ANNUAL REVIEWS

Annual Review of Vision Science

Vision and Action

Mary M. Hayhoe

Center for Perceptual Systems, University of Texas at Austin, Texas 78712;
email: hayhoe@utexas.edu

ANNUAL REVIEWS **Further**

Click here to view this article's online features:
• Download figures as PPT slides
• Navigate linked references
• Download citations
• Explore related articles
• Search keywords

Keywords

sensory-motor decisions, natural behavior, statistical decision theory, saccades, reaching, memory, prediction

Abstract

Investigation of natural behavior has contributed a number of insights to our understanding of visual guidance of actions by highlighting the importance of behavioral goals and focusing attention on how vision and action play out in time. In this context, humans make continuous sequences of sensory-motor decisions to satisfy current behavioral goals, and the role of vision is to provide the relevant information for making good decisions in order to achieve those goals. This conceptualization of visually guided actions as a sequence of sensory-motor decisions has been formalized within the framework of statistical decision theory, which structures the problem and provides the context for much recent progress in vision and action. Components of a good decision include the task, which defines the behavioral goals, the rewards and costs associated with those goals, uncertainty about the state of the world, and prior knowledge.

1. ACTIVE BEHAVIOR INVOLVES SEQUENCES OF SENSORY-MOTOR DECISIONS

The investigations of visual perception and control of movements have historically been treated separately. Indeed, this is appropriate for many questions, but it is also critically important to consider them jointly. This becomes most evident when we consider ordinary behavior such as walking across the street. To accomplish a simple task like this, a person must identify a goal to determine the direction of heading, perhaps establish that the light is green, avoid tripping over the curb, locate other pedestrians or vehicles and their direction of heading so as to avoid bumping into them, and so on. Each of these particular goals requires some visual evaluation of the state of the world to make an appropriate action choice in the moment. Thus, we can see that in the context of normal behavior, humans make continuous sequences of sensory-motor decisions to satisfy behavioral goals, and the role of vision is to provide the relevant information for making good decisions to achieve those goals.

This conceptualization of visually guided behavior as a sequence of sensory-motor decisions has been formalized within the framework of statistical decision theory (Maloney & Zhang 2010, Franklin & Wolpert 2011, Wolpert & Landy 2012). This framework structures the problem and provides the context for much recent progress in the perception and action field. If we think of the components of a good decision, first, we must take into account the behavioral context. For example, walking across the road requires a set of context-specific behaviors. This underlies the effects of tasks on active vision. One prominent aspect of this is the effect of tasks on choice of gaze target, which we review below (Section 2.1). To understand how a particular goal can exert an effect on behavior, we need to address the costs and benefits of the action in bringing about the goal. These costs and benefits are mediated by the underlying neural reward machinery. The idea of reward is a critical component of understanding sensory-motor decisions, and one where there has been tremendous progress over the last 10–15 years. In active behavior, the visual system uses gaze to sample necessary information from the environment to inform the subsequent action. Without up-to-date information about the state of the world, it is difficult to know what the consequences of an action are likely to be. Noisy sensory data and decay of visual memory representations all factor into the subject's uncertainty about world state and into the decision of where to look and what action to take as a consequence. It is widely accepted that evaluating perceptual state can be understood in a Bayesian context (Knill & Richards 1996). This is a particular case of statistical decision theory (Maloney & Zhang 2010). Thus, the probability of a particular world state depends on the likelihood of obtaining that sensory data, given a particular state, weighted by the prior probability of that state. These priors can be thought of as instantiations of memory representations, and the role of memory in determining both action choices and control of actions is another area of recent progress and a critical component for understanding visual control of actions. A final aspect of sensory-motor decisions in the context of behavior is the need for prediction. Body movements are slow and costly, and visual evidence reflects the state of the world a hundred milliseconds or more in the past. Thus, there are high stakes for prediction. Action decisions are often better if they can be made in advance of some relevant change in the scene. This prediction must be based on memory for what has happened on previous occasions, and the role of prediction in vision and control of eye movements is also an area of recent progress.

To return to the example of crossing the road, we see that the behavioral goal of crossing the road determines what visual information is needed and sets the context of relevant actions, such as initiating and choosing walking direction. The value of avoiding obstacles will determine gaze deployment to other pedestrians. Uncertainty about where the other pedestrians are will determine

when the walker searches the scene and checks on their location or looks at the surface of the road to look for the curb. Experience with the dynamics of pedestrian behavior in combination with visual information about their walking direction and speed will allow the walker to choose an avoidance path on the basis of prediction of their likely location in the future. In this review, we consider all these factors, focusing on gaze target selection, reaching, grasping, and locomotion, and we organize the review around effects of task, the role of reward, how gaze is used to reduce uncertainty about properties of the visual scene, the role of memory, and the importance of prediction.

It is useful to contrast this integrative view of vision in its behavioral context with other conceptualizations of the role of vision. For example, in David Marr's (1982) seminal work, the role of vision was seen as identifying the objects and their locations in a scene. This conceptualization persists at the core of fields like computer vision. Although it is clearly a central role for vision, in the broader context of behavior, vision is critical for gathering knowledge about the world to choose rewarding actions as well as for guiding the execution of those actions. Thus, behavior provides a temporal context for vision and highlights its evolution over time. This approach also contrasts with another approach that has dominated the field of perception and action for the past two decades—namely, the approach notably espoused by Goodale & Milner (2005) that sees perception and action as separate systems and focuses on the differences between vision for perception and vision for action, rather than the way they work together. Finally, the approach taken here has many similarities with the ecological approach (Zhao & Warren 2015) that emphasizes the use of specific visual information for control of action. However, that approach emphasizes on-line control, whereas the present approach includes the essential role of memory and prediction. We discuss these alternative approaches in more detail below. Although the review focuses primarily on eye movements, the principles discussed hold true for actions in general. In particular, it seems likely that the representations that guide eye movements also guide the head, hands, and body.

2. IMPORTANCE OF BEHAVIORAL GOALS

2.1. Eye-Tracking in Natural Behavior

A critical advance in allowing the investigation of vision in its behavioral context has been the development of eye trackers mounted on the head (Hayhoe & Ballard 2005, Land & Tatler 2009). This has allowed a much broader range of behavioral contexts in which to measure gaze deployment because the observer is not restricted by a bite bar to a fixed head position in space and is free to make normal movements. In addition, it has allowed the simultaneous measurement of eye, head, hand, and body movements. Because eye movements are revealing about how visual information is collected to serve action decisions, this development has been important in understanding vision in the context of ongoing behavior. Breakthrough work in this domain was done by Land and colleagues, who developed a head-mounted eye tracker and measured a variety of tasks, such as driving, table tennis, cricket, and tea making (Land & Lee 1994, Land & Furneaux 1997, Land et al. 1999, Land & McLeod 2000). This development provided a more fertile empirical base for understanding how gaze is used to gather information to guide behavior. In the subsequent decades, improvements in eye-tracking methodology have allowed a wide variety of natural visually guided behaviors to be explored—for example, walking indoors and outdoors, model building and other tasks using eye and hands, playing squash, and infants and toddlers playing (Pelz & Canosa 2001, Patla & Vickers 2003, 't Hart & Einhäuser 2012, Johansson et al. 2001, Tatler et al. 2011, Land & Tatler 2009, Hayhoe & Ballard 2005, Hayhoe et al. 2012, Sprague et al. 2015, Franchak & Adolph 2010, Yu & Smith 2017). The exploration of natural visually guided action has revealed the extent to which fixations in a scene are tightly linked to behavioral goals, both in space and

time. During performance of tasks like tea or sandwich making, over 95% of the fixations can be accounted for by the task (Land et al. 1999, Land & Hayhoe 2001, Hayhoe et al. 2003). The precise timing relations between eye and hand are also tightly orchestrated, with gaze departing for the next goal as soon as the current goal is achieved (Johansson et al. 2001).

Thus, natural behavior forces consideration of exactly what information is being gathered by the visual system from moment to moment and of its relation to the task. The highly specific nature of the information acquired in a fixation suggested by these experiments was verified in controlled experiments that manipulated image features at specific moments during task performance (Triesch et al. 2003, Droll et al. 2005, Droll & Hayhoe 2007). Thus, a subject instructed to pick up a red block will attend to the color during the pick-up action but not during subsequent actions, when they use the stored memory representation. (See further discussion below in Section 3.2 and in **Figure 2**.) Other aspects of natural behavior point to the need for short-term memory representations in programming movements. For example, many of the reach-to-grasp movements in tasks like sandwich making or tea making are completed while gaze is directed at another target and guiding the other hand. These movements are plausibly guided by spatial memory for the location (Land et al. 1999, Hayhoe et al. 2003, Tatler & Land 2011). This issue is discussed in more detail in Section 4.

2.2. Task Effects Are Mediated by Neural Reward Machinery

How do tasks have their profound effect on sensory-motor decisions? This seems like a challenging problem, given that tasks are intrinsically highly specific. However, over the last 15 years, it has become clear that the brain's internal reward circuitry can provide a mechanism for the role of tasks on both gaze behavior and action choices. The mathematics of reinforcement learning (RL) provides a formalism for how tasks might be learned and how competing goals interact (Sutton & Barto 1998). Much of the brain is involved in representing the different computational elements of RL, and this provides a neural basis for the application of these models to understand sensory-motor decisions (Schultz 2000, Lee et al. 2012, Glimcher 2011). Dopaminergic cells signal the reward expected from an action, and RL models allow an agent to learn what actions or action sequences will lead to reward in the future. Given a set of possible states, and actions that might be associated with those states, RL algorithms allow an agent to learn a policy for selecting actions that will ultimately maximize reward. Sensitivity to reward is manifest throughout the saccadic eye movement circuitry. Saccade-related areas in the cortex [lateral intraparietal cortex (LIP), frontal eye fields (FEF), supplementary eye fields (SEF), and dorsolateral prefrontal cortex (DLPF)] all exhibit sensitivity to reward [Platt & Glimcher 1999, Dorris & Glimcher 2004, Sugrue et al. 2004, Stuphorn & Schall 2006, Seo et al. 2007, Glimcher et al. 2009, Yasuda et al. 2012; see Gottlieb (2012) for a review]. In particular, LIP neurons that are involved in saccade target selection have been implicated in coding the relative subjective value of potential targets (Sugrue et al. 2005, Kable & Glimcher 2009, Trommershäuser et al. 2009). Caudate cell responses reflect both the target of an upcoming saccade and the reward expected after making the movement (Hikosaka et al. 2006). 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. Sensitivity to both these variables is critical for learning, and consequently for linking fixation patterns to task demands. The cortical saccade-related areas 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. Such regulation is a critical requirement for task control of fixations. Thus, for example, in a task where subjects pick up a block and move it past an obstacle, fixation departs from the edge of the obstacle just as the hand

clears the obstacle, at which point vision is no longer needed (Johansson et al. 2001). The precise regulation of the timing of fixations is a critical feature of visually guided behavior and one that has received relatively less attention than the spatial selection of saccade targets.

2.3. Explicit Versus Intrinsic Rewards

Many of the reward effects in neurons have been observed with very simple choice response paradigms where the animal gets rewarded for looking at a particular target. In natural vision, individual eye movements are not directly rewarded but instead are for getting information that allows behavioral goals to be achieved, presumably with associated rewards that are ultimately important for survival. Thus, it is necessary to make the definitive link between the primary rewards used in experimental paradigms and the secondary rewards that operate in natural behavior. In human behavior, it has been shown that saccadic targeting is sensitive to explicit reward (money or points) in simple experiments involving a choice between a small number of targets (Navalpakkam et al. 2010, Schütz et al. 2012). Navalpakkam et al. showed that subjects' saccade behavior in a visual search is consistent with an ideal Bayesian observer, taking into account both rewards and stimulus detectability. It has also been shown that subjects making rapid hand movements quickly adjust their movements to a complicated spatially distributed target reward system and behave in a nearly optimal manner to maximize monetary reward (e.g., Trommershäuser et al. 2003, Seydell et al. 2008). These experiments consider single movements that are directly rewarded. To understand natural behavior, however, it is important to demonstrate the role of implicit reward for particular actions, and this is less clearly established. Jovancevic & Hayhoe (2009) have demonstrated that while walking in a natural environment, subjects looked more frequently at potentially hazardous pedestrians who sometimes veered briefly toward them than at pedestrians who simply stopped and so were visually salient but not hazardous. Because the events were of comparable visual salience, and the eye movements were anticipatory, before the pedestrian actually veered or stopped, this might be interpreted as reflecting the behavioral relevance or intrinsic reward value of the information, as veering signals a potential collision, whereas stopping is less important. Rothkopf et al. (unpublished manuscript) and Tong et al. (2015) have shown that it is possible to recover an estimate of the intrinsic reward value of particular actions, such as avoiding obstacles in a walking task. Matthis & Fajen (2013, 2014) have demonstrated that subjects look at least two steps ahead when walking on irregular terrain and that this allows a minimization of energetic costs. Thus, it seems likely that subjects learn stable values for the costs of particular actions like walking and obstacle avoidance, and these subjective values factor into momentary action decisions.

Not all reward effects are associated with particular action decisions. The visual targets themselves can be associated with rewards, as in experiments by Della Liberia & Chelazzi (2009), where subjects are more likely to attend to targets that have a history of being rewarded when they were selectively attended. These targets are also harder to ignore when they function as distractors. In a similar fashion, Kim et al. (2014) showed that monkeys learn associations between fractal patterns and rewards, and monkeys preferentially saccade to previously rewarded patterns. The tail of the caudate nucleus appears to be involved in coding these memory representations. Thus, learned stimulus–reward associations modify salience, or the ability of a stimulus to automatically bias attention.

2.4. Optimal Feedback Control

Another way that the task is intimately involved in action is in the feedback control of actions (Diedrichsen et al. 2009, Wolpert & Landy 2012). This has been explored primarily in the domain

of reaching actions. Task constraints shape the way sensory feedback acts to control movements (Liu & Todorov 2007, Todorov & Jordan 2002). The optimal control law for any given task tends to minimize variance in task-relevant dimensions at the expense of increasing it in irrelevant dimensions. Thus, Knill et al. (2011) showed that subjects adjusted for apparent perturbations in their reaches to touch targets oriented vertically or horizontally, and the adjustments were greatest for perturbations aligned with the narrow dimension of the target rather than for perturbations along the long dimension. The brain adjusts its feedback control law for individual movements to fit specific momentary task demands. The uncontrolled manifold approach postulates similar task-specific control of variability (e.g., Latash et al. 2002). Thus, specification of costs/value of specific actions is extremely dependent on the task context, and optimal systems take the specific task goals into account, in addition to the reliability of the sensory and motor signals. This latter issue is addressed in the next section.

In summary, exploration of natural behavior has revealed the critical role of tasks in the timing and location of gaze. Tasks specify the context for sensorimotor decisions by defining the cost function. A large body of work has explored the role of the neural reward machinery in action choices, and reinforcement learning algorithms seem well suited to modeling the pervasive effects of reward effects in behavior. Exactly how the effects of intrinsic rewards play out in ongoing natural behavior remains to be explored.

3. BEHAVIORAL GOALS DETERMINE WHAT STATE NEEDS TO BE UPDATED

Unlike most of the neurophysiological experiments on eye movements and reward, individual eye movements are not directly rewarded in natural behavior, and reward alone is not sufficient to understand why something becomes a gaze target (Gottlieb 2012, Hayhoe & Ballard 2014). The fundamental role of vision is to update information about the state of the scene, and the task determines what aspects of that state are relevant to the current goals, and therefore potentially rewarding. What controls this information gathering process?

3.1. Role of Uncertainty

The natural world is highly complex, dynamic, and often unpredictable, so there are many sources of uncertainty about its current state. For example, the lower resolution of the peripheral retina introduces uncertainty in the evaluation of visual evidence and indeed is the reason for fixating a target in many cases. If a subject has fixated a target and subsequently shifts attention, the information will be held in working memory, which decays over time. Further uncertainty will be introduced as the observer moves relative to the environment because, in the absence of overt attention, locations will need to be updated by an estimate of the change in viewer location in the environment. The need to include uncertainty to explain gaze choices stems from the fact that the optimal action choice is unclear if the state is uncertain (Sprague et al. 2007). Not all information needs to be updated because the task-relevant information will be most critical. Thus, we might expect that uncertainty is an important factor in both what gets fixated and when the fixation occurs.

There is mixed evidence for the role of uncertainty reduction in the choice of gaze target. Najemnik & Geisler (2005, 2008) showed that in the context of visual search for a simple pattern in noise, fixations appear to be chosen to reduce uncertainty. Similarly, fixations were governed by entropy reduction in a shape discrimination task (Renninger et al. 2007). Sullivan et al. (2012) and Tong et al. (2017) found that adding noise to a task-relevant visual target increased fixation

frequency, and in the latter case, it was demonstrated that the increased fixations were related to improved avoidance and interception performance. Other work by Verghese and colleagues, however, found that observers do not select targets to minimize uncertainty in a search for multiple targets, especially when under time pressure (Verghese 2012, Ghahghaei & Verghese 2015). In these investigations, subjects performed a search task in noise for multiple targets, and it was observed that subjects choose the most likely location as often as the most uncertain location. There are several factors that might account for the disparate results, such as the way that the information in the fixation is used for subsequent behavior and the role of implicit rewards. Ghahghaei & Verghese noted that the effects of uncertainty become stronger as subjects have longer to make their decisions. [A similar time dependence is observed with reward effects (Schütz et al. 2012)]. Thus, the specific experimental context, including the nature of the behavioral goals and implicit rewards, may need to be taken into account when evaluating the importance of uncertainty in gaze target choice [see also Ackermann & Landy (2013)].

The importance of information acquisition or uncertainty reduction in gaze behavior is supported by evidence for neural coding of the value of information. Bromberg-Martin & Hikosaka (2009) demonstrated that monkeys choose to look at an informative cue that signals upcoming reward when they are given a choice between an informative visual cue that was not itself rewarding and a random cue that provided no information. They recorded from midbrain dopaminergic neurons during this behavior and found large modulations of activity depending on the informativeness of the cue, suggesting that the information itself was rewarding. A more recent behavioral study by Daddaoua et al. (2016) showed that monkeys are intrinsically motivated to view reward-predictive cues independently of the sensory salience or operant rewards associated with the cues. They showed this motivation is shaped by a drive to reduce uncertainty as well as by other factors, such as conditioned reinforcement from positive cues and the cost of sampling information. Recent results suggest that LIP neurons integrate signals of prior reward probability, prior reward uncertainty, and posterior expected value for mediating the active information seeking behavior observed by Daddaoua et al. (Foley et al. 2017).

3.2. Modeling the Effect of Uncertainty and Reward

These ideas about the role of task, reward, and uncertainty have been drawn together in a formal model by Sprague et al. (2007). This model is illustrated in **Figure 1**. They suggest that complex behavior can be broken down into modules, or subtasks, each of which requires specific information from the visual image in order to perform the actions required for those tasks, as shown in **Figure 1a**. Thus, information about obstacle location relative to the observer is required for an avoidance action, for example. Each subtask has an associated reward value that reflects the importance of the behavior for the agent and allows the agent to learn how to arbitrate between the competing tasks to maximize the expected rewards learned using RL. At a given moment, the subject acquires a particular piece of information for a module, using gaze (e.g., locates the nearest obstacle), then takes an action (chooses avoidance path), and then decides what module should get gaze next. When a particular module is updated with information from gaze, as shown for Module 1 in **Figure 1b**, the new sensory information reduces uncertainty about the state of the environment relevant to that module (e.g., location of an obstacle). The next action is chosen on the basis of the learned reward value associated with that action when in a given state by summing the rewards for all modules (❹). If a module's state is not updated, it is assumed that uncertainty about that state grows. For example, if an agent has just looked at an obstacle and updated its location, uncertainty about state relevant to another task, such as location with respect to the goal, is estimated from previous fixations, and uncertainty about that state grows with time. As a

a **Decomposing behavior into subtasks**

b
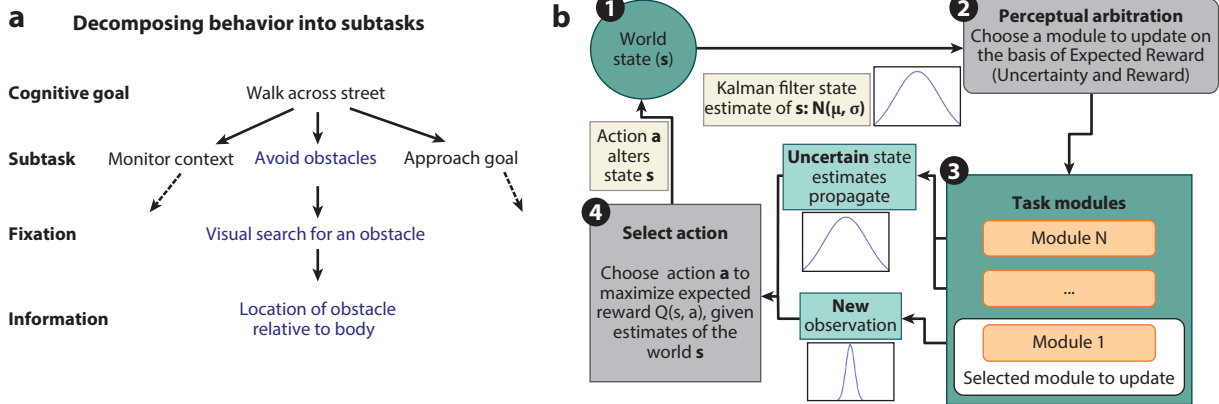
Figure 1

(*a*) Complex behavior may be broken down into subtasks or modules. This is consistent with observations of natural behavior and makes the problem computationally tractable by reducing the size of the state space. Once a subtask is chosen, the subject can search for relevant information, fixate that location, and then acquire task-specific information. (*b*) Flow diagram of the task-module architecture. Each task module keeps an estimate of its task-relevant state. Actions are chosen on the basis of the sum of reward that would be gained by all the modules. Better state estimates lead to better action choices. In the absence of direct observation, state uncertainty grows with time. The system uses gaze to update the state for the module that has the highest expected cost due to uncertainty. Kalman filters propagate state information in the absence of gaze updates. Adapted from Hayhoe & Ballard (2014).

consequence of the action (e.g., moving in a particular direction), the state of the world is changed (❶), and the agent must decide which module's state should be updated next by gaze (❷).

Although this model is undoubtedly an oversimplification, it has some interesting aspects. For example, the proposal that complex behavior can be broken down into modules is quite consistent with the appearance of natural behavior, where subjects appear to step through tasks with task-directed fixations, as described above. One hypothesis incorporated in this model is that the module that is given control of gaze is chosen on the basis of both the current uncertainty about the task-relevant information and the reward associated with the task. In this formalization, one visual subtask or module accesses new information at each time step, and all other tasks rely on noisy memory estimates. As described above, there is evidence that the probability of a change in gaze increases as uncertainty increases (Sullivan et al. 2012, Johnson et al. 2014, Tong et al. 2017). In a more extensive exploration of how gaze probability is modulated by uncertainty, Hoppe & Rothkopf (2016) devised an experiment where subjects had to detect an event occurring at a variable time in either of two locations. The event could not be detected unless the subject was fixating the location, and the subjects learned to adjust the timing of the saccades between the locations in an optimal manner. Subjects readily learned the temporal regularities of the events and traded off event detection rate with the behavioral costs of carrying out eye movements. Thus, humans appear to learn the temporal properties of uncertain environmental events and use these estimates to determine the precise moment to make a gaze change.

Although growth of uncertainty about task-relevant information appears to initiate a gaze change, there is also evidence for the complementary claim that other tasks rely on memory estimates when the associated uncertainty is low. This was shown in experiments by Droll et al. (2005) and Droll & Hayhoe (2007), illustrated in **Figure 2**. In those experiments, subjects picked up virtual blocks on the basis of a feature such as color and then sorted them on the basis of either the same or a different feature. On some trials, the color was changed during the saccade after the block was picked up. When subjects were cued to place the block on the left or right
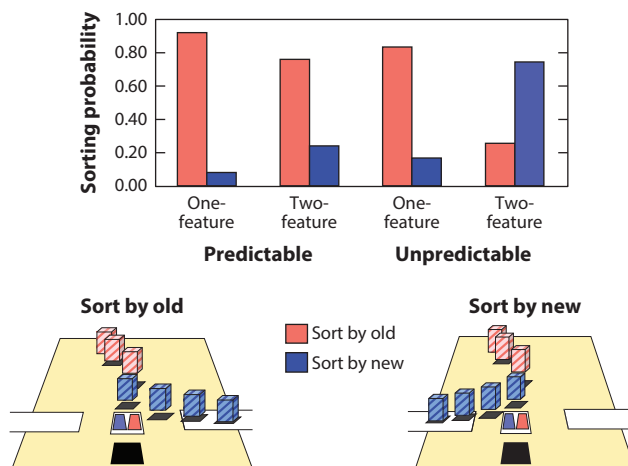
Figure 2

Subjects picked up virtual blocks and sorted them onto the left or right conveyor belt according to their color. On some trials, the block color was changed during a saccade. Despite this, subjects frequently sorted on the basis of the original color rather than the current color, even when directly fixating the block while placing it on the belt. The figure shows the probability of sorting the block on the basis of the original color (*pink bars*) or the new color (*blue bars*) on those trials when the change in the feature was not detected. Subjects frequently sorted by the old color except when there was a greater working memory load in the two-feature, unpredictable condition. Thus, the color information acquired when picking up was frequently not updated to the new state except in the high memory load condition. Adapted from Droll & Hayhoe (2007), figure 6.

depending on its color, they frequently acted as if the block was the original color, presumably held in visual working memory, and not the actual color of the block on the retina. This occurred more frequently in conditions that encouraged subjects to use working memory and less frequently in conditions when subjects made more frequent refixations of the blocks. Thus, in Predictable trials with only one relevant feature, subjects picked up blocks on the basis of their color and also sorted them on the basis of color on every trial. In the Unpredictable trials with two relevant features, subjects always picked up the block on the basis of a feature, such as color, but sorted on the basis on any of four features, and subjects did not know which feature until they looked at a cue following pickup. Consequently, there was a heavier memory load in this condition and subjects frequently waited until after pickup to look at the block in hand to get the relevant information. Given that the increased memory load will also increase uncertainty about the block features, it appears that subjects use memory representations when they have low uncertainty about the state of the information, but use gaze to update state when they are more uncertain. This flexible, context dependent use of memory versus immediately available information is an important feature of natural visually guided behavior.

To summarize, the need to update information about task-relevant, potentially rewarding state is important in determining the location and timing of gaze changes, although it is not the only factor. There is some evidence to suggest that working memory representations are used if they are reliable enough, thus obviating the need for a gaze change. The trade-off between memory and gaze deserves further exploration.

4. ROLE OF MEMORY

Information for action decisions can be made on the basis of current sensory data or a memory representation, or some weighted combination of these. One advantage of investigating natural

behavior is that subjects are immersed in a relatively stable environment where they have the opportunity to develop long-term memory representations. This section reviews evidence that memory is an important factor in targeting eye and body movements and suggests that memory and visual information are combined in a Bayesian manner.

4.1. Visual Search

When subjects search for targets in two-dimensional (2D) images of natural scenes, there is considerable evidence that memory influences the guidance of attention and gaze in the scene (Brockmole et al. 2006, Hollingworth 2009, Võ & Henderson 2010). The information a subject gathers from a scene will depend on existing memory representations, or prior knowledge, of the scene. For example, familiarity with a scene speeds subsequent visual search (Hollingworth 2012). Depending on the particular circumstances, the influence of memory may be from long-term memory encoding of scene semantics, rather than short-term visual memory (Võ & Wolfe 2012). However, there are many differences between ordinary experiences in natural, immersive 3D environments and the stimuli presented in experiments with 2D images on computer monitors, even when those images are taken from realistic scenes. For example, the spatial scale of the images is much compressed relative to real images, and the need to move the body places different constraints on the search process. These factors are likely to be important and give priority to use of memory rather than visually guided search.

Visual search in more ecologically valid conditions also shows facilitation from memory. Kit et al. (2014) showed that visual search performance in an immersive virtual environment improves rapidly over repeated search episodes. Mack & Eckstein (2011) had participants search for objects on the tabletop and demonstrated that co-occurrence of objects, reflecting scene priors, can serve as a contextual cue for guiding search. Jiang et al. (2014) found that in an outdoor environment, whole-body movements influence memory representations by allowing subjects to encode target locations in both egocentric and allocentric frames. Foulsham et al. (2014) also suggest that head movements are important factors in search strategies in real environments. Li et al. (2016) compared search in 2D and 3D environments and found that although the influence of memory was similar in many respects, the costs of moving the body around influenced behavior in the 3D environment. In particular, subjects rapidly encode the global structure of the space in a coarse to fine fashion, so that they can eliminate regions where targets are unlikely to be and confine search to more probable regions.

4.2. Targeting Saccades

The need to orient to regions outside the field of view in natural vision (for example, moving around within a room) provides a rationale for storing information about spatial layout. For example, Land et al. (1999) noted instances when subjects made a number of very large gaze shifts to locations outside the field of view. These gaze shifts involved a combination of eye, head, and body movements and were remarkably accurate. When objects are within the field of view, subjects have the choice of searching for a target on the basis of its visual features so may not need to use memory. Evidence has accumulated that memory is indeed used in this instance. Experiments by Epelboim et al. (1995) provide evidence that saccade targeting is facilitated by memory in tasks such as tapping a sequence of lights in known positions. Aivar et al. (2005) showed that in a task where subjects built a toy model, saccades were sometimes made to the remembered locations of targets that had subsequently been moved to new locations, revealing that subjects often planned saccades on the basis of a memory representation, even in the presence of conflicting visual information,

a Before the room was visible
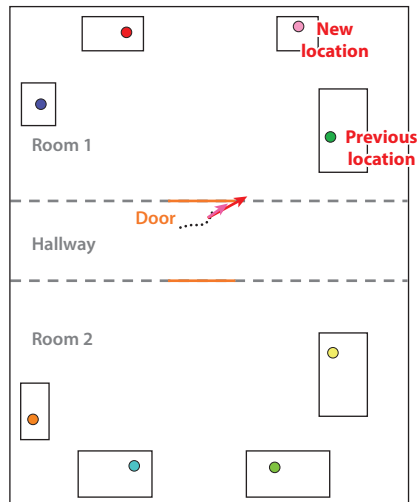
b First fixation after the room was visible
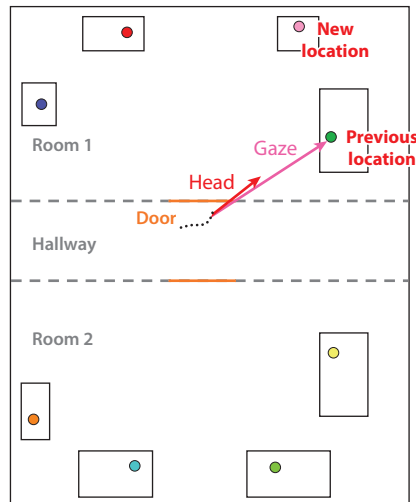
Figure 3

Bird's-eye view of a virtual apartment with two rooms and a hallway separating them. The subject is moving from the corridor into the bedroom to search for a target that has previously been located and whose spatial position has been learned. The black dots show the subject's path as he walks from the hallway into the bedroom at the top of the figure. The pink arrow shows gaze direction, and the red arrow shows head orientation. The green dot is the location of the target during previous search trials. (*a*) The subject orients to the old location of the target even before entering the room even when the target is not visible. (*b*) The subject fixates the old target location after room entry, even though the target is no longer in that location and has been moved to the location indicated by the pink dot. The head orientation and gaze direction must be targeted purely on the basis of memory. Unpublished data by the author, collected in an experiment described in Li et al. (2016).

and then had to make corrective movements. A similar example can be seen in **Figure 3**, where subjects searched for targets in a virtual apartment. After searching for the target on three separate occasions, the target was moved to another location. The figure shows the head and eye directed at the old target location even before the subject entered the room. The data revealed that subjects look at the old location on 58% of trials (Aivar et al. 2016). Individuals with restricted fields of view resulting from damage to early visual cortex (homonymous hemianopia) make accurate saccades into their blind fields when performing real 3D model-building tasks analogous to that used by Aivar et al. (2005) (Martin et al. 2007). Such saccades must be memory based. Subjects do not need to make a sequence of hunting movements in the general region but instead locate targets in the blind field as efficiently as in their normal hemifield.

4.3. Bayesian Weighting of Memory and Vision in Targeting Reaches

A targeting mechanism that relies heavily on spatial memory does not differentiate between targets inside or outside the current field of view. A strategy that uses both visual and memory information, depending on what is available, would ensure a smooth transition between targeting within and outside the field of view. It seems likely that the spatial information from memory and the visual information from the peripheral retina are combined in some way to specify the target location. The relative weights of these two sources of information should indicate the strength of the reliance
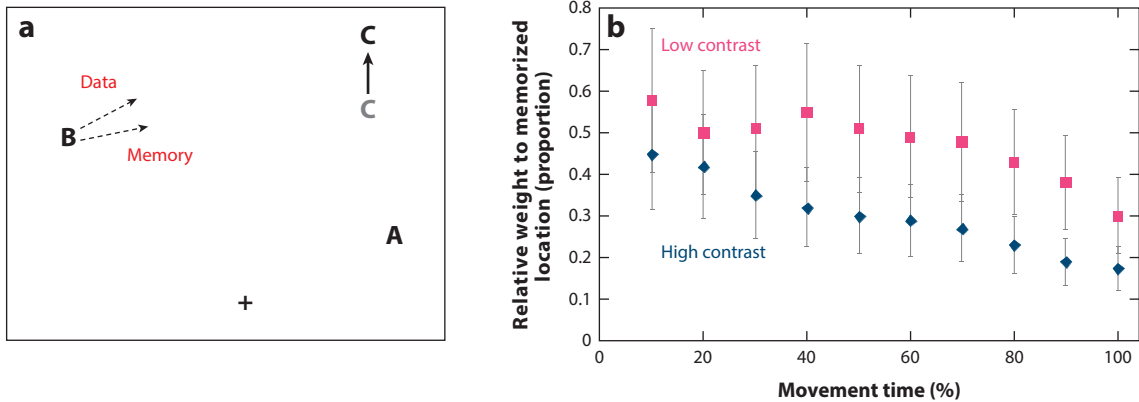
Figure 4

(*a*) Brouwer & Knill's task: The subject picks up a target at location A and moves it to location B. On some trials, a second object previously at location C was shifted during the movement. The movement to pick up the target at C was analyzed to see whether the reaching was directed at the old, remembered location or the new (visual) location, or some combination. (*b*) The weights given to the memory location are plotted for two cases, one where the target at C was low contrast (*top symbols*) and one where it was high contrast (*bottom symbols*). Adapted from Hayhoe (2009).

on the memory information. Evidence for such Bayesian cue combination was demonstrated by Brouwer & Knill (2007, 2009) in programming hand movements. They devised a task illustrated in **Figure 4**, where subjects sequentially picked up and moved two virtual target objects into a virtual trash bin with their index fingers. In some of the trials, the position of the second target was perturbed while the subject was transporting the first target to the trash. Although the new position of the second target was visible in the peripheral retina, subjects' initial movements to pick up the target were biased to the initial remembered position. For high-contrast targets, the initial part of the reach trajectory reflected a substantial weight for the remembered location of the target, as shown in **Figure 4**. Over the course of the movement, the memory weight decreased and the finger ended accurately on the new target position. When the contrast of the target was decreased, the weight given to the remembered location increased substantially. Thus, even when the target was visible in the peripheral retina, the remembered location had a role in programming the reaching movement. Separate measurement of the reliability of the peripheral visual information and the remembered position showed that they were weighted by their reliability in an optimal manner. Thus, humans combine both kinds of information in guiding reaches and take into account their relative reliability. It is also likely that the relative weights depend on the constraints in a particular situation. For example, if there is a need to minimize the time to locate the target or to initiate hand and head movements ahead of the eye, then greater reliance might be found on memory-based targeting. If, however, accuracy is most important, greater weight might be given to the current retinal image.

In summary, spatial memory is likely to be a fundamental component of movement targeting, as it allows more efficient use of attentional resources and can be shared between different effectors allowing more efficient coordination patterns.

5. THE NEED FOR PREDICTION

Another way that the role of memory in action programming is manifest is in prediction, because eye and body movements that occur in advance of the stimulus must be based on stored

knowledge of some kind. The existence of significant sensory-motor delays on the order of 100 to 200 ms presents a particular problem to a reactive organism in a dynamically changing environment. Despite these delays, humans are able to intercept moving targets with extraordinary spatial and temporal precision. It is usually assumed that the solution to this problem is to predict the future state of the environment on the basis of past experience of how the current state is likely to change over time. For example, it is commonly accepted that the proprioceptive consequences of a planned movement are predicted ahead of time using stored internal models of the body's dynamics (Wolpert et al. 1998, Shadmehr et al. 2010), and the comparison of actual and predicted somatosensory feedback is a critical component of the control of movement. Indeed, when somatosensory feedback is severely compromised by somatosensory loss, the consequences for movement can be devastating (Cole & Paillard 1995). Although prediction is often assumed to be a fundamental aspect of vision, clear examples of prediction are somewhat limited, and the extent and nature of prediction are unclear (Zago et al. 2009, Zhao & Warren 2015). It seems likely, however, that prediction plays an important role in vision, as it does in the somatosensory system, both in predicting the visual consequences of self-motion and in controlling the timing of actions in a dynamic world.

Some of the best evidence for prediction in vision comes from the oculomotor system. In this case, both smooth pursuit and saccadic eye movements reveal prediction of the future visual stimulus. Subjects attempting to pursue a target that is briefly occluded often make anticipatory movements to the point of target reappearance (Madelain & Krauzlis 2003, Orban de Xivry et al. 2006). Pursuit also incorporates memory for target motion accumulated over previous experience (Barnes & Collins 2008). In a different paradigm, Spering et al. (2011) showed that pursuit movements facilitate judgments of the future trajectory of a moving target, and in a subsequent paper, Fooken et al. (2016) showed pursuit also improves interceptive performance. Cognitive factors also play a significant role. When pursuing an object moving along a track and toward a junction with two branching target trajectories, subjects continue uninterrupted pursuit along the path implied by visible barriers or auditory cues [see Kowler (2011) for a review]. In the saccadic system, if a target is briefly occluded, subjects make predictive saccades to the expected point of target reappearance even when the occluded target reflects off of an angled line during occlusion (Ferrera & Barborica 2010).

Predictive eye movements are also robust and pervasive in natural behavior, where trajectories are more complex and prediction is presumably more difficult. Athletes playing cricket, table tennis, and squash make predictive eye movements to the ball's future location (Land & Furneaux 1997, Land & McLeod 2000, Hayhoe et al. 2012). Diaz et al. (2013b) investigated a more controlled setting using a virtual racquetball environment, where unskilled subjects intercepted a virtual ball that bounced prior to interception. Subjects made a saccade ahead of the ball, just before it bounced, to a location on the future ball trajectory. Gaze was held in this location during the bounce and until the ball passed within 1–2 degrees of the fixated location about 170 ms after the bounce. The location of the predictive saccade was dependent on the ball's elasticity as well as its velocity. The accuracy of the predictions, both in time and space, despite variation in ball properties, suggest that subjects rely at least in part on their history of experience with balls in order to target the eye movements to the ball's future location.

Eye movements during interception appear to be based on some combination of the visual information specific to a particular trajectory together with a memory-based component. In reaching, there is evidence for the Bayesian integration of current visual information with stored priors reflecting learned statistics of the visual image (Koerding & Wolpert 2004, Tassinari et al. 2006). It seems likely that a similar combination of information sources occurs with both pursuit and saccades. Diaz et al. (2013a) showed that the learned statistics can substitute for visual information

with no obvious cost to pursuit movements. These authors argue that the locus of the information used to guide predictive pursuit and saccadic eye movements is shared. Prediction has also been shown to factor into interceptive steering actions (Diaz et al. 2009).

Although the cerebellum has traditionally been thought to be the locus of internal models used for prediction, it seems likely that cortical areas are also involved in predictive eye movements. The frontal eye fields (FEF) have been implicated in the guidance of both predictive saccades and predictive smooth pursuit (Fukushima et al. 2006, Ferrera & Barborica 2010). The supplementary eye fields (SEF) are also thought to be involved with both predictive pursuit and predictive saccades (Nyffeler et al. 2008, Shichinohe et al. 2009). On the basis of the involvement of both SEF and FEF in pursuit and saccadic eye movements, it has been suggested that the locus of prediction may exist at a higher, so-called supramodal level, consistent with previous results on the role of cognitive factors in predictive pursuit (Kowler 2011). Visual cortical areas might also encode predicted target location. For example, the parietal cortex shows activity corresponding to occluded target motion (Assad & Maunsell 1995). Thus, multiple areas are likely to be involved.

5.1. Prediction Versus On-Line Control

The basis for predictive eye movements or body movements is a somewhat contentious issue. For example, Zhao & Warren (2015) argue that interception depends on some visual parameter that is monitored continuously and controlled by the interceptive action, such as maintaining constant bearing angle, and that the role of spatial memory or prediction is very modest. It may be possible to reconcile these conflicting views. Belousov et al. (2016) show that predictive and reactive strategies may be optimal and operate in different regimes depending on how much time the observer has to catch the ball, the sensory latencies, and noise in both the observation and the stored model. They model catching as a continuous, partially observable Markov decision process, and employing stochastic optimal control theory, they show that the heuristics described in the literature (such as constant bearing angle) are optimal solutions if the catcher has sufficient time to continuously visually track the ball. Specifically, by varying model parameters such as noise, time to ground contact, and perceptual latency, they show that different strategies arise under different circumstances. The catcher's policy switches between generating reactive and predictive behavior on the basis of the ratio of system to observation noise and the ratio between perceptual latency and task duration. De la Malla & López-Moliner (2015) argue for the combination of predictive mechanisms using the early part of a trajectory with later use of on-line mechanisms. They found that the integration of the different estimates of an impending interceptive event was optimal and can help explain the observed high temporal precision in natural interception.

5.2. Prediction and Visual Stability

One indication of the role of prediction is the remapping of visual receptive fields before a saccade (Duhamel et al. 1992, Melcher & Colby 2008). Predictive remapping occurs not only in LIP but also in superior colliculus, FEF, and area V3. Evidence indicates that predictive remapping is mediated by a corollary discharge signal originating in superior colliculus and mediodorsal nucleus of the thalamus. Cicchini et al. (2013) present evidence that this predictive remapping is part of a mechanism for visual stability that relates the pre- and postsaccadic images of a stimulus. They found that the perceived position of flashes around the time of a saccade is strongly influenced by the location of other flashed stimuli in a way consistent with the construction of a response independent of eye position, as required for visual stability. Similar targets that are plausibly seen as the same object are assigned the same spatial location.

Recent evidence suggests that mechanisms for discriminating self-generated motion from externally generated motion may be established very early in visual processing. Roth et al. (2016) found that LP (lateral posterior nucleus) in the mouse thalamus (thought to be the equivalent of the pulvinar in primates) projects diffusely to layer 1 in V1. In an experiment where mice ran on a treadmill and were presented with moving gratings linked to their running speed, they found evidence that the projections from LP carried information to V1 about mismatches between running speed and the speed of the gratings. It is not clear whether a similar mechanism for accounting for self-generated motion exists in primates. There is recent evidence that neurons in V1 distinguish between self-generated and externally generated image motion (Troncoso et al. 2015), and there are eye movement inputs to most visual areas. Thus, the result attests to the fundamental importance of taking account of self-generated motion and the potential role of the thalamus in this function. We now consider how a mechanism might work in order to take account of the complex transformations of the image as observers move both body and eyes through the environment.

5.3. Predicting Effects of Self-Motion and Parsing Optic Flow

A somewhat more complex problem is that of accounting for the effects of body movements. This is important not just for visual stability but for separating externally generated motion from self-motion. When observer motion is accompanied by object motion, as is frequently the case, the optic flow field includes a component due to self-motion and a component due to object motion. For moving observers to perceive the movement of other objects relative to the stationary environment, the visual system needs to recover the object-motion component—that is, it needs to factor out the influence of self-motion (see **Figure 5**). In principle, this could be achieved using visual self-motion information, nonvisual self-motion information, or a combination of both. When stationary subjects view patterns of optic flow that include a moving object, they are able to
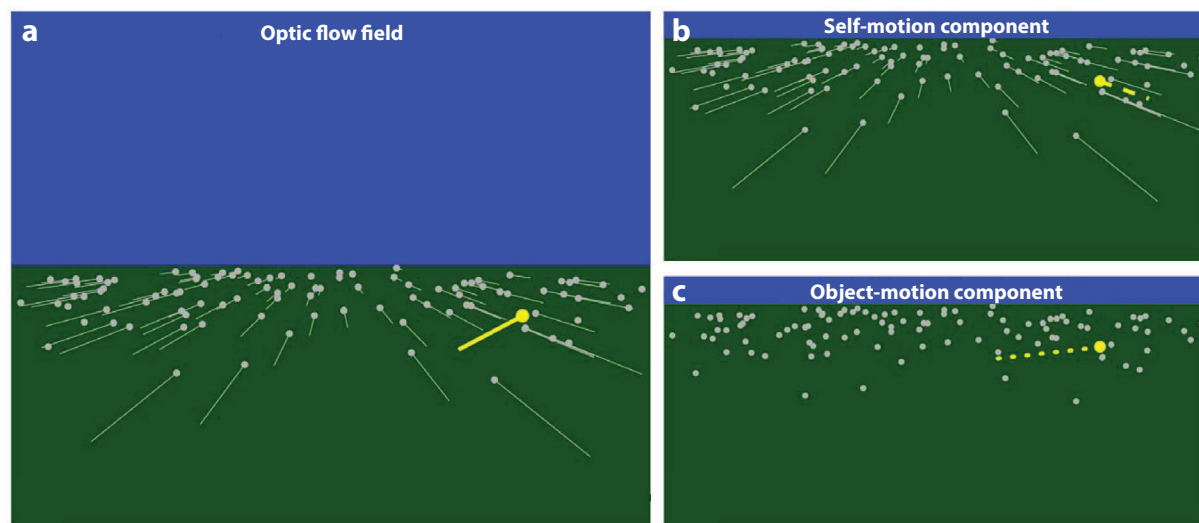
Figure 5

Optic flow field and decomposition into self-motion and object-motion components. (*a*) Optic flow field generated by an observer moving over a ground surface and an object (*yellow dot*) moving from right to left. (*b*) The component of optic flow due to self-motion independent of object motion. (*c*) The component of optic flow due to object motion independent of self-motion. The optic flow field is the vector sum of the self-motion and object-motion components. Adapted from Fajen & Matthis (2013).

parse the flow pattern into a component due to object motion and one potentially caused by self-motion (Warren & Rushton 2008, 2009a,b). However, because the subjects in those studies were not actually moving and self-motion was simulated, none of the sources of nonvisual self-motion information that contribute to the perception of self-motion (Harris et al. 2000, 2002) were present. Fajen & Matthis (2013) and Fajen et al. (2013) used an ambulatory virtual environment viewed in a head-mounted display, so that self-motion cues were veridical, and performed a task that involved judgments about how to avoid moving obstacles. They show that visual information about the speed and direction of self-motion plays a role in recovering the object-motion component even when nonvisual self-motion information is also available. However, the magnitude of the effect was less than one would expect if subjects relied entirely on purely visual self-motion information. They conclude that when self-motion is real and actively generated, both visual and nonvisual self-motion information contribute to the perception of object motion. They also argue that this process may have a role in visually guided interception and avoidance of moving objects. There is considerable evidence that humans and animals use a constant bearing angle strategy to intercept moving objects, and this does not require recovery of object motion in an allocentric reference frame, but only motion of the object relative to the observer. Fajen et al. argue, however, that this does not take into account limits on how fast the observer can move, so to be able to move flexibly with respect to a moving object, recovery of the object-motion component of optic flow may be important.

5.4. Eye Movements that Are Not Task Related

Although many aspects of gaze allocation can be accounted for by the current cognitive goal, there must also be a mechanism for attracting gaze to important information in the environment that is unrelated to the current task. A large literature has investigated the role of stimulus salience in attracting attention (e.g., Itti & Koch 2000, Zhang et al. 2008, Bruce & Tsotsos 2009). However, evidence for the role of salient stimuli in attracting attention is mixed [see reviews by Tatler et al. (2011) and Schütz et al. (2011)]. More importantly, such mechanisms are inherently brittle, as they rely on the properties of the stimulus with respect to the immediate context to attract attention. It is not known whether these mechanisms will be very effective in natural environments in which the sensory milieu and the requirements of visually guided behavior are very different. It is in this context that scene memory may play an important role. If the scene is efficiently coded in long-term memory, subjects may compare the current image with the stored representation, and a mismatch, or so-called residual signal, may serve as a basis for attracting attention to changed regions of scenes. This may allow subjects to be particularly sensitive when there are deviations from familiar scenes, and thus, attention may be drawn to regions that do not match the stored representation. This idea is similar in conception to Itti & Baldi's (2005) model of surprise but uses long-term memory representations as the basis for detecting outliers rather than the immediately preceding scene. In standard saliency models, salient stimuli are statistical outliers in space. Surprising stimuli can be thought of as statistical outliers with respect to learned, expected distributions stored in memory (cf. Rosenholtz 1999). This might be a more robust mechanism of attentional capture than purely spatial saliency. The nature of such memory representation is something of a challenge, but recent models of visual pattern representations using convolution neural nets (Kriegeskorte 2015) suggest that complex scene representations are computationally feasible. This raises the possibility that memory representations of complex scenes are stored and monitored as observers move around, so that scene changes that do not fit the statistics of the stored representations might be used to attract gaze in a task-independent manner. Thus, nontask-driven gaze changes can also be seen as leading to entropy reduction and more accurate representation of state of the environment. This

also suggests that taking account of self-generated scene transformations for visual stability and accounting for attentional capture of gaze are different aspects of the same mechanism.

In summary, the role of predicted visual state in control of actions, and the parallel with somatosensory control, needs further elucidation. There is sound evidence for prediction in both smooth pursuit and saccades and also for comparison of actual and expected effects of self-motion. The role of prediction in interceptive movements is less clear, perhaps because arm and body movements are slower than eye movements and success is critically dependent on the sensory evidence.

6. COMPARISON WITH THE APPROACH THAT PERCEPTION AND ACTION ARE SEPARATE

It is important to compare the approach taken here to the approach that emphasizes the differences between perception and action, espoused notably by Goodale among others (Goodale & Milner 2005, Goodale 2014). This approach has been the focus of a large body of research over the last two decades and is based on observations of patient D.F., who has selective damage to lateral occipital cortex, and also on differences in performance in normal subjects for tasks that are classified as either action or perception. The central claim is that there are two different visual systems, one for perception and one for action, with action being controlled by the dorsal stream, and perception by the ventral stream. This claim is motivated by the finding that D.F. can perform normal reaching and grasping actions despite being unable to make perceptual judgments, such as orientation about the grasped object. Other evidence stems from the apparent immunity of features of the grasp to perceptual size illusions. Although it is not possible to review this large literature here, there are several pertinent issues. First, it is clear that the visual pathways can be broadly segregated into dorsal and ventral streams on the basis of anatomical evidence. It is also clear from a large body of work that the dorsal pathways, including areas in the intraparietal sulcus and neighboring regions, are involved in the initiation and control of actions and that inferotemporal cortex is centrally involved in object recognition. However, it is also well established that there are multiple connections between dorsal and ventral streams. What is unclear is exactly how the two streams function in the planning and execution of actions. Goodale (2008) has defined a set of differences between the two streams, with the ventral stream being associated with perceptual awareness, using allocentric or relative coding, and sustained representations, and the dorsal stream using transient, egocentric representations that are not associated with awareness. There have been a number of critical reviews of this that suggest the pathways are not so clearly separable in their function (Brenner & Smeets 2007, Shenk & McIntosh 2010, Freud et al. 2016). An issue raised in these critiques is that many of the findings that are used to argue for two visual systems are also consistent with the fact that there are multiple kinds of computation in the brain that are highly specific to the task at hand and that computations for different perceptual or motor functions frequently appear inconsistent with each other. Thus, the ability of D.F. to make accurate reaches in spite of lack of awareness should not be interpreted as a categorical distinction between perception and action. Another issue is the existence of multiple interactions between dorsal and ventral streams (Shenk & McIntosh 2010). Although dorsal and ventral streams have different functionality, how those functions are called into play is highly task dependent [see also Spering & Carrasco (2015)].

There are a number of very complex issues involved in evaluating the relationship between dorsal and ventral pathways, as well as the way they interact when controlling movements, and it is not possible to review this literature in any depth here. Although much is known about the posterior parietal regions involved in action targeting and control, the circuit-level organization is still far from clear, and definitive answers probably await elucidation of this organization. The main concern in the present review is this: When we take the point of view outlined above—that

actions should be considered in the context of statistical decision theory—the emphasis changes from the different functionality of the dorsal and ventral streams to one where perception is an integral part of actions. In particular, we see that memory representations are a central component of action planning and execution.

Although there are a number of examples of dissociations between perception and action, it is hard to argue for a strict segregation, because complex perceptual information is essential for the control of actions. Actions require information about objects that we would assume is computed by the ventral stream (e.g., picking up a tool requires memory for experience with the tool, how to hold it and position it for use, etc.). Information about the weight of an object is used to plan grasp and lift forces (Gordon et al. 1991, Flanagan & Beltzner 2000). This can be demonstrated when an object is unexpectedly light and the lifting action overshoots. Similarly, surface properties such as roughness or slipperiness influence both grasp kinematics and grasp configuration (Paulun et al. 2016), and reaches decelerate more for light than for heavy objects at the end of the reach (Pollick et al. 2001). To account for this kind of influence in the context of the independence of actions from ventral stream influence, Goodale distinguishes between visually guided and memory-guided actions (Milner & Goodale 2008, Goodale 2014). The suggestion is that visually guided actions are controlled by the dorsal pathways alone and that only memory-guided actions reflect ventral influence. Whether such a sharp distinction can be made is unclear, and the work reviewed here takes the view that the signals guiding movements can be either purely visual, purely memory-based, or some weighted combination. Thus, the Brouwer & Knill (2007, 2009) experiments described above show an optimally weighted combination of memory and visual information, depending on their relative reliability, with the visual information being weighted more as the hand neared the target. Saccades also show a role for memory information even in the presence of a visible target (Aivar et al. 2005, Li et al. 2016).

Other work also suggests that the perception/action segregation is difficult to maintain. For example, Hesse & Franz (2009, 2010) examined whether kinematics of movements performed after a delay were different from movements performed while targets were visible, as predicted by the hypothesis that the control mechanisms are distinct. They found that grasping kinematics did not change abruptly, depending on whether visual information was present during the movement programming phase. Instead, the authors suggested that there is a gradual (exponential) decay of visuomotor information over time. In addition, a shift from dorsal to ventral pathways is not observed in fMRI. This suggests that changes in kinematics result from increased noise, consistent with the Bayesian cue combination predictions (Hesse & Franz 2009). Thus, it appears that both kinds of information—memory-based and visual—can be intimately linked to on-line control. An argument for the importance of the role of memory is that it allows body, head, and limb movements, which are relatively slow, to be initiated ahead of an eye movement to the target. When the target is foveated, the movement can be refined as needed, but early initiation of the slower effectors presumably leads to greater efficiencies in time or energetic cost (Hayhoe 2009).

In summary, statistical decision theory provides a context for understanding the visual control of action that focuses on the integration of perception and action rather than their segregation. Understanding the circuit-level control mechanisms, how they evolve over time, and how they vary for specific actions, stimulus conditions, and tasks seems like the way forward.

7. CONCLUDING COMMENTS

The review has considered only a subset of the very large literature on perception and action and has focused instead on behavior as a natural organizing principle. Consideration of the task context helps integrate some of the disparate topics under this rubric. I have argued for the

need to situate visually guided action in its behavioral context. By doing this, we can appreciate how common principles govern disparate task contexts—namely, reward, uncertainty reduction, and memory. The approach to understanding visually guided actions in the context of statistical decision theory has been extraordinarily fruitful. Costs and benefits are a critical consideration in action choices and govern whether an eye movement is made, its peak velocity, and how it is coordinated with head and body. These mechanisms work on fast time scales of 100 ms or so to guide action decisions seamlessly and produce extraordinarily stable adaptive behavior. Evaluation of task-relevant state is a central factor necessary for optimal action choices and governs a very large proportion of gaze changes. When reliable information is present in memory, the need for sensory updates is reduced and humans can rely instead on memory estimates, depending on their precision, and combine sensory and memory data according to Bayesian principles. It is suggested that visual memory representations are critically important not only for choosing, initiating, and guiding actions but also for predicting their consequences and separating the visual effects of self-generated movements from external changes.

SUMMARY POINTS

1. In the context of natural behavior, the role of vision can be seen as gathering information about the state of the world to make good action decisions.

2. This role can be formally addressed in the context of statistical decision theory, where the costs and benefits of actions, uncertainty about the state of the world, and prior knowledge all factor into good decisions.

3. A large body of work has explored the neural reward circuitry and its role in action decisions, but the link between the reward circuitry and active visually guided behavior needs to be worked out. Reinforcement learning algorithms may help forge this link.

4. Evidence for the smooth integration of prior knowledge with current sensory data to control actions cuts across previous debates that dichotomize perception and action, and prediction versus on-line control.

5. Learning the statistical structure of scenes and how these statistics vary with self-generated motion may be of central importance in understanding visual space constancy, as well as attentional capture by novel stimuli.

FUTURE ISSUES

1. I have argued for the need to situate the investigation of visual control of action in its behavioral context. This does not preclude controlled experiments. Rather, the challenge is to devise experiments that can capture the essential elements of natural behavior.

2. A pervasive but somewhat neglected factor in visually guided action is the intrinsic reward and cost of the action. We need better strategies for estimating and manipulating the costs of natural actions.

3. One aspect of saccade control that has received relatively little attention is the timing of the movement. The critical nature of saccade timing is evident in activities such as sports (Brenner & Smeets 2015), but it is apparent even in ordinary tasks, where the time of

departure and arrival of gaze is linked with exquisite precision to the information needed at the moment and with the actions of the hands. This aspect of saccade behavior is likely to be an enormously fruitful topic.

4. How are memory representations and prediction smoothly integrated with on-line visual information to control action? This question has been only scantily researched. A related issue is how memory representations modify coordination patterns of the different effectors and whether this minimizes energetic costs.

5. This review has not addressed the critical role of learning in visual control of movements. This involves learning the statistics of visual environments, visuomotor coordination, and task-level control. This is an important topic currently under investigation (Wolpert & Flanagan 2016).

6. As we move through the environment, we generate complex image movement patterns that we effortlessly discount from externally generated motion. Understanding how we might learn the statistics of these patterns now seems possible, given the recent developments in convolution neural net algorithms. This might also help unify our understanding of mechanisms that stabilize the visual scene in the presence of the different kinds of image motion resulting from eye, head, and body movements. These learned statistics might also explain how gaze can be attracted to unexpected stimuli that are not task relevant.

DISCLOSURE STATEMENT

ACKNOWLEDGMENTS

LITERATURE CITED

Ackermann J, Landy M. 2013. Choice of saccade endpoint under risk. *J. Vis.* 13(3):27

Aivar MP, Hayhoe MH, Chizk CL, Mruczek REB. 2005. Spatial memory in saccadic targeting in a natural task. *J. Vis.* 5(3):3

Aivar MP, Li C-L, Tong MH, Kit D, Hayhoe M. 2016. Acquisition and persistence of location information over the time course of natural actions. *J. Vis.* 16(12):353

Assad J, Maunsell J. 1995. Neuronal correlates of inferred motion in primate posterior parietal cortex. *Nature* 373:518–21

Barnes GR, Collins CJS. 2008. Evidence for a link between the extra-retinal component of random-onset pursuit and the anticipatory pursuit of predictable object motion. *J. Neurophysiol.* 100:1135–46

Belousov B, Neumann G, Rothkopf C, Peters J. 2016. Catching heuristics are optimal control policies. *Adv. Neural Inf. Process. Syst. Proc.* 29:1–9

Brenner E, Smeets J. 2007. Insights about seeing. *Cortex* 43:271–74

Brenner E, Smeets J. 2015. How people achieve their amazing temporal precision in interception. *J. Vis.* 15(3):8

Brockmole JR, Castelhano MS, Henderson JM. 2006. Contextual cueing in naturalistic scenes: global and local contexts. *J. Exp. Psychol. Learn. Mem. Cogn.* 32:699–706

Bromberg-Martin ES, Hikosaka O. 2009. Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron* 63:119–26

Brouwer A, Knill D. 2007. The role of memory in visually guided reaching. *J. Vis.* 7(5):6

Brouwer A, Knill D. 2009. Humans use visual and remembered information about object location to plan pointing movements. *J. Vis.* 9(1):24

Bruce NDB, Tsotsos JK. 2009. Saliency, attention, and visual search: an information theoretic approach. *J. Vis.* 9(3):5

Cicchini GM, Binda P, Burr DC, Morrone MC. 2013. Transient spatiotopic integration across saccadic eye movements mediates visual stability. *J. Neurophysiol.* 109:1117–25

Cole J, Paillard J. 1995. Living without touch and peripheral information about body position and movement: studies with a deafferented subject. In *The Body and the Self*, ed. JL Bermúdez, A Marcel, J Eilan, pp. 245–66. Cambridge, MA: MIT Press

Daddaoua N, Lopes M, Gottlieb J. 2016. Intrinsically motivated oculomotor exploration guided by uncertainty reduction and conditioned reinforcement in non-human primates. *Sci. Rep.* 6:20202

de la Malla C, López-Moliner J. 2015. Predictive plus online visual information optimizes temporal precision in interception. *J. Exp. Psychol. Hum. Percept. Perform.* 41:1271–80

Della Liberia C, Chelazzi L. 2009. Learning to attend and to ignore is a matter of gains and losses. *Psychol. Sci.* 20:778–84

Diaz GJ, Cooper J, Hayhoe M. 2013a. Memory and prediction in natural gaze control. *Philos. Trans. R. Soc. B* 368:0064

Diaz GJ, Cooper J, Rothkopf C, Hayhoe M. 2013b. Saccades to future ball location reveal memory-based prediction in a natural interception task. *J. Vis.* 13(1):20

Diaz GJ, Phillips F, Fajen BR. 2009. Intercepting moving targets: a little foresight helps a lot. *Exp. Brain Res.* 195:345–60

Diedrichsen J, Shadmehr R, Ivry R. 2009. The coordination of movement: optimal feedback control and beyond. *Trends Cogn. Sci.* 14:31–39

Dorris MC, Glimcher PW. 2004. Activity in posterior parietal cortex is correlated with the subjective desirability of an action. *Neuron* 44:365–78

Droll J, Hayhoe M. 2007. Deciding when to remember and when to forget: trade-offs between working memory and gaze. *J. Exp. Psychol. Hum. Percept. Perform.* 33:1352–65

Droll J, Hayhoe M, Triesch J, Sullivan B. 2005. Task demands control acquisition and maintenance of visual information. *J. Exp. Psychol. Hum. Percept. Perform.* 31:1416–38

Duhamel J, Colby CL, Goldberg ME. 1992. The updating of the representation of visual space in parietal cortex by intended eye movements. *Science* 255:90–92

Epelboim J, Steinman R, Kowler E, Edwards M, Pizlo Z, et al. 1995. The function of visual search and memory in sequential looking tasks. *Vis. Res.* 35:3401–22

Fajen BR, Matthis JS. 2013. Visual and non-visual contributions to the perception of object motion during self-motion. *PLOS ONE* 8:e55446

Fajen BR, Parade MS, Matthis JS. 2013. Humans perceive object motion in world coordinates during obstacle avoidance. *J. Vis.* 13(8):25

Ferrera VP, Barborica A. 2010. Internally generated error signals in monkey frontal eye field during an inferred motion task. *J. Neurosci.* 30:11612–23

Flanagan J, Beltzner M. 2000. Independence of perceptual and sensorimotor predictions in the size-weight illusion. *Nat. Neurosci.* 3:737–41

Foley NC, Kelly SP, Mhatre H, Lopes M, Gottlieb J. 2017. Parietal neurons encode expected gains in instrumental information. *PNAS* 114:3315–23

Fooken J, Yeo S-H, Pai DK, Spering M. 2016. Eye movement accuracy determines natural interception strategies. *J. Vis.* 16(14):1

Foulsham T, Chapman C, Nasiopoulos E, Kingstone A. 2014. Top-down and bottom-up aspects of active search in a real-world environment. *Can. J. Exp. Psychol.* 68:8–19

Franchak JM, Adolph KE. 2010. Visually guided navigation: head mounted eye tracking of natural locomotion in children and adults. *Vis. Res.* 50:2766–74

Franklin DW, Wolpert D. 2011. Computational mechanisms of sensorimotor control. *Neuron* 72:427–42

Freud E, Plaut DC, Behrmann M. 2016. "What" is happening in the dorsal visual pathway. *Trends Cogn. Sci.* 20(10):773–84

Fukushima J, Akao T, Kurkin S, Kaneko C, Fukushima K. 2006. The vestibular-related frontal cortex and its role in smooth-pursuit eye movements and vestibular-pursuit interactions. *J. Vestib. Res.* 16:1–22

Ghahghaei S, Verghese P. 2015. Efficient saccade planning requires time and clear choices. *Vis. Res.* 113:125–36

Glimcher PW. 2011. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *PNAS* 108(Suppl. 3):15647–54

Glimcher PW, Camerer CF, Fehr E, Poldrack RA. 2009. *Neuroeconomics: Decision Making and the Brain*. London: Acad. Press

Goodale MA. 2008. Action without perception in human vision. *Cogn. Neuropsychol.* 25:891–919

Goodale MA. 2014. How (and why) the visual control of action differs from visual perception. *Proc. R. Soc. B* 281:20140337

Goodale M, Milner D. 2005. *Sight Unseen: An Exploration of Conscious and Unconscious Vision*. Oxford, UK: Oxford Univ. Press

Gordon A, Forssberg H, Johansson R, Westling G. 1991. Visual size cues in the programming of manipulative forces during precision grip. *Exp. Brain Res.* 83:477–82

Gottlieb J. 2012. Attention, learning, and the value of information. *Neuron* 76:281–95

Harris LR, Jenkin M, Zikovitz DC. 2000. Visual and non-visual cues in the perception of linear self motion. *Exp. Brain Res.* 135:12–21

Harris LR, Jenkin M, Zikowitz D, Redlick F, Jaekl P, et al. 2002. Simulating self-motion I: cues for the perception of motion. *Virtual Real.* 6:75–85

Hayhoe M. 2009. Visual memory in motor planning and action. In *Memory for the Visual World*, ed. J Brockmole, pp. 117–39. Hove, UK: Psychol. Press

Hayhoe M, Ballard D. 2005. Eye movements in natural behavior. *Trends Cogn. Sci.* 9(4):188–93

Hayhoe M, Ballard D. 2014. Modeling task control of eye movements. *Curr. Biol.* 24:622–28

Hayhoe M, Chajka K, McKinney T, Pelz J. 2012. Prediction in saccadic eye movements. *Exp. Brain Res.* 217:125–36

Hayhoe M, Shrivastrava A, Mruczek R, Pelz J. 2003. Visual memory and motor planning in a natural task. *J. Vis.* 3(1):6

Hesse C, Franz VH. 2009. Memory mechanisms in grasping. *Neuropsychologia* 47(6):1532–45

Hesse C, Franz VH. 2010. Grasping remembered objects: exponential decay of the visual memory. *Vis. Res.* 50:2642–50

Hikosaka O, Nakamura K, Nakahara H. 2006. Basal ganglia orient eyes to reward. *J. Neurophysiol.* 95:567–84

Hollingworth A. 2009. Two forms of scene memory guide visual search: memory for scene context and memory for the binding of target object to scene location. *Vis. Cogn.* 17:273–91

Hollingworth A. 2012. Task specificity and the influence of memory on visual search: comment on Võ and Wolfe 2012. *J. Exp. Psychol. Hum. Percept. Perform.* 38:1596–603

Hoppe D, Rothkopf C. 2016. Learning rational temporal eye movement strategies. *PNAS* 113:8332–37

Itti L, Baldi P. 2005. Bayesian surprise attracts human attention. In *Advances in Neural Information Processing Systems 18*, ed. Y Weiss, PB Schölkopf, JC Platt, pp. 547–54. Cambridge, MA: MIT Press

Itti L, Koch C. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vis. Res.* 40:1489–506

Jiang YV, Won B-Y, Swallow KM, Mussack DM. 2014. Spatial reference frame of attention in a large outdoor environment. *J. Exp. Psychol. Hum. Percept. Perform.* 40:1346–57

Johansson R, Westling G, Bäckstrom A, Flanagan R. 2001. Eye–hand coordination in object manipulation. *J. Neurosci.* 21:6917–32

Johnson LM, Sullivan BT, Hayhoe MM, Ballard DH. 2014. Predicting human visuo-motor behavior in a driving task. *Philos. Trans. R. Soc. B* 368:1471–2970

Jovancevic J, Hayhoe M. 2009. Adaptive gaze control in natural environments. *J. Neurosci.* 29:6234–38

Kable JW, Glimcher PW. 2009. The neurobiology of decision: consensus and controversy. *Neuron* 63:733–45

Kim HF, Ghazizadeh A, Hikosaka O. 2014. Separate groups of dopamine neurons innervate caudate head and tail encoding flexible and stable value memories. *Front. Neuroanat.* 8:120

Kit D, Katz L, Sullivan B, Snyder K, Hayhoe M, Ballard D. 2014. Searching for objects in a virtual apartment: the effects of experience on scene memory. *PLOS ONE* 9:e94362

Knill DC, Bondada A, Chhabra M. 2011. Flexible, task-dependent use of sensory feedback to control hand movement. *J. Neurosci.* 31:1219–37

Knill DC, Richards W, eds. 1996. *Perception as Bayesian Inference*. New York: Cambridge Univ. Press

Koerding K, Wolpert D. 2004. Bayesian integration in sensorimotor learning. *Nature* 427:244–47

Kowler E. 2011. Eye movements: the last 25 years. *Vis. Res.* 51:1457–83

Kriegeskorte N. 2015. Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* 1:417–46

Land MF, Furneaux S. 1997. The knowledge base of the oculomotor system. *Philos. Trans. R. Soc. B* 352:1231–39

Land MF, Hayhoe M. 2001. In what ways do eye movements contribute to everyday activities? *Vis. Res.* 41:3559–66

Land MF, Lee DN. 1994. Where we look when we steer. *Nature* 369:742–44

Land MF, Mennie N, Rusted J. 1999. The roles of vision and eye movements in the control of activities of daily living. *Perception* 28:1311–28

Land MF, McLeod P. 2000. From eye movements to actions: how batsmen hit the ball. *Nat. Neurosci.* 3:1340–45

Land MF, Tatler B. 2009. *Looking and Acting: Vision and Eye Movements in Natural Behavior*. Oxford, UK: Oxford Univ. Press

Latash ML, Scholz JP, Schöner G. 2002. Motor control strategies revealed in the structure of motor variability. *Exerc. Sport Sci. Rev.* 30:26–31

Lee D, Seo H, Jung MW. 2012. Neural basis of reinforcement learning and decision making. *Annu. Rev. Neurosci.* 35:287–308

Li C-L, Aivar MP, Kit DM, Tong MH, Hayhoe M. 2016. Memory and visual search in naturalistic 2D and 3D environments *J. Vis.* 16(8):9

Liu D, Todorov E. 2007. Evidence for the flexible sensorimotor strategies predicted by optimal feedback control. *J. Neurosci.* 27:9354–68

Mack SC, Eckstein MP. 2011. Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *J. Vis.* 11(9):9

Madelain L, Krauzlis RJ. 2003. Effects of learning on smooth pursuit during transient disappearance of a visual target. *J. Neurophysiol.* 90:972–82

Maloney L, Zhang H. 2010. Decision-theoretic models of visual perception and action. *Vis. Res.* 50:2362–74

Marr D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: MIT Press

Martin T, Riley M, Kelly K, Hayhoe M, Huxlin K. 2007. Visually-guided behavior of homonymous hemianopes in a naturalistic task. *Vis. Res.* 47:3434–46

Matthis JS, Fajen BR. 2013. Humans exploit the biomechanics of bipedal gait during visually guided walking over complex terrain. *Proc. R. Soc. B* 280:20130700

Matthis JS, Fajen BR. 2014. Visual control of foot placement when walking over complex terrain. *J. Exp. Psychol. Hum. Percept. Perform.* 40:106–15

Melcher D, Colby C. 2008. Trans-saccadic perception. *Trends Cogn. Sci.* 12:466–73

Milner AD, Goodale MA. 2008. Two visual systems re-viewed. *Neuropsychologia* 46:774–85

Najemnik J, Geisler WS. 2005. Optimal eye movement strategies in visual search. *Nature* 434:387–91

Najemnik J, Geisler WS. 2008. Eye movement statistics in humans are consistent with an optimal search strategy. *J. Vis.* 8(3):4

Navalpakkam V, Koch C, Rangel A, Perona P. 2010. Optimal reward harvesting in complex perceptual environments. *PNAS* 107:5232–37

Nyffeler T, Rivaud-Pechoux S, Wattiez N, Gaymard B. 2008. Involvement of the supplementary eye field in oculomotor predictive behavior. *J. Cogn. Neurosci.* 20:1583–94

Orban de Xivry J-J, Bennett SJ, Lefèvre P, Barnes GR. 2006. Evidence for synergy between saccades and smooth pursuit during transient target disappearance. *J. Neurophysiol.* 95:418–27

Patla AE, Vickers JN. 2003. How far ahead do we look when required to step on specific locations in the travel path during locomotion? *Exp. Brain Res.* 148:133–38

Paulun V, Gegenfurtner K, Goodale M, Fleming R. 2016. Effects of material properties and object orientation on precision grasp kinematics. *Exp. Brain Res.* 234:2253–65

Pelz JB, Canosa R. 2001. Oculomotor behavior and perceptual strategies in complex tasks. *Vis. Res.* 41:3587–59

Platt M, Glimcher P. 1999. Neural correlates of decision variables in parietal cortex. *Nature* 400:233–38

Pollick F, Chizk C, Hager-Ross C, Hayhoe M. 2001. Implicit accuracy constraints in two-fingered grasps of virtual objects with haptic feedback. In *Haptic Human Computer Interactio*n, ed. M Brewster, R Murray-Smith, pp. 98–107. Lecture Notes Comput. Sci. 2058. Berlin: Springer

Renninger LW, Verghese P, Coughlan J. 2007. Where to look next? Eye movements reduce local uncertainty. *J. Vis.* 7(3):6

Rosenholtz R. 1999. A simple saliency model predicts a number of motion popout phenomena. *Vis. Res.* 39:3157–63

Roth M, Dahmen JC, Muir DR, Imhof F, Martini FJ, Hofer SB. 2016. Thalamic nuclei convey diverse contextual information to layer 1 of visual cortex. *Nat. Neurosci.* 19:299–307

Schultz W. 2000. Multiple reward signals in the brain. *Nat. Rev. Neurosci.* 1:199–207

Schütz AC, Braun DI, Gegenfurtner KR. 2011. Eye movements and perception: a selective review. *J. Vis.* 11(5):9

Schütz AC, Trommershäuser J, Gegenfurtner K. 2012. Dynamic integration of information about salience and value for saccadic eye movements. *PNAS* 109:7547–52

Seo H, Barraclough DJ, Lee D. 2007. Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cereb. Cortex* 17:110–17

Seydell A, McCann BC, Trommershäuser J, Knill DC. 2008. Learning stochastic reward distributions in a speeded pointing task. *J. Neurosci.* 28:4356–67

Shadmehr R, Smith M, Krakauer JW. 2010. Error correction, sensory prediction, and adaptation in motor control. *Annu. Rev. Neurosci.* 33:89–108

Shenk T, McIntosh RD. 2010. Do we have independent visual streams for perception and action? *Cogn. Neurosci.* 1:52–78

Shichinohe N, Akao T, Kurkin S, Fukushima J, Kaneko CRS, Fukushima K. 2009. Memory and decision making in the frontal cortex during visual motion processing for smooth pursuit eye movements. *Neuron* 62:717–32

Spering M, Carrasco M. 2015. Acting without seeing: Eye movements reveal visual processing without awareness. *Trends Neurosci.* 38:247–58

Spering M, Schütz AC, Braun DI, Gegenfurtner KR. 2011. Keep your eyes on the ball: Smooth pursuit eye movements enhance the prediction of visual motion. *J. Neurophysiol.* 105:1756–67

Sprague N, Ballard DH, Robinson A. 2007. Modeling embodied visual behaviors. *ACM Trans. Appl. Percept.* 4:11

Sprague W, Cooper E, Tosic I, Banks M. 2015. Stereopsis is adaptive for the natural environment. *Sci. Adv.* 1:e1400254

Stuphorn V, Schall JD. 2006. Executive control of countermanding saccades by the supplementary eye field. *Nat. Neurosci.* 9:925–31

Sugrue LP, Corrado GS, Newsome WT. 2004. Matching behavior and the encoding of value in parietal cortex. *Science* 304:1782–87

Sugrue LP, Corrado GS, Newsome WT. 2005. Choosing the greater of two goods: neural currencies for valuation and decision making. *Nat. Rev. Neurosci.* 6:363–75

Sullivan B, Johnson L, Rothkopf C, Ballard D, Hayhoe M. 2012. The role of uncertainty and reward on eye movements in a virtual driving task. *J. Vis.* 12(13):19

Sutton R, Barto A. 1998. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press

't Hart BM, Einhäuser W. 2012. Mind the step: complementary effects of an implicit task on eye and head movements in real-life gaze allocation. *Exp. Brain Res.* 223:233–49

Tassinari H, Hudson TE, Landy MS. 2006. Combining priors and noisy visual cues in a rapid pointing task. *J. Neurosci.* 26:10154–63

Tatler B, Hayhoe M, Land MF, Ballard D. 2011. Eye guidance in natural vision: reinterpreting salience. *J. Vis.* 11(5):5

Tatler B, Land MF. 2011. Vision and the representation of the surroundings in spatial memory. *Philos. Trans. R. Soc. B* 366:596–610

Todorov E, Jordan MI. 2002. Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* 5:1226–35

Tong M, Zhang S, Johnson L, Ballard D, Hayhoe M. 2015. Modelling task control of gaze. *J. Vis.* 15(12):784

Tong M, Zohar O, Hayhoe M. 2017. Control of gaze while walking: task structure, reward, and uncertainty. *J. Vis.* 17(1):28

Triesch J, Ballard DH, Hayhoe MM, Sullivan BT. 2003. What you see is what you need. *J. Vis.* 3(1):9

Trommershäuser J, Glimcher PW, Gegenfurtner KR. 2009. Visual processing, learning and feedback in the primate eye movement system. *Trends Neurosci.* 32:583–90

Trommershäuser J, Maloney L, Landy M. 2003. Statistical decision theory and the selection of rapid, goal-directed movements. *J. Opt. Soc. Am. A* 20:1419–33

Troncoso XG, McCamy MB, Jazi AN, Cui J, Otero-Millan J, et al. 2015. V1 neurons respond differently to object motion versus motion from eye movements. *Nat. Commun.* 6:8114

Verghese P. 2012. Active search for multiple targets is inefficient. *Vis. Res.* 74:61–71

Võ ML-H, Henderson JM. 2010. The time course of initial scene processing for eye movement guidance in natural scene search. *J. Vis.* 10(3):14

Võ ML-H, Wolfe JM. 2012. When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *J. Exp. Psychol. Learn. Mem. Cogn.* 38:23–41

Warren PA, Rushton SK. 2008. Evidence for flow-parsing in radial flow displays. *Vis. Res.* 48:655–63

Warren PA, Rushton SK. 2009a. Optic flow processing for the assessment of object movement during ego movement. *Curr. Biol.* 19:1555–60

Warren PA, Rushton SK. 2009b. Perception of scene-relative object movement: optic flow parsing and the contribution of monocular depth cues. *Vis. Res.* 49:1406–19

Wolpert DM, Flanagan JR. 2016. Computations underlying sensorimotor learning. *Curr. Opin. Neurobiol.* 37:7–11

Wolpert DM, Landy MS. 2012. Motor control is decision making. *Curr. Opin. Neurobiol.* 22:1–8

Wolpert DM, Miall RC, Kawato M. 1998. Internal models in the cerebellum. *Trends Cogn. Sci.* 2:338–47

Yasuda M, Yamamoto S, Hikosaka O. 2012. Robust representation of stable object values in the oculomotor basal ganglia. *J. Neurosci.* 32:16917–32

Yu C, Smith LB. 2017. Multiple sensory-motor pathways lead to coordinated visual attention. *Cogn. Sci.* 41(Suppl. 1):5–31

Zago M, McIntyre J, Senot P, Lacquaniti F. 2009. Visuo-motor coordination and internal models for object interception. *Exp. Brain Res.* 192:571–604

Zhang L, Tong MH, Marks TK, Shan H, Cottrell GW. 2008. SUN: a Bayesian framework for saliency using natural statistics. *J. Vis.* 8(7):32

Zhao H, Warren WH. 2015. On-line and model-based approaches to the visual control of action. *Vis. Res.* 110:190–202