

# Approaches to understanding natural behavior

Alexander Goettker

Justus Liebig Universität Giessen, Giessen, Germany  
Center for Mind, Brain and Behavior,  
University of Marburg and Justus Liebig University,  
Giessen, Germany



Nathaniel Powell

The University of Texas, Austin, Texas, USA



Mary Hayhoe

The University of Texas, Austin, Texas, USA



Many important questions cannot be addressed without considering vision in its natural context. How can we do this in a controlled and systematic way, given the intrinsic diversity and complexities of natural behavior? We argue that an important step is to start with better measurements of natural visually guided behavior to describe the visual input and behavior shown in these contexts more precisely. We suggest that, to go from pure description to understanding, diverse behaviors can be treated as a sequence of decisions, where humans need to make good action choices in the context of an uncertain world state, varying behavior goals, and noisy actions. Because natural behavior evolves in time over sequences of actions, these decisions involve both short- and long-term memory and planning. This strategy allows us to design experiments to capture these critical aspects while preserving experimental control. Other strategies involve progressive simplification of the experimental conditions, and leveraging individual differences, and we provide some examples of successful approaches. Thus, this article charts a path forward for developing paradigms for the systematic investigation of natural behavior.

## Introduction

The implicit goal of vision research is to understand the natural function of the visual system and its neural basis. This stems from the pioneering work of (Gibson, 1950, Gibson, 1978), who suggested much of our understanding of human behavior can be explained by understanding the sensory stimulus, and the ways in which humans interact with the natural world (Gibson, 1979). Although much has been learned, there remain many unanswered questions. In particular, there is a relative dearth of information resulting from the direct investigation of natural behavior. A primary

reason for this has been technological. In the past, there have been a number of heroic attempts to design specialized systems for monitoring natural behavior, that have led to important insights (Land & Furneaux, 1997; Lee & Aronson, 1974). However, developments in eye and body tracking, together with numerical representation of the natural visual image, have made many investigations not only possible, but simpler, and much more feasible. As a consequence, there have been a variety of recent position papers arguing that now is the time to use more natural experiments when studying visually guided behavior (Cisek & Pastor-Bernier, 2014; Fookien et al., 2023; Hayhoe, 2017; Maselli et al., 2023). There has been a corresponding increase in articles that call for behavioral context in brain recordings (Cisek & Green, 2024; Krakauer, Ghazanfar, Gomez-Marin, MacIver, & Poeppel, 2017; Miller et al., 2022; Parker, Brown, Smeier, & Niell, 2020; Segraves, 2023). Thus, it seems clear that direct investigation of natural behavior is timely.

Despite these calls, the path forward is not entirely clear. How should one successfully study natural behavior and what are the defining features? One stumbling block has been the huge diversity and complexity of natural behavior. It is not obvious a priori how making a sandwich might generalize to buying groceries, or indeed why one should care to investigate either, so there is no rationale for choosing a particular behavior. Within a given task, there is also a diversity of things to measure and a diversity of questions one might ask. There needs to be some road map for the coherent organization of such work. The goal of this paper was to try to work out some part of that road map.

Historically, the organization of research into vision and action has been in terms of specific stimulus dimensions: for example, vision research

Citation: Goettker, A., Powell, N., & Hayhoe, M. (2025). Approaches to understanding natural behavior. *Journal of Vision*, 25(6):12, 1–17, <https://doi.org/10.1167/jov.25.6.12>.



has focused on color perception, motion perception, and so on. In motor control, research has focused on individual movements, such as reaching or grasping, or specific types of eye movements like saccades. Each of these capabilities seems to be broadly necessary for a successful sensorimotor system. Thus, psychophysical measurements have defined important limits and variations in sensitivity of motion, depth, and color constancy, as well as how individual motor behaviors are controlled. What is missing here is that these visual and motor capabilities are embedded in the context of more complex natural behavior, and the natural environment. What is needed is an understanding of how these individual capabilities contribute to the actions humans need to take, to function in the natural world.

To tackle this problem, we argue that the first step is a more complete description of natural behavior. This process helps to define the problems that vision and visually guided behavior have to solve and helps to specify the sensory input. We then describe the regularities and common features of many natural visually guided actions. Finally, we give some successful examples of targeted and controlled experiments

that retain some of the relevant structures of natural behavior and help us to provide ideas for the future path.

## Describing natural behavior as a first step

Many experiments are inspired by or indirectly related to real-world tasks. However, this does not guarantee that results obtained in such experiments will generalize to results and measurements in the natural world. To understand when controlled experiments generalize, it is necessary to first have a good description of the type of stimuli and behavioral responses we face in the natural world (Hayhoe, 2017; Kingstone, Smilek, & Eastwood, 2008). Advances in technology make a detailed description of natural behavior a timely enterprise (Figure 1). It is now feasible to collect complete data sets in a wide variety of natural contexts. Mobile trackers allow a detailed description of eye, head, and body movements (Foulsham, Walker, & Kingstone, 2011; Greene et al., 2024;

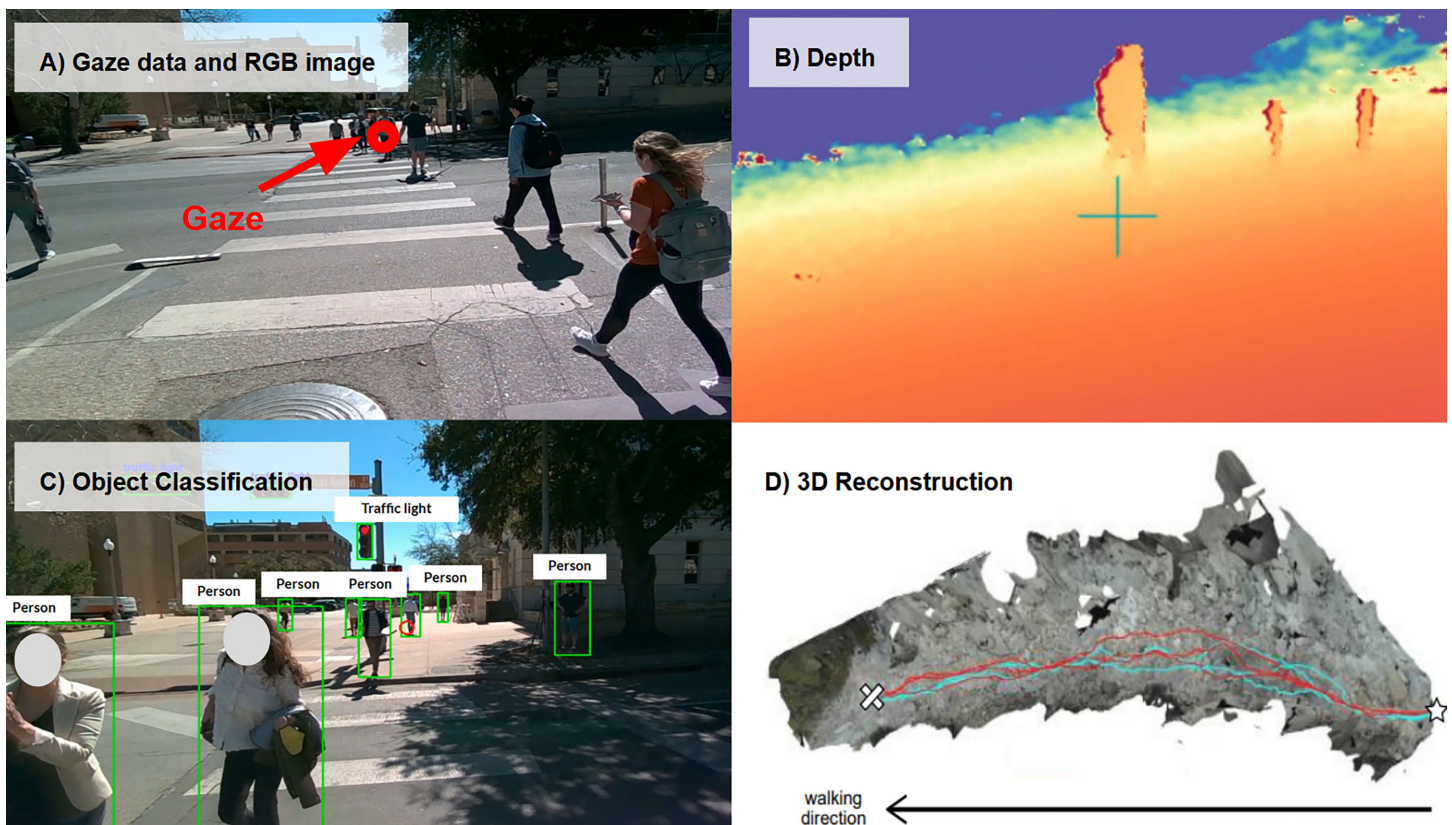


Figure 1. Possibilities with mobile data recordings. (A) Output of mobile eye tracking video with gaze overlay. (B) Output from a stereo camera integrated into mobile eye tracking unit. Crosshair denotes the gaze estimate. (C) Output of example object classification algorithm using the video from mobile eye tracking. (D) A 3D reconstruction of video taken during mobile eye tracking.

Hooge, Niehorster, Nyström, & Hessels, 2024; Matthis, Yates, & Hayhoe, 2018; Valsecchi, Akbarinia, Gil-Rodriguez, & Gegenfurtner, 2020), and computer graphics can also be used to track body movements (Nath et al., 2019). Computer vision algorithms make it possible to recover a numerical description of the images from the head-mounted cameras, which includes both the three-dimensional (3D) and the RGB information (Muller et al., 2023, Muller et al., 2024). The importance of having a reconstruction of the scene is that it makes it possible to investigate directly the image information associated with actions, rather than simply relying on video records of the scene. In addition to scene reconstruction, the availability of stereo cameras and lidar systems allows for additional precision in measuring the 3D structure of the visual scene (Figure 1B). These can be used in conjunction with mobile eye trackers to measure the distance of gaze in the scene without an explicit reconstruction phase, which can be infeasible in the presence of moving objects in the scene. The use of an eye-tracker–integrated lidar sensor directly provides a depth estimate in the reference frame of the eye tracker (DuTell, Gibaldi, Focarelli, Olshausen, & Banks, 2024), bypassing the need for reconstruction to obtain the relative depth of the point of gaze. Also, with advances in machine learning, it is now possible to classify objects in video recordings and automatically process and label the visual context (Kirillov et al., 2023; Wu, Shen, & Van Den Hengel, 2019). This strategy is especially useful for determining what objects people look at from mobile eye-tracking recordings. In addition, to improved measurements, also the available analytical tools have developed (see Maselli et al, 2023 for an overview). We describe some observations that reveal the critical importance of directly measuring natural behavior.

## Measuring behavior in the real world

An important aspect of natural behavior is that observers have to coordinate body, head, and eye movements to sample the visual world. However, this is often not given in an experimental setting. For example, to understand oculomotor control, experiments often use a setup with a computer monitor and a chin-rest to isolate eye movement behavior. In contrast, in natural behavior, the head is rarely stable and the body moves freely (Hayhoe & Lerch, 2022; Land, 1992; Pelz, Hayhoe, & Loeber, 2001). The addition of these degrees of freedom can have a substantial influence on how we move our eyes: even for the small saccade amplitudes between 5° and 10° of visual angle that are typically measured in the lab under head-restrained conditions if tested under unrestrained conditions, observers will naturally use accompanying head movements (Stahl,

1999), which changes the dynamics of the oculomotor system (Morasso, Bizzi, & Dichgans, 1973). For example, a very basic observation for saccadic eye movements is the so-called main sequence, a strong relationship between the amplitude and peak velocity of a saccade (Bahill, Clark, & Stark, 1975), but even this fundamental relationship changes when the eyes move alone or together with the head (Epelboim et al., 1997). In addition, while eye and head often work together to explore scenes (Bischof, Anderson, & Kingstone, 2023), how we control the head can differ from how we control the eyes (David, Beitner, & Vö, 2020; Solman, Foulsham, & Kingstone, 2017). In particular, in unconstrained viewing, multiple saccades are often made during one continuous head movement, with intervening fixations on objects while the eyes counterrotate with respect to the head to keep the gaze stable (Fang, Nakashima, Matsumiya, Kuriki, & Shioiri, 2015). This finding indicates that measurements of isolated eye movements, while being head fixed, might not be a good descriptor of gaze dynamics (which represents the combination of eye and head movements) under unconstrained, natural conditions.

## Measuring sensory input in the real world

Closely related to the unconstrained nature of available movements, is how those movements can change the sensory input. Perhaps the most important reason for looking first at natural behavior is that it allows measurement of the retinal stimulus in the context of that behavior. For example, while many databases of natural scene images exist (Mehrer, Spoerer, Jones, Kriegeskorte, & Kietzmann, 2021; Deng et al., 2009; Patterson & Hays, 2012; Grauman et al., 2022), the natural visual input does not consist of static images. We live in a dynamic world, and even when the head is restrained in front of a computer screen, a static image is transformed by eye movements into a complex spatiotemporal pattern that actively shapes visual perception (Rucci, Ahissar, & Burr, 2018; Intoy & Rucci, 2020; Casile, Victor, & Rucci, 2019).

This factor becomes even more important in the context of a 3D scene. Here, the actual retinal image depends not only on eye movements, but also on what the head is doing. One of the primary functions of eye movements is to stabilize the image in the context of ongoing head and body movements. During forward motion, walkers execute a saccade-and-fixate pattern, where the gaze location is held approximately constant in the scene while the body moves during a step (Muller et al., 2023). Figure 2A shows the vertical angle of the eye in the orbit. Saccades are visible as the rapid jumps and the slow segments show the period when gaze is stable in the world. This stability is achieved by slower counter-rotation of the eye in the orbit. In the case of

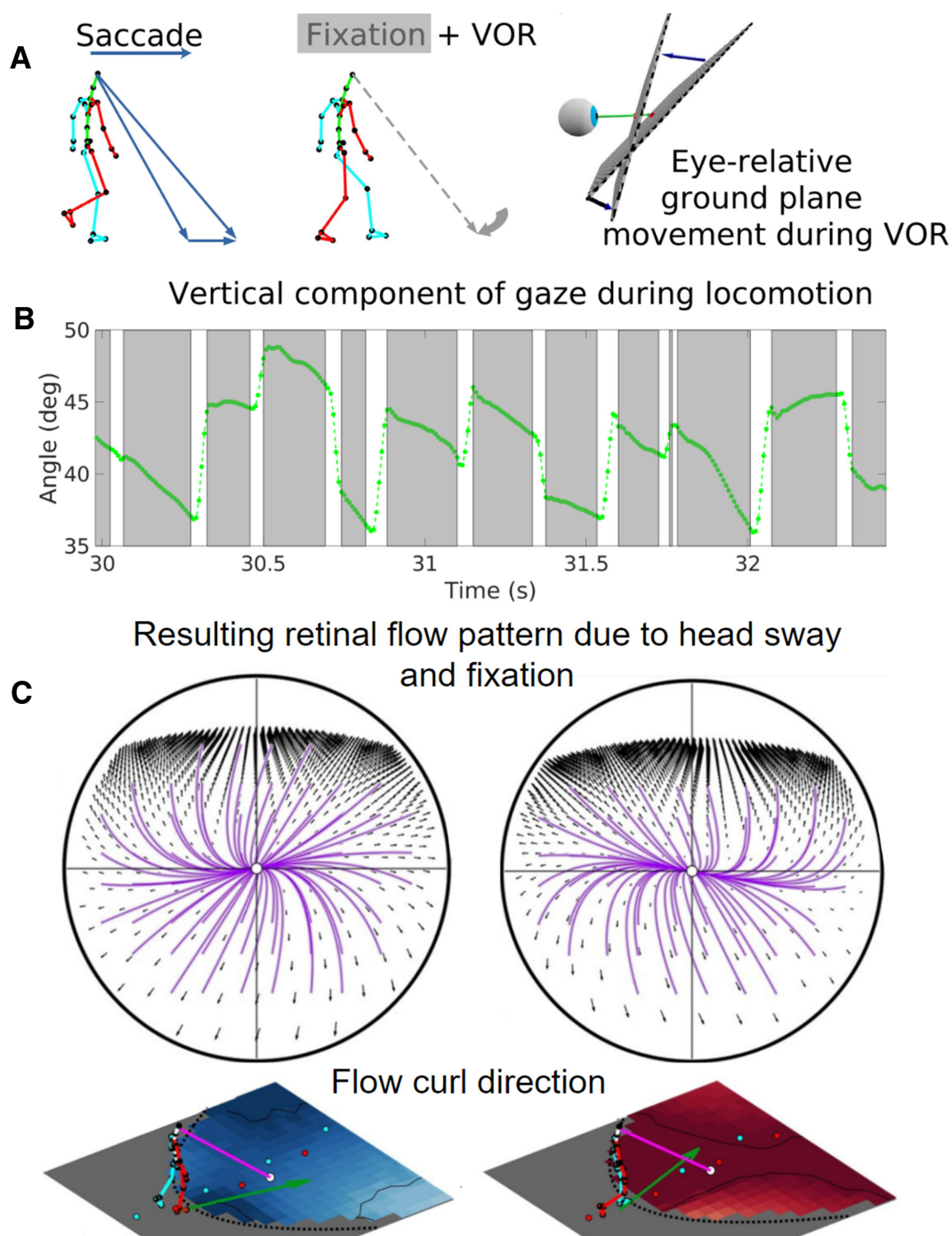


Figure 2. Oculomotor behavior and resulting retinal motion patterns during locomotion. **(A)** Schematic of a saccade and gaze stabilization during locomotion. The top right shows the ground plane relative to the eye expanding and rotating during a fixation. **(B)** Saccades in white panels and the counter rotation of the eye during the subsequent fixation in grey. **(C)** Retinal motion patterns during the gait cycle with stable gaze. The top row shows the retina-centered motion vectors. The bottom shows the gaze location (purple line) and momentary direction of the head (green arrow). The colors indicate the magnitude of the curl. On the bottom left, the curl results from movement of the head to the right of fixation. The bottom right shows the curl pattern resulting from movement of the head to the left of fixation as the body sways in the opposite direction, reversing the direction of the curl pattern. Adapted from (Muller et al., 2023) and (Matthis et al., 2022).

linear motion, the ground plane expands and rotates on the retina for gaze locations off the direction of travel. However, motion is not linear during locomotion. The head executes cyclical gait-induced movements that create complex motion patterns in the retinal regions outside the fovea when gaze is held fixed in the scene (Matthis, Muller, Bonnen, & Hayhoe, 2022). Figure 2B shows the expansion and rotation of the retinal image while the head moves leftward and rightward of the gaze point during a step. This pattern will vary with the location of gaze and the 3D structure of the scene (Muller et al., 2023).

Consequently, forward body motion does not generate a simple expanding retinal motion pattern. Retinal motion results only indirectly from body motion by virtue of the successive stabilizing movements. Matthis et al. (2022) suggested that these retinal motion patterns were most useful for monitoring balance during locomotion. Although this suggestion conflicts with the common interpretation of the role of optic flow to guide heading towards a goal, it is consistent with considerable literature on the role of image motion in postural control (Bardy, Warren, & Kay, 1999; Warren, Kay, & Yilmaz, 1996). Owing to these complexities, the measurement of the retinal image during locomotion is a particularly compelling example of the need to understand the retinal stimulus. In this case, the knowledge is important for interpreting the way image motion might be used in perception. The geometry that leads to these retinal motion patterns is quite general, given the 3D nature of the scene (Glennerster, Hansard, & Fitzgibbon, 2001; Koenderink & Doorn, 1976; Koenderink, 1986). These examples highlight why it is important to consider the dynamics and complexities of real-world sensory input to understand visual processing.

## Understanding what is learned

An important role for the careful evaluation of the retinal images experienced in a natural context is that this is what the visual system adapts to during development, and presumably, what shapes many aspects of visual coding. As described elsewhere in this article, retinal motion patterns are determined by an interconnected set of factors, including gaze location, gaze stabilization, and the structure of the environment. The characteristics of these motion signals likely have important consequences for neural organization.

Using the data collected by Matthis et al. (2022) and Muller et al. (2023), these authors summarized the retinal motion statistics experienced by adults in a range of natural terrains without other moving objects. Features of these statistics such as the fact that motion is always zero at the fovea can explain why

humans appear to adopt a zero velocity motion prior (Weiss, Simoncelli, & Adelson, 2002). Similarly, motion direction is nonuniformly distributed as a consequence of forward motion along the ground plane. Because forward motion across the ground plane covers a substantial segment of visual experience, such statistics might prove useful in understanding the properties of cells in motion-sensitive cortical areas (Beyeler, Dutt, & Krichmar, 2016; Mineault, Bakhtiari, Richards, & Pack, 2021).

Successfully learning how self-motion affects the visual input is also critical in other aspects (Rolfs & Schweitzer, 2022). To successfully interpret the retinal input, we need to differentiate motion in the world from that generated by our own movements (Haarmeier, Thier, Repnow, & Petersen, 1997; Sommer & Wurtz, 2008). This distinction has been extensively discussed in the context of saccadic suppression: To keep a stable percept of the world despite consistent eye movements, visual sensitivity is reduced around the time of a saccade (Binda & Morrone, 2018; Ross, Morrone, Goldberg, & Burr, 2001). This decrease seems to be the result of a combination of multiple different mechanisms that can involve retinal processes (Idrees, Baumann, Franke, Münch, & Hafed, 2020) and extraretinal information about the upcoming eye movement (Diamond, Ross, & Morrone, 2000). However, a recent set of studies indicated that a critical factor is also related to previous learning experiences: the visibility of a high-speed motion stimulus is lawfully linked to the expected motion for an eye movement of a given amplitude (Rolfs, Schweitzer, Castet, Watson, & Ohl, 2023) and with a little bit of training, even new contingencies between the expected visual stimulus and saccadic amplitudes can be learned (Pome, Schlichting, Fritz, & Zimmermann, 2024). This finding suggests that specific predictions based on saccade-contingent learning might explain why we do not perceive retinal stimulation during saccades (Pome et al., 2024; Zimmermann, 2020).

The additional value of collecting the statistics of natural experience is also given in a recent study by Anderson, Candy, Gold, and Smith (2024), who recorded the visual input of young infants and adults at home using head-mounted cameras. By analyzing the statistics of the visual input, they observed that scenes with sparse edge patterns dominate, and this result provided a direct explanation of why young infants show preferences for this type of stimulus in a laboratory setting. In general, Yu et al. have shown that the viewpoint of the adult is quite different from that of the young child, and this has an impact on language learning (Bambach, Crandall, Smith, & Yu, 2018; Smith, Yu, & Pereira, 2011; Yu & Smith, 2012). During toy play, infants' field of view tends to contain fewer objects than their parents', and these objects seem to be larger. Children's actions, such as

object manipulation, shape their field of view and also affect learning. If infants had both hands and eyes engaged with the object as it was labeled, they were more likely to learn the object–label mapping (Schroer & Yu, 2023). Together, these results highlight the importance of understanding the statistics of the visual input learned during development and with experience.

## The role of spatial and temporal context

The importance of knowing the statistics of the visual input also goes beyond a deeper understanding of how these processes are shaping the sensitivity and properties of receptive fields in the visual system during development. We need to know the visual input to understand which statistical regularities and contextual information humans extract to actively shape perception and behavior.

### Spatial context

Although object perception is often studied with isolated objects typically presented at the fixation location, how and where objects are presented can be important (Kaiser, Quek, Cichy, & Peelen, 2019). For example, face and body parts evoke more distinct responses in the occipitotemporal cortex when being presented at their typical locations (e.g., an eye in the upper visual field) (Chan, Kravitz, Truong, Arizpe, & Baker, 2010). Memory for multiple objects is also facilitated when being arranged in accordance with real-world positional regularities (Kaiser, Stein, & Peelen, 2015). Similarly, learned relations between different natural objects (e.g., a toothbrush will be close to a sink) facilitates visual search (Vo, Boettcher, & Draschkow, 2019). Thus, perception in a natural scene is more than just understanding a single object at a time, but is strongly shaped by our previous experiences with natural contexts.

### Temporal context

The statistics of experience are also reflected in the temporal context. For example, gaze is often predictive of physical events such as the future trajectory of a ball (Diaz, Cooper, Rothkopf, & Hayhoe, 2013; Land & Furneaux, 1997), and reflects an abstract understanding of the spatial and temporal context in the scene (Goettker, Pidaparthi, Braun, Elder, & Gegenfurtner, 2021; Kowler, Rubinstein, Santos, & Wang, 2019; Stewart & Fleming, 2023). Conversely, when the natural statistics are violated, subjects quickly learn more adaptive gaze behavior (Jovancevic-Misic & Hayhoe, 2009). In general, predictive movements seem to naturally occur, for example, when trying to intercept a

moving object (Fookien et al., 2023), preparing a peanut butter jelly sandwich (Land & Hayhoe, 2001), or setting up a tent (Sullivan, Ludwig, Damen, Mayol-Cuevas, & Gilchrist, 2021). This finding suggests that, in natural behavior, prediction is the default.

Thus, the investigation of behavior in the real world can tell us what is important in the daily repertoire: The complexity of the sensory input and behavior, the effects of developmental learning, and spatial and temporal context all have profound influences on how we behave, and are hard to address without an exploration of natural, visually guided behavior, and an understanding of visual perceptual learning.

## Natural behavior as sequences of decisions

We have argued that it is necessary to start with an examination of the properties of natural behavior in situ to understand the nature of the visual stimulus and the demands made on the visual system. However, this leaves the problem of what behavior to investigate. How can one generalize from a single behavioral context or task? One approach to the problem is data driven: In recent years there has been seminal work on collecting large open access data sets that measure eye and head movements while people engage in their normal activities (Engel et al., 2023; Greene et al., 2024; Kothari et al., 2020). These datasets can be used to establish fundamental sensory input and behavioral statistics over a wide range of natural contexts. An alternative approach is to recognize that existing studies reveal that there are many commonalities in natural visually guided behavior over a range of different tasks such as driving, making tea and sandwiches, and playing sports. These commonalities help to simplify the issue. Many natural behaviors can be viewed as sequences of decisions, and this helps to provide a unifying theoretical framework for better describing natural behavior. Next, we review some of the critical parts of this continuous decision-making process.

## Inferring world state

To survive, humans and other organisms must make good action decisions, and the role of vision is to provide the information necessary to make those decisions. To choose a suitable action, humans need to estimate the state of the world. Given the time-varying retinal stimulus, the information sampled is not always definitive. It is commonly accepted that sensory data is combined with information stored in memory (the

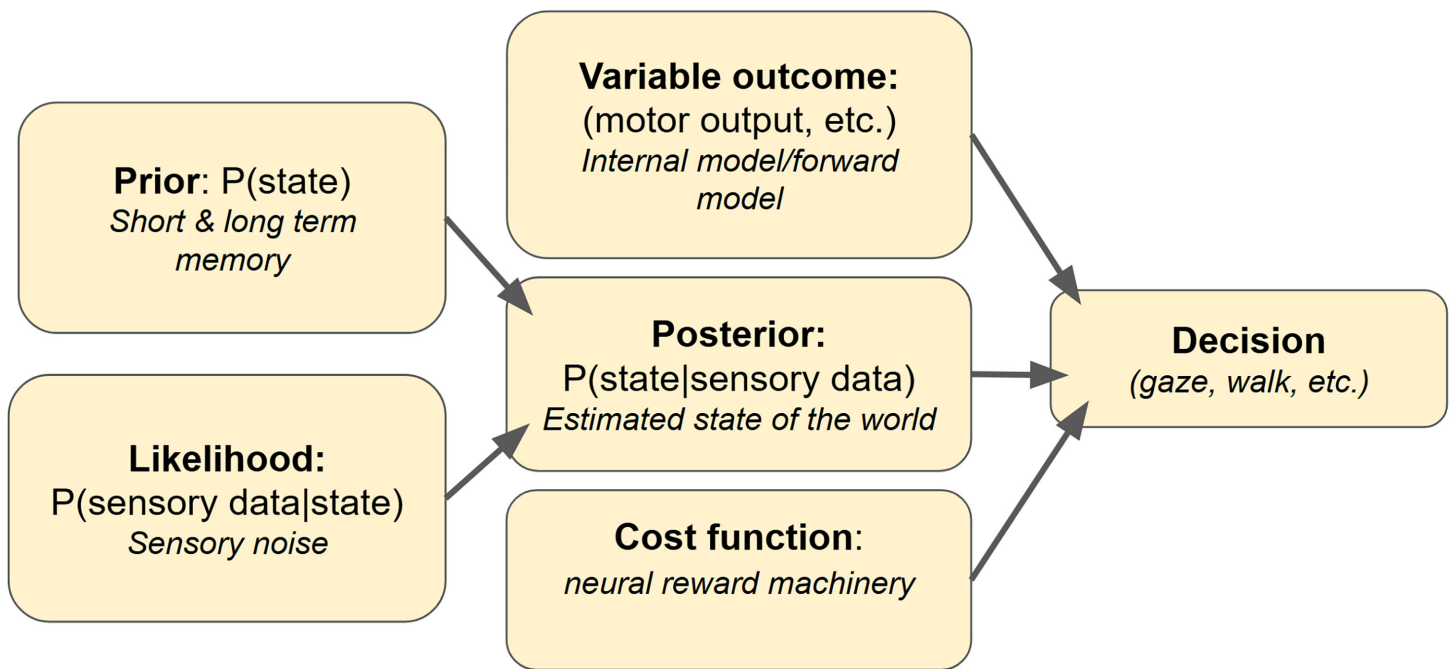


Figure 3. Factors influencing action choices. Sensory data (the likelihood) is combined with information stored in memory (the prior) to compute an estimate of world state (the posterior). The posterior is then used to make action decisions. These action decisions must also take into account noise in the motor system (variable outcomes). The behavioral context determines the costs and benefits, and in consequence, what action should be selected (cost function). This is how a single action decision is made, but it does not take into account the temporal evolution of decision making in the presence of changing visual stimuli.

prior), to compute an estimate of world state (Kersten, Mamassian, & Yuille, 2004; Knill, 1996). This posterior estimate of the probability of the world state, given the data and the prior, can be more or less noisy, depending on the precision of both the memory and of the sensory data. This uncertainty about world state is an important factor that contributes to action choices (see Figure 3 for a schematic of how action choices are made). For example, when walking on flat ground, there is little need to update world state, so walkers can use experienced-based prior knowledge to control locomotion. However, as terrain complexity increases, humans slow down and adjust gaze location to be closer to the body as the complexity of the terrain increases (Matthis et al., 2018; Muller et al., 2023) or if the sensory data is restricted to monocular vision (Bonnen et al., 2021). These nuanced adjustments of the action suggest that uncertainty about the visual image is an integral part of the computations that determine the action. In this context, it becomes clear, also, that although stimulus noise is sometimes a problem, it is often the case that relying on one's prior knowledge is also a way of saving attentional resources, allowing gaze to be devoted to some other task. Although at first blush this variability of the behavior might seem to be problematic, in the context of action choices it can be seen as a stable feature of the decision-making process.

## Behavioral goals

Behavioral goals are a defining feature of action decisions. They define what stimuli are relevant and what actions are appropriate. This is a pervasive constraint in understanding natural behavior, and is clearly demonstrated in the work by (Land & Furneaux, 1997; Land, Mennie, & Rusted, 1999) and (Hayhoe, 2017; Hayhoe & Lerch, 2022; Hayhoe, Shrivastava, Mruczek, & Pelz, 2003). Whether making tea or sandwiches, or playing table tennis, movements of the eye, head, hand, and body are coordinated in space and time to acquire the information needed at that moment to accomplish a step in the task, and sequential execution of those steps complete a larger goal. In addition to defining the relevant stimulus information and the time it is needed, the behavioral context determines the costs and benefits, and in consequence, what action should be selected. For example, walkers usually adopt a preferred gait that minimizes energetic costs (Finley, Bastian, & Gottschall, 2013; Kuo, Donelan, & Ruina, 2005; Lee & Harris, 2018; Selinger, O'Connor, Wong, & Donelan, 2015), but move very differently if under time pressure. These behavioral adjustments to the time-varying costs and benefits are driven by the neural reward machinery (Schultz, 2015) that is a centrally important factor in all action choices. In the above example of walking on rough terrain,

slowing down presumably reflects not only sensory uncertainty but also the increased cost of falling. Rough terrain also puts greater demands on movement accuracy. Consequently, action decisions must take into account noise in the motor system, just as they must take into account sensory noise (Trommershäuser, Maloney, & Landy, 2008). Human subjects can learn to adjust their actions such as fast pointing movements when the variability of action outcomes and the associated rewards are manipulated experimentally (Seydell, McCann, Trommershäuser, & Knill, 2008). This, too, is an integral part of the computation that determines the action choices. Thus, walking speed while crossing the road takes into account both variability in the road surface (perceptual estimate) and the importance of arriving at the destination quickly (the costs), as well as the probability of tripping when walking quickly (motor variability). As mentioned elsewhere in this article, this variability is entirely expected within the framework of choosing good actions that meet behavioral goals, although it might limit the generalization of experimental results when the costs are not known. The advantage of natural behavior in this instance, is that the behavioral goals are known (at least to some extent), so that the potential variables are usually easy to identify. Consequently, behavioral context in many ways simplifies the problems.

## Coordinating actions over time

Finally, behavior evolves through time, on the scale of seconds or minutes. An action such as an eye or body movement changes the sensory input and thus the evaluation of world state. This, in turn, determines the next action decision, and natural behavior evolves as a sequence of actions that are required for behavioral goals such as crossing a road or intercepting an object. This raises a variety of complex issues. For example, there must be some mechanism for choosing the next step in a behavioral sequence. In a task such as making a sandwich, this might be determined by a learnt procedure, but often the next step is probabilistic (Hayhoe & Ballard, 2014). Another concern is that vision, eye movements, and body movements all function at different time scales and must be coordinated appropriately for the task. This would be impossible without working memory and motor planning, both of which are essential components of natural behavior. For example, when approaching an obstacle in the path, walkers increase speed before stepping over the obstacle and then slow down subsequently (Darici & Kuo, 2023). This strategy is most energetically efficient and reveals a plan that spans a sequence of steps. A concomitant of planning is working memory. Working memory is required when a complex world state involving

different sets of information is required for an action decision. For example, changing lanes while driving requires knowledge of the cars in both lanes, both ahead and behind the driver. Addressing the role of working memory in the execution of natural behavior is a centrally important question (Ballard, Hayhoe, & Pelz, 1995; Draschkow, Kallmayer, & Nobre, 2021; Hayhoe, 2009). The temporal continuity of natural behavior means that the neural decision circuits must also operate fluidly over time spans of seconds, in this manner. Although this process seems to be very complex, it can be seen as a critical aspect of behavior that gets left out in many experiments where there is a trial structure. It is also a domain where there has been relatively little experimental focus.

We can see from this analysis of natural behavior that action decisions form a natural organizing structure that includes, in a fairly straightforward manner, the things we want to understand. Therefore, viewing visually guided actions as decisions provides an important unifying principle (Franklin & Wolpert, 2011; Maloney & Zhang, 2010; McNamee & Wolpert, 2019; Wolpert & Landy, 2012). This framework structures the problem and provides the context for much recent progress in the perception and action field. More generally, natural behavior should be thought of not simply as individual decisions, but as sequences of decisions (Hoppe & Rothkopf, 2019; Hoppe & Rothkopf, 2016; Kessler, Frankenstein, & Rothkopf, 2024) because this process allows the consideration of prediction, planning, and memory. Within this framework, experiments can then probe individual parts of the decision process to see how behavior adapts. For example, by experimentally manipulating sensory and motor noise, studies were able to demonstrate that these variables trade-off in a flexible manner to optimize performance (Battaglia & Schrater, 2007; Sims, Jacobs, & Knill, 2011). This flexibility reveals the existence of well-specified internal models of both sensory and motor noise.

## The controlled investigation of essential elements of natural behavior

Given the diversity and complexities of natural behavior, it is unclear how to design experiments that capture the essence of it. We argued that, as a path forward, we should start to see natural behavior as a sequence of decisions and systematically investigate parts of this decision process. By considering the commonalities and complexities of natural behavior, it is possible to devise paradigms for more controlled experiments that exhibit some of the features of natural behavior while constraining the experiment to allow

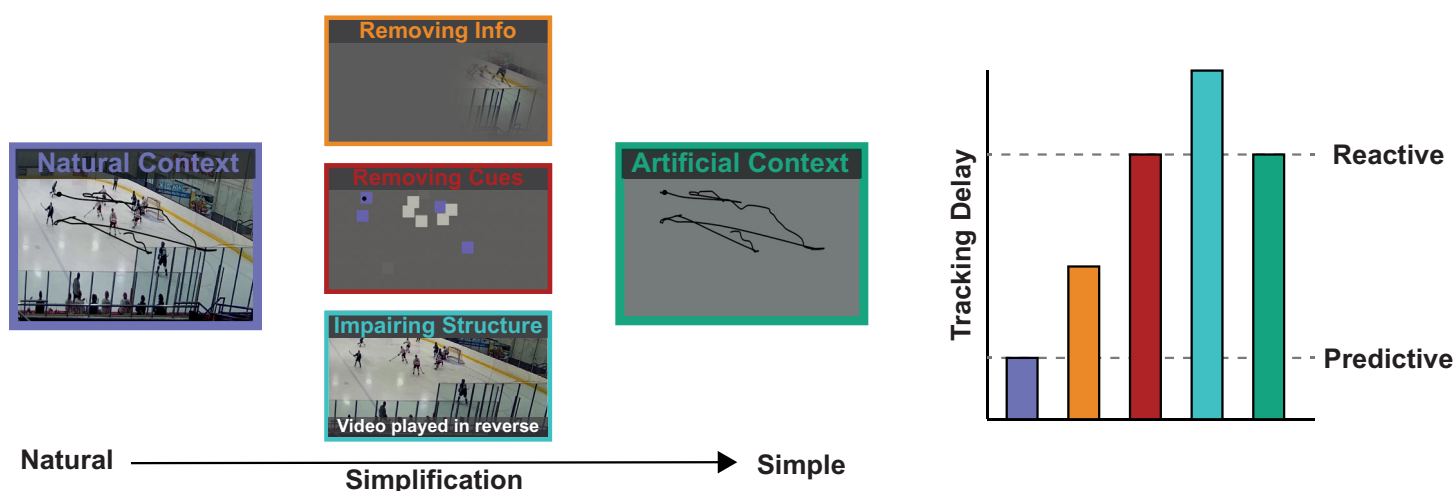


Figure 4. Examples for experiments systemically varying the naturalness of sensory input. In the natural condition, observers viewed ice hockey videos and had to track the puck. The simplest condition only showed the puck movement without any context. In intermediate conditions, the amount of visual information was varied, kinematic cues were removed by replacing players with squares, or the causal structure of the scene was impaired by playing the video in reverse. The results showed that observers showed predictive tracking behavior only under natural conditions, and tracking was reactive when the possibility of scene understanding was impaired. Figures and results are adapted from (Goettker et al., 2021; Goettker et al., 2023).

strong conclusions. In what follows, we give some examples of how to move from simple observational studies to more controlled paradigms that seem likely to generalize beyond the specific experiment.

### Systematically varying the sensory input

A successful strategy to investigate how observers try to infer the world state from sensory input is by systematically manipulating the stimulus. Here one can use a paradigm that starts with more complex natural visual input and then systematically removes information to find the critical factors underlying performance. For example, this strategy was taken in a recent set of studies where Goettker et al. tried to understand which cues and factors are critical for predictive eye movements often observed in natural behavior (Goettker et al., 2021; Goettker, Agtzidis, Braun, Dorr, & Gegenfurtner, 2020; Goettker, Borgerding, Leeske, & Gegenfurtner, 2023). In these experiments, observers saw videos of an ice hockey game and needed to track a specific target (the puck). In the fully natural videos, observers pursued the puck effortlessly with no tracking delay, and even made saccades ahead of the puck when there were passes between players. In contrast, when subjects saw the same target trajectory (the isolated puck) with no context, they showed purely reactive behavior with substantial tracking delays. Intermediate conditions were then created by systematically manipulating the videos to vary the amount of available information, removing cues about player movements

by replacing them with boxes, or impairing the causal understanding by playing the video in reverse. The results demonstrated that observers showed predictive tracking only if scene understanding was possible. When scene understanding was impaired, for example, when the video was played in reverse, or only boxes were visible indicating the position of the players but not their kinematics, the behavior was reactive or even disturbed when the video was played in reverse (Figure 4). Thus, these results demonstrate that predictive tracking based on expectations and previous experience is the default in complex dynamic natural situations. This result is to be expected if the critical feature of behavior is to use the current world state to anticipate future states. Experimental paradigms where the stimulus is systematically manipulated to take away potential cues for these predictions are an ideal testbed for how they are formed under natural circumstances.

### Translating natural behavior to controlled experiments

A related approach to varying the stimulus is to simply observe natural behavior and use the resulting insights to develop more controlled experiments that directly test hypotheses that arise about natural behavior. An example of this is a study by Foulsham et al. (2011), where people walked to a cafe to get a cup of coffee while wearing a mobile eye-tracking device. They observed that observers avoid fixating faces when they got close, suggesting some social constraints on

looking at others. They then compared how frequently people fixated on faces in the real-world task with videos of first-person recording of someone getting coffee. They observed that, especially when other people got close, observers kept fixating on them in the video condition, much more than in the real-world situation. This supported the hypothesis of social regulation of interpersonal gaze in this situation.

In another example, [Matthis et al. \(2022\)](#) calculated the effects of gait on the patterns of retinal motion when walking over various outdoor terrain types. They showed that retinal flow patterns were influenced substantially by movements of the head over the gait cycle, and suggested that the retinal signal is most likely used for controlling balance and posture and may not be useful for heading toward a goal unless it is possible to integrate direction of heading over the gait cycle. The role of retinal flow patterns in controlling balance while walking has some support in the literature (see review in [Matthis et al., 2022](#)). The direct link between self-generated flow and walking is somewhat unclear, since retinal motion is determined by idiosyncratic head movement and gait, it seems likely that individuals need to learn their own characteristic retinal flow patterns. Therefore, deviations from their expected flow patterns may change where walkers place their feet to preserve balance. In a follow-up study, [Powell, Oh, Panfili, and Hayhoe \(2023\)](#) examined this hypothesis and showed that the retinal flow from the ground plane, where the influence of gait on the flow is most prominent, is important for guiding walking trajectories and foot placement. This evidence supports Matthis's suggestion that retinal flow patterns control foot placement. This finding shows an example of how collecting data on a well-defined natural behavior can lead to important questions about visual control of actions. Thus, the measurement of natural behavior can provide a starting point for the development of new hypotheses that can be tested systematically in more controlled laboratory settings.

## Experiments that contain features of sequential decision-making

Another strategy is to devise experiments that involve some of the complexities of sequential decision making, but are constrained enough to focus on one particular aspect. Leveraging the possibilities of new recording techniques, there have been some recent studies explicitly focusing on the coordination of different movements over time, showing how flexibly people adopt their behavioral sequences to the dynamics and requirements of a task ([Fookan, Johansson, & Flanagan, 2024](#); [Keshava et al., 2024](#); [Schroeger, Goettker, Braun, & Gegenfurtner, 2024](#)). One specific example of this strategy is an experiment

by ([Hoppe & Rothkopf, 2016](#)), who investigated how people learn environmental event statistics and use them as priors to guide behavior. To do this, they devised a temporal event detection task where subjects looked to the left and right of fixation to detect events that varied in duration and were drawn from different distributions of duration. Over trials, subjects learned these distributions and varied their viewing times to improve performance. In this respect, the temporal event detection task ([Hoppe & Rothkopf, 2016](#)) is comparable with natural vision. An advantage of their carefully sculpted task was that they were able to predict performance if they also accounted for motor noise in the duration of a fixation as well as the cost of a saccade. By taking account of all these decision variables in the context of a model, they were able to explain performance in detail. Their paradigm could then be used to estimate how the cost of a saccade might vary with amplitude, for example, and allow estimation of some of the internal parameters associated with action decisions. Other work by Rothkopf et al. adopts comparable methodology. For example, [Hoppe and Rothkopf \(2019\)](#) used similar models to predict sequences of saccades in a visual search task, and [Kessler et al. \(2024\)](#) accurately predicted navigation performance by taking into account both sensory and motor noise. These papers demonstrate that it is possible to devise paradigms that can manipulate and estimate the complex factors that influence natural behavior. Thus, it is possible to make progress by explicitly taking into account the variety of factors in natural decisions, and either holding them constant, or explicitly manipulating selected variables.

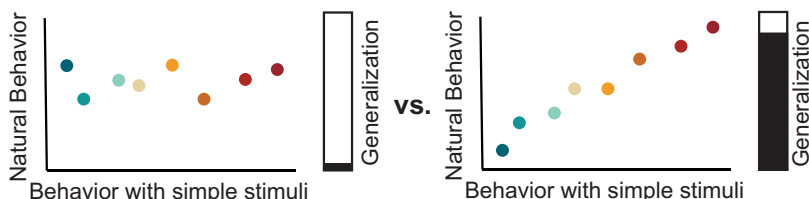
## Leveraging individual differences to test generalization

Another tool to bridge the gap between simplified experiments and more natural behavior can be the use of individual differences (see [Figure 5](#)). Individual variability is often treated as measurement noise ([Mollon, Bosten, Peterzell, & Webster, 2017](#); [Wilmer, 2008](#)). However, we know from a wide range of tasks that there are stable and reliable individual differences both in perception ([Cretenoud, Grzeczowski, Kunchulia, & Herzog, 2021](#); [Grzeczowski, Clarke, Francis, Mast, & Herzog, 2017](#)) and in motor control: reliable individual differences in accuracy and latency of single movements responses ([Bargary et al., 2017](#); [Ettinger et al., 2003](#)), as well as when continuously sampling visual information during free-viewing images of natural scenes ([Andrews & Coppola, 1999](#); [Castelhano & Henderson, 2008](#); [Zangrossi, Cona, Celli, Zorzi, & Corbetta, 2021](#)) have been observed. We know from previous research that these stable individual differences are more than noise, and can

## Tasks with different complexity



## Potential outcomes:



## Example:

Experiment 1: Naturalistic Visual Search Experiment 2: Classic Search



Task: Find Object

Task: Find red T

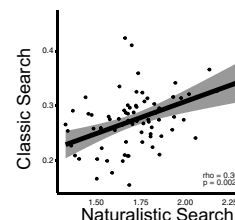


Figure 5. Illustration of the individual difference approach. (Left) By measuring individual differences in varying settings ranging from single dots in a screen-based experiment, to unconstrained natural behavior, it is possible to test for a relation between these scenarios. Low correlations between these different tasks suggest a lack of possible generalization, high correlations demonstrate the importance of the mechanisms and behavior identified in the simpler experiment. (Right) Example adapted from a study by Botch and colleagues (Botch et al., 2023). Visual search performance was measured either for naturalistic stimuli or in a classical abstract visual search task. Below is the correlation between these two tasks, which indicates around 10 percent of shared variance between the tasks.

provide insights into sensorimotor processing (De Haas, Iakovidis, Schwarzkopf, & Gegenfurtner, 2019; Moutsiana et al., 2018). Based on this logic, we can use individual differences to test whether results from simpler experiments successfully generalize to describe more natural behavior. If the same observers who show better performance in a simplified experiment when an isolated mechanism is tested also show better performance in more natural conditions, critical features of natural behavior are captured by the simplified experiment and the results can be generalized. The strength of this approach is that the inverse holds as well: If there is no relationship between the individual differences across different behavioral or stimulus complexities, results in the simpler task do not generalize. We provide some examples that used this approach.

In recent work, Goettker and Gegenfurtner (2024) showed that while saccades and pursuit are often investigated in isolation (Goettker & Gegenfurtner, 2021), in natural behavior they work closely together. Here, a large group of observers performed multiple tasks that measured the performance of isolated saccadic and pursuit eye movements as well as tasks that measured their interaction. The links across these tasks showed that the coordination of saccades and pursuit is tailored to the strength of individual observers. Observers with more accurate saccades to moving targets also rely on catch-up saccades more frequently when tracking moving targets. Thus, the sequential behavior of different types of eye movements

observed in more natural tasks, can be successfully explained by the individual differences in the simpler experiments.

In a similar approach, Botch, Garcia, Choi, Feffer, and Robertson (2023) compared visual search performance of the same group of observers both in a classical visual search task and also in a complex 3D environment in virtual reality (VR). They observed that individual differences in search efficiency for the simpler task could predict search efficiency in the complex environment, suggesting that similar mechanisms are involved (Figure 5). As mentioned elsewhere in this article, the strength of this approach is that it is also possible to quantify the level of similarity. The magnitude of the correlation between search efficiency in the two tasks suggests that roughly 10% of the variance in more natural search behavior could be explained by the variance in the simpler task. This suggests that, although behavior across these tasks shows similarities, only a smaller amount of the explainable variance is captured by the simplified search task. This in turn can then point to the relative importance of mechanisms that are not captured in the simpler search task.

## Using VR

An emerging technological approach to help with the typical trade-off between experimental control and natural validity is the use of VR. VR allows the

presentation of complex naturalistic environments (Scarfe & Glennerster, 2015) while at the same time keeping experimental control high and measuring behavior at the same time. It has many potential use cases in research.

VR allows us to simulate typical real-world situations to test behavior there. An example of this is from Shinoda, Hayhoe, and Shrivastava (2001), who looked at the allocation of attention while driving in a simulator. When environments are only partly predictable, such as in driving, it is not clear what controls the allocation of attention to make sure that drivers see important events. This is one of the unsolved problems of vision that can only be addressed in circumstances where behavior is monitored over an extended time period and the allocation of attention is controlled by the observer, as in normal life. They found that drivers frequently miss stop signs if it replaced an innocuous sign such as a no parking sign, demonstrating that awareness of the sign required active search. The stimulus itself did not attract attention. Similarly, Stop signs at intersections were detected more frequently than those mid-block. This finding suggests that the visibility of the signs requires an active search and that the frequency of this search is influenced by learned knowledge of the probabilistic structure of the environment.

VR also has the advantage that it allows a flexible and systematic exploration of natural behavior by controlled manipulation of the stimulus conditions in complex behavior. For example, Draschkow et al. (2021) investigated trade-offs in the use of working memory versus just-in-time representations (Ballard et al., 1995). They followed up on earlier work by allowing subjects to move in a 3D virtual environment. Using the block-copying task where a model, workspace, and supply of blocks were separated by a variable distance they could manipulate the cost of memory versus the cost of large body movements over a large range. Interestingly, despite the typically high cost of body movements, working memory remained quite expensive relative to body movements. The authors were also able to show how memory use increased as viewing time was made longer. Thus, VR in a task context can allow parametric exploration of critical factors that influence visually guided behavior.

However, it is important to note that it is important to validate whether the current research question can be successfully tackled in a virtual environment. On the one hand, when comparing gaze behavior during locomotion in a virtual and the real world, Drewes, Feder, and Einh user (2021) observed that adjustments in gaze behavior between different terrains (e.g., flat surfaces or stair cases) are similar in VR and in the real world, indicating that VR is a good substitute real-world behavior in this case. On the other hand, a recent study by Lavoie, Hebert, and Chapman

(2024) compared eye–hand coordination during object interaction in the real world with the same task in VR. They observed that, in the real world, observers looked away from objects they interacted with once their hand made contact, whereas in VR observers fixated on these objects for much longer. Thus, the lack of haptic feedback about the object in VR changed gaze behavior indicating that it is a critical aspect of natural behavior. Together, these findings indicate that, although VR can provide opportunities to study perceptual questions (David et al., 2020; Rodriguez et al., 2024) and potentially eye movement behavior, for manual interactions with objects it might not be an ideal testing bed.

## Conclusions

It has become increasingly clear that we need to consider vision in its natural context. In this article, we address the problem of how to do this in a controlled and systematic fashion, given the intrinsic diversity and complexities of natural behavior. We argue that it is first necessary to describe both the sensory input and behavioral responses in natural conditions, where ongoing movements of the body profoundly affect the retinal stimulus, and the spatial and temporal context of complex scenes shape perception and actions. Sequential decision making provides a unifying principle for understanding natural visually guided behavior. This helps to identify the variety of factors that might influence behavior in any given experiment and devise paradigms that allow more controlled investigation of these factors. It also highlights the importance of developmental learning in understanding both perception and action. Although it is clear that there is not a simple answer to tackling the complexities of natural behavior, we have given a variety of examples of strategies that seem promising, such as systematically simplifying the stimulus or by leveraging individual differences. We hope that these insights can provide a path forward for the controlled investigation of natural behavior.

*Keywords: natural behavior, eye movements, decisions, visually guided actions*

## Acknowledgments

The authors dedicate this paper to Eileen Kowler. Eileen has been a source of inspiration to all of us, and in particular to me (Mary Hayhoe), since our careers have overlapped for more than 40 years. Eileen was one of the first investigators who highlighted the importance

of looking at eye movements in the natural world. Her work with Julie Epelboim and Bob Steinman in the 1990s was brilliant and groundbreaking, and showed decisively how important memory and task were for eye and head movements. She subsequently organized a memorable meeting in Amsterdam on the topic of natural vision. Vision research has lost an iconic leader, and we will miss her.

A.G. was supported by the Deutsche Forschungsgemeinschaft (Project No. 222641018–SFB/TRR 135 Project A1). N.P. and M.H. were supported by NIH grant EY05729.

Commercial relationships: none.

Corresponding author: Alexander Goettker.

Email: Alexander.Goettker@psy.jlug.de.

Address: Justus-Liebig-Universität Gießen, Alter Steinbacher Weg 38, Giessen 35394, Germany.

## References

- Anderson, E. M., Candy, T. R., Gold, J. M., & Smith, L. B. (2024). An edge-simplicity bias in the visual input to young infants. *Science Advances*, 10(19), ead8571.
- Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, 39(17), 2947–2953.
- Bahill, A. T., Clark, M. R., & Stark, L. (1975). The main sequence, a tool for studying human eye movements. *Mathematical Biosciences*, 24(3–4), 191–204.
- Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7(1), 66–80.
- Bambach, S., Crandall, D., Smith, L., & Yu, C. (2018). Toddler-inspired visual object learning. *Advances in Neural Information Processing Systems* (p. 31). Cambridge, MA: MIT Press.
- Bardy, B. G., Warren, W. H., & Kay, B. A. (1999). The role of central and peripheral vision in postural control during walking. *Perception & Psychophysics*, 61(7), 1356–1368.
- Bargary, G., Bosten, J. M., Goodbourn, P. T., Lawrance-Owen, A. J., Hogg, R. E., & Mollon, J. D. (2017). Individual differences in human eye movements: An oculomotor signature? *Vision Research*, 141, 157–169.
- Battaglia, P. W., & Schrater, P. R. (2007). Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. *Journal of Neuroscience*, 27(26), 6984–6994.
- Beyeler, M., Dutt, N., & Krichmar, J. L. (2016). 3d visual response properties of mstd emerge from an efficient, sparse population code. *Journal of Neuroscience*, 36(32), 8399–8415.
- Binda, P., & Morrone, M. C. (2018). Vision during saccadic eye movements. *Annual Review of Vision Science*, 4(1), 193–213.
- Bischof, W. F., Anderson, N. C., & Kingstone, A. (2023). Eye and head movements while encoding and recognizing panoramic scenes in virtual reality. *PloS One*, 18(2), e0282030.
- Bonnen, K., Matthis, J. S., Gibaldi, A., Banks, M. S., Levi, D. M., & Hayhoe, M. (2021). Binocular vision and the control of foot placement during walking in natural terrain. *Scientific Reports*, 11(1), 20881.
- Botch, T. L., Garcia, B. D., Choi, Y. B., Feffer, N., & Robertson, C. E. (2023). Active visual search in naturalistic environments reflects individual differences in classic visual search performance. *Scientific Reports*, 13(1), 631.
- Casile, A., Victor, J. D., & Rucci, M. (2019). Contrast sensitivity reveals an oculomotor strategy for temporally encoding space. *Elife*, 8, e40924.
- Castelhano, M. S., & Henderson, J. M. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance*, 34(3), 660.
- Chan, A. W., Kravitz, D. J., Truong, S., Arizpe, J., & Baker, C. I. (2010). Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nature Neuroscience*, 13(4), 417–418.
- Cisek, P., & Green, A. M. (2024). Toward a neuroscience of natural behavior. *Current Opinion in Neurobiology*, 86, 102859.
- Cisek, P., & Pastor-Bernier, A. (2014). On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655), 20130479.
- Cretenoud, A. F., Grzeczowski, L., Kunchulia, M., & Herzog, M. H. (2021). Individual differences in the perception of visual illusions are stable across eyes, time, and measurement methods. *Journal of Vision*, 21(5), 26.
- Darici, O., & Kuo, A. D. (2023). Humans plan for the near future to walk economically on uneven terrain. *Proceedings of the National Academy of Sciences of the United States of America*, 120(19), e2211405120.
- David, E., Beitner, J., & Vö, M. L.-H. (2020). Effects of transient loss of vision on head and eye movements during visual search in a virtual environment. *Brain Sciences*, 10(11), 841.

- De Haas, B., Iakovidis, A. L., Schwarzkopf, D. S., & Gegenfurtner, K. R. (2019). Individual differences in visual salience vary along semantic dimensions. *Proceedings of the National Academy of Sciences of the United States of America*, 116(24), 11687–11692.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. June 20–25, 2009, Miami, Florida (pp. 248–255).
- Diamond, M. R., Ross, J., & Morrone, M. C. (2000). Extraretinal control of saccadic suppression. *Journal of Neuroscience*, 20(9), 3449–3455.
- 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.
- Draschkow, D., Kallmayer, M., & Nobre, A. C. (2021). When natural behavior engages working memory. *Current Biology*, 31(4), 869–874.
- Drewes, J., Feder, S., & Einhäuser, W. (2021). Gaze during locomotion in virtual reality and the real world. *Frontiers in Neuroscience*, 15, 656913.
- DuTell, V., Gibaldi, A., Focarelli, G., Olshausen, B. A., & Banks, M. S. (2024). High-fidelity eye, head, body, and world tracking with a wearable device. *Behavior Research Methods*, 56(1), 32–42.
- Engel, J., Somasundaram, K., Goesele, M., Sun, A., Gamino, A., & Turner, A. et al. (2023). Project aria: A new tool for egocentric multi-modal ai research. *arXiv preprint arXiv:2308.13561*.
- Epelboim, J., Steinman, R. M., Kowler, E., Pizlo, Z., Erkelens, C. J., & Collewyn, H. (1997). Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision Research*, 37(18), 2597–2607.
- Ettinger, U., Kumari, V., Crawford, T. J., Davis, R. E., Sharma, T., & Corr, P. J. (2003). Reliability of smooth pursuit, fixation, and saccadic eye movements. *Psychophysiology*, 40(4), 620–628.
- Fang, Y., Nakashima, R., Matsumiya, K., Kuriki, I., & Shioiri, S. (2015). Eye-head coordination for visual cognitive processing. *PloS One*, 10(3), e0121035.
- Finley, J. M., Bastian, A. J., & Gottschall, J. S. (2013). Learning to be economical: the energy cost of walking tracks motor adaptation. *Journal of Physiology*, 591(4), 1081–1095.
- Fooker, J., Baltaretu, B. R., Barany, D. A., Diaz, G., Semrau, J. A., Singh, T., . . . Douglas Crawford, J. (2023). Perceptual-cognitive integration for goal-directed action in naturalistic environments. *Journal of Neuroscience*, 43(45), 7511–7522.
- Fooker, J., Johansson, R. S., & Flanagan, J. R. (2024). Adaptive gaze and hand coordination while manipulating and monitoring the environment in parallel. *bioRxiv*, 2024–10.
- Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, 51(17), 1920–1931.
- Franklin, D. W., & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, 72(3), 425–442.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin
- Gibson, J. J. (1978). The ecological approach to the visual perception of pictures. *Leonardo*, 11(3), 227–235.
- Gibson, J. J. (1979). The theory of affordances:(1979). In *The people, place, and space reader* (pp. 56–60). New York: Routledge.
- Glennerster, A., Hansard, M. E., & Fitzgibbon, A. W. (2001). Fixation could simplify, not complicate, the interpretation of retinal flow. *Vision Research*, 41(6), 815–834.
- Goettker, A., Agtzidis, I., Braun, D. I., Dorr, M., & Gegenfurtner, K. R. (2020). From gaussian blobs to naturalistic videos: Comparison of oculomotor behavior across different stimulus complexities. *Journal of Vision*, 20(8), 26.
- Goettker, A., Borgerding, N., Leeske, L., & Gegenfurtner, K. R. (2023). Cues for predictive eye movements in naturalistic scenes. *Journal of Vision*, 23(10), 12.
- Goettker, A., & Gegenfurtner, K. R. (2021). A change in perspective: The interaction of saccadic and pursuit eye movements in oculomotor control and perception. *Vision Research*, 188, 283–296.
- Goettker, A., & Gegenfurtner, K. R. (2024). Individual differences link sensory processing and motor control. *Psychological Review* Jun 13. doi: [10.1037/rev0000477](https://doi.org/10.1037/rev0000477). Online ahead of print.
- Goettker, A., Pidaparthi, H., Braun, D. I., Elder, J. H., & Gegenfurtner, K. R. (2021). Ice hockey spectators use contextual cues to guide predictive eye movements. *Current Biology*, 31(16), R991–R992.
- Grauman, K., Westbury, A., Byrne, E., Chavis, Z., Furnari, A., Girdhar, R., . . . Malik, J. (2022). Ego4d: Around the world in 3,000 hours of egocentric video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. New Orleans, Louisiana, June 18–24, 2022 (pp. 18995–19012).
- Greene, M. R., Balas, B. J., Lescroart, M. D., MacNeilage, P. R., Hart, J. A., Binaee, K., . . . Weissmann, E. (2024). The visual experience dataset: Over 200 recorded hours of integrated eye

- movement, odometry, and egocentric video. *Journal of Vision*, 24(11), 6.
- Grzeczowski, L., Clarke, A. M., Francis, G., Mast, F. W., & Herzog, M. H. (2017). About individual differences in vision. *Vision Research*, 141, 282–292.
- Haarmeier, T., Thier, P., Repnow, M., & Petersen, D. (1997). False perception of motion in a patient who cannot compensate for eye movements. *Nature*, 389(6653), 849–852.
- Hayhoe, M., & Ballard, D. (2014). Modeling task control of eye movements. *Current Biology*, 24(13), R622–R628.
- Hayhoe, M. M. (2009). Visual memory in motor planning and action. *Memory for the Visual World*, 117–139.
- Hayhoe, M. M. (2017). Vision and action. *Annual Review of Vision Science*, 3, 389–413.
- Hayhoe, M. M., & Lerch, R. A. (2022). Visual guidance of natural behavior. In *Oxford research encyclopedia of psychology*. Oxford, UK: Oxford University Press.
- Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1), 6.
- Hooze, I. T., Niehorster, D. C., Nyström, M., & Hessels, R. S. (2024). Large eye–head gaze shifts measured with a wearable eye tracker and an industrial camera. *Behavior Research Methods*, 56, 1–14.
- Hoppe, D., & Rothkopf, C. A. (2016). Learning rational temporal eye movement strategies. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8332–8337.
- Hoppe, D., & Rothkopf, C. A. (2019). Multi-step planning of eye movements in visual search. *Scientific Reports*, 9(1), 144.
- Idrees, S., Baumann, M. P., Franke, F., Münch, T. A., & Hafed, Z. M. (2020). Perceptual saccadic suppression starts in the retina. *Nature Communications*, 11(1), 1977.
- Intoy, J., & Rucci, M. (2020). Finely tuned eye movements enhance visual acuity. *Nature Communications*, 11(1), 795.
- Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *Journal of Neuroscience*, 29(19), 6234–6238.
- Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences*, 23(8), 672–685.
- Kaiser, D., Stein, T., & Peelen, M. V. (2015). Real-world spatial regularities affect visual working memory for objects. *Psychonomic Bulletin & Review*, 22, 1784–1790.
- Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, 55(1), 271–304.
- Keshava, A., Nezami, F. N., Neumann, H., Izdebski, K., Schäuler, T., & Köonig, P. (2024). Just-in-time: Gaze guidance in natural behavior. *PLoS Computational Biology*, 20(10), e1012529.
- Kessler, F., Frankenstein, J., & Rothkopf, C. A. (2024). Human navigation strategies and their errors result from dynamic interactions of spatial uncertainties. *Nature Communications*, 15(1), 1–19.
- Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, 99(3), 317–340.
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., . . . Girshick, R. (2023). Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 4015–4026) October 1–6, 2023, Paris, France.
- Knill, D. (1996). *Perception as Bayesian inference*. Cambridge, UK: Cambridge University Press.
- Koenderink, J. J. (1986). Optic flow. *Vision Research*, 26(1), 161–179.
- Koenderink, J. J., & Doorn, A. J. van. (1976). Local structure of movement parallax of the plane. *JOSA*, 66(7), 717–723.
- Kothari, R., Yang, Z., Kanan, C., Bailey, R., Pelz, J. B., & Diaz, G. J. (2020). Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Scientific Reports*, 10(1), 2539.
- Kowler, E., Rubinstein, J. F., Santos, E. M., & Wang, J. (2019). Predictive smooth pursuit eye movements. *Annual Review of Vision Science*, 5(1), 223–246.
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, 93(3), 480–490.
- Kuo, A. D., Donelan, J. M., & Ruina, A. (2005). Energetic consequences of walking like an inverted pendulum: step-to-step transitions. *Exercise and Sport Sciences Reviews*, 33(2), 88–97.
- Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11), 1311–1328.
- Land, M. F. (1992). Predictable eye-head coordination during driving. *Nature*, 359(6393), 318–320.
- Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1358), 1231–1239.

- Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25–26), 3559–3565.
- Lavoie, E., Hebert, J. S., & Chapman, C. S. (2024). Comparing eye–hand coordination between controller-mediated virtual reality, and a real-world object interaction task. *Journal of Vision*, 24(2), 9.
- Lee, D. N., & Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Perception & Psychophysics*, 15, 529–532.
- Lee, D. V., & Harris, S. L. (2018). Linking gait dynamics to mechanical cost of legged locomotion. *Frontiers in Robotics and AI*, 5, 111.
- Maloney, L. T., & Zhang, H. (2010). Decision-theoretic models of visual perception and action. *Vision Research*, 50(23), 2362–2374.
- Maselli, A., Gordon, J., Eluchans, M., Lancia, G. L., Thiery, T., Moretti, R., . . . Pezzulo, G. (2023). Beyond simple laboratory studies: developing sophisticated models to study rich behavior. *Physics of Life Reviews*, 46, 220–244.
- Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, 28, 1224–1233.
- Matthis, J. S., Muller, K. S., Bonnen, K. L., & Hayhoe, M. M. (2022). Retinal optic flow during natural locomotion. *PLoS Computational Biology*, 18(2), e1009575.
- McNamee, D., & Wolpert, D. M. (2019). Internal models in biological control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2(1), 339–364.
- Mehrer, J., Spoerer, C., Jones, E. C., Kriegeskorte, N., & Kietzmann, T. (2021). An ecologically motivated image dataset for deep learning yields better models of human vision. *Proceedings of the National Academy of Sciences of the United States of America*, 118(8), 1–9.
- Miller, C. T., Gire, D., Hoke, K., Huk, A. C., Kelley, D., Leopold, D. A., . . . Niell, C. M. (2022). Natural behavior is the language of the brain. *Current Biology*, 32(10), R482–R493.
- Mineault, P., Bakhtiari, S., Richards, B., & Pack, C. (2021). Your head is there to move you around: Goal-driven models of the primate dorsal pathway. *Advances in Neural Information Processing Systems*, 34, 28757–28771.
- Mollon, J. D., Bosten, J. M., Peterzell, D. H., & Webster, M. A. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision Research*, 141, 4–15.
- Morasso, P., Bizzi, E., & Dichgans, J. (1973). Adjustment of saccade characteristics during head movements. *Experimental Brain Research*, 16, 492–500.
- Moutsiana, C., Soliman, R., De Wit, L., James-Galton, M., Sereno, M. I., Plant, G. T., . . . Schwarzkopf, D. S. (2018). Unexplained progressive visual field loss in the presence of normal retinotopic maps. *Frontiers in Psychology*, 9, 1722.
- Muller, K. S., Bonnen, K., Shields, S. M., Panfili, D. P., Matthis, J., & Hayhoe, M. M. (2024). Analysis of foothold selection during locomotion using terrain reconstruction. *eLife*, 12, RP91243.
- Muller, K. S., Matthis, J., Bonnen, K., Cormack, L. K., Huk, A. C., & Hayhoe, M. (2023). Retinal motion statistics during natural locomotion. *eLife*, 12, e82410.
- Nath, T., Mathis, A., Chen, A. C., Patel, A., Bethge, M., & Mathis, M. W. (2019). Using deeplabcut for 3d markerless pose estimation across species and behaviors. *Nature Protocols*, 14(7), 2152–2176.
- Parker, P. R., Brown, M. A., Smear, M. C., & Niell, C. M. (2020). Movement-related signals in sensory areas: roles in natural behavior. *Trends in Neurosciences*, 43(8), 581–595.
- Patterson, G., & Hays, J. (2012). Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2751–2758). Providence, Rhode Island, June 16–21, 2012.
- Pelz, J., Hayhoe, M., & Loeber, R. (2001). The coordination of eye, head, and hand movements in a natural task. *Experimental Brain Research*, 139, 266–277.
- Pome, A., Schlichting, N., Fritz, C., & Zimmermann, E. (2024). Prediction of sensorimotor contingencies generates saccadic omission. *Current Biology*, 34(14), 3215–3225.
- Powell, N., Oh, Y., Panfili, D., & Hayhoe, M. (2023). Is optic flow used for steering to a goal? *Journal of Vision*, 23(9), 5790.
- Rodriguez, R. G., Hedjar, L., Toscani, M., Guarnera, D., Guarnera, G. C., & Gegenfurtner, K. R. (2024). Color constancy mechanisms in virtual reality environments. *Journal of Vision*, 24(5), 6.
- Rolfs, M., & Schweitzer, R. (2022). Coupling perception to action through incidental sensory consequences of motor behaviour. *Nature Reviews Psychology*, 1(2), 112–123.
- Rolfs, M., Schweitzer, R., Castet, E., Watson, T. L., & Ohl, S. (2023). Lawful kinematics link eye movements to the limits of high-speed perception. *bioRxiv*, 2023–07.

- Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, 24(2), 113–121.
- Rucci, M., Ahissar, E., & Burr, D. (2018). Temporal coding of visual space. *Trends in Cognitive Sciences*, 22(10), 883–895.
- Scarfe, P., & Glennerster, A. (2015). Using high-fidelity virtual reality to study perception in freely moving observers. *Journal of Vision*, 15(9), 3.
- Schroeger, A., Goettker, A., Braun, D. I., & Gegenfurtner, K. R. (2024). Keeping your eye, head, and hand on the ball: Rapidly orchestrated visuomotor behavior in a continuous action task. *bioRxiv*, 2024–12.
- Schroer, S. E., & Yu, C. (2023). Looking is not enough: Multimodal attention supports the real-time learning of new words. *Developmental Science*, 26(2), e13290.
- Schultz, W. (2015). Neuronal reward and decision signals: from theories to data. *Physiological Reviews*, 95(3), 853–951.
- Segraves, M. A. (2023). Using natural scenes to enhance our understanding of the cerebral cortex's role in visual search. *Annual Review of Vision Science*, 9, 435–454.
- Selinger, J. C., O'Connor, S. M., Wong, J. D., & Donelan, J. M. (2015). Humans can continuously optimize energetic cost during walking. *Current Biology*, 25(18), 2452–2456.
- Seydell, A., McCann, B. C., Trommershäuser, J., & Knill, D. C. (2008). Learning stochastic reward distributions in a speeded pointing task. *Journal of Neuroscience*, 28(17), 4356–4367.
- Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research*, 41(25–26), 3535–3545.
- Sims, C. R., Jacobs, R. A., & Knill, D. C. (2011). Adaptive allocation of vision under competing task demands. *Journal of Neuroscience*, 31(3), 928–943.
- Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, 14(1), 9–17.
- Solman, G. J., Foulsham, T., & Kingstone, A. (2017). Eye and head movements are complementary in visual selection. *Royal Society Open Science*, 4(1), 160569.
- Sommer, M. A., & Wurtz, R. H. (2008). Brain circuits for the internal monitoring of movements. *Annual Review of Neuroscience*, 31(1), 317–338.
- Stahl, J. S. (1999). Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research*, 126, 41–54.
- Stewart, E. E., & Fleming, R. W. (2023). The eyes anticipate where objects will move based on their shape. *Current Biology*, 33(17), R894–R895.
- Sullivan, B., Ludwig, C. J., Damen, D., Mayol-Cuevas, W., & Gilchrist, I. D. (2021). Look-ahead fixations during visuomotor behavior: Evidence from assembling a camping tent. *Journal of Vision*, 21(3), 13.
- Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences*, 12(8), 291–297.
- Valsecchi, M., Akbarinia, A., Gil-Rodriguez, R., & Gegenfurtner, K. R. (2020). Pedestrians egocentric vision: Individual and collective analysis. In *ACM Symposium on Eye Tracking Research and Applications* (pp. 1–5). Denver, Colorado, June 2–5, 2020.
- Vo, M. L.-H., Boettcher, S. E., & Draschkow, D. (2019). Reading scenes: How scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, 29, 205–210.
- Warren, W. H., Kay, B. A., & Yilmaz, E. H. (1996). Visual control of posture during walking: Functional specificity. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 818.
- Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5(6), 598–604.
- Wilmer, J. B. (2008). How to use individual differences to isolate functional organization, biology, and utility of visual functions; with illustrative proposals for stereopsis. *Spatial Vision*, 21(6), 561.
- Wolpert, D. M., & Landy, M. S. (2012). Motor control is decision-making. *Current Opinion in Neurobiology*, 22(6), 996–1003.
- Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or deeper: Revisiting the resnet model for visual recognition. *Pattern Recognition*, 90, 119–133.
- Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244–262.
- Zangrossi, A., Cona, G., Celli, M., Zorzi, M., & Corbetta, M. (2021). Visual exploration dynamics are low-dimensional and driven by intrinsic factors. *Communications Biology*, 4(1), 1100.
- Zimmermann, E. (2020). Saccade suppression depends on context. *Elife*, 9, e49700.