distance errors (rather than noise characteristics), position data was collected with the default filter configuration ('AC wide' and 'DC' filters on) and the average of 100 samples is reported (see the next section for noise characteristics). Accuracy was measured in a 3-dimensional volume representing the task workspace by moving a test plane in depth through the space. Figure A. 1 shows the test plane used to calibrate the performance of the magnetic tracking system's position output. The 1.02 x 0.64 m array consisted of 54 target locations forming a 9 x 6 test grid, 12.8 cm on center.

Figure A. 2 shows the target volume consisting of 270 points. The plane was placed at five depth planes; -11 cm, 1 cm, 13 cm, 25 cm, and 37 cm (distances are reported with respect to the transmitter's center).

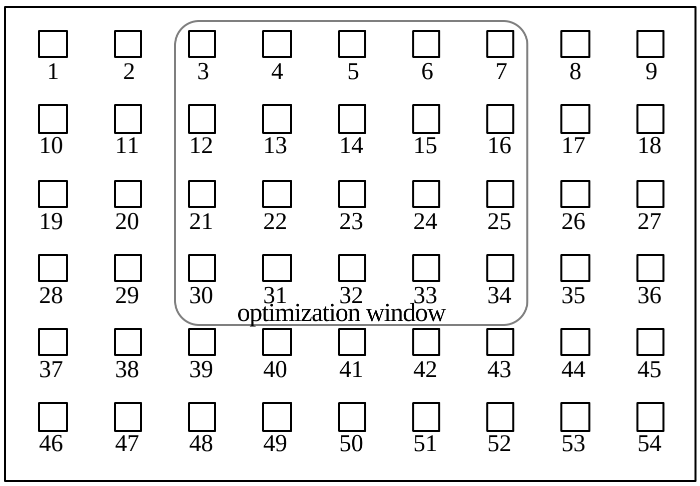


Figure A. 1 A 9 x 6 target grid was used in each of 5 depth planes to calibrate the Ascension 6DFOB magnetic tracking system.

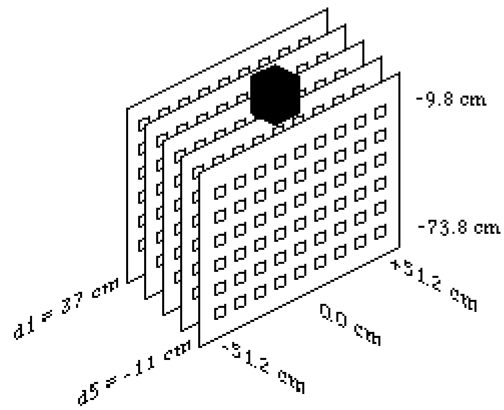


Figure A. 2 The 54 target test plane was positioned at five depths, giving a total of 270 test positions in the test volume.

The magnetic field transmitter was mounted above the target planes (9.8 cm above the top row of the target plane #4). Plane #4 was below the transmitter. The smallest distance between transmitter and receiver (9.8 cm) was position 5 on plane #4. Points 46 and 54 were the farthest on that plane at 89.8 cm. The maximum range on any plane was 97.6 cm (points 46 and 54 on plane #1, 37 cm behind the transmitter). The Ascension 6DFOB is scaled to +/-36" (91 cm). The head sensor was always within 25 cm of the transmitter, and in most cases the hand sensor was within 60 cm.

Vertical and horizontal position were measured at each of the 270 points in the volume. The mean of 100 samples was calculated at each point in the volume. Three runs of the full 3D data set were completed. Before beginning the measurements, the test plane was adjusted in an attempt to make it's axes collinear with the transmitter's. Because the goal was to determine the limits of the Ascension 6DFOB's performance, errors due to misalignment of the test plane with respect to the transmitter were eliminated by finding the best fit (allowing translation and rotation of the target points) within the 5x4 'optimization window' shown in Figure A. 1. Figure A. 3 shows a sample calibration plane with the mean measured positions ('mean_x') and the best fit after translation and rotation transformation overlaid on the target positions. Figure A. 4 shows the remaining root-squared error $[((x_{target} - x_{reported})^2 + (y_{target} - y_{reported})^2)^{0.5}]$ in the plane for three trials with the test plane at position 4 (1 cm behind the transmitter center). Note that the errors rise dramatically beyond 50 cm. Figure A. 5 shows the same data plotted at 10x vertical scale to better show the magnitude of the errors at small distances. The mean error below 50 cm is approximately 1.5 mm.

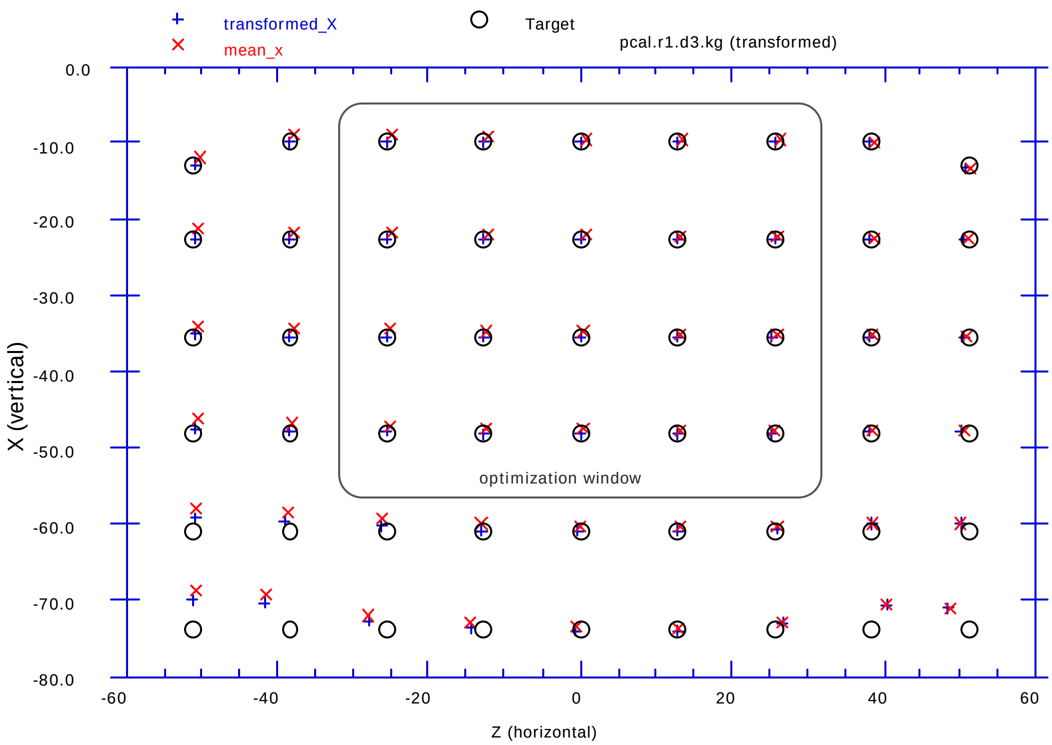


Figure A. 3 Mean position values (in cm) as reported by Flock ('mean-x'), and best-fit ('transformed_x') values after rotation and translation.

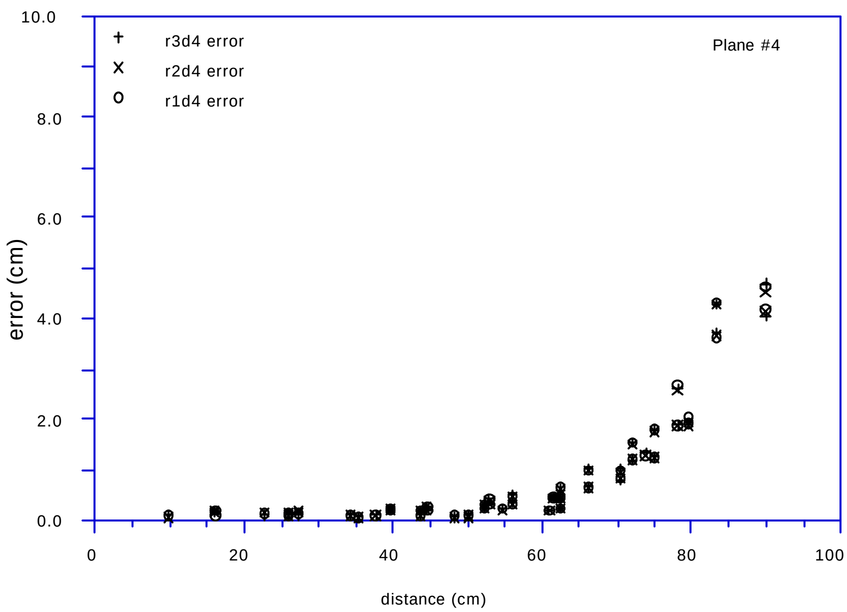


Figure A. 4 Root-square error as a function of distance from the transmitter for three trials at depth plane #4.

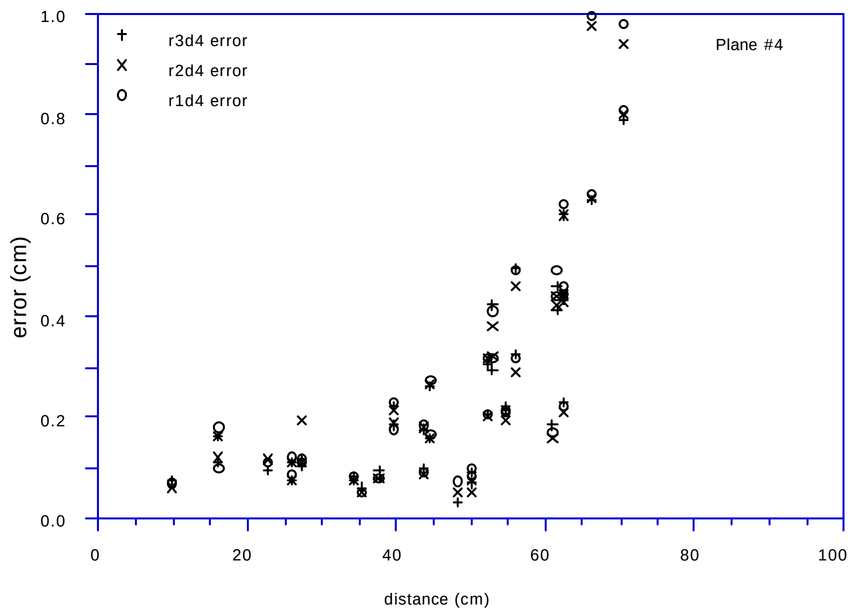


Figure A. 5 Data from Figure A. 4 plotted at 10X vertical scale to show errors at smaller distances.

Errors in the other planes were similar at small transmitter/ receiver distances, but increased with distance. Figure A. 6 shows the errors for plane #1, the most distant. The maximum error nearly doubled, from 5 cm to 9 cm. The performance below 50 cm is very similar to that shown in Figure A. 5.

Each data point in Figure A. 4 through Figure A. 6 represents the average of 100 samples collected with the default 'AC' and 'DC' filters on. These data give a clear view of the accuracy of the magnetic tracking system, but not about the noise in the signal. The top trace in Figure A. 7 shows raw data collected over a 1 cm range at a distance of 65 cm with the default filters on. The lower trace shows data collected with all filters disabled (the vertical offset in the plot was introduced to allow both data sets to be seen). The effect of the filters is clear, though there is still residual noise present with the default filers engaged. The range of the unfiltered data (max - min over 100 samples at a constant distance) in Figure A. 7 is approximately 1.0 cm. The standard deviation of the 100 samples is approximately 3 the range, as expected for a large sample from a normal distribution.

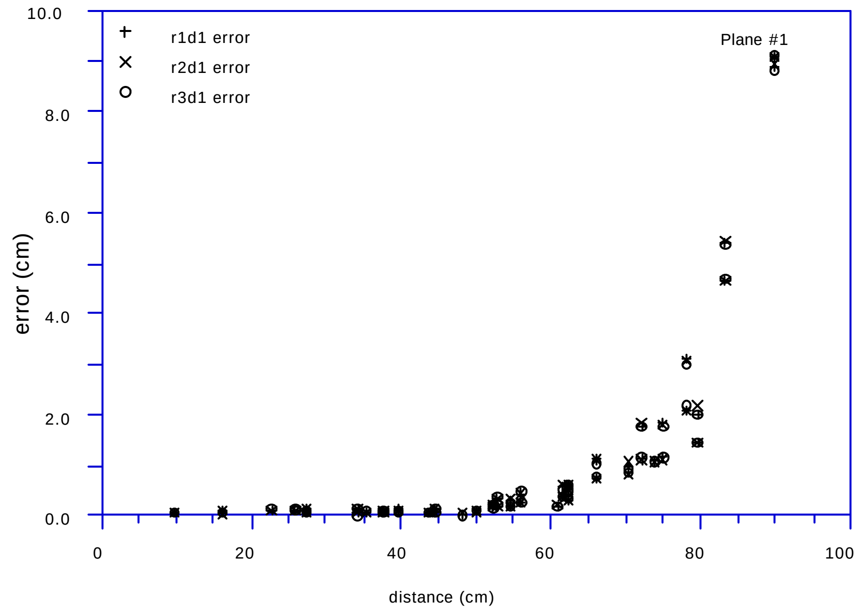


Figure A. 6 Root-square error as a function of distance from the transmitter for three trials at the most distant plane (#1).

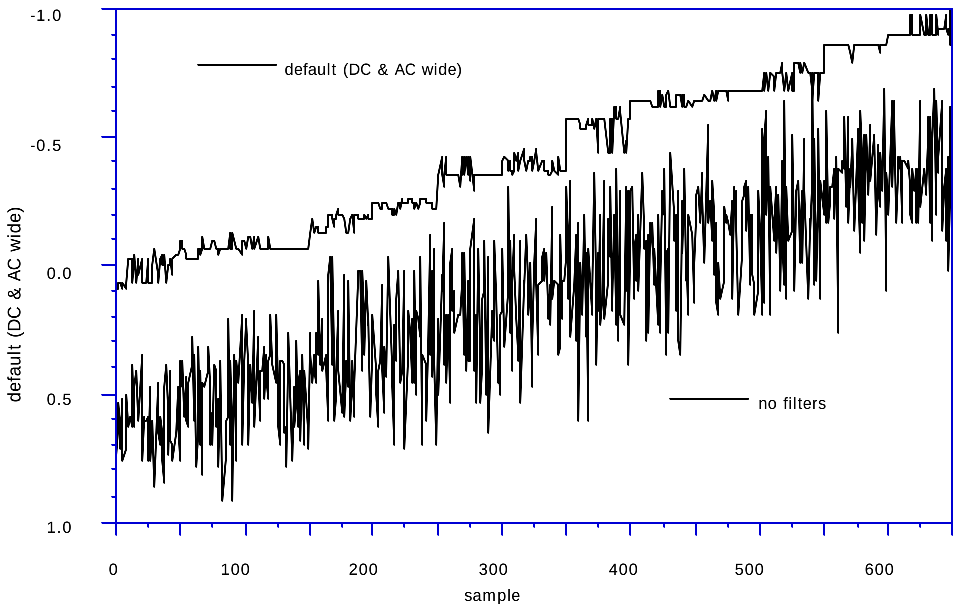


Figure A. 7 Position samples from the Ascension 6DFOB with default (top trace) and no filters.

Figure A. 8 shows how standard deviation and range values vary over transmitter to receiver distances of 10 to 85 cm. The local maxima evident in the graphs are due to the tracker's auto-ranging system, which adjusts the transmitter's power in discrete steps as the distance between transmitter and receiver changes.

The filters do not simply add a constant 'delay' to the signal; the adaptive nature of the 'DC' filter changes the effective time constant of the filter based on recent velocity measurements. If the sensor has been stable (or moving at very low velocity) for a period of time, the time constant is increased, on the assumption that any apparent movement is due to noise. After several samples of higher velocity, the time constant is reduced, on the assumption that apparent movements are real, and not due to noise. Figure A. 9 illustrates the effect. To examine the temporal delay introduced by the filters, two receivers were attached together, and moved rapidly from rest to the right. In the example shown, the receivers were moved approximately 60 cm, at a peak velocity of approximately 5 m/s. Figure A. 10 shows the best-fit cumulative Gaussians to the 'default filters' and 'no filters' data. The temporal difference at half height (i.e., the difference in the Gaussians' means) is 20 msec, and the difference at 10% of maximum is 38 msec.

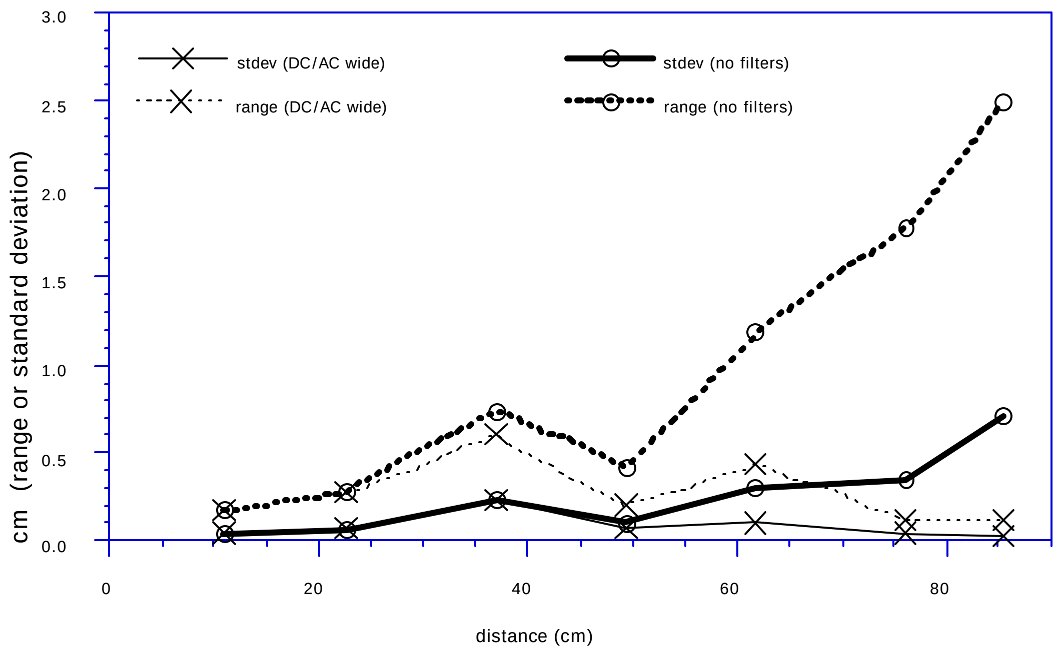


Figure A. 8 Standard deviation and range of sample data with transmitter to receiver distances of 10 to 85 cm. Local extremes are due to the Flock's autoranging system.

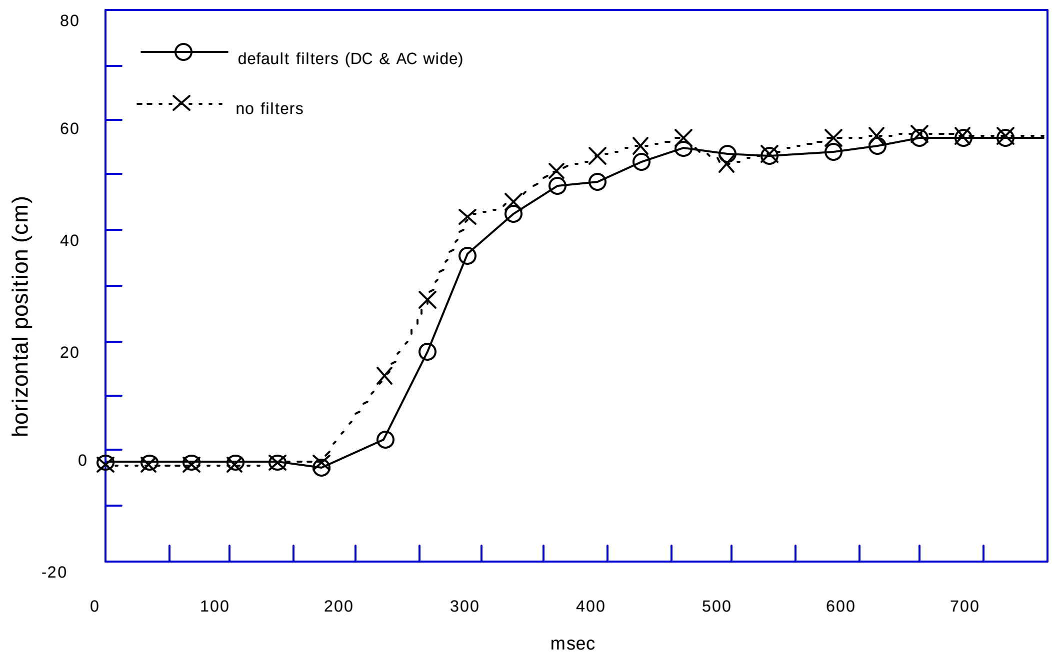


Figure A. 9 Variable delay caused by the adaptive 'DC' filter when the receiver is moved suddenly.

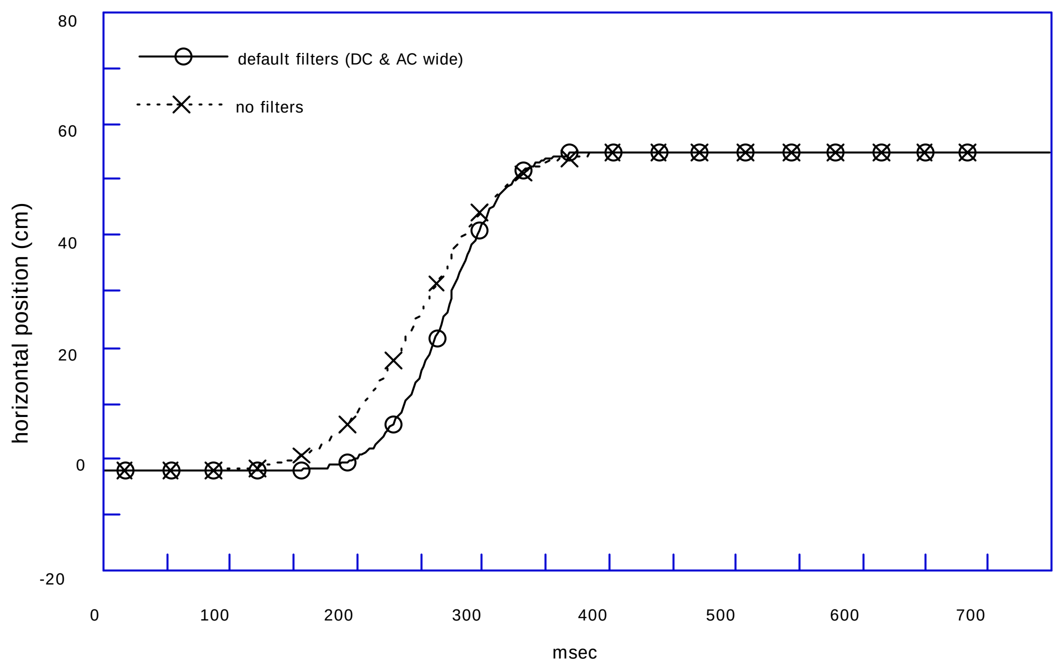


Figure A. 10 Best-fit cumulative Gaussian to data shown in Figure A. 9

Orientation Calibrations

The Ascension 6DFOB sensor measures position and orientation, so in addition to measuring the Flock's positional accuracy, it was necessary to determine the accuracy and resolution of the Flock's orientation measurements. The 6DFOB can output orientation in several formats. In this series of experiments, and for all calibrations, the system was configured to "point/angle" mode, in which the device reports a 12 byte record consisting of 3 2-byte integers representing (x, y, & z) position, and 3 2-byte integers representing (azimuth, elevation, and roll). A jig was constructed to allow the 6DFOB receiver to be precisely oriented along a single rotational axis (Figure A. 11). A vernier scale allowed the orientation to be set in 5 arc minute increments.

Figure A. 12 shows the absolute error (in minutes of arc) as the receiver was rotated through 180° in 5° steps at a distance of 40 cm. Figure A. 13Figure A. 14 show the mean error magnitude, and the noise characteristics at 40 and 65 cm. As was true for the position measurements, noise increased at greater transmitter to receiver distances, though enabling the filters reduced the noise level.

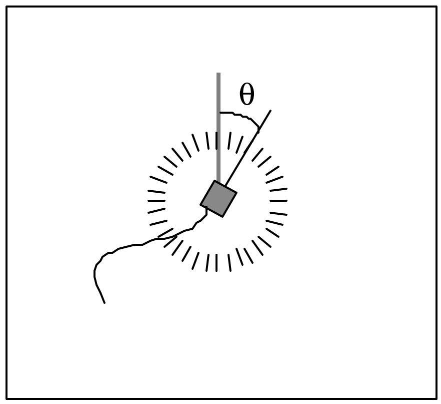


Figure A. 11 The Flock's orientation calibration was measured using a jig with a resolution of 5 minutes of arc.

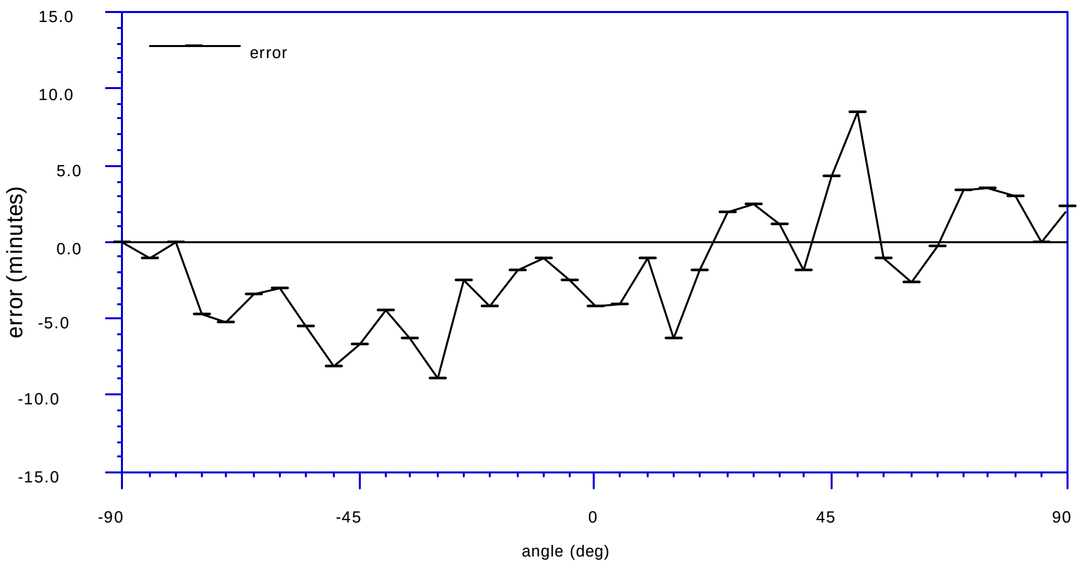


Figure A. 12 Absolute orientation error over 180° rotation with the receiver at a distance of 40 cm.

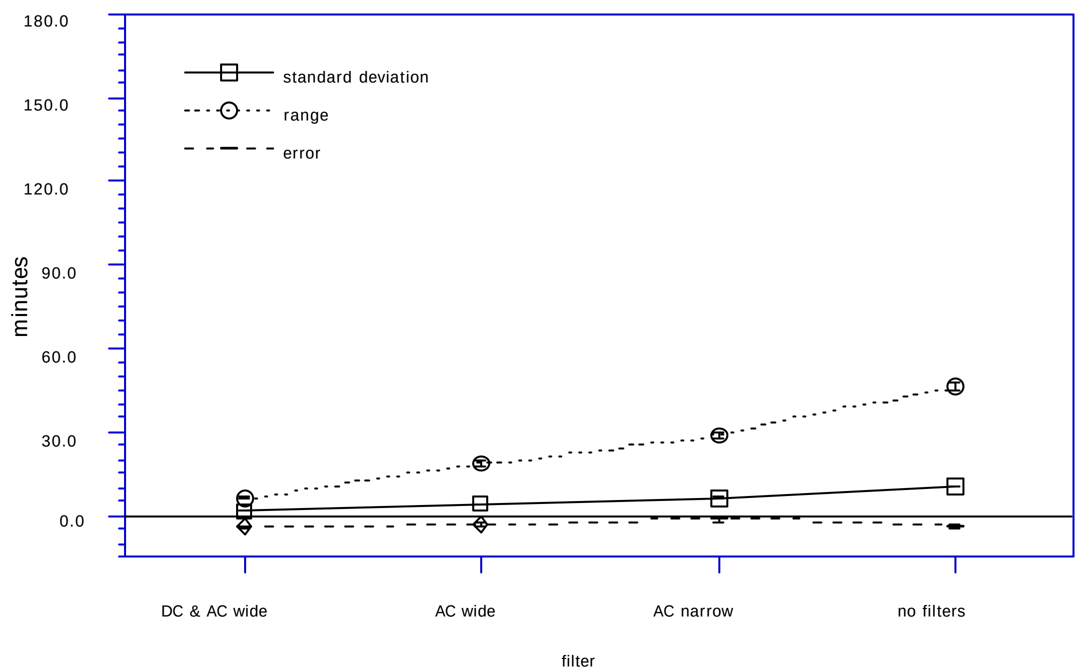


Figure A. 13 Standard deviation, range, and error for all filter settings at a distance of 40 cm from the transmitter.

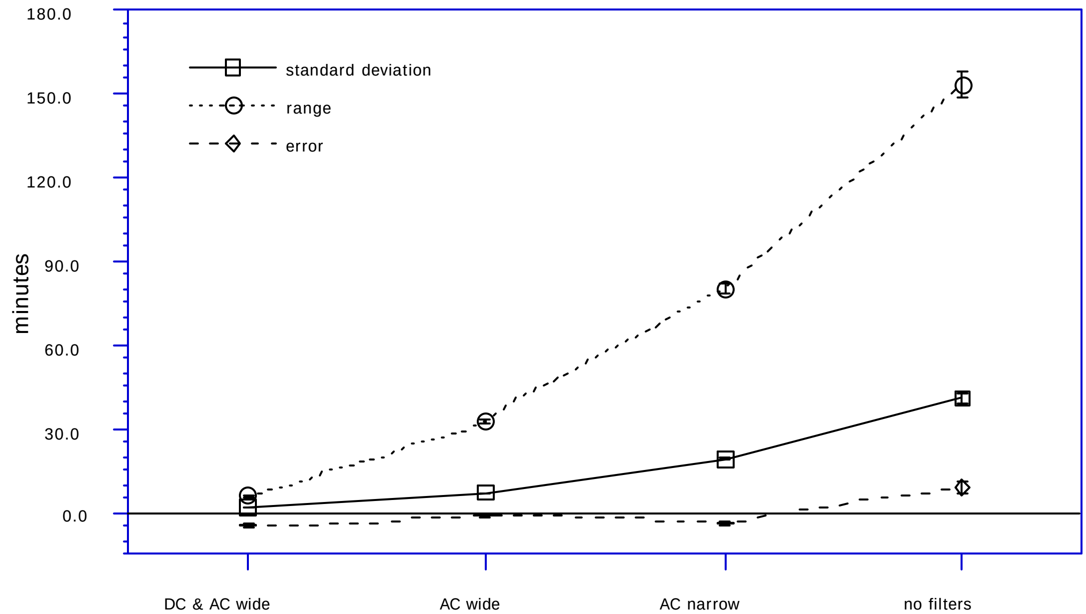


Figure A. 14 Standard deviation, range, and error for each filter setting, at a distance of 65 cm from the transmitter.