

Review



Cite this article: Hayhoe MM, Matthis JS. 2018 Control of gaze in natural environments: effects of rewards and costs, uncertainty and memory in target selection. *Interface Focus* 8: 20180009. <http://dx.doi.org/10.1098/rsfs.2018.0009>

Accepted: 8 May 2018

One contribution of 12 to a theme issue 'Understanding images in biological and computer vision'.

Subject Areas:
systems biology


Keywords:
eye movements, natural vision, sensory-motor decisions

Author for correspondence:
Mary M. Hayhoe
e-mail: hayhoe@utexas.edu

Control of gaze in natural environments: effects of rewards and costs, uncertainty and memory in target selection

Mary M. Hayhoe and Jonathan Samir Matthis

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

 JSM, 0000-0003-3683-646X

The development of better eye and body tracking systems, and more flexible virtual environments have allowed more systematic exploration of natural vision and contributed a number of insights. In natural visually guided behaviour, humans make continuous sequences of sensory-motor decisions to satisfy current goals, and the role of vision is to provide the relevant information in order to achieve those goals. This paper reviews the factors that control gaze in natural visually guided actions such as locomotion, including the rewards and costs associated with the immediate behavioural goals, uncertainty about the state of the world and prior knowledge of the environment. These general features of human gaze control may inform the development of artificial systems.

1. Introduction

Vision in both natural and artificial systems must deal with problems in the real world, where the environment is often unpredictable and the sensing agent is mobile. Despite the profound differences between artificial and human visual systems, it is clearly advantageous to understand how human vision accomplishes its goals with such apparent ease. However, given the historical development of vision science, our understanding of natural vision is still limited, largely because of limitations in both display technology and movement monitoring. The investigation of visual guidance of natural behaviour requires immersion in natural environments, choice of an appropriate behavioural paradigm, monitoring of eye and body movements, and recording the image sequence while the behaviour is executed. All of these are technically challenging. Until the 1970s, the primary device for presenting visual stimuli was the Maxwellian view optical system, which 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 behaviours that the subject can engage in is limited, and the stimulus conditions are also limited by the absence of observer movements. 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 (VR) are now cheap and comfortable, eye tracking within VR displays has vastly improved, and realistic environments easier to generate (e.g. [1]). Body movement monitoring has also improved. These technical developments have made it possible to observe in much greater detail the dynamic interplay between observer and environment than in the past (e.g. [2]). In this paper, we give a brief overview of some of the recent results documenting this dynamic interplay, and attempt to outline some of the broader implications that might

be relevant to the development of artificial vision systems. A more extensive review is given in Hayhoe [3].

The complexity of human visually guided behaviour poses a challenge, because for any given context we do not know what information is being acquired or how it is being used to control behaviour. One way to approach the problem is to assume that complex behaviours can be broken down into a set of specific sub-goals, each of which requires specific visual information [4–6]. Thus, when walking on rough ground the walker needs to select suitable footholds, view the surrounding scene and perhaps commit it to memory, control direction of heading and make global navigation decisions. Each of these particular goals requires some evaluation of the relevant visual information in the world in order to make an appropriate action choice at the moment. We do not, in fact, know precisely what this set of tasks is, or how they might change as the terrain gets smoother or the scene is more structured like a city street. Knowledge about what the momentary requirements of the visual system are is necessary for a complete characterization of a behaviour such as walking outdoors, and remains a challenge for understanding many natural tasks.

The simple example of walking outdoors reveals that normal behaviour can be characterized as continuous sequences of sensory-motor decisions that satisfy behavioural goals and the role of vision is to provide the relevant information for making decisions in order to achieve those goals. In a given context, the particular behavioural goal specifies the costs and benefits of an action in bringing about the goal. These costs and benefits are mediated by the underlying neural reward machinery [7]. The idea of reward is a critical component of understanding sensory-motor decisions, and one where there has been tremendous progress over the last decade. In active behaviour, the visual system uses gaze to sample necessary information from the environment in order to inform the subsequent action. Up to date information about the task-relevant state of the world, such as the location of an obstacle with respect to the body, is necessary to know what the consequences of an action are likely to be. Noisy sensory data and visual memory representations contribute to the subject's uncertainty about the visual scene, and factor into the decision of where to look and what action to take as a consequence. As many properties of natural environments have been learned through development and experience, it also seems likely that memory is a critical component in understanding visual control of actions. In this paper, we will consider how these factors—costs and benefits, memory and uncertainty—shape the choice of gaze location in natural behaviour.

2. Costs and benefits

One fundamental cost that is probably a pervasive influence on action choices is an energetic expenditure. Human locomotion is remarkably energy efficient due to our ability to exploit the inverted pendulum-like passive dynamics of the body. On flat obstacle-free surfaces, walkers naturally adopt a preferred gait that is maximally efficient for their individual passive dynamics comprising an optimal combination of step length, width and duration that minimizes energetic cost [8–10]. In this case, vision is barely needed. When the terrain is more complex and walkers must use vision to choose

footholds, Fajen & Matthis [11] have shown that it is necessary to select a foothold two or more steps ahead in order to preserve this energetic minimum. This strategy allows the walker to initialize the mechanical state of their body before the beginning of each step so that the ballistic trajectory of the centre of mass will facilitate stepping on the targeted footholds [12]. If the necessary visual information arrives after that, walkers need to brake or alter direction, which incurs an energetic cost [13]. In indoor environments, it has been shown that walkers do indeed direct gaze 1–2 steps ahead [14–16].

Although these findings are informative about the visual control of foot placement, it remains an open question how gaze is used to acquire the information necessary for locomotion in a natural environment. In the studies described above, obstacles and footholds were patches of light on the flat ground, and the visibility and location of the patches were artificially manipulated. We asked how walkers gather the information needed to guide foot placement in natural terrain, where stability is another important constraint. How does the benefit of moving the body efficiently trade-off against the need to place the foot in stable locations? To answer these questions during natural walking, we developed a novel system that integrates three-dimensional gaze and full-body motion capture of subjects walking in unconstrained outdoor environments. This work is described in more detail in Matthis *et al.* [2].

Subjects walked in a range of different terrains, with a Positive Science mobile eye tracker and a Motion Shadow Inertial Measurement Unit motion capture system for monitoring body movements. This allowed reconstruction of the body skeleton with the gaze vector intersecting the ground plane as shown in the inset of figure 1. To examine the location of subjects' gaze position relative to upcoming footholds, we first calculated the intersections between the three-dimensional gaze vector and a two-dimensional ground plane at the location of the subject's planted foot. Using the planted foot as the origin, we then accumulated the gaze distribution relative to the planted foot for each foot plant repeated across the entire length of the walk. This gives the gaze density relative to the planted foot (bottom right part of figure 1, top row). The distributions for all the terrains showed that subjects spend most of their time in each of the terrains looking at the ground roughly two to four steps ahead. However, if subjects direct gaze at the locations of their upcoming footholds then some of the spread of the distribution of gaze on the ground will be related to spatial variability in the footholds themselves because the path the subjects take is not straight, especially when the terrain is very rough. To remove this source of variability, we re-centred each gaze point on the locations of upcoming footholds. That is, rather than plotting gaze relative to the location of the planted foot (Foothold N), gaze location was plotted relative to the location where the foot was going to be placed next ($N + 1$), or to the footholds that were coming up after that ($N + 2$, 3 and 4); figure 1 shows the results for flat, medium and rough terrain. The top part of the figure is for medium terrain, and we found that the peak of the gaze density distribution was at step $N + 2$, that is the peak of the distribution in the $N + 2$ reference frame is centred over the origin. The gaze distribution for the rough terrain is a little different. Although subjects' gaze in that condition was most highly clustered around Foothold $N + 2$, a considerable

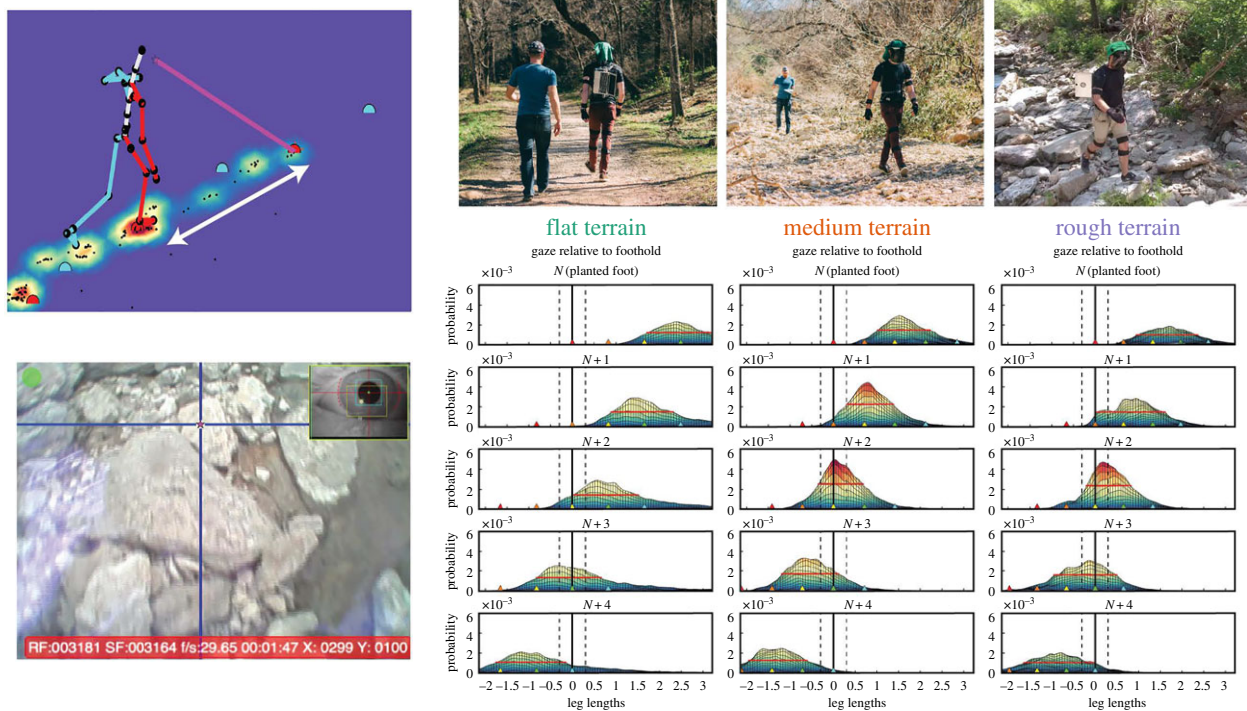


Figure 1. The right part of the top row shows the three terrains used. Top left shows the reconstructed data from the integration of eye and body measurements, with the gaze vector shown in pink. The white arrow indicates the gaze location relative to the planted foot. These values were cumulated for each footplant along the path and the resulting distributions are shown in the bottom right of the figure. Bottom left is the view of the path from the scene camera mounted on the subject's head. Bottom right shows a side-view of gaze distributions on the groundplane plotted relative to either the planted foot, foothold N in the top row, or foothold $N + 1$, $N + 2$, $N + 3$ or $N + 4$ in subsequent rows. Adapted from [2].

portion of the gaze density falls around foothold $N + 3$. By contrast, in the flat terrain, the gaze is only loosely linked to foot placement resulting in the only minimal change in the peak of the distribution when it is plotted in the different foothold-centred reference frames. In that condition, subjects only looked at the ground roughly 58% of the time, as opposed to 94% and 96% in the medium and rough terrains, respectively.

How do we interpret this behaviour? For the flat terrain, subjects adopt their preferred gait cycle, with stable step lengths and durations, and did not appear to use gaze to locate specific footholds. In medium terrain, subjects indeed choose to look two steps ahead as predicted from energy minimization. To achieve this, however, they needed to slow down so that the necessary information could be gathered in time. Thus, their step length shortened and step duration increased [2]. However, subjects still maintained a regular relationship between these parameters similar to that seen in the flat terrain. This indicates a desire to maintain an energetically efficient gait while still satisfying the demands on stability. The rough terrain is similar, in that subjects distribute gaze at step $N + 2$, but, in addition, they looked ahead towards foothold $N + 3$ sooner than they do in the medium terrain. The increased attention on Step $N + 3$ seems to be related to the greater need for path planning in the rough terrain, where stable footholds are sparser, and advance planning ensures that there will be a feasible path. Gait was also slower, with shorter steps, as with the medium terrain, and these parameters were also more variable. Steps take about half a second, and subjects often made two or three fixations in this period. Thus unsuitable footholds must be rejected and new saccade targets found on a time scale of 100 or 200 ms. This is achieved in

a tightly linked fashion that preserves energy efficiency while respecting stability, indicating that both factors contribute to the gaze and gait behaviour that emerges in the various terrains. We interpret this to mean that action decisions are dynamically regulated by costs and benefits on the time scale of single steps and single gaze changes.

A variety of previous findings also attest to the importance of energetic costs. Ballard *et al.* [17] investigated a scenario where subjects copied a model made up of eight coloured Duplo blocks. They were required to pick up appropriate blocks from a resource area and place them in a nearby area to make a copy. In this paradigm, subjects do not appear to hold the model pattern in memory, but instead make frequent looks back to the model pattern in the course of copying it. However, if the model pattern was located further 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 *et al.* [18] and Solman & Kingstone [19] has found similar trade-offs.

There are other intrinsic costs that are revealed in natural behaviour. For example, Jovancevic & Hayhoe [20] measured gaze distribution while subjects walked around a room in the presence of other walkers. Some of the walkers behaved in an unexpected and potentially hazardous manner, by briefly heading towards 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, 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 sensory-motor decisions that are being made on a time scale of a single gaze change. 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 & Ballard [21] and Tong *et al.* [22] 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 these subjective values factor into momentary action decisions. The unexpectedly low variability between subjects in many natural behaviours may be the result of a common set of costs and optimization criteria. By looking at natural behaviour that extends over time scales of seconds, we can gain insight into the factors that affect momentary action choices, what the task structure might be and what the subjective value of different actions is.

A large body of work over the last 15 years or so has explored the neural basis of the effects of task costs/rewards. Neurons at many levels of the saccadic eye movement circuitry are sensitive to reward (see [23], for a review). In particular neurons, in the Lateral Intraparietal Cortex that is likely involved in saccade target selection have been implicated in coding the relative subjective value of potential targets [24]. However, reward effects in neurons have been observed with very simple choice response paradigms where the animal gets rewarded for looking at a particular target, whereas in natural vision individual eye movements are not directly rewarded, but instead are for getting information that allows behavioural goals to be achieved, presumably with associated rewards [25]. Thus, the examples described above reveal how the underlying reward circuitry is manifest in natural behaviour. It is of interest to note that in the natural behaviour, eye and body reflect these reward effects in a linked manner, whereas the neurophysiology has focused on eye movements, with little work on reward effects on combined eye and body movements. Another point of interest is that the value of different behavioural constraints, such as energetic efficiency and stability must somehow be expressed in the same neural currency in order to modulate an action decision. It has been suggested that an area in the ventromedial prefrontal cortex may be responsible for computing a common currency for reward [26].

3. The role of memory in search

When visual search is conducted with a subject moving around in a real or virtual immersive environment, different factors come into play that are present in situations where the subject searches for targets on a screen. In the context of searching for an object in a scene, the decision where to

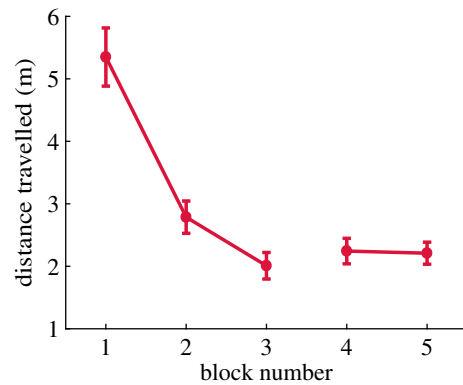


Figure 2. Total distance travelled during a visual search task where subjects walked around in a virtual three-room apartment as a function of experience in the virtual environment. Each block consisted of eight trials. Subjects rapidly learn the layout of the environment and adopt shorter paths. Adapted from Li *et al.* [29].

direct gaze can be made on the basis of the results of a visual search operation together with some visual memory representation of target location in a scene. How important the factor is visual memory? In natural behaviour, subjects are immersed in a relatively stable environment where they have the opportunity to develop precise long-term memory representations of object locations. As an individual moves around in the environment, it is necessary to store information about spatial layout in order to orient to regions outside the immediate field of view. This allows planning ahead, and this presumably leads to more energetically efficient strategies. 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 [27]. In a real-world search task, Foulsham *et al.* [28] 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. Li *et al.* [29] performed an experiment 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. Following the displacement, subjects frequently look at the old location on entering the room where the target had been located (58% of trials). This means that the signal used to programme the eye movement to that location depended on a memory representation, not on the visual information currently on the retina in that location. In addition, the subject's head and eye were directed at the old target location even before entering the room, when the target was not yet visible. By rapidly encoding, the global structure of the space subjects was able to reduce the total path walked. This is shown in figure 2, which plots the total path travelled during the search as a function of experience in the environment. By encoding the spatial layout subjects were able to reduce path length by eliminating regions where targets were unlikely to be and confining search to more probable regions. Thus, the 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. Thus, the energetic cost of moving around may lead subjects to depend more heavily on memory

rather than visually guided search, in a similar manner to the block-copying example described above.

4. The role of state uncertainty

The natural world is complex, dynamic and often unpredictable, so there are many sources of uncertainty about its current state. Consider the previously described example of walking in an outdoor environment. At any moment, there are a number of behavioural needs competing for gaze/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 relevant information acquired previously will also need to be held in working memory and 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 relevant image state is uncertain [4]. Thus, the probability of a change in gaze to update state increases as uncertainty increases. Evidence for the role of uncertainty in gaze shifts has been demonstrated by Sullivan *et al.* [30] and by Tong *et al.* [6]. Sullivan *et al.* used a virtual driving environment, and Tong *et al.* used a virtual walking environment. In both cases, gaze shifts reflected both the importance of the target (for example, the speedometer in driving or an obstacle when walking) and also the uncertainty about the state of the task-relevant information. Noise was added to the speedometer reading in the driving environment and obstacles drifted about randomly in the walking environment. Both these manipulations lead to increased gaze allocation.

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 & Rothkopf [31] 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 easily learned the temporal regularities of the events and traded off event detection rate with the temporal and energetic 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.

5. Self-motion and the role of visual memory

As we move through the environment we generate complex movement patterns that we effortlessly discount from externally generated motion. This is a pervasive feature of natural vision and one that must be addressed also by artificial vision systems. Recent evidence suggests that mechanisms for discriminating self-generated motion from

externally generated motion may be established very early in visual processing. In an experiment where mice ran on a treadmill and were presented with moving gratings linked to their running speed, Roth *et al.* [32] found evidence that the projections from the lateral posterior nucleus in the mouse thalamus (thought to be the equivalent of the pulvinar in primates) carried information to primary visual cortex 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 early visual cortex distinguish between self-generated and externally generated image motion [33], 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.

How does the visual system take account of the complex transformations of the image as observers move both body and eyes through the environment? 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. When stationary subjects view patterns of optic flow that include a moving object, they are able to parse the flow pattern into a component due to object motion and one potentially caused by self-motion [34–37]. However, because the subjects in those studies were not actually moving and self-motion was simulated, none of the sources of non-visual self-motion information that contribute to the perception of self-motion [38,39] were present. Fajen & Matthis [11] and Fajen *et al.* [40] used an ambulatory virtual environment viewed in a head-mounted display, so that self-motion cues were veridical, and performed a task that involved judgements about how to avoid moving obstacles. They conclude that when self-motion is real and actively generated, both visual and non-visual self-motion information (efferent motor commands and sensory reafference) 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. MacNeilage *et al.* [41] also showed that vestibular inputs can facilitate image parsing. These experiments underscore the fundamental nature of the perceptual processes that take self-motion into account.

In this context, scene memory may also play an important role, not in scene parsing, but in providing a mechanism for attracting gaze to unexpected features in the visual scene. If the scene is efficiently coded in long-term visual memory, subjects may compare the current image with the stored representation, and a mismatch, or 'residual' signal may serve as a basis for attracting attention to changed regions of scenes [42]. 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. The nature of such memory representation is something of a challenge, but recent models of visual pattern representations using Convolution Neural Nets [43] 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, non-task driven gaze changes can also be seen as leading to entropy reduction and more accurate representation of the state of the environment. Such a mechanism can also greatly reduce the computational load of mobile autonomous agents.

6. Conclusion

Because natural behaviour unfolds in time, we need to consider sequences of actions chosen by the subject. This is in contrast to the traditional trial structure where repeated sensory-motor decisions are made in response to experimental manipulations. 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. By looking at natural behaviour, it becomes clear that knowing the immediate behavioural goals is critical, as it defines the costs and benefits of momentary action decisions, which turns out to be a fundamentally important factor. Another important factor that emerges when subjects move around in the world is the visual memory. While it has long been accepted that memory can guide movements, it

is only in a behavioural context that we can evaluate how important a factor memory actually is. Not only is memory essential for navigating in natural environments, it allows the more efficient use of energetic resources, and is a fundamental component of momentary decisions of where to look and how to move. Finally, vision systems must take into account the effects of self-generated movements. While biological systems do this with apparent ease, the mechanisms are far from clear. Recent work using Convolution Neural Nets demonstrate that learning the statistics of self-generated movement patterns is computationally feasible and biologically plausible, and may form the substrate for visual guidance in mobile vision systems. As visual guidance of action is a ubiquitous property of animal behaviour, these principles might apply broadly to a range of different species. In addition, these features of human vision systems may provide guidance for artificial systems.

Data accessibility. Data from Matthis *et al.* [2] may be found here <https://doi.org/10.6084/m9.figshare.6130850.v1>. Datasets from papers discussed may be accessed via the relevant citation.

Authors' contributions. M.H. and J.M. wrote the manuscript.

Competing interests. We declare we have no competing interests.

Funding. Supported by NIH grant nos. R01 EY05729 and T32 EY021462.

References

1. Draschkow D, Vö ML-H. 2017 Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Sci. Rep.* **7**, 16471. (doi:10.1038/s41598-017-16739-x)
2. Matthis JS, Yates JL, Hayhoe MM. 2018 Gaze and the visual control of foot placement when walking over real-world rough terrain. *Curr. Biol.* **28**, 1224–1233. (doi:10.1016/j.cub.2018.03.008)
3. Hayhoe M. 2017 Perception and action. *Annu. Rev. Vis. Sci.* **3**, 4.1–4.25. (doi:10.1146/annurev-vision-102016-061437)
4. Sprague N, Ballard DH, Robinson A. 2007 Modeling embodied visual behaviors. *ACM Trans. Appl. Percept.* **4**, 11. (doi:10.1145/1265957.1265960)
5. Rothkopf C, Ballard D, Hayhoe M. 2007 Task and scene context determines where you look. *J. Vis.* **7**, 16., 1–20. (doi:10.1167/7.14.16)
6. Tong MH, Zohar O, Hayhoe M. 2017 Factors influencing gaze behavior in a walking task: rewards, uncertainty, and task modules. *J. Vis.* **17**, 28. (doi:10.1167/17.1.28)
7. Glimcher P, Fehr E. 2014 *Neuroeconomics: decision making and the brain*. 2nd ed. London, UK: Academic Press.
8. Holt KG, Jeng SF, Ratcliffe R, Hamill J. 1995 Energetic cost and stability during human walking at the preferred stride frequency. *J. Motor Behav.* **27**, 164–178. (doi:10.1080/00222895.1995.9941708)
9. Kuo AD. 2001 A simple model of bipedal walking predicts the preferred speed–step length relationship. *J. Biomech. Eng.* **123**, 264. (doi:10.1115/1.1372322)
10. Kuo AD. 2007 The six determinants of gait and the inverted pendulum analogy: a dynamic walking perspective. *Hum. Mov. Sci.* **26**, 617–656. (doi:10.1016/j.humov.2007.04.003)
11. Fajen B, Matthis J. 2013 Visual and non-visual contributions to the perception of object motion during self-motion. *PLoS ONE* **8**, e55446. (doi:10.1371/journal.pone.0055446)
12. Matthis JS, Barton SL, Fajen BR. 2015 The biomechanics of walking shape the use of visual information during locomotion over complex terrain. *J. Vis.* **15**, 1–13.
13. Matthis JS, Barton SL, Fajen BR. 2017 The critical control phase for the visual control of walking over complex terrain. *Proc. Natl Acad. Sci. USA* **114**, E6720–E6729. (doi:10.1073/pnas.1611699114)
14. Hollands M, Marple-Horvat D. 1995 Human eye movements during visually guided stepping. *J. Mot. Behav.* **27**, 155–163. (doi:10.1080/00222895.1995.9941707)
15. Patla AE, Vickers JN. 1997 Where and when do we look as we approach and step over an obstacle in the travel path? *Neuroreport* **8**, 3661–3665. (doi:10.1097/00001756-199712010-00002)
16. Hollands M, Marple-Horvat D. 2001 Coordination of eye and leg movements during visually guided stepping. *J. Mot. Behav.* **33**, 205–213. (doi:10.1080/00222890109603151)
17. Ballard D, Hayhoe M, Pelz J. 1995 Memory representations in natural tasks. *J. Cogn. Neurosci.* **7**, 66–80. (doi:10.1162/jocn.1995.7.1.66)
18. Hardiess G, Gillner S, Mallot HA. 2008 Head and eye movements and the role of memory limitations in a visual search paradigm. *J. Vis.* **8**, 7. 1–13.
19. Solman GJF, Kingstone A. 2014 Balancing energetic and cognitive resources: memory use during search depends on the orienting effector. *Cognition* **132**, 443–454. (doi:10.1016/j.cognition.2014.05.005)
20. Jovancevic, Hayhoe 2009 Adaptive gaze control in natural environments. *J. Neurosci.* **29**, 6234–6238.
21. Rothkopf C, Ballard DH. 2013 Modular inverse reinforcement learning for visuomotor behavior. *Biol. Cybern.* **107**, 477–490. (doi:10.1007/s00422-013-0562-6)
22. Tong M, Zhang S, Johnson L, Ballard D, Hayhoe M. 2015 Modelling task control of gaze. *J. Vis.* **15**, 784. (doi:10.1167/15.12.784)
23. Gottlieb J. 2012 Attention, learning, and the value of information. *Neuron* **76**, 281–295. (doi:10.1016/j.neuron.2012.09.034)
24. Trommersha J, Glimcher PW, Gegenfurtner K. 2009 Visual processing, learning and feedback in the primate eye movement system. *Trends Neurosci.* **32**, 583–590. (doi:10.1016/j.tins.2009.07.004)
25. Gottlieb J. 2018 Understanding active sampling strategies: empirical approaches and implications for attention and decision research. *Cortex* **102**, 150–160. (doi:10.1016/j.cortex.2017.08.019)

26. Levy DJ, Glimcher PW. 2012 The root of all value: a neural common currency for choice. *Curr. Opin Neurobiol.* **22**, 1027–1038. (doi:10.1016/j.conb.2012.06.001)
27. Hayhoe M. 2009 Visual memory in motor planning and action. In *Memory for the visual world* (ed. J Brockmole), pp. 117–139. Hove, UK: Psychology Press.
28. 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. Can. Psychol. Exp.* **68**, 8–19. (doi:10.1037/cep0000004)
29. Li CL, Aivar MP, Tong MH, Hayhoe MM. 2018 Memory shapes visual search strategies in large-scale environments. *Sci. Rep.* **8**, 4324. (doi:10.1038/s41598-018-22731-w)
30. Sullivan BT, Johnson LM, Rothkopf C, Ballard D, Hayhoe M. 2012 The role of uncertainty and reward on eye movements in a virtual driving task. *J. Vis.* **12**, 19. (doi:10.1167/12.13.19)
31. Hoppe D, Rothkopf C. 2016 Learning rational temporal eye movement strategies. *Proc. Natl Acad. Sci. USA* **113**, 8332–8337. (doi:10.1073/pnas.1601305113)
32. 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. (doi:10.1038/nn.4197)
33. Troncoso X, McCamy M, Iazi AN, Cui J, Otero-Millan J, Macknick S, Costela F, Martinez-Conde S. 2015 V1 neurons respond differently to object motion versus motion from eye movements. *Nat. Commun.* **6**, 8114. (doi:10.1038/ncomms9114)
34. Rushton SK, Warren PA. 2005 Moving observers, relative retinal motion and the detection of object movement. *Curr. Biol.* **15**, R542–R543. (doi:10.1016/j.cub.2005.07.020)
35. Warren PA, Rushton SK. 2008 Evidence for flow-parsing in radial flow displays. *Vision Res.* **48**, 655–663. (doi:10.1016/j.visres.2007.10.023)
36. Warren PA, Rushton SK. 2009 Optic flow processing for the assessment of object movement during ego Movement. *Curr. Biol.* **19**, 1555–1560. (doi:10.1016/j.cub.2009.07.057)
37. Warren PA, Rushton SK. 2009 Perception of scene-relative object movement: optic flow parsing and the contribution of monocular depth cues. *Vision Res.* **49**, 1406–1419. (doi:10.1016/j.visres.2009.01.016)
38. Harris LR, Jenkin M, Zikowitz DC. 2000 Visual and non-visual cues in the perception of linear self motion. *Exp. Brain Res.* **135**, 12–21. (doi:10.1007/s002210000504)
39. Harris LR, Jenkin M, Zikowitz D, Redlick F, Jaekl P, Jasiobedzka UT, Jenkin HL, Allison RS. 2002 Simulating self-motion I: cues for the perception of motion. *Virtual Real.* **6**, 75–85. (doi:10.1007/s100550200008)
40. Fajen B, Parade M, Matthis J. 2013 Humans perceive object motion in world coordinates during obstacle avoidance. *J. Vis.* **13**, 25. (doi:10.1167/13.8.25)
41. MacNeilage PR, Zhang Z, DeAngelis GC, Angelaki DE. 2012 Vestibular facilitation of optic flow parsing. *PLoS ONE* **7**, e40264. (doi:10.1371/journal.pone.0040264)
42. 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. (doi:10.1371/journal.pone.0094362)
43. Kriegeskorte N. 2015 Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* **1**, 417–446. (doi:10.1146/annurev-vision-082114-035447)