Davida Teller Award Lecture 2017: What can be learned from natural behavior?

Mary M. Hayhoe

Center for Perceptual Systems,
University of Texas Austin, Austin, TX, USA ✉

The essentially active nature of vision has long been acknowledged but has been difficult to investigate because of limitations in the available instrumentation, both for measuring eye and body movements and for presenting realistic stimuli in the context of active behavior. These limitations have been substantially reduced in recent years, opening up a wider range of contexts where experimental control is possible. Given this, it is important to examine just what the benefits are for exploring natural vision, with its attendant disadvantages. Work over the last two decades provides insights into these benefits. Natural behavior turns out to be a rich domain for investigation, as it is remarkably stable and opens up new questions, and the behavioral context helps specify the momentary visual computations and their temporal evolution.

Introduction

Research in vision has always been strongly influenced by the technology available at the time. Until the 1970s, the primary device for presenting visual stimuli was the Maxwellian view optical system, which allowed precise control of stimulus size, duration, color, and luminance of patches of light. However, with only these basic parameters to control, the kinds of questions that could be asked were somewhat restricted. Vision research at the time therefore focused on early visual mechanisms, in step with the breakthroughs in retinal neurophysiology, with recording from photoreceptors and retinal ganglion cells. Maxwellian view systems required that the head be stabilized by a bite bar in order to control the retinal illuminance. Eye-tracking devices also required that the head be stabilized, and this constraint persists to a large extent in modern eye-tracking experiments, where the head is frequently stabilized with a forehead rest. The drawback of having the head fixed in space is that the repertoire of behaviors that the subject can engage in is limited. Vision is designed to function in the context of a constantly moving observer, executing goal-directed actions. While this has long been recognized, for example, in the context of the ecologically focused perception and action tradition, designing experiments to investigate vision in the context of active behavior has been quite challenging. Experimental convenience has always been a strong influence, and as display technology has become more sophisticated and eye and body monitoring in unconstrained observers has become easier, so too has the range of convenient experiments broadened. Head-mounted eye trackers have become lighter and less expensive, with higher spatial and temporal resolution. Head-mounted displays for virtual reality are now cheap and comfortable, eye tracking within virtual-reality displays has vastly improved, and realistic environments are easy to generate. Body-movement monitoring has also improved.

These technical developments lead to a variety of exciting possibilities. However, it is important to analyze just what difference it makes to investigate vision in the context of ongoing behavior, given its attendant complexities and the reduction in experimental control. What insights can be gained from doing this? I will review some of the work in my lab and others over the last two decades to gain perspective on this question. I will focus in particular on situations involving ongoing natural behavior, extending over periods of several seconds or more, where there is only limited experimental intervention. This means that we are looking at sequences of actions chosen by the subject, in contrast to the traditional trial structure controlled by the experimenter. This means that we can examine the factors that influence the transitions from one action to the next, something which is harder to get at in more controlled paradigms. Natural behavior also allows us to ask just what information is available to vision and what computations or tasks need to be performed within a given context. Again, these questions are important but hard to answer without looking at natural behavior. Of necessity, there are many large gaps in this review, and a lot of important
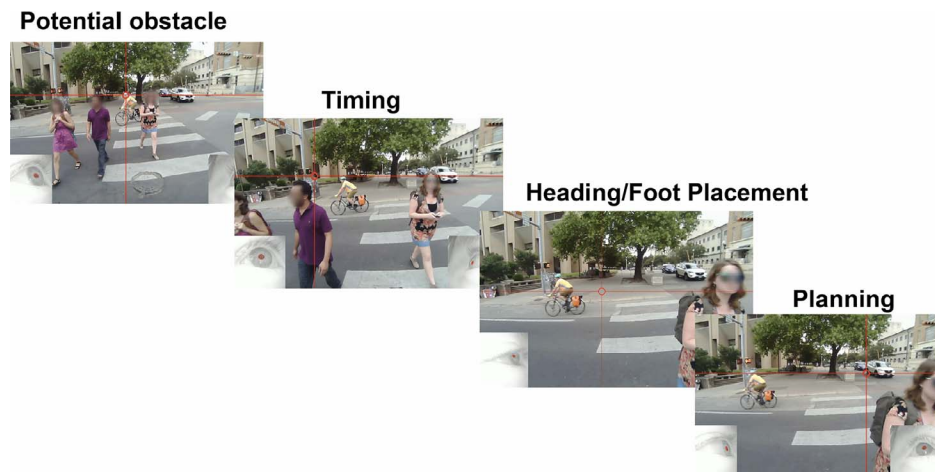
Figure 1. A sequence of gaze locations when walking across an intersection recorded using a head-mounted portable eye tracker. Gaze is shown by the red crosshairs. A possible function of each fixation is indicated above each frame.

work is not covered. A more extensive review can be found in Hayhoe (2017).

Decomposing complex behavior into tasks

I first consider how to simplify the understanding of complex behavior by breaking it down into specific task components. I will focus on gaze control, since it is a central aspect of active vision. In natural behavior, gaze is used to acquire information about the world to choose and control actions. Looking at behaviors extended in time over periods of seconds or more, different sets of questions emerge, and the behavioral context provides clues to the answers. Consider ordinary behavior such as walking across the street, illustrated in Figure 1. 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 in order to make an appropriate action choice in the moment. We can think of this as a sequence of decisions about where to look and what direction to walk. How are these decisions made? What is controlling the gaze changes? Why does gaze move from one location to another so that the walker gets the visual information she needs at the right time? This example is more challenging in some ways than a task context such as making tea or sandwiches (Land, Mennie, & Rusted, 1999; Hayhoe, Shrivastrava, Myruczek, & Pelz, 2003), where there is presumably a remembered task sequence that can guide the next

action. Thus, when one has put peanut butter on the knife, the next action would be to look at the bread, then guide the knife to the bread, and so on. These tasks clearly reveal the extent to which fixations in a scene are tightly linked to momentary behavioral goals, in both space and time. During performance of tasks like making tea or a sandwich, over 95% of the fixations can be accounted for by the task (Land et al., 1999; Land & Hayhoe, 2001; Hayhoe et al., 2003).

Tasks like walking across the street are more difficult, because the scene is less predictable and there is no obvious predetermined sequence of gaze locations. The default strategy has been to consider how properties of the visual image might attract gaze, leading to a large body of work on saliency (Borji, Itti, Liu, Musialski, & Wonka, 2013). While this will account for some fraction of gaze changes, it is not known how much of the visual information accrued in the course of everyday experience is the result of looking at salient stimuli, because much of the important information may not be particularly salient, and salient information might not be important (Tatler & Land, 2011). The approach taken here, instead, is to consider what information is needed in a task such as crossing an intersection and how gaze targets are chosen to gather that information. What determines task-driven changes in gaze location? Analysis of visually guided behavior in this way focuses on how selective attention is sequentially controlled to gather behaviorally relevant visual information.

The first issue to consider is what the particular subasks in natural behavior are and what information is required for them. This is something that we typically make assumptions about, but it requires a natural behavioral context to answer. Until we look at natural behavior, we do not know the properties of the stimulus milieu that the visual system must deal with. For example, optic flow is typically presented as a

constant velocity pattern, but recent measurements of the stimulus array during locomotion reveal complex time-varying optic-flow patterns with rhythmic accelerations and decelerations linked to gait (Matthis, Muller, Bonnen, & Hayhoe, 2017). In a similar vein, W. Sprague, Cooper, Tosic, and Banks (2015) have shown that natural-image statistics depend on the convergence distances humans choose. We need to examine both the stimulus and the linked behavior in order to be confident about what visual information is required for different aspects of locomotion, such as control of heading direction, foot placement, way finding, and so on (e.g., Fajen & Warren, 2003). Many of these questions remain unresolved. The second issue is which of several tasks to choose. That is, should a walker execute a visual search for an obstacle or check the traffic light at a particular moment? This question has been investigated more directly, and I will examine several of the factors influencing task choice in what follows.

Rewards and costs

An important factor that influences choice of gaze location is the value of the information for the current behavioral goal. It has been demonstrated that primary rewards, in the form of money or points in humans or juice in monkeys, influences eye movements in a variety of experiments (Navalpakkam, Koch, Rangel, & Perona, 2010; Gottlieb, 2012; Schütz, Trommershäuser, & Gegenfurtner, 2012). It remains to be established how to make the link between the primary rewards used in experimental paradigms and the secondary rewards that operate in natural behavior, where eye movements are for the purpose of acquiring information (Tatler & Land, 2011; Hoppe & Rothkopf, 2016; Tong, Zohar, & Hayhoe, 2017). In principle, the neural reward machinery provides an evaluation mechanism by which gaze shifts can ultimately lead to primary reward, and thus potentially allows us to understand the role that gaze patterns play in achieving behavioral goals. A general consensus is that this accounting is done by a secondary reward estimate, and a huge amount of research implicates dopamine in this role. It is now well established that cells in many of the regions involved in saccade target selection and generation are sensitive to expectation of reward, in addition to coding the movement itself (e.g., Platt & Glimcher, 1999; Sugrue, Corrado, & Newsome, 2005; Gottlieb, 2012; Yasuda, Yamamoto, & Hikosaka, 2012). There is also good evidence that the neural reward machinery acts in ways predicted by reinforcement-learning models (Schultz, 2000; Lee, Seo, & Jung, 2012). The challenge is to understand just how the rewards modulate
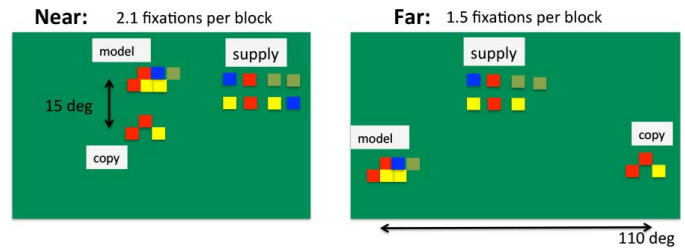

Figure 2. The layout for the block-copying task studied by Ballard et al. (1995). Subjects pick up blocks from the Supply area and make a copy of the Model pattern. When Model and Copy are separated by a distance of 110° and looking between them thus entails a head movement as shown on the right in the Far condition, the number of fixations on the Model pattern goes down from 2.1 to 1.5 fixations per block.

momentary action selection in the context of ongoing behavior.

One factor that is probably a pervasive influence on action choices is energetic cost. Matthis, Barton, and Fajen (2015) controlled the visibility of future footholds and showed that walkers need to have visual information from two steps ahead to take advantage of passive dynamics of the body, which acts like an inverted pendulum. Information from two or more steps ahead avoids braking, and so allows optimal energetic efficiency (Matthis, Barton, & Fajen, 2017). Further observations in natural outdoor walking have shown that walkers naturally choose to fixate locations that are two steps ahead, allowing minimization of energetic cost. When the terrain becomes rough, walkers also spend time looking three steps ahead, a strategy that may reflect the need to balance energetic costs with other needs such as choosing stable footholds (Matthis, Yates, & Hayhoe, 2017).

Earlier work also attests to the importance of energetic costs. Ballard, Hayhoe, and Pelz (1995) investigated a scenario where subjects copied a model made up of eight colored Duplo blocks, as shown in Figure 2. Typically, subjects make frequent looks back to the model pattern in the course of copying it. However, if the model pattern was located farther away from the location where the copy was made, separated so that a head movement was required in order to look at the model, subjects made fewer fixations on the model. This suggests that fixations on the model were more costly when a combined eye-and-head movement was required, so now memory was used more. Thus, the choice to fixate the model depended on the cost of the fixation. Subsequent work by Hardiess, Gillner, and Mallot (2008) and Solman and Kingstone (2014) has found similar results.

There are other intrinsic costs that are revealed in natural behavior. For example, Jovancevic and Hayhoe (2009) measured gaze distribution while subjects walked around a room in the presence of other
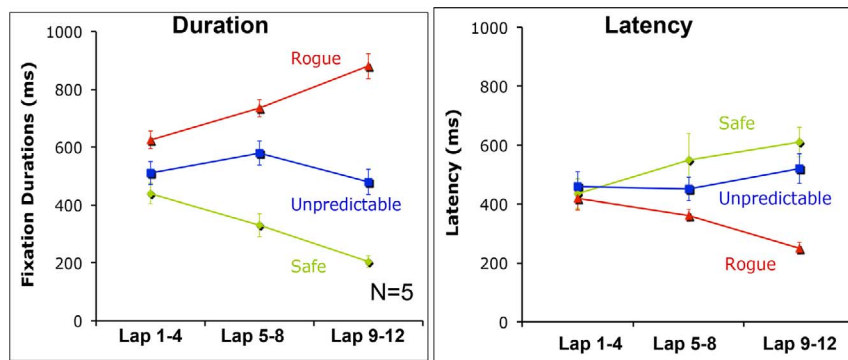
Figure 3. Fixation durations and latencies as a function of circuits around a room, for pedestrians exhibiting different behaviors. Rogues briefly veered toward the subject, Safe walkers behaved normally, and Unpredictable walkers veered 50% of the time. Error bars are ±1 standard error of the mean across five subjects. Adapted from "Adaptive Gaze Control in Natural Environments" by J. Jovancevic and M. M. Hayhoe, 2009, *Journal of Neuroscience*, *29*(19), p. 6236. Copyright 2009 by Society for Neuroscience.

walkers. Some of the walkers behaved in an unexpected and potentially hazardous manner, by briefly heading toward the subject on a collision course before reverting to a normal avoidance path. Subjects rapidly modified their gaze-allocation strategies, and the probability of fixations on these pedestrians was increased. Perhaps more importantly, the latencies and durations of these fixations also changed, as shown in Figure 3, so that fixations on the veering walkers became longer and occurred sooner after the walker appeared in the field of view. This tightly orchestrated aspect of gaze distribution suggests an underlying adaptive gaze-control mechanism that learns the statistics of the environment and allocates gaze in an optimal manner as determined by potential costs.

The point of all these examples is that the momentary costs of actions factor into sensorimotor decisions that are being made on a timescale of tens of milliseconds. Thus, whether to step to the right or left of an obstacle, how to allocate attention, and exactly when to make the movement are flexibly adjusted to satisfy global task constraints. Rothkopf and Ballard (2013) and Tong, Zhang, Johnson, Ballard, and Hayhoe (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. Thus, it seems likely that subjects learn stable values for the costs of particular actions like walking and obstacle avoidance, and that these subjective values factor into momentary action decisions. The unexpectedly low variability between subjects in many natural behaviors may be the result of a common set of costs and optimization criteria. By looking at natural behavior that extends over timescales of seconds, we can gain insight into the factors that affect momentary action choices, what the task structure might be, and what the subjective values of different actions are.

The role of state uncertainty in gaze transitions

The natural world is complex, dynamic, and unpredictable, so there are many sources of uncertainty about its current state. Consider the previously described example of crossing the street, illustrated in Figure 4. At any moment there are a number of behavioral needs competing for gaze or attention. Suppose a walker is currently looking at the location of an obstacle in order to gather information to execute an avoidance action. The previous fixation might have been in the direction of the goal, to control heading. This information will be in the peripheral retina with poor spatial resolution, so goal position with respect to the body will probably be stored in working memory, which will decay over time and will also need to be updated as the observer moves in the scene, introducing additional uncertainty. Other rele-
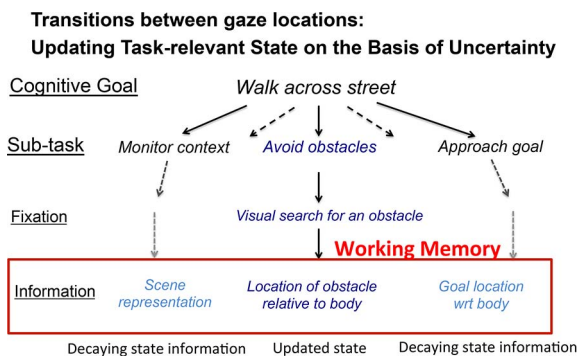

Figure 4. Schematic of task decomposition for walking across an intersection, as in Figure 1, illustrating information held in working memory following a fixation on an obstacle. Other task-relevant information is also held in working memory and decays over time.

vant information acquired previously will also need to be held in working memory and will decay over time. The choice of the next gaze location will be determined by these various uncertainties. 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 (N. Sprague, Ballard, & Robinson, 2007). Thus, the probability of a change in gaze to update state increases as uncertainty increases (Sullivan, Johnson, Rothkopf, Ballard, & Hayhoe, 2012; Johnson, Sullivan, Hayhoe, & Ballard, 2014; Tong et al., 2017).

Examination of precisely when a gaze change occurs can be revealing about the underlying mechanisms. In an exploration of how gaze probability is modulated by uncertainty, Hoppe and 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, subjects learn the temporal properties of uncertain environmental events and use these estimates to determine the precise moment to make a gaze change.

While 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 has been shown in experiments by Droll, Hayhoe, Triesch, and Sullivan (2005) and Droll and Hayhoe (2007), illustrated in Figure 5. 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 feature (color) or a different feature (e.g., size). On some trials, the color was changed during the saccade after the block was picked up, as illustrated in the figure. When subjects were cued to place the block on the left or right depending on its color, they frequently acted as if the block was the original color that it was when they picked it up. This information was presumably held in visual working memory, and it was this information—not the actual color of the block on the retina—that was used for sorting. 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. Trials when subjects picked up blocks on the basis of their color and also sorted them on the basis of color on every trial are labeled Predictable One-feature trials in Figure 5, and on these trials subjects used memory for sorting on over 90% of trials. In the trials
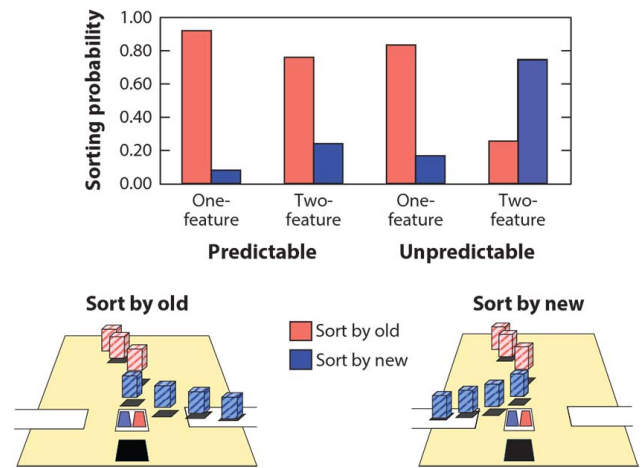

Figure 5. Subjects picked up virtual blocks and sorted them onto the left or right "conveyor belt" according to their color. On some trials the block color was changed during a saccade. Despite this, subjects frequently sorted on the basis of the original color rather than the current color, even when directly fixating the block while placing it on the belt. Thus, the color information acquired when picking up was not updated to the new state. Adapted from "Deciding When to Remember and When to Forget: Trade-Offs Between Working Memory and Gaze," by J. Droll and M. Hayhoe, 2007, *Journal of Experimental Psychology: Human Perception and Performance, 33*(6), p. 1360. Copyright 2007 by American Psychological Association.

labeled Unpredictable Two-feature, 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 did not know which feature would be needed until they looked at the placement cue after they had picked up the block. 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, so in this case they sorted on the basis of memory on only 21% of trials. Given that the increased memory load will also increase uncertainty about the block features, is 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-
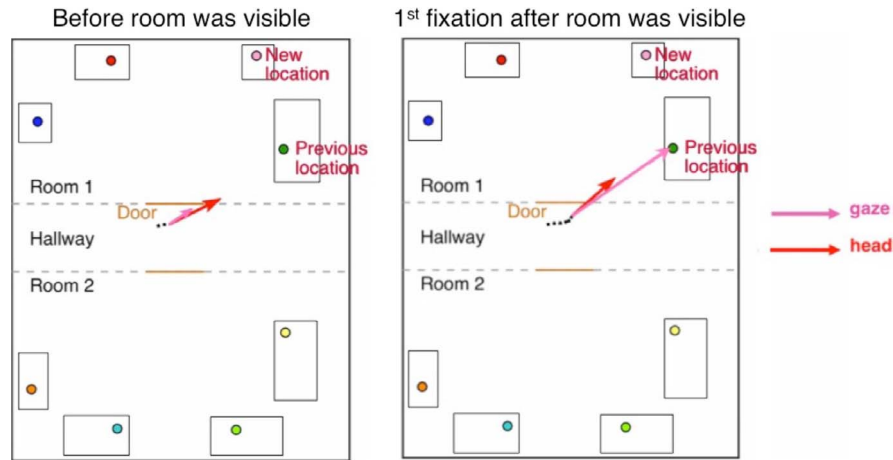
Figure 6. Bird's-eye schematic of the layout of a virtual apartment with two rooms and a hallway separating them. The subject is moving from the corridor into the bedroom to search for a target that has previously been located and whose spatial position has been learned. The black dots show the subject's path 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. The subject orients to the old location of the target even before entering the room, when the target is not visible (left), and 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 primarily on the basis of memory.

off between memory and gaze deserves further exploration.

The role of memory in gaze targeting

Another insight that is made possible by investigating natural behavior is the role of memory in action decisions and control. Information for action decisions can be made on the basis of current sensory data, a memory representation, or some weighted combination of these. In natural behavior, subjects are immersed in a relatively stable environment where they have the opportunity to develop long-term memory representations, and the use of memory in targeting eye and body movements may allow more energetically efficient strategies. Thus, natural behavior introduces constraints that are not evident in standard paradigms.

As an individual moves around in the environment, it is necessary to store information about spatial layout. One need for this information arises when orienting to regions outside the field of view. Land et al. (1999) noted instances when subjects made a number of very large gaze shifts to locations outside the field of view in a tea-making task. 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 choice of searching for a target on the basis of its visual features, so may not

need to use memory. However, it appears that memory is indeed typically 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. In a task where subjects built a toy model, Aivar et al. (2005) showed that 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, and then had to make corrective movements. The most likely reason for choosing memory-based targeting over visual targeting is that it allows planning ahead, and this presumably leads to more efficient movements. For example, eye–head–hand coordination patterns to known target locations appear to be designed so that all the effectors arrive at about the same time, which is presumably optimal in terms of executing the next action (Hayhoe, 2009).

Another advantage of planning movements based on spatial memory is that it allows more efficient use of body movements. In a real-world search task, Foulsham, Chapman, Nasiopoulos, and Kingstone (2014) found that 60%–80% of the search time was taken up by head movements, so there is an advantage to minimizing the cost of these movements. Whole-body movements can also be minimized using spatial memory. An example can be seen in Figure 6, where subjects searched for targets in a virtual apartment. After they searched for the target on three separate

occasions, it 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 (Li, Aivar, Tong, & Hayhoe, 2017). In addition, subjects rapidly encoded the global structure of the space and reduced the total path walked by eliminating regions where targets were unlikely to be, confining search to more probable regions. Memory of the large-scale spatial structure allows more energetically efficient movements, and this may be an important factor that shapes memory for large-scale environments.

Another aspect of natural behavior is that it provides different sensorimotor information and may change the nature of the memory structures. Chrastil and Warren (2012) argue that idiothetic information deriving from efferent motor commands and sensory reafference generated by observer movements aid the development of spatial memory, and Draschkow and Võ (2016) found that active object manipulation influenced memory. Thus, 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 movement patterns.

Prediction

Examination of natural behavior immediately makes apparent another factor, namely the central importance of prediction. Body movements are slow, so any action decisions need to be appropriate for the state of the scene hundreds of milliseconds in the future. 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, Miall, & Kawato, 1998; Mulliken & Andersen, 2009), 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).

Perhaps not surprisingly, it is in the context of movements that prediction is most apparent, since movements generate a time-varying visual input. One clear-cut demonstration of prediction is in the context of visual stability, where the need to predict the consequences of one's own movements is readily apparent. These predictions appear to be revealed in the remapping of visual receptive fields before a saccade (Duhamel, Colby, & Goldberg, 1991; Melcher &

Colby, 2008). Predictive remapping occurs not only in lateral intraparietal cells, but also in superior colliculus, frontal eye fields, and area V3. Evidence indicates that predictive remapping is mediated by a corollary discharge signal originating in the superior colliculus and the mediodorsal nucleus of the thalamus. Cicchini, Binda, Burr, and Morrone (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.

Other evidence for prediction also comes from the oculomotor system. Both smooth pursuit and saccadic eye movements reveal prediction of the future visual stimulus in a variety of experimental paradigms (Madelain & Krauzlis, 2003; Orban de Xivry, Missal, & Lefèvre, 2008; Ferrera & Barborica, 2010; Kowler, 2011; Spering, Schütz, Braun, & Gegenfurtner, 2011). Predictive eye movements are also robust and pervasive in natural behavior, where trajectories are complex and predictions are 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, Cooper, Rothkopf, and Hayhoe (2013) 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° 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 in space, despite variation in ball properties, suggests 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.

The evidence for prediction in the visual system is not entirely clear. Zhao and Warren (2015) argue that actions are planned on the basis of current state using a mapping that has been found as a result of learning to be effective for future state. It may be necessary to take into account a variety of factors in order to understand any one situation. Belousov, Neumann, Rothkopf, and Peters (2016) have shown that predictive and reactive strategies may be optimal and operate in different regimes depending on how much time the observer has, the sensory latencies, and noise both in the observation and in the stored model. Within the framework of optimal probabilistic control, they show that the optimal policy depends on perceptual and internal prediction uncertainties, time to ground contact, and perceptual latency, and switches between generating reactive and predictive
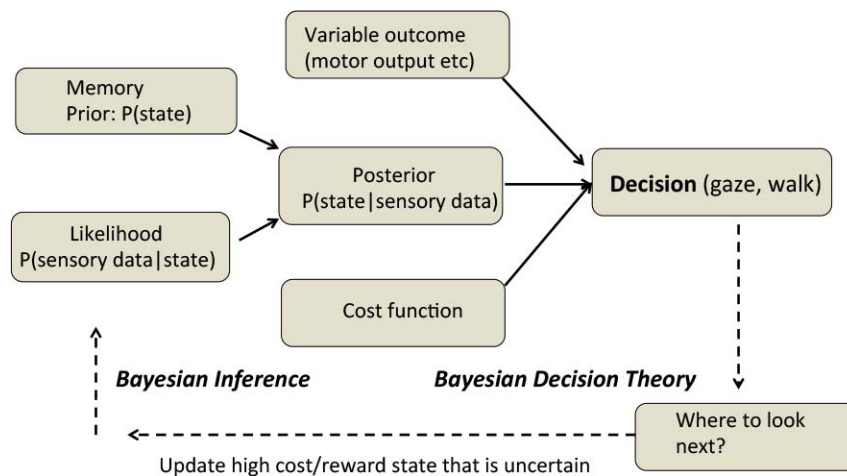
Figure 7. Schematic showing how memory and costs influence sensorimotor decisions.

behavior based on the ratio of system to observation noise and the ratio between perceptual latency and task duration.

theoretic framework provides a useful structure for conceptualizing at least some aspects of natural behavior.

Decision making

Recent approaches to sensorimotor decisions formalize the process within statistical decision theory (Maloney & Zhang, 2010; Wolpert & Landy, 2012). This provides a useful framework for understanding natural visually guided behavior and shows how the various factors so far discussed relate to one another. Wolpert and Landy (2012) have reviewed a large body of work over the last 10–15 years within this framework, which is illustrated in Figure 7. To make a good decision, the actor needs to evaluate the task-relevant state, and this requires both sensory data and a prior, as shown in the figure. 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, as already described. In order to understand how a particular goal affects behavior, we need to address the costs and benefits of the action in bringing about the goal. Sensorimotor decisions in the context of behavior reveal the pervasive effects of these costs and benefits in momentary decisions of where to look or walk. The framework is not strictly applicable for describing sequences of decisions in behavior, where we also need to consider the transitions from one decision to the next, leading to the reinforcement-learning framework. For simplicity this has been represented as the dotted arrows in the figure indicating where to look next, and I have discussed how uncertainty and the need to update state information factor into that decision. However, the decision-

What can be learned from natural behavior?

The work reviewed here shows that investigation of natural behavior has contributed a number of insights to our understanding of visual guidance of actions. Natural behavior forces consideration of exactly what information is being gathered by the visual system from moment to moment. First, it allows a more accurate specification of exactly what the spatiotemporal properties of the visual stimulus are, as experienced by the observer in the context of active behavior. In addition, looking at behavior in situ, it becomes clear that knowing the immediate behavioral goals is critical, as it provides the rationale for momentary action decisions. Knowledge of the current behavioral context allows us to understand how various factors are integrated and how they might be modulated in different contexts. Analysis of natural behavior allows an evaluation of the importance of particular factors in behavior. For example, while it has long been accepted that memory can guide movements, it is only in a behavioral context that we can evaluate how important a factor memory actually is. Similarly, the critical role of costs and benefits emerges as a fundamentally important factor. The commonality of the stimulus milieu that humans experience, and the well-defined optimality criteria of much natural behavior, means that the behavioral measures are unexpectedly stable and similar between different individuals. This stability points to the lawfulness of the underlying principles. Finally, in contrast to standard paradigms—where the focus is on

events during a single experimental trial—natural behavior focuses attention on behavior over timescales of seconds or minutes, so new questions emerge, such as what factors control the transition from one gaze location to the next within a larger-scale behavioral goal. Thus, while there are many daunting challenges in analysis of natural behavior, it allows the opportunity for exceptional insights.

Acknowledgments

Commercial relationships: none.
Corresponding author: Mary M. Hayhoe.
Email: hayhoe@utexas.edu.
Address: Center for Perceptual Systems, University of Texas Austin, Austin, TX, USA.

References

Aivar, M. P., Hayhoe, M. M., Chizk, C. L., & Mruczek, R. E. B. (2005). Spatial memory and saccadic targeting in a natural task. *Journal of Vision*, *5*(3):3, 177–193, https://doi.org/10.1167/5.3.3. [PubMed] [Article]

Belousov, B., Neumann, G., Rothkopf, C., & Peters, J. (2016). Catching heuristics are optimal control policies. *Neural Information Processing Systems Proceedings*, *29*, 1–9.

Borji, A., Itti, L., Liu, J., Musialski, P., & Wonka, P. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185–207.

Ballard, D., Hayhoe, M., & Pelz, J. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, *7*, 66–80.

Chrastil, E. R., & Warren, W. H. (2012). Active and passive contributions to spatial learning. *Psychonomic Bulletin & Review*, *19*, 1–23.

Cicchini, G. M., Binda, P., Burr, D., & Morrone, C. (2013) Transient spatiotopic integration across saccadic eye movements mediates visual stability. *Journal of Neurophysiology*, *109*, 1117–1125.

Cole, J., & Paillard, J. (1995). Living without touch and peripheral information about body position and movement: Studies with a deafferented subject. In J. L. Bermudez, A. Marcel, & J. Eilan (Eds.), *The body and the self* (pp. 245–266). Cambridge, MA: MIT Press.

Diaz, G., Cooper, J., Rothkopf, C., & Hayhoe, M. (2013). Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of Vision*, *13*(1):20, 1–14, https://doi.org/10.1167/13.1.20. [PubMed] [Article]

Draschkow, D., & Võ, M. L.-H. (2016). Of "what" and "where" in a natural search task: Active object handling supports object location memory beyond the object's identity. *Attention, Perception, & Psychophysics*, *78*, 1574–1584.

Droll, J., & Hayhoe, M. (2007). Deciding when to remember and when to forget: Trade-offs between working memory and gaze. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1352–1365.

Droll, J., Hayhoe, M., Triesch, J., & Sullivan, B. (2005). Task demands control acquisition and maintenance of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1416–1438.

Duhamel, J., Colby, C. L., & Goldberg, M. E. (1991, January 3). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, *255*, 1989–1991.

Epelboim, J., Steinman, R., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C., & Collewijn, H. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research*, *35*, 3401–3422.

Fajen, B. R., & Warren, W. H. (2003). The behavioral dynamics of steering, obstacle avoidance, and route selection. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 343–362.

Ferrera, V. P., & Barborica, A. (2010). Internally generated error signals in monkey frontal eye field during an inferred motion task. *The Journal of Neuroscience*, *30*, 11612–11623.

Foulsham, T., Chapman, C., Nasiopoulos, E., & Kingstone, A. (2014). Top-down and bottom-up aspects of active search in a real-world environment. *Canadian Journal of Experimental Psychology/Revue Canadienne De Psychologie Expérimentale*, *68*, 8–19.

Gottlieb, J. (2012). Attention, learning, and the value of information. *Neuron*, *76*, 281–295.

Hardiess, G., Gillner, S., & Mallot, H. A. (2008). Head and eye movements and the role of memory limitations in a visual search paradigm. *Journal of*

Vision, 8(1):7, 1–13, https://doi.org/10.1167/8.1.7. [PubMed] [Article]

Hayhoe, M. (2009). Visual memory in motor planning and action. In J. Brockmole (Ed.), *Memory for the visual world* (pp. 117–139). Hove, UK: Psychology Press.

Hayhoe, M. (2017). Perception and action. *Annual Review of Vision Science, 3*(4), 389–413.

Hayhoe, M. H., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research, 217*(1), 125–136.

Hayhoe, M., Shrivastrava, A., Myruczek, R., & Pelz, J. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3*(1):6, 49–63, https://doi.org/10.1167/3.1.6. [PubMed] [Article]

Hoppe, D., & Rothkopf, C. (2016). Learning rational temporal eye movement strategies. *Proceedings of the National Academy of Sciences, USA, 113*, 8332–8337.

Johnson, L. M., Sullivan, B. T., Hayhoe, M., & Ballard, D. H. (2014). Predicting human visuomotor behavior in a driving task. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*:20130044, https://doi.org/10.1098/rstb.2013.0044.

Jovancevic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *J Neurosci, 29*(19), 6234–6238.

Kowler, E. (2011). Eye movements: The last 25 years. *Vision Research, 51*(13), 1457–1483.

Land, M., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society B: Biological Sciences, 352*, 1231–1239.

Land, M., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*, 3559–3566.

Land, M. F., & McLeod, P. (2000). From eye movements to actions: How batsmen hit the ball. *Nature Neuroscience, 3*, 1340–1345.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*, 1311–1328.

Lee, D., Seo, H., & Jung, M. W. (2012). Neural basis of reinforcement learning and decision making. *Annual Review of Neuroscience, 35*, 287–308.

Li, C.-L., Aivar, M., Tong, M., & Hayhoe, M. (2017). Visual search in large-scale spaces: Spatial memory and head movements. *Journal of Vision, 17*(10): 926, https://doi.org/10.1167/17.10.926. [Abstract]

Madelain, L., & Krauzlis, R. J. (2003). Effects of learning on smooth pursuit during transient disappearance of a visual target. *Journal of Neurophysiology, 90*, 972–982.

Maloney, L. & Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research, 50*, 2362–2374.

Matthis, J. S., Barton, S. L., & Fajen, B. R. (2015). The biomechanics of walking shape the use of visual information during locomotion over complex terrain. *Journal of Vision, 15*(3):10, 1–13, https://doi.org/10.1167/15.3.10. [PubMed] [Article]

Matthis, J. S., Barton, S. L., & Fajen, B. R. (2017). The critical control phase for the visual control of walking over complex terrain. *Proceedings of the National Academy of Sciences, USA, 114*(30), e6720–e6729, https://doi.org/10.1073/pnas.1611699114.

Matthis, J. S., Muller, K. S., Bonnen, K., & Hayhoe, M. M. (2017). Optic flow and self-motion information during real-world locomotion. *Journal of Vision, 17*(10): 211, https://doi.org/10.1167/17.10.211. [Abstract]

Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (in press). Gaze and the visual control of foot placement when walking over real-world rough terrain. *Current Biology*.

Melcher, D., & Colby, C. (2008). Trans-saccadic perception. *Trends in Cognitive Science, 12*, 466–473.

Mulliken, G. H., & Andersen, R. A. (2009). Forward models and state estimation in posterior parietal cortex. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences IV* (pp. 599–611). Cambridge, MA: MIT Press.

Navalpakkam, V., Koch, C., Rangel, A., & Perona, P. (2010). Optimal reward harvesting in complex perceptual environments. *Proceedings of the National Academy of Sciences, USA, 107*, 5232–5237.

Orban de Xivry, J. J., Missal, M., & Lefèvre, P. (2008). A dynamic representation of target motion drives predictive smooth pursuit during target blanking. *Journal of Vision, 8*(15):6, 1–13, https://doi.org/10.1167/8.15.6. [PubMed] [Article]

Platt, M., & Glimcher, P. (1999, July 15). Neural correlates of decision variables in parietal cortex, *Nature, 400*, 233–238.

Rothkopf, C. & Ballard, D. H. (2013). Modular inverse reinforcement learning for visuomotor behavior *Biological Cybernetics, 107*(4), 477–490.

Schultz, W. (2000). Multiple reward signals in the brain. *Nature Reviews Neuroscience, 1*, 199–207.

Schütz, A., Trommershäuser, J., & Gegenfurtner, K.

(2012). Dynamic integration of information about salience and value for saccadic eye movements. *Proceedings of the National Academy of Sciences, USA*, *109*, 7547–7552.

Solman, G. J. F., & Kingstone, A. (2014). Balancing energetic and cognitive resources: Memory use during search depends on the orienting effector. *Cognition*, *132*, 443–454.

Spering, M., Schütz, A. C., Braun, D. I., & Gegenfurtner, K. R. (2011). Keep your eyes on the ball: Smooth pursuit eye movements enhance the prediction of visual motion. *Journal of Neurophysiology*, *105*, 1756–1767.

Sprague, N., Ballard, D. H., & Robinson, A. (2007). Modeling embodied visual behaviors. *ACM Trans. Appl. Perception*, *4*(2):11, https://doi.org/10.1145/1265957.1265960.

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

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2005). Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nature Reviews Neuroscience*, *6*, 363–375.

Sullivan, B. T., Johnson, L. M., Rothkopf, C., Ballard, D., & Hayhoe, M. (2012). The role of uncertainty and reward on eye movements in a virtual driving task. *Journal of Vision*, *12*(13):19, 1–17, https://doi.org/10.1167/12.13.19. [PubMed] [Article]

Tatler, B., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society B*, *366*, 596–610.

Tong, M., Zhang, S., Johnson, L., Ballard, D., & Hayhoe, M. (2015). Modelling task control of gaze. *Journal of Vision*, *15*(12): 784, https://doi.org/10.1167/15.12.784. [Abstract]

Tong, M. H., Zohar, O., & Hayhoe, M. M. (2017). Control of gaze while walking: Task structure, reward, and uncertainty. *Journal of Vision*, *17*(1): 28, 1–19, https://doi.org/10.1167/17.1.28. [PubMed] [Article]

Wolpert, D., & Landy, M. (2012). Motor control is decision making. *Current Opinion in Neurobiology*, *22*, 1–8.

Wolpert, D. M., Miall, R. C., & Kawato, M. (1998). Internal models in the cerebellum. *Trends in Cognitive Science*, *2*, 338–347.

Yasuda, M., Yamamoto, S., & Hikosaka, O. (2012). Robust representation of stable object values in the oculomotor basal ganglia. *The Journal of Neuroscience*, *32*, 16917–16932.

Zhao, H., & Warren, W. H. (2015). On-line and model-based approaches to the visual control of action. *Vision Research*, *110*, 190–202.