Overview

Vision in the natural world

Mary M. Hayhoe[1]* and Constantin A. Rothkopf[2]

Historically, the study of visual perception has followed a reductionist strategy, with the goal of understanding complex visually guided behavior by separate analysis of its elemental components. Recent developments in monitoring behavior, such as measurement of eye movements in unconstrained observers, have allowed investigation of the use of vision in the natural world. This has led to a variety of insights that would be difficult to achieve in more constrained experimental contexts. In general, it shifts the focus of vision away from the properties of the stimulus toward a consideration of the behavioral goals of the observer. It appears that behavioral goals are a critical factor in controlling the acquisition of visual information from the world. This insight has been accompanied by a growing understanding of the importance of reward in modulating the underlying neural mechanisms and by theoretical developments using reinforcement learning models of complex behavior. These developments provide us with the tools to understanding how tasks are represented in the brain, and how they control acquisition of information through use of gaze. © 2010 John Wiley & Sons, Ltd. *WIREs Cogn Sci* 2011 2 158–166 DOI: 10.1002/wcs.113

INTRODUCTION

The study of human vision has traditionally been performed in simple, highly controlled laboratory experiments. This is a natural consequence of the historical development of the study of sensory systems, which must begin with an analysis of the sensory end organs. Another factor has been the general strategy within the field, of investigating the simplest components of perception, with the goal of understanding complex behavior in terms of its elemental components. Although this has been extraordinarily valuable, there are compelling reasons to explore the use of vision in its natural context. The visual system is used to acquire information that is needed to guide a variety of behaviors. Until we examine vision *in situ*, we have no way of knowing how this is done, or even if the results in simpler laboratory experiments are relevant for natural vision. Many important issues arise in natural contexts that are absent, or difficult to address, in standard paradigms. Natural behavior is distinctive, in that eye movements, attention, working memory, sensory decisions, and control of actions are all involved in the generation of even the simplest behaviors. However, in standard paradigms, these issues are typically studied in isolation. Only by addressing vision in its natural context can we attempt to understand how these factors jointly contribute to the composition of natural behavioral sequences.

A central feature of natural vision is that information is dynamically acquired from the environment to guide ongoing actions and behavioral goals. Information from a scene is actively sampled by a sequence of rapid gaze changes that bring the high acuity foveal region at the center of the retina, onto different regions or objects of interest in a scene. These rapid shifts of the eye from one place to another are called saccadic eye movements, and are one of the primary types of eye movement involved in gathering information from different locations in a visual scene. We are largely unaware of the importance of these gaze changes in gathering the visual information that is needed to perform even the simplest everyday actions. Almost without exception, whenever we pick up an object in the course of everyday life, we direct our eyes to the object in order to guide our grasp. Similarly, we almost always direct our eyes at the things we are most interested in at the moment. Thus

*Correspondence to: mary@mail.cps.utexas.edu

[1]Department of Psychology, University Station, A8000 Austin, TX 78750, USA

[2]Frankfurt Institute for Advanced Studies, Goethe-University, Frankfurt, Room 1.404, Ruth-Moufang-Str. 1, 60438 Frankfurt, Germany

DOI: 10.1002/wcs.113

eye movements are a critical component of everyday visually guided behavior. Moreover, they reveal what information is selected from the image and when it is selected. Consequently, observation of eye movements in natural behavior can provide a window into the underlying visual computations, on a time scale of a few hundred milliseconds.

MEASURING EYE MOVEMENTS

Eye movements are an overt manifestation of the momentary deployment of attention in a scene. Attention can also be distributed to regions off the line of sight. Information such as large-scale spatial layout or the optic flow generated by body movement is inherently global in nature and requires attention to be distributed across the entire visual field. However, the direction of gaze carries a tremendous amount of information about the current focus of attention, and provides an entrée to studying information acquisition.[1] Until recently, measuring eye movements in an unconstrained observer was very difficult. In order to measure the tiny rotations of the eye, early eye position measurement devices involved awkward and often uncomfortable attachments to the eye that necessitated stabilizing the head. However, the last 10–15 years has seen the advent and refinement of eye trackers mounted on the head, which allow considerable freedom of movement. Michael Land built one of the first of these, and a recent version by Babcock and Pelz is completely portable.[2] The new eye trackers allow the study of eye movements over extended tasks in natural settings, where a wide variety of natural coordinated behaviors are possible. As well as providing eye position in the head, or in space, these trackers also provide a video from the subject's viewpoint, with eye position superimposed. This is particularly useful in interpreting the function of the eye movements as it shows the stage of the task corresponding to a particular gaze point.

TASK DEPENDENCE OF GAZE PATTERNS

Since the development of head mounted eye trackers, eye movements during a variety of natural behaviors such as driving, walking, various sports, and making tea or sandwiches have been investigated. The central finding of all these investigations is that gaze locations are very tightly linked to the task. Observers rarely look at regions or objects that are irrelevant to the task at hand.[3–5] The period in between the saccadic gaze shifts, when the eye is relatively stationary, is called a 'fixation'. When viewing a video replay of fixations made during a natural task, from the subject's viewpoint, with eye position superimposed, it is clear that fixations are tightly linked, moment-by-moment, to the actions, such as grasping and manipulating objects, and move on to the next object when the needs of the current action have been met.[6–8]

To understand this better, imagine a small segment of behavior that might be observed during sandwich making.[9] A view of fixations made in this task is shown in figure 1. Suppose that the subject has finished making the sandwich and is about to cut it in half. Gaze will be directed at the knife handle to guide the hand to pick it up. As the hand closes on the knife, the eye will move to the corner of the sandwich where the knife tip will be placed in order to begin cutting. Gaze then moves slowly along the bread with the knife following shortly afterwards. After the sandwich has been cut, the eye will move to a location on the table where the knife is to be placed. As the knife nears the table, gaze will move on to the next object of interest, such as a glass.

Thus, the basic structure of the task allows one to link the visual operations fairly closely in time with the occurrence of eye and hand movements. Not only is the sequence of fixations tightly linked to the task, but also many of the fixations appear to have the purpose of obtaining quite specific information. For example, subjects fixate the middle of the jar for grasping with the hand in a vertical posture and the rim for putting on the lid, with the hand in a horizontal posture. This suggests that the visual information being extracted

FIGURE 1 | The yellow circles show the fixations made while a subject makes a peanut butter and jelly sandwich. Views from a video camera mounted on the subject's head have been superimposed to make a composite mosaic to compensate for the subject's head movements during the task using the method described by Rothkopf & Pelz (2004). The diameter of the yellow circles indicates the duration of the fixations.

 © 2010 John Wiley & Sons, Ltd.

controls the pre-shaping of the hand in one case and the orientation of the lid in the other. This has been called a 'just-in-time' strategy, where observers acquire the specific information they need just at the moment it is required in the task.[10] Interestingly, while subjects are in general aware that they are looking at the objects involved in the task, they are not aware of the speed and specificity of the eye movements, which occur more or less automatically. Thus, the machinery for moving the eyes in a scene appears to be engaged in the service of the immediate cognitive goal.

TWO DIMENSIONAL IMAGES VERSUS THE REAL WORLD

To appreciate the significance of these observations, we should note that a natural expectation might be that gaze would roam haphazardly around the scene, fixating 'salient' points that stand out from the surroundings, such as bright regions, regions of saturated color, or edges between regions. A natural intuition is that such stimuli might automatically attract our attention. A large body of modeling work by Itti and colleagues has attempted to predict fixation locations by analyzing the content of the image for regions that differ from the surrounds in this way. If subjects are inspecting still images of scenes, these saliency calculations can predict some of the fixations.[11,12] In addition, some stimuli such as sudden onsets have considerable ability to capture attention even if the observer's attention is directed elsewhere.[13] There is a large body of work that attempts to predict the fixations in images displayed on computer monitors. However, inspecting such pictures is a very different situation from using vision in the natural world, and we cannot assume that picture-viewing captures what happens in the broader visual context. One reason to expect a difference is that the stimulus itself is different in significant ways. The visual input in the real world is three-dimensional (3D) and varies constantly as a consequence of the observer's movements in the scene. Images of scenes typically do not respect the actual scale of the real image, so that a panoramic view of a scene is often compressed into a relatively small region on a monitor, and viewers must make an implicit inference about the size of the objects in the image in order to interpret the scene. Another reason to expect a difference is that we must take into account the tasks the observer is performing. In a 2D image, subjects are most likely recognizing and remembering the objects in the scene. In the real world, however, subjects need specific information for navigation, obstacle avoidance, etc., and this information changes as a function of time. Paradigms where subjects view 2D images cannot capture these aspects of attention and gaze control. Thus, immersion in an actual scene that captures the exigencies of the real world is critical for understanding the principles that control gaze and the selective acquisition of visual information from scenes.

SPECIALIZED COMPUTATIONS DURING FIXATIONS

There is far too much information in visual scenes to process at once. Even at the point of fixation, multiple kinds of information are available. It seems likely that, in the context of natural behavior, the task controls the specific information that is selected within a given fixation. For example, when driving around a bend in the road, drivers fixate the tangent point of the curve.[14] When making a sandwich, subjects will fixate the handle of the knife when picking it up, but the tip of the knife when spreading the peanut butter. In the first case, the subject needs the location and orientation of the handle to guide the pick up action. In the latter case, the tip needs to be fixated to control the spreading action. When first viewing the table-top scene, subjects make a series of short duration fixations on the relevant objects, such as the peanut butter jar. In this case, the fixation on the peanut butter is presumably for the purpose of recognition and perhaps locating it for future use. Subsequent fixations will be for guiding the grasping action or for removing the lid. In the absence of ongoing behavior, we are inclined to think that the job of vision is primarily object recognition, but these examples remind us of the complexity of the information one can get while fixating an object and the variety of operations that vision must perform. This specificity is indicated not only by the ongoing actions and the point in the task, but also by the durations of the fixations, which may vary over a range from less than 100 millisecond to several seconds.[9] A large component of this variation appears to depend on the particular information required for that point in the task, fixation being terminated when the particular information is acquired.[12]

In order to conceptualizing how visual information acquisition is organized by the task, it is useful to adopt a hierarchical framework as shown in figure 2. At the most general level, acquisition is driven by a larger cognitive goal, such as making a sandwich. To accomplish this goal, the observer must perform a sequence of microtasks, such as grasping the peanut butter, grasping the lid, removing the lid, and so on. A microtask, such as grasping the peanut butter, involves a fixation on the jar. During that fixation,

 © 2010 John Wiley & Sons, Ltd.

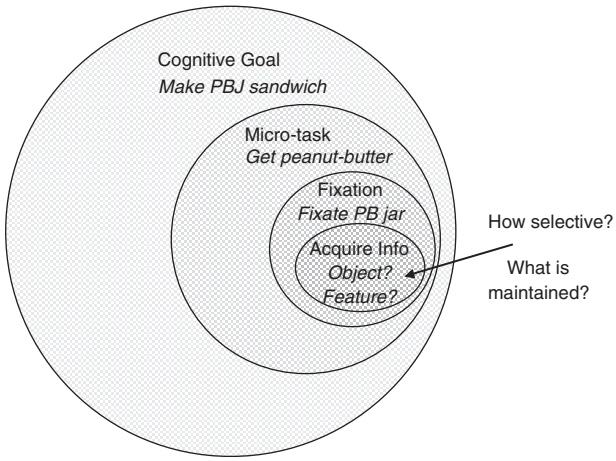

FIGURE 2 | Illustration of the task control of fixations. A particular behavioral goal, such as eating a peanut butter and jelly sandwich, can be broken down into sub-tasks such as picking up the peanut butter jar. To accomplish this the subject must fixate the jar to guide the pickup action. The fixation in turn is used to acquire specific kinds of information such as location of the jar, size, orientation, surface texture, and so on.

specific visual information will be acquired, such as the size of the jar, which is necessary in order to plan the grasp. These task-specific computations have been referred to as 'visual routines'.[15–17] The idea of visual routines is that even quite low-level visual information, such as color, requires specialized computation of some kind and is not done automatically by the visual system. The idea is that the visual computations operating while the subject is guiding a grasp do not necessarily involve computing other information, such as the color of the peanut butter, that are not necessary at that moment. Such information might of course be stored in working memory or in some longer term memory representation of the scene.

The hypothesis that very specific information is acquired within a fixation is, strictly, an hypothesis about neural mechanisms. That is, cortical state (including V1) is different when subjects are involved in different tasks, even when the retinal input is the same. However, it is possible to get supporting evidence for this psychophysically. Recent work by Droll, Triesch, and others, using virtual environments, has revealed that subjects use vision to extract information very selectively from an image.[18,19] In a task where subjects picked up and sorted colored virtual blocks of different sizes, properties of the blocks, for example, the block color, were artificially changed at unpredictable moments during the task. Subjects were twice as likely to notice these changes when that property was needed in order to perform the task as when it was irrelevant. For example, if subjects were instructed to pick up a tall brick, they would be more sensitive to a change in size than if they were required to pick up a red brick, when only the color is relevant. Thus subjects are more aware of some visual properties of the blocks than others, even though the block is the focus of attention throughout the trial. Given the subjective coherence of an object's features, such selectivity or preferential representation is non-intuitive, but nonetheless it highlights the specialized nature of visual computations, even for such simple visual properties like size or color. It is commonly thought that the effect of fixating and attending to an object is to bind object properties together into a coherent representation, called an 'object file'.[20] Droll et al.'s experiment suggests that such binding does not always happen, even when the object is the focus of attention. Thus 'seeing' is not a unitary process, despite our subjective impression that we effortlessly perceive everything at once. Rather, it is an active process and is inextricably linked to the observer's behavioral goals.

This result also sheds some light on how to think about attention. The nature of selective attention has long been an issue of great concern to experimental psychologists and more recently to neurophysiologists and computer scientists. Despite decades of research and the fundamental nature of the constraints that attention imposes on brain function, we have little understanding of exactly what attention is, and why it is limited. A common metaphor is that attention enhances sensory processing in some way, just as an increase in the brightness or contrast in an image makes vision easier.[21,22] Droll et al.'s results show, in contrast, that the visual effects of attention cannot be described in a way that is independent of the task. Perhaps, a better way to think about attention is as the mechanism that determines whether and when a particular computation will be performed not simply as the enhancement of that computation. This interpretation is consistent with the findings of Mack & Rock,[23] who demonstrate many instances of 'inattentional blindness', where subjects appear to be completely unaware of unattended stimuli. For any particular visual task or visual context, a circumscribed set of computations will be needed. These will define what information is attended or computed. The task-irrelevant information will be unattended. Thus attention is most naturally studied in the context of ongoing behavior, where the task structure defines the computations that need to be performed.

LEARNING WHERE TO LOOK

Implicit in much of the research on natural tasks is the finding that eye movement patterns must

 © 2010 John Wiley & Sons, Ltd.

be learned.[24,3,25] By implication, the sequential deployment of attention must also be learned. For example, in tea making and sandwich making, observers must have learnt what objects in the scene are relevant, as almost no fixations fall on irrelevant objects. In driving, Shinoda and colleagues showed that approximately 45% of fixations fell in the neighborhood of intersections. As a consequence of this, subjects were more likely to notice 'Stop' signs located at intersections as opposed to signs in the middle of a block. Thus, it is likely that subjects have learnt that traffic signs are more likely around intersections. At a more detailed level, subjects must learn the optimal location for the information they need. For example, when pouring tea, fixation is located at the tip of the teapot spout. Presumably, flow from the spout is best controlled by fixating this location. Similarly, in walking, observers must learn where and when to look at locations critical for avoiding obstacles while controlling direction and balance. Subjects must learn not only the locations at which relevant information is to be found, but also the order in which the fixations must be made in order to accomplish the task. Thus, a subject must locate the peanut butter and the bread before picking them up, pick up the knife before spreading, and so on. This means that a complete understanding of vision in natural behavior will require an understanding of the way that tasks are learnt and represented in the brain.

Another way in which learning is critical for deployment of gaze and attention is that observers must learn the dynamic properties of the world in order to distribute gaze and attention where they are needed. When making tea or sandwiches, items remain in stable locations with stable properties, for the most part. In a familiar room, the observer need only update the locations of items that are moved, or monitor items that are changing state (e.g., water filling the kettle). In dynamic environments, such as driving, walking, or in sports, more complex properties must be learnt. In walking, humans need to know how pedestrians typically behave and how often to look at them. The fact that humans do indeed learn such statistics was demonstrated by Jovancevic and Hayhoe.[26] In a real walking setting, they were able to actively manipulate gaze allocation by varying the probability of potential collisions. Manipulation of the probability of a potential collision by a risky pedestrian was accompanied by a rapid change in gaze allocation. Subjects learn new priorities for gaze allocation within a few encounters and look both sooner and longer at potentially dangerous pedestrians. Other evidence for such learning is the fact that saccades are often pro-active, that is, they are made to a location in a scene in advance of an expected event. In walking, subjects looked at risky pedestrians before they veered onto a collision course. In cricket, batsmen anticipated the bounce point of the ball, and more skilled batsmen arrived at the bounce point about 100 millisecond earlier than less skilled players.[27] The ability to predict where the ball will bounce depends on previous experience of the cricket ball's trajectory. These saccades were always preceded by a fixation on the ball as it left the bowler's hand, showing that batsmen use current sensory data in combination with prior experience of the ball's motion to predict the location of the bounce. This suggests that observers have learnt models of the dynamic properties of the world that can be used to position gaze in anticipation of a predicted event.

SIMILARITY BETWEEN DIFFERENT INDIVIDUALS

Another important and unexpected feature of eye movements in ordinary behavior is the reproducibility of the eye movement patterns and the similarity between observers. For example, when cutting a sandwich, subjects always fixate the initial point of contact with the knife and move their gaze along the locus of the cut, just ahead of the knife. Similarly, subjects always fixate the neck of the bottle when beginning to pour a liquid. Given the role of eye movements in gathering information to accomplish tasks, it makes sense that fixation patterns of different individuals should be similar, as they will reflect the physical and dynamic properties of the environment as well as common behavioral goals. This stability in fixation patterns makes the investigation of natural behavior unexpectedly accessible to experimental investigation.

NEURAL SUBSTRATE FOR LEARNING GAZE ALLOCATION IN TASK EXECUTION

We have seen that the selective acquisition of information from the environment is largely controlled by the observer's immediate behavioral goals. We have also seen the critical role of learning in this process. How does this learning come about? Recent developments in understanding the neural machinery controlling eye movements, together with developments in the mathematics of reinforcement learning are making this problem increasingly tractable.[28]

 © 2010 John Wiley & Sons, Ltd.

It has become clear the brain's internal reward mechanisms are intimately linked to the neural machinery controlling eye movements. Schultz and colleagues have shown that dopaminergic neurons in the basal ganglia (an important mid-brain structure involved in the generation of movements) signal the reward expected from an action. In these experiments, typically the reward is juice or food of some kind, but dopamine release is linked to reinforcing events in general. Sensitivity to reward is manifest throughout the saccadic eye movement circuitry. Caudate cell responses reflect both the target of an upcoming saccade and the reward expected after making the movement.[29] Supplementary eye fields and anterior cingulate cortex also play an important role in monitoring the behavioral significance of saccades.[30] Saccade-related areas in the cortex (LIP, FEF, SEF, and DLPF) all exhibit sensitivity to reward.[31–35] 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 at both behavioral and neural levels, the target selection mechanisms appear to be intimately related to behavioral outcomes.

MODELING EYE MOVEMENTS USING REWARD

The reward sensitivity of the eye movement circuitry provides the neural underpinnings for reinforcement learning models of behavior.[36] The mathematics of reinforcement learning is potentially useful for understanding how complex gaze patterns might be generated. There have been few attempts to model the eye movements observed in complex behavior. Indeed, given the intrinsic complexity of the problem and the need to understand the underlying task control, this is not surprising. However, one such model, by Sprague et al.[37] exemplifies an approach that is potentially powerful. They show how a simulated agent in a virtual environment can learn to allocate gaze to avoid obstacles and control direction in walking. The model assumes that visual computations required in a real world behavior such as walking can be broken down into a set of subtasks, such as controlling direction, avoiding obstacles, and so on. The Sprague et al. model learns how to distribute attention and gaze between these visual subtasks in a dynamic environment. The model agent walks along a virtual path with three simultaneous tasks: stay on the path, avoid obstacles, and pick up litter as shown in figure 3. The model assumes that the agent can only attend to one location at any moment in time. To choose between ongoing competing tasks, uncertainty increases (together with an attendant cost) when gaze is withheld from an informative scene location. The model assumes that eye movements are selected to maximize reward by reducing uncertainty that could result in suboptimal actions.

It may seem strange that reward plays a role in something as seemingly automatic as eye movements. In neurophysiological paradigms, usually a primary reward such as juice or a raisin is delivered after the animal performs an action. This of course does not happen in real life when one makes an eye movement. However, consider that eye movements are for the purpose of obtaining information, and this information, as we have seen, is used to achieve behavioral goals such as making a sandwich, that are ultimately important for survival. Thus visual information acquired during a fixation can be thought of as a secondary reward, and mediate learning of gaze patterns by virtue of its ultimate significance for adaptation and survival. Indeed, several researchers have quantified the intrinsic reward associated with looking at particular visual stimuli. Platt and colleagues[38,39] measured how much liquid reward monkeys were willing to give up in order to obtain visual information about members of their social group. In this case, liquid is the measurable, external equivalent of an internal reward resulting from gaze.

Reward is also a central component of recent applications of statistical decision theory to understanding control of body movements. In this approach, the concepts of *reward* (costs and benefits of the outcome of the action), *uncertainty* (of both sensory state and outcome), and *prior knowledge* (probability distributions associated with world states) are central to understanding sensory-motor behavior.[40] When reward is externally defined (e.g. by monetary reward), it has been shown that subjects making rapid hand movements learn a complicated spatially distributed target reward system and behave in a nearly optimal manner to maximize reward.[41,42] Note, however, that these experiments considered single trial behavior in which fast reaches to targets are executed. The power of reinforcement learning, on the other hand, is that it addresses optimality in sequential tasks, where the goal state

 © 2010 John Wiley & Sons, Ltd.

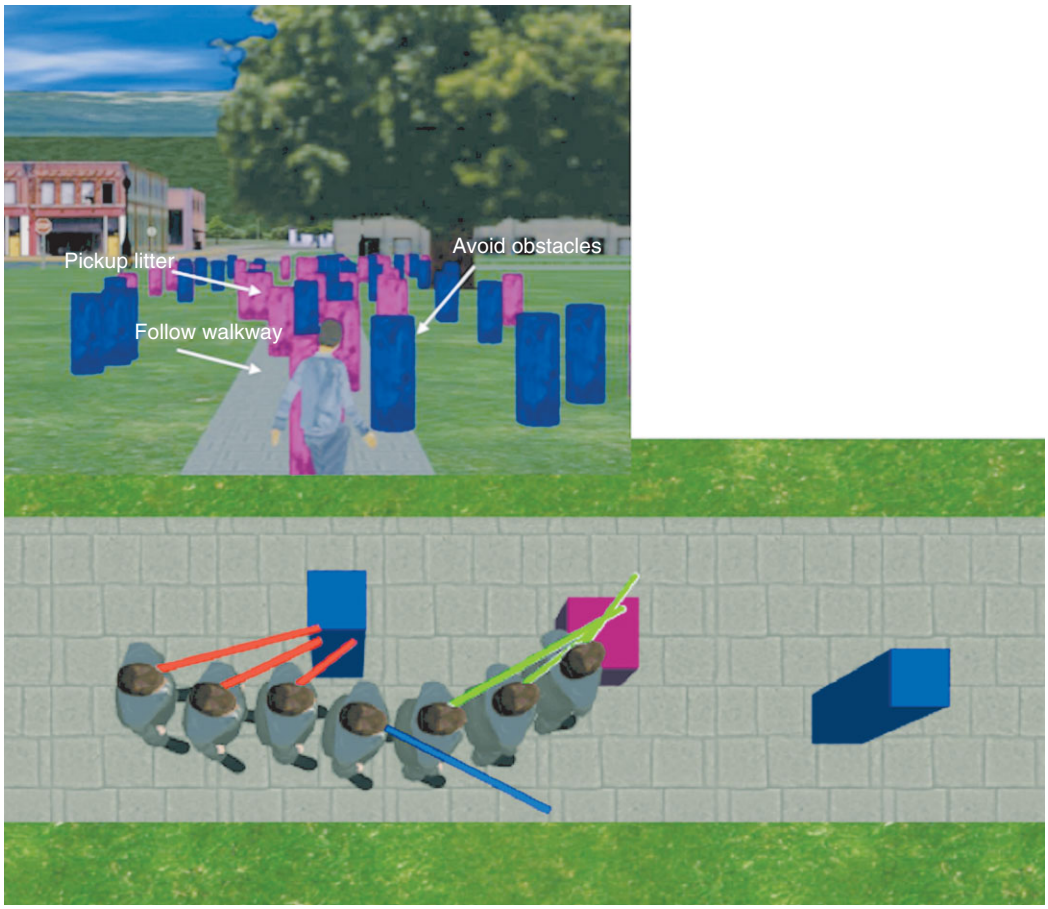

FIGURE 3 | The top part of the figure shows a virtual agent in a simulated walking environment. The agent must extract visual information from the environment in order to do three sub-tasks: staying on the sidewalk, avoiding blue obstacles, and picking-up purple litter objects (achieved by contacting them). The model agent learns how to deploy attention/gaze at each time step. The bottom panel shows seven time steps after learning. The red lines indicate that the agent is using visual information to avoid the obstacle. The blue line indicates that the agent is using information about position on the sidewalk, and the green lines show the agent using vision to intersect the purple litter object.

may be reached after a long sequence of actions. According to this framework, it becomes interesting to ask how rewards are distributed among different task components and task alternatives in everyday behavior. The technical difficulty here is to extract the rewards implicit in human sequential behavior. Recent work by Rothkopf[43] has addressed this question in the context of walkway navigation. By considering human walking trajectories in different task combinations, it was possible to find the relative reward weights associated with staying on a walkway, avoiding obstacles, and picking up litter. The value of having an estimate of internal reward values is that it allows specific predictions to be made for other aspects of behavior and other task contexts. Thus, it is plausible that a primary determinant of the patterns of eye movements observed in the natural world is the reward structure of the environment, and that this is mediated by the importance of reward in the execution of tasks.

APPLICATIONS

A particular advantage of investigating eye movements in everyday behavior is that the outcomes are much more directly related to both applied and clinical problems than work in simpler, more controlled experiments. For example, if we wish to know the consequences for everyday life of visual losses from disease or injury, it is necessary to understand how information in the visual world is used for particular tasks, such as navigation, obstacle avoidance, or preparing a meal. There is a long history of investigation of reading, but more recently investigators[44,45] have examined gaze patterns in individuals with visual impairments in a wider range of tasks. Such work can reveal not only potential mechanisms for compensating for visual losses, but also indicate potential rehabilitation strategies. This is true for visual losses caused by central damage

 © 2010 John Wiley & Sons, Ltd.

such as that caused by stroke, as well as losses from peripheral disease. The clinical applications of eye movements in everyday behavior is an important area of research that has been greatly facilitated by the development of more 'user friendly' eye trackers as we can expect to see exciting developments in this area.

CONCLUSION

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

REFERENCES

1. Findlay J, Gilchrist I. *Active Vision*. Oxford: Oxford University Press; 2003.
2. Babcock J, Pelz J. Building a lightweight eyetracking headgear, Proceedings, ACM SIGCHI Eye Tracking Research & Applications. San Antonio, Texas; 2004, 109–114.
3. Land M. Eye movements in daily life. In: Chalupa L, Werner J, eds. *The Visual Neurosciences*, vol 2. Cambridge, MA: MIT Press; 2004, 1357–1368.
4. Hayhoe M, Ballard D. Eye movements in natural behavior. *Trends Cogn Sci* 2005, 9:188–193.
5. Pelz JB, Canosa R. Oculomotor behavior and perceptual strategies in complex tasks. *Vision Res* 2001, 41:3587–3596.
6. Land M, Mennie N, Rusted J. The roles of vision and eye movements in the control of activities of daily living. *Perception* 1999, 28:1311–1328.
7. Johansson R, Westling G, Backstrom A, Flanagan JR. Eye-hand coordination in object manipulation. *J Neurosci* 2001, 21:6917–6932.
8. Rothkopf C, Ballard D, Hayhoe M. Task and scene context determines where you look. *J Vision* 2007, 7:16, 1–20.
9. Hayhoe M, Shrivastrava A, Mruczek R, Pelz J. Visual memory and motor planning in a natural task. *J Vision* 2003, 3:49–63.
10. Ballard DH, Hayhoe M, Pelz JB. Memory representations in natural tasks. *J Cogn Neurosci* 1995, 7:66–80.
11. Itti L, Koch C. Computational modeling of visual attention. *Nat Rev Neurosci* 2001, 2:194–203.
12. Henderson J. Human gaze control during real-world scene perception. *Trends Cogn Sci* 2003, 7:498–504.
13. Theeuwes J, Godisn R. Attentional and oculomotor capture. In: Folk C, Gibson B, eds. *Attraction, Distraction, and Action: Multiple Perspectives on Attentional Capture*. Amsterdam: Elsevier; 2001, 121–150.
14. Land MF, Lee DN. Where we look when we steer. *Nature (Lond)* 1994, 369:742–744.
15. Ullman S. Visual routines. *Cognition* 1984, 18:97–157.
16. Ballard D, Hayhoe M, Pook P, Rao R. Deictic codes for the embodiment of cognition. *Behav Brain Sci* 1997, 20:723–767.
17. Roelfsema P. The implementation of visual routines. *Vision Res* 2000, 40:1385–1411.
18. Droll JA, Hayhoe M, Triesch J, Sullivan BT. Task demands control acquisition and storage of visual information. *J Exp Psychol Hum Percept Perform* 2005, 31:1416–1438.
19. Triesch J, Ballard DH, Hayhoe MM, Sullivan BT. What you see is what you need. *J Vision* 2003, 3:86–94.
20. Treisman A. The perception of features and objects. In: Baddeley A, Weiskrantz L, eds. *Attention: Selection, Awareness, and Control*. A tribute to Donald Broadbent, 5–35. Oxford, UK: Clarendon Press; 1993.
21. Liu T, Abrams J, Carrasco M. Voluntary attention enhances contrast appearance. *Psychol Sci* 2009, 20:354–362.
22. Reynolds JH, Pasternak T, Desimone R. Attention increases sensitivity of V4 neurons. *Neuron* 2000, 26:703–714.

 © 2010 John Wiley & Sons, Ltd.

23. Mack A, Rock I. *Inattention Blindness*. Cambridge, MA: MIT Press; 1998.
24. Land M, Furneaux S. The knowledge base of the oculomotor system. *Phil Trans R Soc B* 1997, 352:1231–1239.
25. Chapman P, Underwood G. Visual search of dynamic scenes: event types and the role of experience in viewing driving situations. In: Underwood G, ed. *Eye Guidance in Reading and Scene Perception*. Oxford: Elsevier; 1998, 369–394.
26. Jovancevic-Misic J, Hayhoe M. Adaptive gaze control in natural environments. *J Neurosci* 2009, 29:6234–6238.
27. Land MF, McLeod P. From eye movements to actions: how batsmen hit the ball. *Nat Neurosci* 2000, 3:1340–1345.
28. Sutton R, Barto A. *Reinforcement Learning: An Introduction*. Cambridge, MA, Bradford: MIT Press; 1998.
29. Hikosaka O, Nakamura K, Nakahara H. Basal ganglia orient eyes to reward. *J Neurophysiol* 2006, 95:567–584.
30. Stuphorn V, Schall JD. Executive control of countermanding saccades by the supplementary eye field. *Nat Neurosci* 2006, 9:925–931.
31. Platt M-L, Glimcher P-W. Neural correlates of decision variables in parietal cortex. *Nature* 1999, 400:233–238.
32. Glimcher PW. The neurobiology of visual-saccadic decision making. *Ann Rev Neurosci* 2003, 26:133–179.
33. Dorris M-C, Glimcher P-W. Activity in posterior parietal cortex is correlated with the subjective desirability of an action. *Neuron* 2004, 44:365–378.
34. Sugrue L-P, Corrado G-S, Newsome W-T. Matching behavior and the representation of value in the parietal cortex. *Science* 2004, 304:1782–1787.
35. Stuphorn V, Taylor T, Schall J. Performance monitoring by the supplementary eye field. *Nature* 2000, 408:857–860.
36. Schultz W. Multiple reward signals in the brain. *Nat Rev Neurosci* 2000, 1:199–207.
37. Sprague N, Ballard D, Robinson A. Modeling embodied visual behaviors. *ACM Trans Appl Percept* 2007, 4:11.
38. Deaner RO, Khera AV, Platt ML. Monkeys pay per view: adaptive valuation of social images by rhesus macaques. *Curr Biol* 2005, 15:543–548.
39. Shepherd SV, Deaner RO, Platt ML. Social status gates social attention in monkeys. *Curr Biol* 2006, 16:119–120.
40. Tassinari H, Hudson TE, Landy MS. Combining priors and noisy visual cues in a rapid pointing task. *J Neurosci* 2006, 26:10154–10163.
41. Trommershauser J. Statistical decision theory and the selection of rapid, goal-directed movements. *J Opt Soc Amer A* 2003, 20:1419–1433.
42. Seydell A, McCann B, Trommershauser J, Knill D. Learning to behave optimally in a probabilistic environment. *J Vision* 2008, 8:544.
43. Rothkopf C. Modular models of task based visually guided behavior, University of Rochester Ph D thesis; 2009.
44. Martin T, Riley M, Kelly K, Hayhoe M, Huxlin K. Visually-guided behavior of homonymous hemianopes in a naturalistic task. *Vis Res* 2007, 47:3434–3446.
45. Sullivan B, Jovancevic-Misic J, Hayhoe MM, Sterns G. Use of multiple preferred retinal loci in Stargardt's disease during natural tasks: a case study. *Ophthal Phys Optics* 2008, 28:168–177.
46. Rothkopf CA, Pelz JB. Head movement estimation for wearable eye tracker. *Proceedings ACM SIGCHI: Eye Tracking Research and Applications*. ACM Press; 2004, 123–130.

 © 2010 John Wiley & Sons, Ltd.