

Task Constraints in Visual Working Memory

Mary M. Hayhoe, David G. Bensinger, and Dana H. Ballard

Center for Visual Science

University of Rochester

Rochester, NY 14627, USA

Abstract

This paper examines the nature of visual representations that direct ongoing performance in sensorimotor tasks. Performance of such natural tasks requires relating visual information from different gaze positions. To explore this we used the technique of making task relevant display changes during saccadic eye movements. Subjects copied a pattern of colored blocks on a computer monitor, using the mouse to drag the blocks across the screen. Eye position was monitored using a dual-purkinje eye tracker, and the color of blocks in the pattern was changed at different points in task performance. When the target of the saccade changed color during the saccade, the duration of fixations on the model pattern increased, depending on the point in the task that the change was made. Thus different fixations on the same visual stimulus served a different purpose. The results also indicated that the visual information that is retained across successive fixations depends on moment by moment task demands. This is consistent with previous suggestions that visual representations are limited and task dependent. Changes in blocks in addition to the saccade target led to greater increases in fixation duration. This indicated that some global aspect of the pattern was retained across different fixations. Fixation durations revealed effects of the display changes that were not revealed in perceptual report. This can be understood by distinguishing between processes that operate at different levels of description and different time scales. Our conscious experience of the world may reflect events over a longer time scale than those underlying the substructure of the perceptuo-motor machinery [Ballard et al., 1996].

Visual function naturally occurs in the context of ongoing goal directed behavior, involving movements of the eyes, head, and hand. A central issue emerging from this is the nature of the brain's internal visual representations that direct this ongoing behavior. This paper examines aspects of visual processing that arise when vision is considered in its normal behavioral context. The experiments investigate the extent to which elemental visual representations are computed moment by moment explicitly for the immediate task, and the extent to which a more general purpose visual representation is maintained and updated as the observer moves within a scene.

Some of the consequences of head and body motion are dealt with by ocular mechanisms for stabilizing gaze with respect to the world. However, changes in gaze introduce problems in both the spatial and temporal domains: observers must maintain some kind of constancy of visual direction as a basis of their motoric interaction with the world and some kind of visual memory for the information in previous views. Thus one aspect of visual processing in its normal context concerns the way in which visual information from different gaze positions is related.

The primary class of eye movements involved in gaze changes are saccades. These movements structure the visual input into a sequence of retinotopic images sampling different parts of the scene, interposed by brief periods of blur. Although some kind of visual memory across these successive retinotopic images is clearly necessary, the form of this memory is poorly understood. There is a fairly general consensus that the memory is limited. Experiments by Irwin, examining sensitivity to changes in a random check pattern, reveal rather strict capacity limits on memory for such patterns across saccadic eye movements [Irwin, 1991]. He suggests that only information which has been the

focus of attention will be retained across saccades and that this has the capacity limits associated with short term visual memory. There are a number of compelling demonstrations in support of this, that show limited awareness of major changes in visual displays when the change is made between successive fixations [Grimes and McConkie, 1995, McConkie and Currie, 1996] or even during steady viewing of natural images [Rensink et al., 1996, O'Regan et al., 1996, Simons, 1996].

A series of studies by Rayner, Pollatsek, and colleagues have looked for some facilitation of perception (usually during reading) by presenting a stimulus in a peripheral location before the saccade. These studies suggest rather limited interactions between the peripheral, pre-saccadic view and the post-saccadic stimulus, for example [Pollatsek and Rayner, 1990, Pollatsek and Rayner, 1992, McConkie and Zola, 1979]. It is possible that only a sparse semantic or 'post categorical' description of the objects in a scene and their approximate locations is preserved, with other information being actively acquired by gaze changes [O'Regan and Lévy-Schoen, 1983, O'Regan, 1992, Irwin, 1991]. Nakayama has made an even more extreme suggestion, that perception is essentially serial and limited to the information which can be attended to in a single fixation [Nakayama, 1990]. Thus it seems unlikely that anything like a complete viewer-independent reconstruction of the visual scene is built up from successive gaze locations, as is often thought to be the job of vision [Marr, 1982].

One appealing possibility to explain the way vision works across a sequence of gaze positions would be to suppose that the ongoing visual goals determine both what is computed within a gaze position and across different gaze positions. This is a computationally efficient strategy, as demonstrated by recent theoretical work known as *Active Vision*. These models take advantage of

observer movements to model a range of tasks using limited, task-specific memory representations. Complex internal representations are avoided in these models by allowing frequent access to the sensory input during performance of the task [Brooks, 1986, Bajcsy, 1985, Ballard, 1991]. This allows construction of transient representations as needed for the current operation using information actively acquired from the world. This raises the possibility that human vision may be similarly efficient, taking advantage of successive gaze locations to serialize visual tasks and minimize the complexity of visual representations. It also means that it is important to examine this issue in the context of a well-defined task. This should indicate the nature of the ongoing visual computations, and what information is needed from prior gaze positions.

An additional motivation for examining this question in the context of an ongoing task is that it provides concrete suggestions as to how attention is being deployed from moment to moment. In many standard experimental paradigms it is not clear how attention is being distributed, and this makes it hard to generalize beyond the particular paradigm. This is a particularly intractable problem in experiments aimed at understanding how unattended information is processed. For example, in many experiments on the nature of visual processing of unattended information subjects covertly distribute their attention across the visual field while performing the primary task [Rock et al., 1992] because they are aware they will need to respond to occasional 'unexpected' stimuli.

Experimental Approach

Our previous work [Ballard et al., 1995] provides support for minimal, task dependent visual representations. We developed a paradigm reflecting basic sensory, cognitive, and motor operations involved in a wide range of human performance. The task is constrained enough to allow fairly precise identification of the current cognitive state at each point in the task. The task was to copy a pattern of colored blocks, as shown in Figure 1. The model pattern is at the top left, the workspace for building the copy directly below it, and blocks for use in making the copy are in the resource area on the right. Subjects move the blocks with the mouse.¹

Examination of performance in this task revealed that subjects use frequent eye movements to the model pattern to acquire information about the pattern just as it is needed, in preference to using visual memory, even for very short intervals. Subjects frequently looked at a block in the model pattern twice in the process of copying it, first presumably to acquire the color of the block, and then after picking up a block of that color they returned to fixate the pattern again. This second fixation is presumably to find the relative location information necessary for positioning the block in the copy. It appeared that even the color and location of a single block were acquired in separate fixations, rather than being automatically bound together as object properties during a single fixation. Thus the easily observable aspects of performance in this task pointed to extremely scant visual representations and minimal visual information carried over between views.

¹The eye and hand traces shown in Figure 1 are not seen by the subject, just recorded for later analysis.

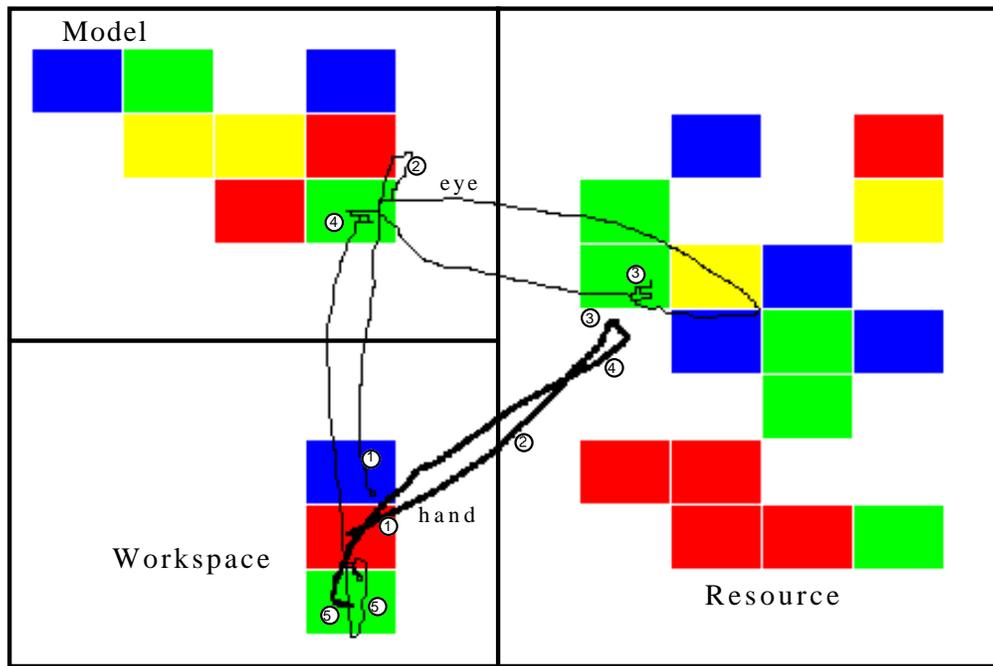


Figure 1: Copying a single block within the task. The eye position trace is shown by the dotted line. The cursor (hand) trace is shown by the dark line. The numbers indicate corresponding times in the task.

However, it is necessary to test this claim more carefully, since the frequent fixations on the model may not be obligatory, but rather, reflect the fact that the eye must wait on the slower hand movements to catch up. The approach we use here is to change various aspects of the display during saccadic eye movements and observe if there are any consequences on task performance. Fixation duration, for example, should be a sensitive measure of an effect of altering information critical for performance at that moment. If the display change leads to longer fixations this would suggest that the current visual representation is no longer consistent with some aspect of the recent visual history. By making a range of changes of different kinds of visual information we can find out what information is actually being used in the task, whether it is preserved across different fixations, and also whether information that is not obviously or immediately relevant is in fact being represented. This is not necessarily the same as what the subject is aware of, which may reveal quite different estimates of residual visual information. By exploring the types of change that affect performance, we can identify the information that is retained.

Experiment 1: Single Block Changes

One of the observations we have previously made about performance of the block copying task is that subjects behave in a very stereotypical way [Ballard et al., 1995]. Since only one block can be moved at a time, we break down performance into a description of single block moves. These can be categorized into the sequences diagrammed in Figure 2a. The sequence on the top left, labeled

MPMD or "Model-Pickup-Model-Drop" is the most common. In this case, following placement of the previous block, subjects fixated the model pattern to begin copying the next block. This is followed by a saccade to the resource area to guide the hand movement for pickup, and then by a saccade back to the model before fixating the workspace to guide block placement. Thus the model is fixated twice while copying the block. In the MPD (model-pickup-drop) and PMD (pickup-model-drop), the next two strategies, the subject makes only one model reference, either before or after pickup. In the PD (pickup-drop) strategy the subject copies a block without reference to the model. In this case the color and placement must come entirely from visual memory. Only about 16 percent of the block moves used the pickup-drop strategy. The other strategies involve either one or two fixations on the pattern. Thus subjects showed a very strong preference for using eye movements to acquire information from the display as needed, rather than using visual memory. (A variety of other aspects of performance have been investigated in our previous work, but are not relevant for the current investigation [Ballard et al., 1992, Ballard et al., 1995, Pelz, 1995].

The first experiment to probe the extent of the memory across different fixations examines the consequences of changing the color of blocks in the model area during saccades from the Workspace to the Model. The cyclic and stereotyped nature of performance in this task makes it a convenient starting point for investigating natural behavior. The subject's behavior is quite predictable, so this allows well controlled manipulations. Since fixations in the Workspace area are almost invariably for the purpose of guiding the placement of a block in the partially completed copy, if the next saccade is to the model area, then the implication is that the subject is fixating the model to acquire

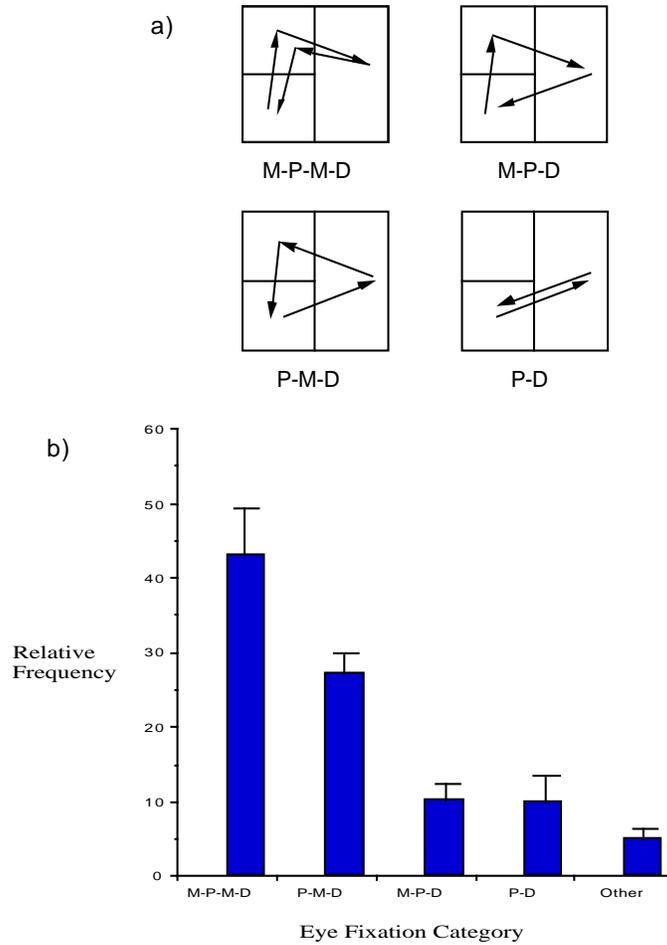


Figure 2: (a) Schematic for eye movement patterns for copying a block. “M” means that the eyes are directed to the model; “P” and “D” mean that the eyes and cursor are coincident at the pickup point and drop-off point, respectively. Thus for the PMD strategy, the eye goes directly to the resource for pickup, then to the model area, and then to the resource for drop-off. (b) The relative frequency of the different strategies for seven subjects.

color or location information. In this experiment we changed the color of one or more blocks in the unworked part of the model during the saccade to the model, and observed the consequences of this change on task performance. This should reveal whether color information in the model is preserved from either the immediately preceding peripheral stimulus or from previous fixations on the model pattern.

A second condition was also examined, where the color change was made following block pickup, during the return saccade to the model area from the resource area. In this case, the subject is targeting the block fixated previously, and a block of the same color has been picked up, so we expect that changing the color of the targeted block would interfere with performance. Thus we can compare the effects of the manipulation for the same fixation position at a different stage in task performance.

Methods

Apparatus

The experiments were performed using a 14" Macintosh color monitor viewed at about 0.8m. Thus the display subtended about 16 deg of visual angle. Individual blocks subtended about 1.5 deg horizontally and 1 deg vertically. Blocks were shifted using the mouse. Eye position was monitored by a Dual Purkinje Image Generation V eye tracker, which provides horizontal and vertical eye position signals sampled every msec, with an accuracy of 10-15 min arc over about 15 deg range. The

experiments were controlled by a MacintoshIIIfX computer. Saccades were detected and the display updated within the 17 msec limit set by the refresh rate of the monitor. This was accomplished by using 1 msec interrupts generated by a National Instruments analog data acquisition board. Display updates were performed seamlessly through video look up table changes which occurred during interrupt handling. All events were timed with an accuracy of 1 msec. The saccades in this experiment typically lasted about 50 msec, and all changes occurred before the end of the saccade, with 88 per cent occurring at least 20 msec before the end of the saccade. This was determined by measuring the display change with a fast photodetector and comparing photodetector output with the eye position signal from the tracker [Karn, 1995].

Procedure

Ss copied the block pattern as described above, and were asked only to complete the task as quickly and accurately as possible. No other instructions were given, so as not to bias Ss towards particular strategies. (Subjects generally finish the task in about 15 sec, and make only occasional errors.) A display change was triggered by a transition from either the workspace area or the resource area to the model pattern. One of the uncopied blocks adjoining the block just placed was randomly chosen and its color changed to one of the 3 remaining colors. On some trials this corresponded to the next block copied. The experiment is cartooned in Figure 3a. The arrow shows the saccade, the zig-zag shows when the change is made. In another condition shown in Figure 3b, the change was made after the S had picked up a block and was returning to the model, presumably to check

its location. In both conditions the changed block was chosen randomly from the unworked blocks. Trials were subsequently sorted into those where the change was made to the next block copied, or to the adjoining block, diagonally adjacent, or remote. The probability that a block will be changed on any given saccade to the model was between 0.18 and 0.26 for the different subjects. (This varied between subjects because different subjects had slightly different distributions of the strategies shown in Figure 2.) Patterns where no changes were made were randomly interleaved with those with saccade-contingent display changes. Since a change occurs only with a probability of about 0.24 in the experimental patterns, in order to get enough trials where the change is made to the targeted block, between 183 and 280 patterns consisting of 8 blocks each were presented. This gave from 308 to 563 individual color changes. The relatively low frequency of changes was designed to avoid a situation where Ss consciously adopt a different distribution of strategies for the experimental trials.

Results

For preliminary analysis we developed software to replay a segment of any length of a trial and retrace the subject's eye and hand movements (see Figure 1). A graphical timeline window displays the events that occurred during a trial. Selection of a point on the timeline displays the screen configuration of the trial starting at that moment in time. It is then possible to scroll up and down to select any segment of the data stream. The program then sets up the blocks and replays whatever eye and hand movements the subject made during the selected period and shows any color

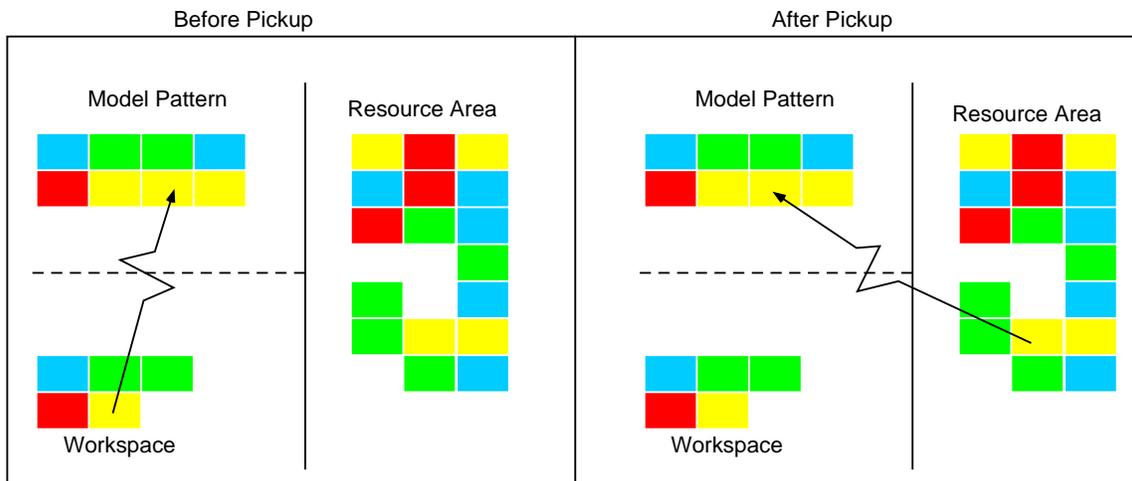


Figure 3: Cartoon of the two conditions in Experiment 1. The jagged arrow represents a saccade to the model area. When the saccade crosses the dashed line a block in the model pattern changes color. In (a) the color of an uncopied model block is changed during a saccade from the workspace to the model, before a pickup. In (b) the color of an uncopied model block is changed during the return saccade from the resource to the model, after the subject has picked up a block.

changes that occurred. Figure 1 shows what a typical selection might display. At the beginning of the selection the subject has already placed two blocks in the workspace area and is in the process of placing the third. The thin trace represents the eye movements, the thick trace the hand movement. To identify the saccades and fixations, the raw eye velocity data was filtered using a 17 point velocity filter, and a criterion of 35 deg per sec set for identifying a saccade. It was also necessary to set a minimum duration of 30 msec for defining saccade termination, to avoid erroneously labeling the overshoots in the DPI tracker as a separate fixation. These overshoots result from inertial motion of the lens following a high velocity movement and are discussed in [Duebel and Bridgeman, 1995].

Our primary focus was on the duration of the fixations in the model area following a color change, since changing information critical for task performance should be disruptive and lead to longer fixations. Following placement of a block in the workspace it seems plausible that the next fixation in the model area is for the purpose of acquiring the color of the next block to be moved. If subjects really have no record of the color of the blocks in the model from previous model fixations or from the peripheral pre-saccadic view, then changing the colors of blocks in the model pattern should have no effect. Similarly, if the color of the block currently being carried is changed during the return saccade, we expect longer fixations as a consequence of task interference.

Data for this experiment is shown in Figure 4 for the individual subjects. Each plot shows the total time the subject spent fixating in the model area in several cases. That is, if the subject made two different fixations in the model area these were summed. This was done because interference with the task might reveal itself as either a fixation on another block or longer fixations on the same block. On the right of each plot is the summed fixation duration for the control trials, where no color changes occurred. On the left is the average fixation duration on trials when the changed block was the one the subject was about to copy. Other categories are when the changed block was adjoining the current block, diagonally adjacent, or remote. This allows us to compare the effects of changes in the attended areas with more remote locations. The lower line is for trials when the change was made during the workspace to model transition, following placement of a block in workspace, and before visiting the resource area to pick up the next block. The upper line shows data for trials when the change was made following a pickup in the resource area, before placement

in the copy.

Figure 5 shows the average across the five subjects in Figure 4. When the changed block was the target of the saccade, on average there was a small (43 msec) increase in fixation duration for changes preceding pickup (Before Pickup condition). All subjects showed an increase of about this magnitude, although it was not statistically significant for any of the individual subjects. There was no reliable effect when the changed block was adjacent to the targeted block, although again, all subjects showed an increase. Thus it appears that task performance is little affected by changing the color of the target block during the saccade, at least to a first approximation. This supports the speculation that the color of the next block is acquired during this fixation.

In the *After Pickup* condition, when the change was made at a different stage of task performance, when subjects were carrying a block, changing the color of the saccade target produces an increase in fixation duration of 104 msec, on the average. Again, all subjects showed an increase that was statistically significant for four of the five subjects. Comparing the Before and After Pickup conditions suggests that a color change may be more disruptive after pickup, although the interaction term is not statistically reliable. A priori, such a difference might be expected, since the color of the current block must be held in memory at this stage of the task, in order to get information about block position. When one of the neighboring blocks is changed, there is an increase in fixation duration. None of these increases was statistically significant.

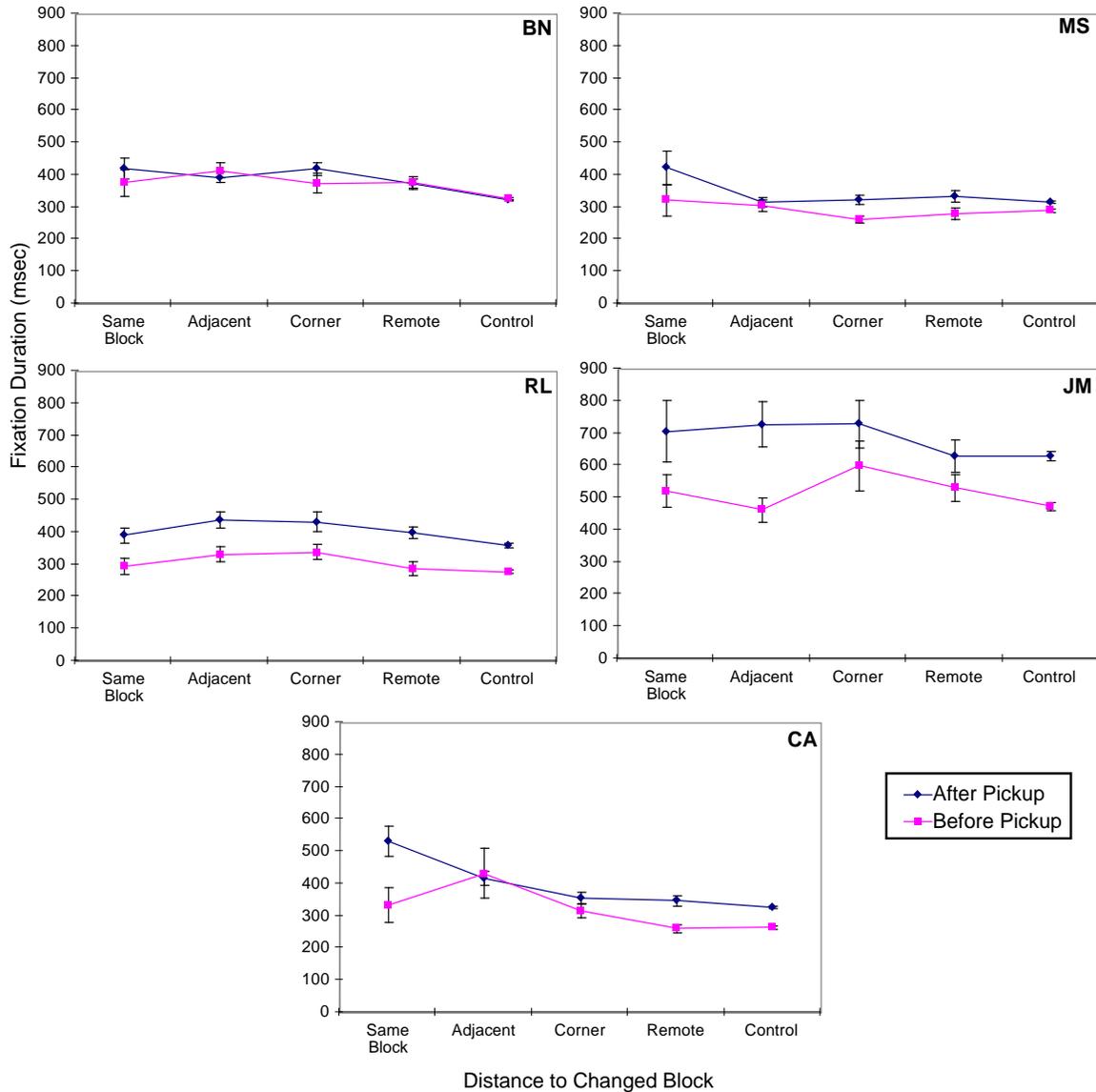


Figure 4: Summed fixation time in the Model area for the Before Pickup and After Pickup conditions for each of five subjects. In each plot the leftmost data points show the fixation time when the saccade target was the block currently worked on. The other data points show the cases when the changed block was adjacent or remote from the current block. The rightmost points are the control. The error bars represent +/- one standard error of the mean difference between experimental and control conditions. Thus the SEMs around the control conditions are always zero. The data points for the control conditions are based on approximately 1000 trials; for the Same Block conditions, between 15 and 45 trials; for the other conditions the number of trials ranged from 40 to 130.

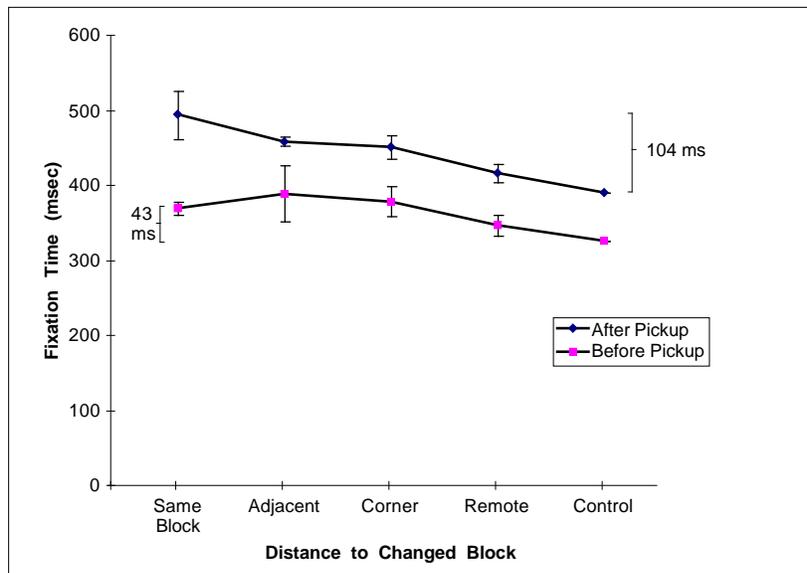


Figure 5: Summed fixation time in the Model area for the *Before Pickup* and *After Pickup* conditions, averaged over the five subjects. The error bars represent +/- one standard error of the mean difference between experimental and control conditions across subjects.

Experiment 2: Multiple Changes

Changing the color of a single 1.5 deg block in the context of a random pattern of colored blocks is from some perspectives a rather small change. In order to get a better idea of the representation of the visual information maintained over a period of a few seconds across different fixations, we need to explore a wider range of changes. The next step was to change several blocks in the model instead of just a single block. One extreme possibility is that only the immediately task relevant information is computed. On this view, changes in the any of the uncopied blocks should have no effect on task performance, and this should give results similar to those in the first experiment. This is not likely to be strictly true, because in the normal performance of the task subjects copy some proportion of the blocks without reference to the model pattern. However, it is quite possibly the case on the trials where subjects make two fixations on the model pattern. The next experiment explores the consequences of these larger changes.

The experiment was repeated, but now the colors of all unworked blocks were changed on a proportion of the saccades into the model area, either following block placement in the workspace, or following block pickup in the resource, as in the previous experiment. When the change occurred near the beginning of the copying task, as many as seven blocks were changed. When the change was made near the end of the task, a smaller number of blocks changed. A change was made on about six to ten percent of transitions into the model area. All other aspects of the experiment were the same as the first one.

Results

The data are plotted in Figure 6 for the individual subjects. The lower line shows fixation durations in the model area following placement of the previous block and before pickup of the next block. The upper line shows the data when the change was made following pickup. In this experiment, since all the unworked blocks change, the block the subject is targeting in the model area will always change color, along with a variable number of other blocks. All trials where a change occurred were combined in the figure and compared with the control trials.

Figure 7 shows the data averaged across subjects. Color changes before a pickup give a statistically reliable increase in summed fixation duration of 127 msec averaged across four subjects. Changes made following pickup have a more profound effect. Summed fixations in the model are increased by 275 msec, averaged over subjects. This increase is significantly greater in the After Pickup condition than Before Pickup. This provides further evidence that subjects retain the color of the current block at this stage of the task, since it is necessary for getting position information. Thus it appears that the color information from the pre-saccadic peripheral view or, more plausibly, from the prior model fixation is retained when it is currently task-relevant.

It appears that there is a greater disruption of task performance when several blocks are changed, than in the first experiment where only the target block changes (127 vs. 43 msec before pickup, and 275 vs. 104 msec after pickup). This is true for each of the three of the subjects who performed both experiments. Changing the color of the uncopied blocks is not obviously relevant to the ongoing

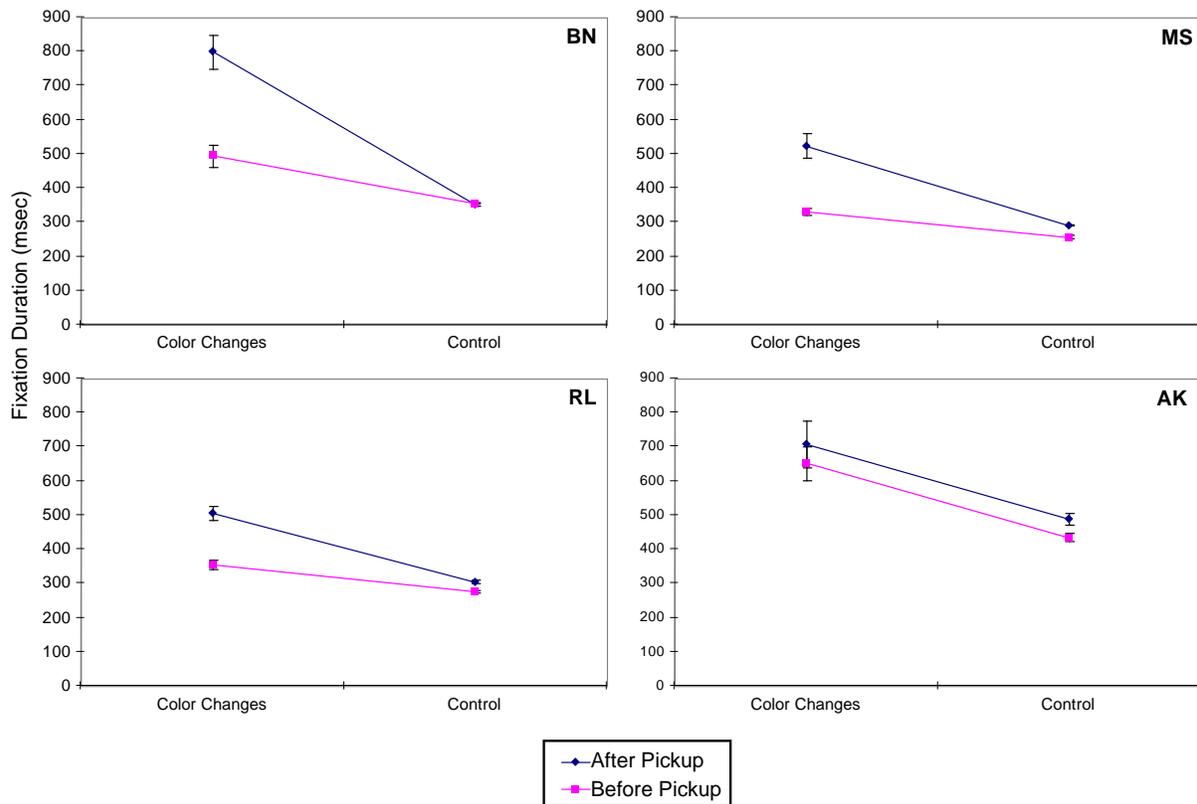


Figure 6: Summed fixation time in the Model area for the *Before Pickup* and *After Pickup* conditions for the individual subjects in Experiment 2. Error bars represent +/- one standard error of the mean difference between experimental and control conditions. Data points for the control trials are based on approximately 1000 measurements; for the color change trials, the number of measurements ranged from 30 to 170.

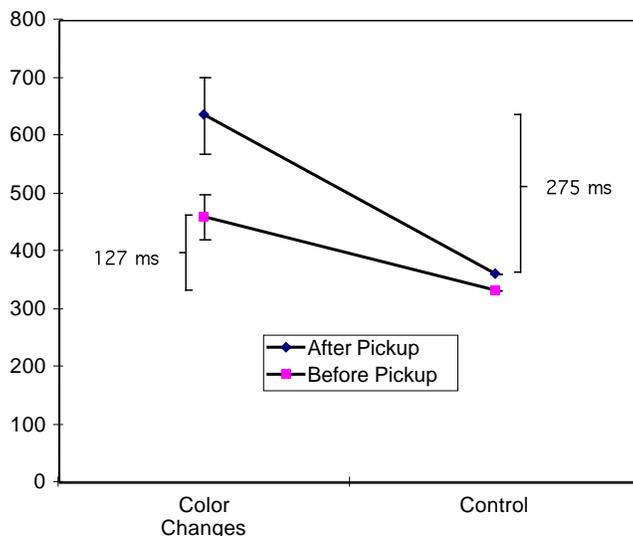


Figure 7: Summed fixation time in the Model area for the *Before Pickup* and *After Pickup* conditions in Experiment 2. The data are averaged across the four subjects. Error bars represent +/- one standard error of the mean difference between experimental and control conditions, across subjects.

computation of acquiring color of the fixated block. Thus some global aspect of the model pattern must be preserved across fixations, since changing irrelevant blocks interferes with performance.

As mentioned above, the data plotted in Figure 5 and Figure 7 show the total time spent fixating in the model area, and does not differentiate between individual fixations. Thus the added time might be accounted for by another fixation or by longer individual fixations. To investigate this question, we separated the data where only one fixation was made from that where two (or more) fixations were made. When the block the observer has just picked up changes color, the observer must either pick up a block of the new color at that location, or alternatively, find a model block of the appropriate color in a different position. Subjects almost never picked up a new block, but instead found a new place to put the block they were holding. This was frequently accompanied

by an additional fixation.

Figure 8 shows histograms of fixation durations for subject RL in the second experiment. The four distributions on the left show data from the Before Pickup condition, the four on the right, the After Pickup condition. In both conditions the upper two graphs are for the trials in which subjects made only one fixation in the model area. The top graph is for the fixations that occurred after color changes, the one below it for control fixations in which no colors change. The bottom two graphs correspond to trials where the subject made more than one fixation in the model area. In these graphs fixation duration refers to sum of the consecutive fixations in the model area. The distributions for the experimental conditions are biased to longer durations even when only one fixation was made. Thus individual fixations are longer following a change. Other subjects' data showed a similar pattern.

This increase in the duration of individual fixations does not account for the entire effect, however, as shown in Table 1 and Table 2. The Tables show the proportion of visits to the model area that consisted of more than one fixation. Table 1 is for Experiment 1, and Table 2 is for Experiment 2. Data for the Before Pickup condition are on the left, and for the After Pickup condition on the right. In the After Pickup condition, the proportion of double fixations increased in the experimental conditions. Thus the color change increases the number of fixations within the model area. In the first experiment the increase is more modest and is not observed at all in the Before Pickup condition. In the second experiment the increase is greater, particularly in the After Pickup condition. Thus the increases in time spent in the model area in Figure 5 and Figure 7

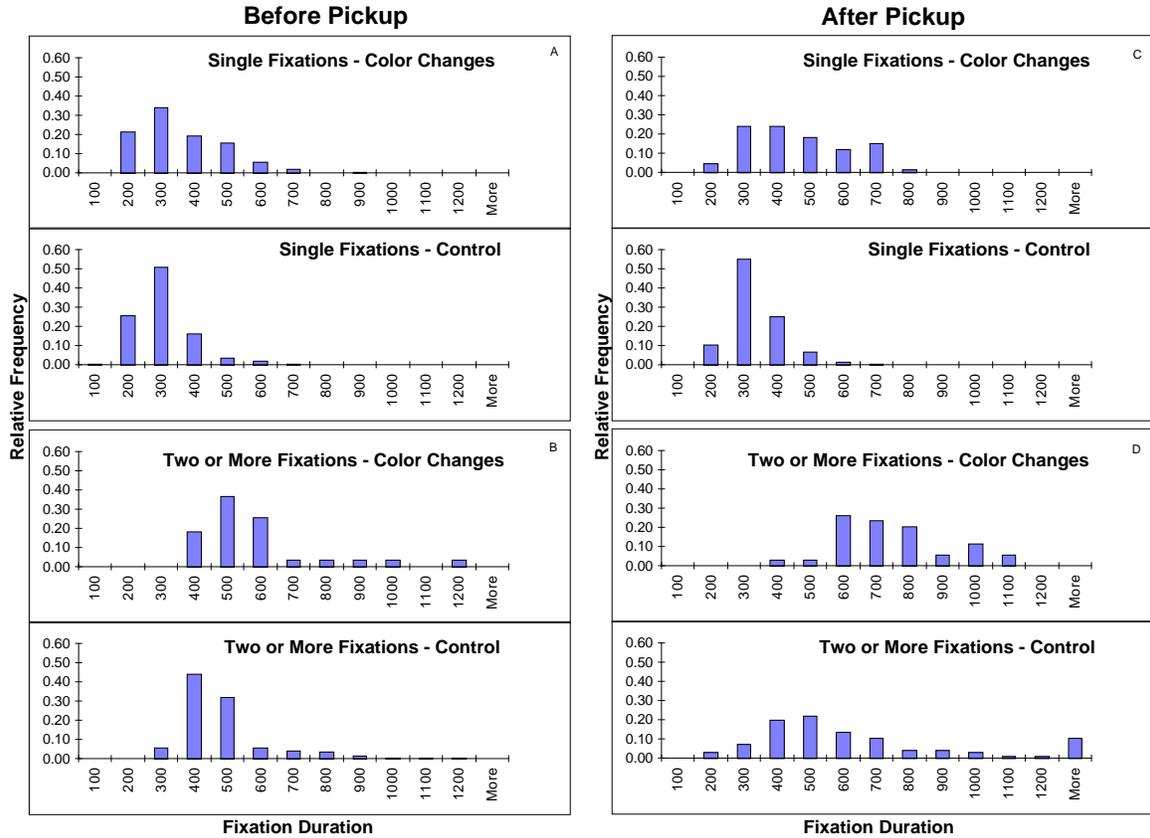


Figure 8: Histograms of fixation durations for subject RL. Histograms A and B are for model fixations Before Pickup. C and D are for fixations After Pickup. In A and C, the distributions of fixation durations are for those instances in which a single fixation was made in the model area. In B and D, the distributions are for the remaining instances, where two or more fixations were made during a visit to the model area. Within each box the top distribution is for the experimental trials, and the bottom one is for the control trials. Fixation durations are in msec, and the frequencies scaled relative to the total number of trials per histogram.

Subject	Before Pickup		After Pickup	
	Changes	Control	Changes	Control
1.	0.14	0.17	0.26	0.18
2.	0.36	0.40	0.72	0.50
3.	0.12	0.12	0.08	0.13
4.	0.38	0.25	0.26	0.12
5.	0.46	0.32	0.59	0.52

Table 1: Proportion of double looks for five subjects in Experiment 1.

Subject	Before Pickup		After Pickup	
	Changes	Control	Changes	Control
1.	0.11	0.09	0.26	0.07
2.	0.66	0.47	0.71	0.34
3.	0.17	0.12	0.34	0.07
4.	0.35	0.38	0.64	0.22

Table 2: Proportion of double looks for four subjects in Experiment 2.

come partly from longer individual fixations and partly from additional fixations.

Strategies

Another way of evaluating the effect of the display changes is at a coarser level of description of task performance. As described above (see Figure 2), individual block moves can be categorized into four different strategies: Model-Pickup-Model-Drop, Model-Pickup-Drop, Pickup-Model-Drop, and Pickup-Drop. The Pickup-Drop strategy clearly reveals that some memory from previous looks at the model is used in completing the task. We can collapse these strategies into three: Two or More

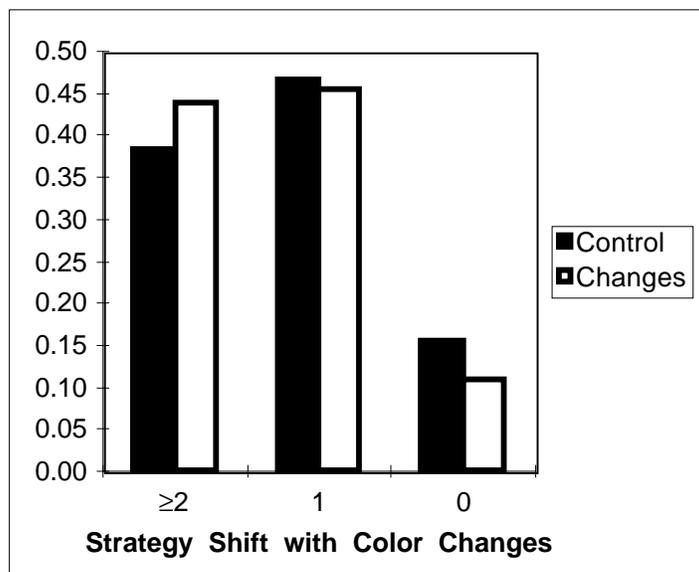


Figure 9: The relative frequency of the different strategies involving either zero (Pickup-Drop), one (Pickup-Model-Drop or Model-Pickup-Drop), or two or more (Model-Pickup-Model-Drop) looks to the model area, averaged over subjects. Black bars represent strategies used in experimental conditions, white bars show strategies used in the control conditions.

Looks, One Look, No Looks, depending on how many times the S inspects the model pattern in the process of copying a block. Figure 9 shows the difference between experimental and control conditions. There are small but consistent changes in strategy as a result of the manipulation. All subjects reduce the number of Zero Looks slightly and increase Two or More Look strategies. These goes up on the average from 38 to 44 percent of the time, $t(8) = 5.27$, $p < .01$, in the model area, and there is a drop in Zero Look strategies from 15 percent to 11 percent, $t(8) = -3.26$, $p < .01$. These changes in the frequency of model inspections are small but are statistically reliable and presumably reflect the fact that subjects normally depend to some extent, on memory for model items from previous fixations.

Perceptual Awareness

Except for experiments in a reading paradigm, it is more common to ask the subject what changes can be detected during a saccade (eg how big a displacement is detectable), rather than examining the consequences of a change in an ongoing task. Thus attention is being deployed in a different manner. We were concerned in this experiment that the subject's primary task be copying the blocks rather than detecting changes, so they were not informed that changes were being made. At the end of the first experiment, subjects were asked if they noticed any changes. Based on their verbal report, it appeared that subjects noticed only occasional changes. This is remarkable given that changes after pickup were often associated with a second fixation onto a neighboring block. Thus conscious reports of the experimental manipulations appear to be a less sensitive measure than fixation durations or strategy effects. In the second experiment subjects were more aware of the display changes, but still greatly underestimated their extent. In this experiment, after observers MS and RL became aware that changes were occurring, they were asked to report how many blocks changed color after completing a pattern. These responses are plotted in Figure 10, which shows the frequency distribution for the number of blocks the subject reported as changing. Also plotted on this graph is the frequency distribution of the number of blocks which actually changed. Note that in most cases the subject typically saw only one block change, although the modal change was seven blocks. Subject RL showed a similar pattern, and usually reporting seeing two of six actual changes. This seems particularly significant given that the subject is presumably

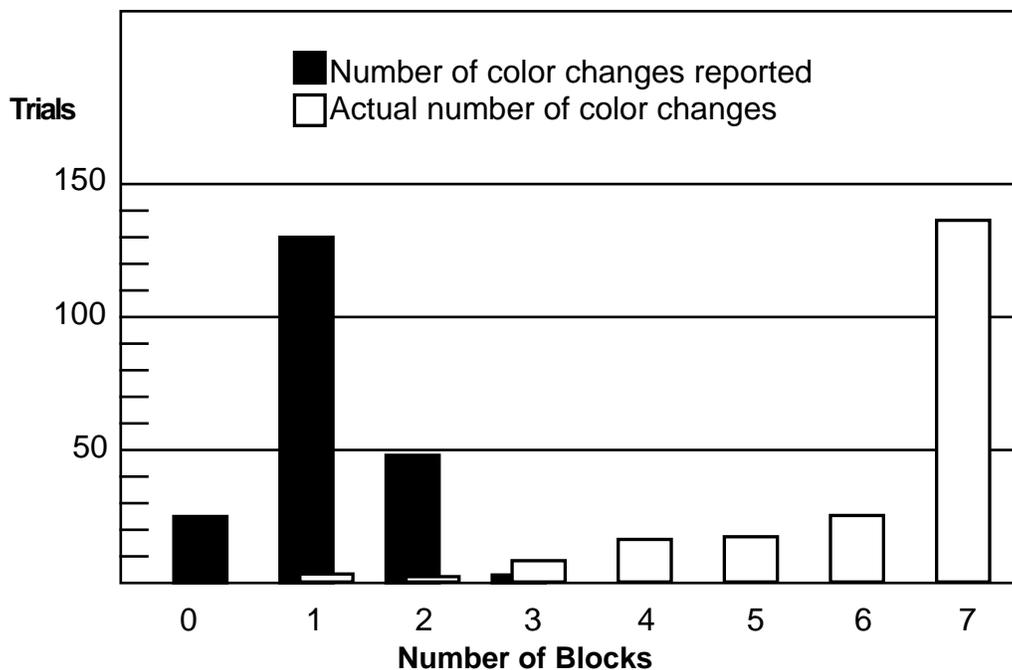


Figure 10: . The frequency of the number of blocks (per pattern) seen to change (black bars) compared with number actually changing (white bars) for subject MS. If two changes occurred while copying a single pattern, the sum of the blocks changed is reported here. (e.g. a trial in which first 4 blocks changed color and then 3 more changed color is reported as a 7). This is because the S only reported the total number of blocks that changed color during the trial after the end of the trial.

attending to only a single block. Note also that on 25 of the 206 experimental trials, no change was apparent to the observer. Not surprisingly 22 of these 25 occasions were in the Before Pickup condition in which little memory is required on the part of the subject.

Discussion

In summary, if a color change is made in a single block during the saccade to the model at the beginning of a block move, there is at most a small increase in fixation duration (43msec). When

the change occurred following block pickup, however, a clear increase in fixation duration was observed (106msec). In the second experiment, the effect of the change also depends on when it is done in the task. Fixation durations increased by 275 msec when changes were made following pickup, versus 127 msec before pickup. In this experiment all the uncopied blocks were changed in addition to the target block. This had generally greater effects than the first experiment, both before and after pickup, even though the non-targeted blocks were not immediately relevant to the task. In both experiments, the frequency of the different copying strategies was affected slightly, with the proportion of double look strategies increasing at the expense of the zero look strategies. In addition, an analysis of the source of the longer average time spent fixating the model pattern revealed that display changes increased the likelihood that two or more individual fixations would be made within the model area on any given visit. This was true for both experiments, but was largest in the second experiment and affected only the fixations following pickup.

An important feature of these results is that the same display change affects performance in a different way, depending on when it comes in the task. This supports the suggestion that the first and second fixations in the model area serve a different purpose. After block pickup, the color of the current block in the model is needed to identify the right location for placement in the copy, whereas at the beginning of a block move it has little consequence unless the color has been retained from prior exposures. This result therefore provides strong evidence that block color is acquired in the first fixation, and its relative location in the second. Thus it appears that the color information is preserved across changes in gaze when the immediate task requires it. This might come from the

immediately preceding peripheral view or from the prior model fixation.

Although it is fairly clear that display changes slow down task performance by about 100-200 msec when the change follows a pickup, it is difficult to evaluate the 40 msec effect when only one block is changed before pickup. All observers show a comparable slowing, but this was small relative to the within subject variance. Even if we suppose that the increase is reliable it would be difficult to argue that this increase implied carry over of the color from the prior view. When the observer is fixating in the model area, the average time spent is about 350 msec. During this period the observer must get the color, select a corresponding color in the resource area, and initiate the saccade. These processes go in parallel to some extent, and estimates of the time required to acquire simple feature information are in the range of 100-300 msec [Eriksen and Eriksen, 1971, T.Salthouse et al., 1981], so 40 msec seems unlikely to be long enough for complete acquisition of a new color. If we estimate color acquisition as needing 200msec, say, then an average value of 40 msec might result from this extra time being required on about 20 percent of the experimental trials. This would mean that on the majority of trials, there is no record of the color of the target block from prior model fixations, and color is indeed acquired during the fixation. This is interesting since the peripheral location must be selected as the saccade target somehow, and it appears that this target selection process involves a different kind of information from that acquired during foveation. This is consistent with a special computational role for foveation [Ballard et al., 1995].

The increased effectiveness of the display changes in the second experiment is also important. This increase is observed both before and after pickup, even though the additional blocks that were

changed were not the ones being worked on. Thus changes in irrelevant blocks appear to affect performance. This suggests that there is some global aspect of the configuration that is carried over between fixations. One possibility for explaining this would be if the location of the current block in the model area is encoded relative to the neighboring blocks. Some mechanism like this seems to be necessary for keeping track of one's place in the task. In this case, changes in the color of neighboring blocks changes the contextual information and this may be necessary for programming the saccade to the target block before pickup and for the return saccade to the model after pickup. One possible mechanism for realizing this is described in [Rao and Ballard, 1996]. This work is also discussed in [Ballard et al., 1996].

One important issue in considering the copying task is the extent to which the fixations actually reflect the immediate visual representation. For example, the relative slowness of the hand movements might allow extra fixations that are not really necessary. In previous work we have examined this question by manipulating various aspects of the task in ways that vary the need for fixations. The frequency of model fixations can be reduced by stimulus manipulations that allow visual chunking, for example, and by display configurations that require large head movements [Ballard et al., 1995, Pelz, 1995]. If subjects are obliged to use memory by removing the display, however, only 2 blocks can be copied without error [Ballard et al., 1992]. These observations suggest that the fixations are indeed diagnostic of information-gathering state. In addition, the task dependency we observe in the present experiments, supports our earlier suggestion [Ballard et al., 1995] that the two fixations in the model area serve a different purpose: the first

fixation in the model area is to acquire color information, and that the second fixation is for relative location of the current block. Thus the computational role of fixation is to acquire information just prior to its use. It also indicates that the information that is retained from the immediately prior fixation depends on the ongoing cognitive operations, and is not for the purpose of maintaining a general purpose spatial representation independent of the observer. This is an important result for a number of reasons. The first is that it suggests that visual processing may be much more task driven, or top down, than is commonly supposed, even for simple feature information such as color. This view has also been presented by [Churchland et al., 1994] as well as in earlier work [O'Regan and Lévy-Schoen, 1983, O'Regan, 1992, Nakayama, 1990]. The idea of an elaborate scene model is perhaps clearest in the computer vision literature, where until recently the goal of the models has been primarily one of reconstruction of a general purpose, task independent representation of the visual world [Marr, 1982]. However, the result that the same display changes affect fixation durations depending on the observer's momentary place in the task suggests that human vision may create only those perceptual descriptions that are currently necessary, and that fixation plays a crucial role in this process [Ballard et al., 1995, Ballard et al., 1996].

The task dependency observed here provides a context for interpreting results of a range of experiments which use the technique of changing the visual stimulus during a saccade. For example, experiments in reading show that certain changes in the parafoveal word that is the target for the next saccade, such as case changes, go undetected [McConkie and Rayner, 1976, McConkie and Zola, 1979]. Such results are usually interpreted in terms of the level of the pre-

served information (features/ semantic content etc). In a reading paradigm, the ongoing task is to extract meaning. Thus a consistent interpretation is that it is the immediately task relevant information which is retained. It may also explain the wide variation of estimates of sensitivity to shifts in position of an isolated stimulus during a saccadic eye movement (from about 7 to 30 per cent of the saccade size, [Li and Matin, 1990, Bridgeman et al., 1975, Bridgeman et al., 1994, McConkie and Currie, 1996]. It seems likely that some of this variation may result from the subject's attentional deployment in response to the implicit task called for by the instructions. Thus in experiments on changing the gain of saccadic eye movements, the subject's task is to simply fixate the stimulus as it jumps around rather than detecting changes during a saccade. In this case changes of stimulus position during a saccade of as much as 40 per cent can go unnoticed by the subject[Albano and King, 1989].

The issue of the complexity of the visual representation is often confused with the issue of whether visual information is integrated across different eye positions. An internal scene representation is usually assumed to reflect information acquired from several fixations, so evidence that visual information can be integrated across eye movements has been seen as important evidence for this kind of view. The issues are separate, however. Humans can clearly integrate visual information across eye movements when they are required to do so [Hayhoe et al., 1992, Hayhoe et al., 1991], and some ability to relate information across time and space is necessary for coordinated action.

An important feature of the results in the experiments described here is that the observers were only occasionally aware of the change, even on those trials when the current block was

changed and some disruption of performance was observed. A number of recent investigations also reveal considerable insensitivity in the perceptual report of quite striking display changes [Rensink et al., 1996, O'Regan et al., 1996, McConkie and Currie, 1996]. In the present experiment however, evidence of task interference was observed for all but the single block change preceding pickup. Thus it appears that fixation duration, strategy frequency, and the proportion of double fixations are all more sensitive indicators than perceptual report in revealing how the brain's representations control ongoing behavior. This implies that reported detection of such changes may not reflect the effect of display manipulations on visual function. Thus multiple block changes led to reports that only one or two had changed. This awareness presumably results from the effects of attention at the location of the current block. Changes in the other blocks affected fixation durations but were not reported. One way of understanding this difference is to distinguish between the two different levels of description relevant to task performance. Our everyday experience of the world may reflect events over a longer time scale than those revealed by individual fixations [Ballard et al., 1996]. Perceptual insensitivity to display changes makes sense if conscious experience corresponds to brain state at a time scale of the task, using task relevant variables such as 'Next block', 'Pickup' and 'Putdown'. This time scale and variables describe short term memory. The perceptuo-motor machinery that governs the fixations and operations within a fixation presumably runs at a shorter time scale with different primitives. The longer and more frequent fixations and probabilistic shifts in strategies observed in this experiment reflects the effect of the changes at this time scale. In this experiment, for example, (and indeed in most natural performance) observers

are not conscious of the eye movements themselves, but primarily of events described in task terms, such as seeing a red block, picking it up and putting it in the correct place in the copy. This is the appropriate time scale for goal directed behavior. If awareness corresponded to the act of moving the eyes themselves (as one can do if instructed) this would mean representing the eye movement explicitly as a variable in short term memory, rather than as an autonomous process. This would compete with events pertinent to the task, if we view short term memory as a (capacity limited) system that functions with a small number of variables [Ballard et al., 1996].

Acknowledgements

This research was supported by the National Institutes of Health under grants EY05729 and 1-P41-RR09283.

References

- [Albano and King, 1989] Albano, J. and King, W. (1989). Rapid adaptation of saccadic amplitude in humans and monkeys. *Investigative Ophthalmology and Visual Science*, 30:883–893.
- [Bajcsy, 1985] Bajcsy, R. (1985). Active perception vs. passive perception. In *Proc., Workshop on Computer Vision*, pages 55–59.

- [Ballard, 1991] Ballard, D. (1991). Animate vision: An evolutionary step in computational vision. *J. of the Inst. of Electronics, Information, and Communication Engineers*, 74(4):343–348.
- [Ballard et al., 1992] Ballard, D., Hayhoe, M., Li, F., and Whitehead, S. (1992). Eye hand coordination in a sequential task. *Proc. Royal Soc Lond B*, 337:331–339.
- [Ballard et al., 1995] Ballard, D., Hayhoe, M., and Pelz, J. (1995). Memory representations in natural tasks. *J Cognitive Neuroscience*, 7:66–80.
- [Ballard et al., 1996] Ballard, D., Hayhoe, M., Pook, P., and Rao, R. ((in press) 1996). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*.
- [Bridgeman et al., 1975] Bridgeman, B., Hendry, D., and Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, 15:719–722.
- [Bridgeman et al., 1994] Bridgeman, B., van der Heijden, A., and B.Velichkovsky (1994). A theory of visual stability across saccadic eye movements. *Behavioral and Brain Sciences*, 17:247–292.
- [Brooks, 1986] Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE J. Robotics and Automation*, 2:14–22.
- [Churchland et al., 1994] Churchland, P., Ramachandran, V., and Sejnowski, T. (1994). A critique of pure vision. In Koch, C. and Davis, J., editors, *Large-Scale Neuronal Theories of the Brain*, pages 23–60. MIT Press (A Bradford Book), Cambridge, MA.

- [Duebel and Bridgeman, 1995] Duebel, H. and Bridgeman, B. (1995). Fourth purkinje image signals reveal eye-lens deviations and retinal image distortions during saccades. *Vision Research*, 35:529–538.
- [Eriksen and Eriksen, 1971] Eriksen, C. and Eriksen, B. (1971). Visual perception processing rates and backward and forward masking. *Journal of Experimental Psychology*, 89:306–313.
- [Grimes and McConkie, 1995] Grimes, J. and McConkie, G. (to appear, 1995). On the insensitivity of the human visual system to image changes made during saccades. In Akins, K., editor, *Problems in Perception*. Oxford University Press, Oxford, UK.
- [Hayhoe et al., 1991] Hayhoe, M., Lachter, J., and Feldman, J. (1991). Integration of form across saccadic eye movements. *Perception*, 20:393–402.
- [Hayhoe et al., 1992] Hayhoe, M., Lachter, J., and Möller, P. (1992). Spatial memory and integration across saccadic eye movements. In Rayner, K., editor, *Eye Movements and Visual Cognition: Scene Perception and Reading*, pages 130–145. Springer-Verlag, New York.
- [Irwin, 1991] Irwin, D. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, 23:420–456.
- [Karn, 1995] Karn, K. (1995). *Spatial Representations for Programming Saccadic Eye Movements*. Doctoral Dissertation (Dept. of Brain and Cognitive Science) The University of Rochester, Rochester, New York.

- [Li and Matin, 1990] Li, W. and Matin, L. (1990). Saccadic suppression of displacement: influence of postsaccadic exposure duration and of saccadic stimulus elimination. *Vision Research*, 30:945–955.
- [Marr, 1982] Marr, D. (1982). *Vision*. W.H. Freeman and Co., Oxford.
- [McConkie and Currie, 1996] McConkie, G. and Currie, C. ((in press) 1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*.
- [McConkie and Rayner, 1976] McConkie, G. and Rayner, K. (1976). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In Singer, H. and Ruddell, R., editors, *Theoretical Models and Processes of Reading*, pages 137–162. International Reading Association, Newark, DE.
- [McConkie and Zola, 1979] McConkie, G. and Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception and Psychophysics*, 25:221–224.
- [Nakayama, 1990] Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. In Blakemore, C., editor, *Vision: Coding and Efficiency*, pages 411–422. Cambridge University Press.
- [O’Regan, 1992] O’Regan, J. (1992). Solving the ‘real’ mysteries of visual perception: The world as an outside memory. *Canadian J. Psychology*, 46:461–488.

- [O'Regan and Lévy-Schoen, 1983] O'Regan, J. and Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: Does trans-saccadic fusion exist? *Vision Research*, 23:765–769.
- [O'Regan et al., 1996] O'Regan, J. K., Rensink, R. A., and Clark, J. (1996). Mud splashes render picture changes invisible. *Investigative Ophthalmology and Visual Science (Suppl)*, 37:213–213.
- [Pelz, 1995] Pelz, J. (1995). Visual representations in a natural visuo-motor task. Technical Report PhD thesis, University of Rochester.
- [Pollatsek and Rayner, 1990] Pollatsek, A. and Rayner, K. (1990). Eye movements and lexical access in reading. In Balota, D., d'Arcais, G. F., and Rayner, K., editors, *Comprehension Processes in Reading*, pages 143–164. Lawrence Erlbaum Associates, Publisher, Hillsdale, NJ.
- [Pollatsek and Rayner, 1992] Pollatsek, A. and Rayner, K. (1992). What is integrated across fixations? In Rayner, K., editor, *Eye Movements and Visual Cognition: Scene Perception and Reading*, pages 161–199. Springer-Verlag, New York.
- [Rao and Ballard, 1996] Rao, R. and Ballard, D. (1996). An active vision architecture based on iconic representations. *Artificial Intelligence*, 78:461–505.
- [Rensink et al., 1996] Rensink, R., O'Regan, J., and Clark, J. (1996). To see or not to see: the need for attention to perceive changes in scenes. *Investigative Ophthalmology and Visual Science (Suppl)*, 37:213–213.

[Rock et al., 1992] Rock, I., Linnett, E., Grant, P., and Mack, A. (1992). Perception without attention: results of a new method. *Cognitive Psychology*, 24:502–534.

[Simons, 1996] Simons, D. (1996). In sight, out of mind: When object representations fail. *Psychological Science*.

[T.Salthouse et al., 1981] T.Salthouse, C.Lewis, Diener, D., and Somberg, B. (1981). Stimulus processing during eye fixation. *Journal of Experimental Psychology:HPP*, 7:611–623.