Chapter 30

LEARNING WHERE TO LOOK

MARY M. HAYHOE
University of Rochester, USA

JASON DROLL
University of California, USA

NEIL MENNIE
University of Nottingham, UK
Abstract

How do the limitations of attention and working memory constrain acquisition of information in the context of natural behavior? Overt fixations carry much information about current attentional state, and are a revealing indicator of this process. Fixation patterns in natural behavior are largely determined by the momentary task. The implication of this is that fixation patterns are a learnt behavior. We review several recent findings that reveal some aspects of this learning. In particular, subjects learn the structure and dynamic properties of the world in order to fixate critical regions at the right time. They also learn how to allocate attention and gaze to satisfy competing demands in an optimal fashion, and are sensitive to changes in those demands. Understanding exactly how tasks exert their control on gaze is a critical issue for future research.
A central feature of human cognition is the strict limitation on the ability to acquire visual information from the environment, set by limitations in attention. Related to this are the limits in retaining this information, set by the capacity of working memory. We are far from understanding how the organization of the brain leads to these limitations. We also have little understanding of how they influence the way that visual perception operates in the natural world, in the service of everyday visually guided behavior. Consideration of how the limited processing capacity of cognition influences acquisition of visual information leads us to the problem of how such acquisition is controlled. It is not really possible to address the question of precisely what information is selected from the image, and when it is selected, in the context of traditional experimental paradigms, where the trial structure is designed to measure a particular visual operation over repeated instances, each of short duration. In natural behavior, on the other hand, observers control what information is selected from the image and when it is selected. By observing natural behavior, knowledge of the task structure often allows quite well constrained inferences about the underlying visual computations, on a time scale of a few hundred milliseconds.

1. Eye movements and task structure

How can we study the acquisition of information in the natural world? Although incomplete, eye movements are an overt manifestation of the momentary deployment of attention in a scene. Covert attentional processes, of course, mean that other information is processed as well, but overt fixations carry a tremendous amount of information about current attentional state, and provide an entrée to studying the problem (Findlay & Gilchriest, 2003). Investigation of visual performance in natural tasks is now much more feasible, given the technical developments in monitoring eye, head, and hand movements in unconstrained observers, as well as the development of complex virtual environments. This allows some degree of experimental control while allowing relatively natural behavior. In natural behavior, the task structure is evident, and this allows the role of individual fixations to be fairly easily interpreted, because the task provides an external referent for the internal computations. In contrast, when subjects simply passively view images, the experimenter often has little control of, and no access to, what the observer is doing. When viewing pictures, observers may be engaged in object recognition, remembering object locations and identity, or performing some other visual operation. Immersion in a real scene probably calls for different kinds of visual computations, because observers may be interacting with the objects in the scene. When viewing images of scenes, some regularities in fixation patterns can be explained by image properties such as contrast or chromatic salience. However, these factors usually account for only a modest proportion of the variance (Itti & Koch, 2001; Mannan, Ruddock & Wooding, 1997; Parkhurst, Law, & Neibur, 2002).

Over the past ten years, a substantial amount of evidence has accumulated about deployment of gaze during ongoing natural behavior. In extended visuomotor tasks such as driving, walking, sports, playing a piano, hand-washing, and making tea or sandwiches,
the central finding is that fixations are tightly linked to the performance of the task (Hayhoe, Shrivastava, Mruczek & Pelz, 2003; Land & Furneaux, 1997; Land & Lee, 1994; Land, Mennie, & Rusted, 1999; Patla & Vickers, 1997; Pelz & Canosa, 2001; Turano, Geruschat, & Baker, 2003). Subjects exhibit regular, often quite stereotyped fixation sequences as they step through the task. Very few irrelevant areas are fixated. Figure 1 shows an example of the clustering of fixations on task-specific regions when a subject makes a sandwich. This is hard to capture in a still image, but can be clearly appreciated in video sequences such as those in Hayhoe et al., (2003). A feature of the relationship of the fixations to the task is that they are tightly linked, in time, to the actions (Land et al., 1999; Hayhoe et al., 2003). The temporal linkage has been demonstrated clearly by Johansson, Westling, Bäckström, & Flanagan (2001), who measured fixation locations and hand path while a subject picked up a bar and maneuvered the tip past an obstacle, to contact a switch. Fixations were made at critical points such as the tip of the obstacle while the bar was moved around it, and then on the switch once the bar had cleared the obstacle. Gaze arrived at the critical point just before the action, and departed just as the action was accomplished. This is illustrated in Figure 2.

This aspect of natural behavior, where observers acquire the specific information they need just at the point it is required in the task, was called a “just-in-time” strategy (Ballard, Hayhoe, & Pelz, 1995). In their experiment, subjects copied a pattern of colored blocks (the Model) using pieces in a Resource area, which they picked up and placed in

Figure 1. Fixations made by an observer while making a peanut butter and jelly sandwich, indicated by yellow circles. Images were taken from a camera mounted on the head, and a composite image mosaic was formed by integrating over different head positions using a method described in Rothkopf and Pelz (2004) et al. (The reconstructed panorama shows artifacts because the translational motion of the subject was not taken into account.) Fixations are shown as yellow circles, with a diameter proportional to fixation duration. The red lines indicate the saccades. Note that almost all fixations fall on task relevant objects. (See Color Plate 10.)
the area where the copy was being made. When subjects copied a particular block, they typically fixated a block in the Model, then looked at a block of the same color in the Resource while they picked it up, then looked back at the Model block, presumably to get information about location for placement, and then finally to the copy area where the block was placed in the appropriate location. Thus subjects appeared not to memorize the relatively simple model patterns, but simply to fixate individual blocks to get the information they need at that moment.

2. Learning where to look

Implicit in much of the research on natural tasks is the finding that the observed pattern of eye movements is a consequence of learning at several levels (Land & Furneaux, 1997; Land, 2004; Chapman & Underwood, 1998). For example, in tea making and sandwich making, observers must have learnt what objects in the scene are relevant, and how to locate them in visual search, since almost no fixations fall on irrelevant objects. In driving, Shinoda, Hayhoe, & Shrivastava (2001) showed that approximately 45% of fixations fell in the neighborhood of intersections. As a consequence of this, subjects were more likely to notice Stop signs located at intersections as opposed to signs in the middle of a block, suggesting that subjects have learnt that traffic signs are more likely to be located around intersections. At a more detailed level, subjects must learn the optimal location for the information they need. For example, when pouring tea, fixation is located at the tip of the teapot spout (Land et al., 1999). Presumably, flow from the spout is best controlled by fixating this location. Similarly, in walking, observers must learn where and when to look at locations critical for avoiding obstacles while controlling direction and balance (Patla et al., this volume). Subjects must learn not only the locations at which relevant information is to be found, but also the order in which the fixations must be made in order to accomplish the task. Thus, a subject must locate the peanut butter and the bread before picking them up, pick up the knife before spreading, and so on. This means that a complete understanding of fixations in natural behavior will require an understanding of the way tasks are learnt and represented in the brain.
Another way in which learning is critical for deployment of gaze and attention is that observers must learn the dynamic properties of the world in order to distribute gaze and attention where they are needed. When making tea or sandwiches, items remain in stable locations with stable properties, for the most part. In a familiar room, the observer need only update the locations of items that are moved, or monitor items that are changing state (for example, water filling the kettle). In dynamic environments, such as driving, walking, or in sports, more complex properties must be learnt. Evidence for such learning is the fact that saccades are often pro-active; that is, they are made to a location in a scene in advance of an expected event. For example, in Land & MacLeod’s investigation of cricket, batsmen anticipated the bounce point of the ball, and more skilled batsmen arrived at the bounce point about 100 ms earlier than less skilled players (Land & McLeod, 2000). The ability to predict where the ball will bounce depends on previous experience of the cricket ball’s trajectory. These saccades were always preceded by a fixation on the ball as it left the bowler’s hand, showing that batsmen use current sensory data in combination with prior experience of the ball’s motion to predict the location of the bounce. This suggests that observers have stored internal models of the dynamic properties of the world that can be used to position gaze in anticipation of a predicted event.

There is considerable evidence for the role of internal models of the body’s dynamics in the control of movement (e.g. Wolpert, Miall & Kawato, 1998). Such models predict the internal state of the body as a consequence of a planned movement, and help mitigate the problem of delays in sensory feedback about body posture. Similar delays in processing visual information about events in the world suggest a similar need for models of the environment, particularly in dynamic situations. However, the need for internal models of the environment is less well established. Indeed, the body of evidence in the past has suggested the contrary, that observers construct only minimal representations of the world (Ballard, Hayhoe, Pook, & Rao, 1997; O’Regan, 1992; Simons, 2000). To build internal models of the visual environment, observers must be able to accumulate visual information over the time-varying sequence of visual images resulting from eye and body movements. Visual representations that span fixations are typically thought to be very impoverished. It is generally agreed that, following a change in gaze position, observers retain in memory only a small number of items, consistent with the capacity limits of visual working memory, together with information about scene “gist,” and other higher level semantic information (Irwin & Andrews, 1996; Hollingworth & Henderson, 2002). However, some kind of internal model of the environment, such as memory for spatial structure, seems necessary to ensure coordinated movement (Chun & Nakayama, 2000; Loomis & Beall, 2004).

Hayhoe, Mennie, Sullivan, and Gorgos (2005) provide further evidence of the existence of sophisticated internal models of the structure of the environment. Such models may be used to predict upcoming events and plan movements in anticipation of those events. In this study, eye, head, and hand movements were recorded while subjects caught balls thrown with a bounce. Three participants stood in a triangular formation, and threw a ball around the circle. Initially, subjects threw a tennis ball around the circle of three participants. Each throw was performed with a single bounce approximately mid-way
between the participants. One of the throwers then changed the ball without warning, to one with greater elasticity (bounciness).

Similar to batsmen in cricket, when catching a ball, subjects initially fixated the hands of the thrower, then made a saccade to the anticipated bounce point, and then pursued the ball until it was close to the hands. Average departure time of gaze from the hands of the thrower was 61 ms after the ball left the hands. Gaze then arrived at a point a little above the anticipated bounce location an average of 53 ms before the bounce. (Note that the ASL 501 eyetracker used in these experiments has a real-time delay of approximately 50 ms. This value was used to correct the latency measurements.) Subjects maintained gaze at this location until the ball came into the fovea, and then made a smooth pursuit movement, maintaining gaze on the ball until the catch. Since the minimum time to program a saccadic eye movement is 200–250 ms (in the absence of any kind of anticipation or preparation), the saccade from the hands to the bounce point must be at least partially underway prior to the release of the ball. The landing points of the saccades relative to the actual bounce point clustered within about 5° laterally, and about 15° vertically above the bounce point. Thus subjects appeared to be targeting a region just above the bounce point, rather than the bounce point itself. This presumably facilitates the subsequent tracking movement by allowing time to capture the ball’s trajectory after the bounce. The tight lateral clustering of the saccade landing points relative to the bounce point suggested that subjects were using information from the early part of the throw to target the likely location of the bounce.

2.1. Adjusting to the ball’s dynamic properties

Ability to pursue the ball depended on experience with the ball’s dynamic properties. When the tennis ball was unexpectedly replaced with a bouncier ball, subjects were unable to track the ball, and instead made a series of saccades. Within a few trials, subjects were once again able to accurately pursue the ball. A crude evaluation of pursuit accuracy was made by measuring the proportion of time gaze was close to the ball, in the period between bounce and catch. Improvement in pursuit performance over six trials is shown in Figure 3, which shows the pursuit accuracy improving rapidly over the first three trials, close to the performance level with the tennis ball. The ability to make accurate pursuit movements in this context therefore depends on the knowledge of the dynamic properties of the new ball. The adjustment in performance was quite rapid, and uniform across subjects, suggesting that adjusting to such changes in the environment is an important feature of natural behavior. (The ability to pursue the tennis ball accurately on the first trial presumably reflects either its slower speed, or that its motion is closer to subjects’ prior expectations.) The latency of the first saccade from hands to bounce point also changed over the course of a few trials. Arrival time at the bounce point advanced by about 100 ms over the first six trials following the change from tennis to bouncy ball. The earlier arrival of the eye at the bounce point is accompanied by earlier departure from the hands at the point of release. Thus, anticipatory saccades and pursuit movements reveal that acquisition of visual information is planned for a predicted state of the world.
Such predictions must be based on a stored memory representation of some kind. The precision of the predictions reveals the quality of the information in the stored memory, or internal model. The spatial and temporal precision of the anticipatory saccades, and the fine-tuning of these movements following a change in the ball’s dynamic properties indicate that subjects have an accurate internal model of the ball’s spatiotemporal path, and rapidly update this model when errors occur. Rapid adjustment of performance suggests that such prediction is a ubiquitous feature of visually guided behavior.

3. Neural substrate for learning where to look

Control of eye movements, visual acquisition, and memory use by the immediate task entails several different kinds of learning. As mentioned above, observers must learn the structure of tasks such as making tea, they must learn where to look to get the information they need, they must learn the properties of the world, and how those properties change, and they must learn how to allocate attention and fixations in an optimal manner. Recent developments in neurophysiology help us understand how some of this learning might come about. Much research supports a reward-based learning mechanism involving dopamine. Considerable evidence for this comes from experiments by Schultz that show that dopaminergic neurons in the substantia nigra pars compacta in the basal ganglia behave in ways predicted by mathematical models of reinforcement (Montague, Hyman, & Cohen, 2004; Schultz, 2000). This reward system is integral to the generation of saccadic eye movements. Saccade-related areas in the cortex (frontal eye fields, dorso-lateral pre-frontal, and lateral intra-parietal) all converge on the caudate nucleus in the basal ganglia, and the cortical-basal ganglia-superior colliculus circuit appears to regulate the control of fixation and the timing of planned movements. This is
achieved by regulation of tonic inhibition exerted by the substantia nigra pars reticulata on the superior colliculus, the mid-brain saccade generator. Such regulation is a critical component of task control of fixations. Hikosaka and colleagues have demonstrated that caudate cell responses reflect both the target of an upcoming saccade and the reward expected after making the movement (Hikosaka, Takikawa, & Kawagoe, 2000; Watanabe, Lauwereyns, & Hikosaka, 2003). Since some kind of sensitivity to reinforcement is necessary for learning, and saccadic eye movements demonstrate such sensitivity, the neural substrate for learning where to look in the context of a task is present in the basal ganglia.

Other areas involved in saccade target selection and generation also exhibit sensitivity to reward. In the lateral intra-parietal area (LIP), the neurons involved in saccadic targeting respond in a graded manner to both the amount of expected reward and the probability of a reward, in the period prior to execution of the response (Dorris & Glimcher, 2004; Glimcher, 2003; Platt & Glimcher, 1999; Sugrue, Corrado, & Newsome, 2004). Sensitivity to both these variables is critical for linking fixation patterns to task demands. Cells in the supplementary eye fields also signal the animal’s expectation of reward and monitor the outcome of saccades (Stuphorn, Taylor, & Schall, 2000). Sensitivity to stimulus probability is also revealed in build-up neurons in the intermediate layers of the superior colliculus (Basso & Wurtz, 1998). This parallels psychophysical observations showing that saccade reaction time is similarly influenced by stimulus probability (He & Kowler, 1989).

The neural data showing context-specific responses and the role of reward form a critical substrate for explaining task-directed eye movement patterns. The reward sensitivity of the eye movement circuitry provides a basis for reinforcement learning models that are necessary for understanding how these elemental processes are organized to compose the complex gaze patterns observed in everyday behaviors. Theoretical work such as that by Sprague & Ballard (2003) shows how a graphical agent in a virtual environment can learn to allocate gaze sequentially to areas in the environment important for walking and avoiding obstacles. To choose between ongoing competing tasks such as avoiding obstacles and controlling direction of locomotion, in their model, uncertainty increases (together with an attendant cost) when gaze is withheld from an informative scene location. Fixation is allocated to the task that would have the greatest cost if the relevant information were not updated. They show that such a cost is calculable within the reinforcement learning framework described by Schultz and others. In the context of a complex behavioral sequence, a single eye movement does not, of course, generate a primary reward such as a drop of juice, as in the neurophysiological experiments; but all complex behaviors involve secondary reward of some kind, and the acquisition of information is always a critical step in achieving behavioral goals.

4. Specialized computations during fixations

There is far too much information in visual scenes to process at once. Even at the point of fixation, multiple kinds of information are available. It seems likely that, in the context
of natural behavior, the task controls the specific information that is selected within a given fixation. For example, when driving around a bend in the road, drivers fixate the tangent point of the curve (Land & Lee, 1994). Interestingly, the angle of gaze with respect to the body is a measure of the necessary change in the steering angle required for navigating the bend. Thus gaze directly provides the control variable needed for the momentary task. When making a sandwich, subjects will fixate the handle of the knife when picking it up, but the tip of the knife when spreading the peanut butter. In the first case, the subject needs the location and orientation of the handle to guide the pick up action. In the latter case, the tip needs to be fixated to control the spreading action. When first viewing the tabletop scene, subjects make a series of short duration fixations on the relevant objects, such as the peanut butter jar. In this case the fixation on the peanut butter is presumably for the purpose of recognition, and perhaps locating it for future use (Hayhoe et al., 2003). Subsequent fixations will be for guiding the grasping action or for removing the lid. In the absence of ongoing behavior, we are inclined to think that the job of vision is primarily object recognition, but these examples remind us of the complexity of the information one can get while fixating an object, and the variety of operations that vision must perform. This specificity is indicated not only by the ongoing actions and the point in the task, but also by the durations of the fixations, which may vary over a range from less than 100 ms to several seconds (Hayhoe et al., 2003; Pelz et al., 2000). A large component of this variation appears to depend on the particular information required for that point in the task, fixation being terminated when the particular information is acquired (Hayhoe Bensinger, & Ballard, 1998; Henderson, 2003; Pelz et al., 2000). For example, in a task where subjects were required to tap a pre-determined sequence of lights on a table top, fixations in the search phase of the experiment, while subjects are locating the lights, are much shorter than fixations while subjects are tapping, where gaze is used for guiding the hand (Epelboim et al., 1995).

Figure 4 illustrates our conception of the way visual information acquisition is organized by the task (see also Land et al., 1999; Schwartz, Reed, Montgomery, Palmer, & Mayer, 1991). At the most general level, acquisition is driven by a larger cognitive goal, such as making a sandwich. To accomplish this goal, the observer must perform a sequence of micro-tasks, such as grasping the peanut butter, grasping the lid, removing the lid, and so on. A micro-task, such as grasping the peanut butter, involves a fixation on the jar. During that fixation, specific visual information will be acquired, such as the size of the jar, which is necessary in order to plan the grasp. These task-specific computations have been referred to as “visual routines” (Ballard et al., 1997; Hayhoe, 2000; Roelfsema, Lamme, & Spekreijse, 2000; Ullman, 1984). The idea of visual routines is that even quite low-level visual information, such as color, requires specialized computation of some kind, and is not done automatically by the visual system. We hypothesize that the visual computations operating while the subject is guiding a grasp do not necessarily involve computing other information, such as the color of the peanut butter, that are not necessary at that moment. Such information might of course be stored in working memory, or in some longer-term memory representation of the scene.
The hypothesis that very specific information is acquired within a fixation is, the strictly, a hypothesis about neural mechanisms. That is, the cortical state (including V1) is different when subjects are involved in different tasks, even when the retinal input is the same. However, it is possible to get supporting evidence psychophysically. Two studies have provided evidence that highly task-specific information is extracted in different fixations (Hayhoe et al., 1998; Triesch, Ballard, Hayhoe, & Sullivan, 2003). Both these studies used the technique of making changes to task relevant information in the scene, and looking for an influence of the changes on task performance, as well as at sensitivity to the changes. In the Triesch et al. experiment, subjects picked up virtual bricks and placed them on one of two virtual “conveyor belts”, as illustrated in Figure 5. The bricks were of two different heights, and subjects sorted them onto the conveyor belts according to different rules that varied the point in the task at which the brick height was relevant. In the first case, subjects picked up the bricks according to their location, and size was irrelevant. In the second, the bricks were picked up on the basis of their size, and then all placed on the same conveyor belt, so size was relevant for pickup only. In the third, bricks were placed on different belts depending on their size, so size was relevant for both pickup and placement. On a small proportion of trials, the brick changed size by about 20% while it was being moved to the belt. Although the change in size was very obvious when subjects were attending to it, when they were performing the task, subjects rarely reported seeing the changes, either during the task or upon subsequent questioning. They were most unaware of the change when size was irrelevant, suggesting that information about size might not have been present in the subject’s visual representation of the brick.
in these trials. Size changes were noticed more often when it was relevant for pickup or putdown (conditions 2 and 3). Interestingly, when size was relevant only for pickup (condition 2), subjects were less likely to notice changes than when it was relevant for placement as well (condition 3), suggesting that in the second condition they did not retain a representation of size in working memory when it was no longer needed after they had picked it up. Thus the experiment supported the idea that specific information about the brick’s properties, in this case its size, was only perceptually encoded in conditions 2 and 3, when it was needed for pickup, and only retained in working memory in condition 3, when it was needed for putdown as well. In each case, the subject is “attending” to the brick as he/she fixates it, picks it up, moves, and places it. This experiment reveals that a general undifferentiated concept of attention is too crude to reveal the subtleties of the ongoing visual computations. Another finding in the experiment was that subjects were often unaware of the change in size even when they were directly fixating the brick when the change occurred. This happened on some trials when subjects tracked the brick while they moved it across the workspace, so that they were fixating the brick at the point when it changed. (Note, this differs from the fixation pattern shown in Figure 5, where the subject saccades directly to the conveyor belt after pickup.) This suggests that subjects may not represent particular stimulus features such as size when they are performing other computations such as guiding the arm. This appears to be a form of “inattentive blindness”, described extensively by Mack & Rock (1998). The pervasive nature of inattentive blindness in perception is taken advantage of by magicians. The strength of this phenomenon in the context of magical tricks is described by Tatler and Kuhn in this volume.

4.1. Task-specific representations and Object Files

The idea that task relevance guides top-down selection of even simple feature information contrasts with the idea that visual information is represented and maintained in the form of object files (Kahneman, Treisman, & Gibbs, 1992). This theory posits that when attention is directed to an object in a scene, a temporary representation called an “object file” is
created and held in visual short-term memory, and about three or four object files and their spatial locations can be held in memory at any one time (Gordon & Irwin, 1996; Irwin & Andrews, 1996). Object file theory is consistent with claims that the units of short-term visual memory are integrated objects, not simple features (Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001). The Treisch et al. experiment suggests that the concept of object files might not extend to natural behavior, where task demands are dynamic and specific to the immediate needs of the observer. To explore this question further, Droll, Hayhoe, Triesch, & Sullivan (2005) conducted an experiment similar in concept to that of Treisch et al. The goal of the experiment was to identify more precisely the information acquired and held in memory and provide a more definitive test of the hypothesis that the visual information acquired in a fixation may be limited to the specific feature required by the task. In the Triesch et al. (2003) experiment the only feature of the bricks that was used in the task was height. Is it the case that brick features that were not required, such as color and shape, were never encoded? This would mean that fixating an object and attending to it would not necessarily bind the features of the brick into some object representation, or object file.

The basic task used by Droll et al. was to select one brick from an array, and to sort this brick onto one of two conveyor belts. The bricks were defined by several features (color, height, width, and texture), and a pickup cue indicated which feature value was relevant for a particular trial. After picking up the brick, a put-down cue was displayed to guide the sorting decision. The brick was placed on the appropriate conveyor belt, removing the brick from the scene, and initiating a new trial with a new pickup cue and array of bricks. Thus, because the put-down cue was presented after pickup, the put-down decision was separated in time and space from pickup, and the representations of the relevant object feature needed to be maintained until the put-down decision was made. In one condition, subjects performed a task in which only one feature dimension was relevant for both pickup and put-down (e.g. color). In another condition, different features were used for pickup and put-down (e.g. color for pickup, height for put-down). The task sequence is illustrated in Figure 6.

This experiment also used the strategy of changing the bricks on a small proportion of trials, but in this case it was possible to change either the feature that was relevant for pick up or putdown, or one of the other features that was not relevant in that block of trials. Subjects indicated whether they saw a change by placing the brick in a “trash can” (the black square in Figure 6). Subjects were about twice as likely to notice the feature change when that feature was relevant to the task as when it was irrelevant (either for pickup or for put-down, or for both). This supports the hypothesis that subjects preferentially represent the task-specific features of the objects. Objects are not necessarily stored in working memory as bound entities. Thus, understanding what visual operations are occurring during a fixation requires knowledge of the current task.

Memory vs Just-in-time representations: Another important finding in Droll et al.’s experiment was subjects’ behavior on trials when a task-relevant feature was changed, but subjects still failed to notice that a brick had changed. On these trials, subjects sorted the bricks onto one of the conveyor belts instead of placing it in the trash. How do the subjects
sort the bricks? Figure 7 shows the two possibilities. Either the subject can treat the brick as if it retained its old feature, or else he/she can sort it according to its current feature. Given that the subject invariably fixates the brick either before or during placement, one might expect that the new feature state would be clearly visible, and that the subjects would sort the brick on the basis of its current (changed) feature. Instead, they almost always sorted it by its old feature, despite the fact that they fixated the brick, with its new feature state, for an average of 750 ms after the change. This indicates that subjects are using
their memory of the brick feature, rather than its actual current state, to make the sorting decision. (Note that the put-down cue is not available until the block has been picked up, so subjects cannot plan the put-down movement immediately after pickup, but must fixate the put-down cue first.) This finding is important because failure to detect changes to items in scenes has traditionally been interpreted as a failure to retain information from the pre-change image in memory (O'Regan, 1992; Rensink, 2000). Although this is undoubtedly the case in many situations, in the current context, sorting by the old feature reveals that subjects do indeed maintain a memory representation of the previous state of the brick, but simply fail to update their representation with the new information following a change. Change blindness as a consequence of failure to update the internal representation is consistent with the suggestion of Henderson & Hollingworth (1999) and Simons & Rensink (2005). It seems likely that in many situations the information in the visual scene is stable. Changes like those in the Droll et al. experiment would be impossible in real scenes. Subjects presumably take advantage of prior experience about the stable properties of normal scenes to accumulate information about scenes in internal representations. Such a strategy makes sense given the limited bandwidth that is set by attention on the accrual of information. As long as the observer has accurate knowledge of the stable properties of scenes, there is no point in constantly updating the information in the image if it takes up attentional resources. It appears that even such low level information as color or size takes computational resources. The fact that subjects were fixating the brick while sorting onto the wrong conveyor belt suggests that attentional resources were devoted to the placement action, rather than to updating information about the brick. This phenomenon may be related to “inattentional blindness,” similar to the phenomena mentioned above (Mack & Rock, 1998). Note that the strategy of retaining information in memory in the brick sorting task is in contrast to the “just-in-time” strategy exhibited by observers in the block copying task (Ballard et al., 1995). In one case, observers opt to retain information in memory, and in the other they minimize memory load by fixating the block just at the point when the information is needed. Both are ways that the human cognitive system can deal with attentional and memory limitations. An important question for future research is: What determines which strategy is used?

In Droll et al.’s experiment, subjects treated the block properties as stable. An interesting observation in their experiment was subjects’ behavior in subsequent trials after a feature change was successfully detected. Figure 8 shows the total time spent fixating the brick on trials following a noticed change, and also the total duration of the hand movement between lifting and placement on the conveyor belt. Both fixations and hand movements were significantly longer on the trial immediately following a detected change, by as much as 400 ms. This effect fell off sharply over the next few trials. Thus subjects were taking more time to perform the task after they observed the unlikely event where a brick changed its properties. This suggests that they reallocated attentional resources to the brick on the next trial, when events violated their expectations. However, since the event was not repeated, they reverted to the prior strategy. This suggests that internal models of the environment are dynamic, and may often be more strongly influenced by the most
recent events. Recall that similar dynamic updating was required as subjects learned to track a ball with a different elasticity.

5. Summary

The last decade has seen tremendous advances in our knowledge of the way gaze is used in everyday behavior. This chapter attempts to capture some of the insights from this work in order to point toward areas of future exploration. Perhaps the most important insight is that gaze patterns reflect extensive learning at several levels. At the most general level, observers must learn the sequence of operations required to perform tasks. They
must learn where to look in a scene to get the information they need for component sub-tasks. They must learn the structure and dynamic properties of the world in order to fixate critical regions at the right time. They must learn how to allocate attention and gaze to satisfy competing demands in an optimal fashion and be sensitive to changes in those demands. There are many questions about the precise way that learning affects gaze patterns. Perhaps the most critical issue is understanding exactly how tasks exert their control on gaze. A growing understanding of the importance of reward in modulating the underlying neural mechanisms and theoretical developments using reinforcement learning models of complex behavior provides us with the tools to understanding how tasks are represented in the brain, and how they control acquisition of information through use of gaze.

Acknowledgements

This work was supported by NIH grants EY 05729 and RR 09283. The authors wish to thank Brian Sullivan, Keith Gorgos and Jennifer Semrau for assistance with the experiments and Keith Parkins for programming support.

References


