

Toward a Theory of Visual Information Acquisition in Driving

Benjamin Wolfe¹, Massachusetts Institute of Technology, Cambridge, USA,
Ben D. Sawyer, University of Central Florida, Orlando, USA, and **Ruth Rosenholtz**,
Massachusetts Institute of Technology, Cambridge, USA

Objective: The aim of this study is to describe information acquisition theory, explaining how drivers acquire and represent the information they need.

Background: While questions of what drivers are aware of underlie many questions in driver behavior, existing theories do not directly address how drivers in particular and observers in general acquire visual information. Understanding the mechanisms of information acquisition is necessary to build predictive models of drivers' representation of the world and can be applied beyond driving to a wide variety of visual tasks.

Method: We describe our theory of information acquisition, looking to questions in driver behavior and results from vision science research that speak to its constituent elements. We focus on the intersection of peripheral vision, visual attention, and eye movement planning and identify how an understanding of these visual mechanisms and processes in the context of information acquisition can inform more complete models of driver knowledge and state.

Results: We set forth our theory of information acquisition, describing the gap in understanding that it fills and how existing questions in this space can be better understood using it.

Conclusion: Information acquisition theory provides a new and powerful way to study, model, and predict what drivers know about the world, reflecting our current understanding of visual mechanisms and enabling new theories, models, and applications.

Application: Using information acquisition theory to understand how drivers acquire, lose, and update their representation of the environment will aid development of driver assistance systems, semiautonomous vehicles, and road safety overall.

Keywords: information acquisition, vision, visual attention, peripheral vision, driving, surface transportation

Address correspondence to Benjamin Wolfe, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 32 Vassar St 32-D416, Cambridge, MA 02139, USA; e-mail: bwolfe@mit.edu

HUMAN FACTORS

2022, Vol. 64(4) 694–713

DOI:10.1177/0018720820939693

Article reuse guidelines: sagepub.com/journals-permissions

© The Author(s) 2020.

INTRODUCTION

Imagine driving down an urban road. A cyclist riding at your right starts to turn left, without signaling, taking them across your lane. What information do you need in order to avoid a collision? Critically, you need to be aware of how the cyclist and all other actors (e.g., other vehicles, pedestrians) are moving within the environment. Additionally, you need to predict what each agent will do in the immediate future, so you can act accordingly. How do you acquire the information you need in time to respond? One way these sorts of questions have been framed has been using Endsley's theory of situation awareness (Endsley, 1995, 1988), which has been widely embraced by the fields of human factors, driver behavior, and traffic safety. However, situation awareness speaks to the cognitive processes that apply *after* the information has been acquired, rather than how this information is acquired and represented. In this paper, we aim to describe how the driver's visual system acquires and represents the information that they need and to speak to the details captured (or not) by these processes. To do this, we set forth our theory of information acquisition, describing how key mechanisms in the visual system work together to provide the information that is the foundation on which all driver awareness and behavior rest.

While questions of driver knowledge and awareness have been asked in the context of theories like situation awareness, the question of how the driver acquires the information they need is outside its scope (Endsley, 2015). That said, understanding how drivers build a mental representation of their operating environment is both essential and is, fundamentally, a question of visual perception and visual mechanisms.



Therefore, our objective is to describe our theory of information acquisition—how the driver becomes aware of their environment, in terms of how their visual system acquires information, drawing from research and theory in vision science. In our assessment, information acquisition is a necessary precondition for later, more cognitive processes, like situation awareness, but is not a replacement for them. In doing so, we will exclusively focus on visual perception, because driving is primarily a visual task, and while other sensory modalities are useful, vision is central to safe driving (c.f., Sivak, 1996; Spence & Ho, 2015).

A central goal of information acquisition theory is to provide an explanation of how the driver acquires the visual information they need. More broadly, information acquisition theory is intended to be generalizable beyond driving to any visual task, because the underlying mechanisms and processes are not exclusive to driving. Specifically, understanding how drivers acquire visual information, particularly what they know from peripheral vision and what information they must search for in the scene, will enable researchers to determine what the driver might be expected to know in a given situation. This is not to say that drivers will always know everything that information acquisition predicts they will know, rather that it provides an upper bound on what they can know, given the realities of visual perception. However, testing with this upper bound in mind allows researchers to understand what drivers know and how they come to know it. One approach to testing this draws from techniques in vision science, where the observers' gaze is tracked and the stimulus is gaze-contingently manipulated based on where they are looking. For example, if one wants to determine whether foveal information (from the highest resolution portion of the retina, the focus of gaze) is required, this information can be blocked wherever the observer looks, and their ability to do the task with only peripheral vision assessed. The inverse, where only foveal information is provided, can test the role of foveal information and, together, can illuminate their respective contributions to the driver's mental representation of the environment and their ability to drive

safely. While these techniques are easier to use in the laboratory, these laboratory findings generalize to the road environment. For example, if a driver keeps looking at an infotainment display when there is nothing operationally relevant there, it suggests they are acquiring sufficient information peripherally to change their behavior from what we think would be their ideal. However, information acquisition, while being directly applicable to driving, is not limited to it. Looking at the theory more broadly, what we learn in the laboratory by teasing apart how drivers acquire information, and whether or not they can use it, speaks not only to questions in driver behavior and traffic safety, but to other visual tasks in daily life, whether that is walking down the sidewalk, finding your keys on your desk, or any other task that requires visual information. By looking to tasks such as driving to understand how we acquire the information we need, we can better understand not just what drivers need to know, but how we use vision in the world.

Given that, we argue that information acquisition begins with the information available at a glance from the entire visual field, which is often called scene gist (Greene & Oliva, 2009b; Navon, 1977; Oliva & Torralba, 2006). However, only considering information from across the visual field, and ignoring where the driver is looking, and how and why their focus of gaze can shift, would be to ignore a critical set of acquisitive mechanisms. In considering information acquisition, it is essential to realize that peripheral information guides eye movements, since there would be no way to determine where the eye should move without it (Kowler et al., 1995; Wolfe & Whitney, 2014). For that matter, this same peripheral information is essential in the context of shifting covert attention, as in visual search (c.f., Treisman & Gelade, 1980; Wolfe, 1994). However, by emphasizing the broad visual field, rather than where the driver is presently attending or looking (in line with Hochstein & Ahissar, 2002), we are not setting search and attention aside, rather, we are merely emphasizing a different mechanism.

It is worth noting, at this point, that while these mechanisms are universal—the ability to acquire information from the periphery and

the ability to use this information to shift eye movements and attentional focus—they are profoundly influenced by top-down factors and, specifically, the driver's expertise. An expert driver will be better able to use these mechanisms to their advantage in a given situation and will be better able to interpret, say, a vague impression of motion in the periphery as something they might want to attend to or make an eye movement to, than a novice driver. In this paper, however, our focus is the basic perceptual processes universal to typical drivers and observers, rather than how they are modified by expertise, training, age, and disorder. To summarize, information acquisition operates at both global (peripheral vision and gist) and local (attentional and gaze shifts) scales and is the process by which the driver constructs or updates their representation of the environment.

In describing information acquisition, we note that the driver's need for information is inherently time limited: it would do the driver no good to acquire information too slowly and only realize that they were going to collide with the cyclist long after the collision. How long might the driver have to get the information they need? One estimate of how much time drivers might have available to do this can be deduced from Green's meta-analysis of brake reaction time which reports a mean brake reaction time of 1,300 ms for unanticipated events (Green, 2000). That reaction time includes drivers' need to understand the situation, plan, and initiate their response, which suggests that information acquisition must complete in a fraction of that time, in order for the driver to do something with the knowledge they have acquired. Understanding a dynamic driving environment is, inherently, a difficult visual task, but it is made easier by the fact that you probably do not need to know every detail of the scene, nor is your visual system capable of acquiring and representing that level of detail in the time you have to respond. Since drivers do, by and large, respond adequately to unanticipated events (although the amount of time they have to do so varies widely based on speed, environment, event, and other factors), this suggests that the mechanisms in play must be fast and, usually, sufficient to the drivers' needs. Merely being

able to acquire the information quickly is, of course, not sufficient in its own right, but it is a necessary precondition for the driver to perform well in the situation. Critically, our focus here is the acquisition and representation of the visual environment in driving, rather than how this representation is used to plan action. To put it another way, information acquisition describes the processes by which the driver obtains the information that they then use to achieve situation awareness.

Our work here has two major inspirations: prior attempts in driver behavior to describe the problem of what the driver knows and how they know it in the context of existing theories like situation awareness, and vision science research which illuminates particular visual mechanisms. We agree with Endsley that situation awareness is not the right tool for understanding "process measures," that is, how drivers acquire and represent information (Endsley, 2015, Endsley, 2019). However, it is worth noting that researchers have tried to answer the question of what the driver knows and how they know it using situation awareness precisely because it seemed like the best tool available. In the absence of a theory like information acquisition, situation awareness is used inappropriately for questions that are more perceptual than they are cognitive. Reflecting this, we will start by describing information acquisition and how it engages with questions that have been asked in the context of situation awareness, followed by a review of key topics in the vision science literature, specifically peripheral vision, visual attention, and eye movement planning. We will conclude by discussing how information acquisition can be used to study key questions at the intersection of visual perception and driving and its implications for development of driver assistance systems, semiautonomous vehicles and safer roadways overall.

Review Methodology and Logic

Information acquisition, as a theory, comes out of persistent questions that we encountered at the intersection of traffic safety and vision science. In building our theory and this work, our goal was to illuminate connections between

three topics in vision science research (peripheral vision, visual attention, and eye movement planning) that, when understood as part of an integrated whole, can better explain outstanding questions in driver behavior. Each of these areas reflects a vast literature, with the study of peripheral vision originating in the development of visual perimetry in the 1840s, ideas on visual attention emerging around the same time, and questions of eye movement planning going back at least to Javal's original work on saccades in 1878. Therefore, it is not practical to give a list of keywords and exclusion criteria, as the terms we focus upon for each topic have each evolved and changed across their respective histories. Our goal in this work is to show how our understanding of each of these topic areas has changed in recent decades and why it has changed and to give a sense of the current state of each area in a way that is relevant to readers. Since the key motivation for this work is to show the connections between these areas, and, in particular, how understanding them together is key to the integrative understanding that information acquisition is built on, we have taken a selective approach, since a keyword-based approach would not capture these essential connections. That said, we do not claim to have exhaustively reviewed even the recent work on these topics and point interested readers at recent reviews for each topic at the end of this section.

For our review of peripheral vision, which is the core of information acquisition, we wanted to provide a grounding in the basics of visual anatomy and begin by discussing the anatomy and physiology of the visual system. More broadly, we wanted to address specific misconceptions about peripheral vision, and we do so by first discussing the anatomical and physiological underpinnings and how they constrain perception. We then move on to more perceptual questions, particularly the problem of visual crowding and how it lead to our work on modeling peripheral vision. Considering peripheral vision as informative in its own right leads us, in turn, to work on scene gist, which is a topic we return to in the context of visual attention and search in the following section, as gist and peripheral vision more generally have

been instrumental in changing how we think about attention and search.

We follow our review of peripheral vision by reviewing key highlights of the last 40 years of research on visual attention, using Treisman and Gelade (1980) as our starting point, not because it is the origin of visual attention research, but because much of the last four decades of work has reacted to it, and many of Treisman's ideas are seen throughout traffic safety research. In particular, our goal here was to show how Treisman's ideas changed over the decades, and how the theories which reacted to Treisman's also changed over that same time. Our goal in doing this is to show why thinking of attention the way Treisman did in 1980 does researchers in 2020 a disservice and to show our readers how and why these changes occurred. This work is closely linked to our review on peripheral vision which precedes it, in so much as the work on scene gist and ensemble perception were a major driving force in changing how vision science thinks about attention and visual search.

Our last review, on eye movement planning, is the key link between questions of peripheral vision and visual attention, and the core of information acquisition. To put it briefly, the processes by which eye movements are planned, why they are made, and what information underlies the process explain a great deal of behavior and are instrumental to understanding both what drivers know and how they know it. However, our focus here is on eye movements in the larger context of our paper, rather than everything that has been learned about the planning process in recent decades. In particular, in the context of driving, eye movements and fixations are often used as a proxy for what the driver knows, and by discussing how eye movements are planned and how this process links back to questions of peripheral vision and visual attention, we can enable traffic safety researchers to understand eye movements in a new way.

By structuring our work in this way, we are not claiming to have exhaustively reviewed the entire breadth of each topic area; rather, we are trying to show how our understanding of each of these areas has changed in the last several decades and, critically, how all of

Information Acquisition: How the driver acquires the visual information they need for a task

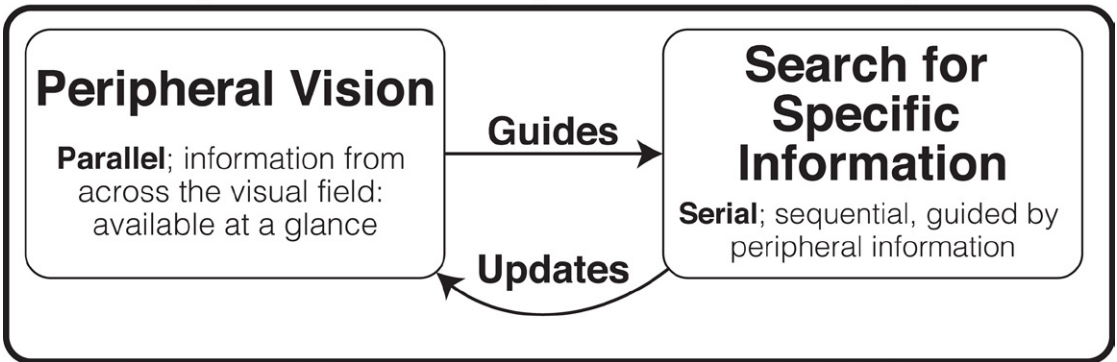


Figure 1. Diagram of information acquisition theory, beginning with the information available across the visual field in a single glance, which guides visual search for specific information.

these areas are connected in the context of real-world behavior like driving. There exist deeper reviews on each of these individual topics, and we point them out here for readers who may find them useful. For peripheral vision, see Rosenholtz (2016), for visual attention, see Carrasco (2011) and Kristjánsson and Egeth (2020) and, more broadly, the pair of Treisman Memorial Issues in *Attention, Perception and Psychophysics*, published in 2020. For saccade planning, Kowler (2011) covers a key window in this work (1985–2010), overlapping with the increased availability of eye-tracking equipment and the changes in our understanding that has brought about.

A Theory of Information Acquisition

Our theory of information acquisition is, fundamentally, a way to explain how the driver acquires the information they need, based on current state-of-the-art theories and empirical results on fundamental questions in vision science and visual perception. Rather than framing this question around what the driver is looking at now, we consider what information the driver can acquire from the entire visual field, based on where they are currently looking, and how they augment this full-field information as necessary for the task(s) they are performing. We start by considering what information the driver can acquire from their entire visual field (Figure 1, left), rather than only focusing

on what element of the environment they might currently be looking at and perceiving. This, broadly speaking, is the question of what information is available from peripheral vision, or what you can perceive without attending to specific elements or locations? We argue that peripheral input provides much of the information the driver needs, both at a global level (the gist of the scene, acquired in parallel) and at a local level (providing information to guide search processes and eye movements more generally; Figure 1, right).

Of course, drivers need to search for specific objects in the scene, and making eye movements to relevant objects is useful, but to understand *why* drivers look where they do, we must consider how attention and eye movements are guided, and how else the visual system might use this information. Critically, focusing on shifts of gaze and attention alone discounts how and why those shifts are made. Therefore, information acquisition emphasizes what the driver can acquire from peripheral vision because this information is necessary to plan any subsequent shift of gaze or attention, although it may not be as detailed as the information available from attending to a specific object. Information acquisition also deemphasizes the idea that attention is required for perception, but does not negate it: attending to specific objects or elements is useful, but not always essential. A critical addition to this is the idea that these processes are

Images and Results from Wolfe et.al, 2019

**Hazard Detection at a Glance**

Hazards detected in video with
220 ms stimulus duration

**Hazard Detection and Localization**

Hazards localized in video with
400 ms stimulus duration

Figure 2. In a recent study by Wolfe and colleagues (2020), participants could detect hazards in real road scenes in a single glance (left panel) and could represent these hazards with sufficient fidelity to plan an evasive response with only enough time for a single eye movement (right).

fast and continuous: these processes are perpetual and essentially mandatory; therefore, the driver's perception of the world, while dictated by the speed of neural processing, is inherently fast, lest the world catch up to them.

Our theory of information acquisition owes a great deal to several areas of research in vision science. Our focus on peripheral vision arises in large part from work on scene gist, the ability to perceive the essence of a scene in the blink of an eye (100 ms stimulus duration; Greene & Oliva, 2009b), as well as work on modeling what information is available from the visual periphery (Balas et al., 2009; Rosenholtz, Huang, Raj et al., 2012; Rosenholtz, 2016). However, drivers cannot acquire enough information to be safe with peripheral vision alone, and to address this, we turned to ideas in visual attention, both historically and more recently, particularly Treisman's feature integration theory (Treisman, 2006; Treisman & Gelade, 1980) and Wolfe's guided search (Wolfe, 1994; Wolfe & Horowitz, 2017). Along with this work on attention and search, we are deeply influenced by work on the role of presaccadic attention in eye movement planning (Deubel, 2008; Kowler et al., 1995; Wolfe & Whitney, 2014) and what our ability to plan shifts of gaze implies about the information available to the visual system. Since we know that drivers shift their gaze and the focus of their attention around the scene,

how useful is the information that allows these shifts?

If peripheral vision is so important, what kind of driving-relevant tasks might be possible at a single glance? How quickly might drivers be able to acquire enough information to do tasks similar to those they would do on the road? We addressed this question in a recent study (Wolfe, Seppelt et al., 2020; Figure 2) where we asked participants to watch brief clips of dashcam video which had a 50% chance of containing a precollision scenario and to report whether or not they perceived the scene as containing a hazard. Participants in their twenties detected hazards 80% of the time with a mean video duration of 220 ms, suggesting they could extract the information necessary in a single glance. When we asked the same participants to evade the hazard they perceived (by steering left or right), they required 400 ms, suggesting they acquired enough information on an initial glance to guide an eye movement to a useful region or object in the scene, and thereby understood the scene well enough to plan an evasive response. Critically, participants performed a task which probed their ability to use the information they acquired, rather than probing details of the scene which they may not have needed.

These results suggest that participants in our study are understanding the environment

very quickly in the face of rapid change, which speaks to how drivers on the road are able to respond quickly when hazards suddenly appear. Their representation of the environment reflects these emerging changes very quickly, and while this representation is likely both imperfect and facilitated by their degree of experience on the road, it is broadly sufficient. While our results suggest information acquisition is a fast process, other recent work, framed in the context of situation awareness, asking a similar perceptual question, came to a very different conclusion. Lu and colleagues, in an experiment assessing the development of situation awareness, report that participants needed a minimum of 7 s of video to achieve it in a given task (Lu et al., 2017). They came to this conclusion based on an experiment in which participants watched between 1 and 20 s of simulated road video (forward-facing) and were then asked to indicate, on a top-down diagram, where other vehicles were. They interpret the long durations required as evidence for slow acquisition of situation awareness. How might we reconcile these two sets of results?

In part, the contrast of the subsecond scale of information acquisition (Wolfe, Seppelt et al., 2020) with the tens of seconds required to achieve situation awareness (Lu et al., 2017) speaks to the need to consider what a task is measuring. Both our task and theirs asked participants to acquire and represent information about a driving scene, but the differences in our respective results are considerable, suggesting just how critical it can be to understand what information a given task probes. As a result, these sets of results cannot be so much reconciled as understood to be probing different elements of a question, in spite of their considerable conceptual similarities. That being said, Lu and colleagues' statement "sub-second viewing times are probably too short for processing dynamic traffic scenes" (Lu et al., 2017) may be true for their task, but is certainly not true for many real-world driving tasks, or ours. A critical consideration in information acquisition, and in studying driver behavior overall, is to carefully assess the tasks we ask participants to perform, how well these tasks probe the information drivers actually need, and how the visual system might enable (or hinder) acquiring

the information required in the time available. Specifically, expecting participants (or drivers) to be able to recall considerable detail is at odds with both what drivers might actually represent for a given task or circumstance on the road and with the information they need. Naturalistic driving is a process of information acquisition, with a subset of all gathered information feeding situation awareness; driving decisions can, as shown by our work, be rendered using only the former.

In the context of both information acquisition and situation awareness, eye tracking has significant potential to help explain how drivers acquire information and what they know about the situation. That said, eye movements are often viewed as a proxy for what the observer has attended to and are sometimes interpreted with the assumption that drivers fixate objects in order to perceive them. In many laboratory studies, this is a reasonable assumption; for example, when the task requires discriminating fine details or reading, and the eye movement recording is sufficiently precise to know where the observer actually looked relative to what they were shown. In contrast, eye tracking in the vehicle is a far harder problem, and eye movements may not signal attention because the drivers' task may not require fixation. Technical limitations mean that in-vehicle data are captured more coarsely, both spatially and temporally, than in the laboratory. Collapsing large areas of space (e.g., windshield, side mirrors, instrument cluster, center stack) means data are not coregistered with what the driver could see. Coarse temporal classification of eye movements has similar challenges, with classifications such as "driving," "urban," or "highway" subtending long, complex periods of information acquisition, often interleaved between road and in-vehicle information systems (see Sawyer, Mehler, 2017). While questions about situation awareness can be asked to acquire understanding about what information the driver has acquired, questions about how they acquired that information are lost to averaging over space and time.

Broad area-and-time-of-interest-based approaches are sufficient only if the goal is an approximate measure of where the driver directed their

eyes. This is especially true of approaches which leverage large regions of interest, which only give a rough measure of where the driver looked and, critically, do not incorporate what the driver looked at. As a result, these approaches cannot say what the driver actually saw or what information they acquired, either at the focus of gaze or across the field of view more broadly. Heuristics are therefore often used; for example, assuming that a driver who is looking at the road is less distracted than if they are looking away (the core of Kirchner and Ahlström's *AttenD*; Kirchner & Ahlstrom, 2009), which underpins some driver monitoring systems (Braunagel et al., 2017). Longitudinal eye position recordings of "appropriate" and "inappropriate" gaze similarly underpin attempts to measure situation awareness using models like *AttenD* (c.f., Seppelt et al., 2017). Indeed, such models show some predictive power in aggregate datasets, but it is difficult to argue that eye position relative to the roadway speaks to information acquisition. Consider the argument that position of your hands relative to your body is predictive of baking ability; there might be sufficient evidence for this in a vast dataset of hand position in natural environments, but attempting to harness it for "baking excellence systems" would be unlikely to capture information about how a given baker made a given cake or what they needed to know in order to do so. Indeed, the ways in which these driving assumptions are fragile is well studied, with phenomena like cognitive tunneling (Reimer, 2009) and inattention blindness (Simons & Chabris, 1999) clearly outlining times when "appropriate" gaze is indicative of insufficient information acquisition. Note what this means for the goal of measuring drivers' situation awareness, which is that eye tracking has strong limits (Endsley, 2019). It is quite striking that new research continues using this manner of eye tracking to uncover the "causes" of distraction and engineers systems which purport to detect it.

The critical issue here is that decoupling eye movements from their targets—that is, separating *what* subjects looked at from *where* they looked (as Seppelt and colleagues did)—renders the eye movements themselves fundamentally noninformative. Since the eye movement is made for a reason (e.g., to bring the target

to the fovea, or to optimize the information available from across the visual field, or to track multiple moving objects at once), discounting the reason ignores the purpose of the eye movement. In order to understand what information is incorporated into the driver's representation of their environment, studies using eye movement data should consider why the driver made the eye movements they did, which requires considering the information that informed the eye movement and how the visual field changed thereafter. To understand what drivers know about their environment, which is intrinsically linked to where they look and where they attend, we need to consider what drivers look at, what lead them to look there, and what tasks they were performing.

Information From Across the Visual Field: Peripheral Vision and Scene Gist

However, focusing only on where drivers look discounts their ability to use peripheral information. The importance of the totality of visual information, from across the visual field, is the core of information acquisition. While it is tempting to focus only on where the driver is looking, doing so exclusively discounts the extent of visual input the driver receives at every moment. The visual field covers approximately 180° horizontally and 70° vertically (Traquair, 1927) and needs to be considered as a whole entity, rather than just focusing on the fovea (anatomically, the central 1.7° of the visual field, less than 1% of the area of the retina), corresponding to the current focus of gaze. While the fovea is the area of highest photoreceptor density, and its input is significantly magnified in visual cortex (accounting for nearly half of early visual cortex; Tootell et al., 1982), the remaining 99% of the retina must be useful. Additionally, there is a rapid decrease in photoreceptor density with increasing eccentricity or distance on the retina away from the fovea (from ~200k cones per mm² at the fovea to between 2 and 4k rods and cones per mm² throughout the peripheral retina; Curcio et al., 1990), this is a reduction in resolution and sensitivity, not an elimination. Peripheral visual input is essential for developing awareness of

the world; imagine trying to drive (or walk) with only a limited tunnel (e.g., a 15° radius around the fovea, as suggested by some interpretations of the useful field, which we have previously discussed; Wolfe et al., 2017)—it is possible, but far harder and less safe than it is with peripheral vision (Fishman et al., 1981; Vargas-Martín & Peli, 2006).

Specifically, peripheral vision is often misunderstood in several ways, all of which discount its utility and impair our understanding of how we use visual input to acquire information about the world (see Rosenholtz, 2016 for an in-depth discussion). Perhaps the most common misconception is that peripheral vision is a very “blurry” version of foveal vision. In fact, while there is a drop in acuity (Anstis, 1974, 1998), this is comparatively small (and this alone is not enough to explain why we have difficulty with some peripheral tasks). Along similar lines, peripheral vision in driving is often thought of as primarily a source of motion information and little else (Owsley, 2011). While optic flow is undoubtedly essential as a cue to the driver (Higuchi et al., 2019; McLeod & Ross, 1983), our ability to understand static scenes and make predictions about how they might change suggests that peripheral vision provides considerably more information (c.f., Blättler et al., 2011; Freyd & Finke, 1984; Wolfe, Fridman et al., 2019).

The primary difference between foveal and peripheral vision is that of visual crowding (Bouma, 1970; Pelli, 2008; Whitney & Levi, 2011): objects close to each other in the periphery become difficult to identify when the visual world is complex or cluttered, but can still be detected. In the world, and, critically, in driving, crowding is universal. The driver can perceive other vehicles on the road around them, but becomes progressively less aware of the details of each vehicle as its eccentricity increases. One implication of crowding, when studying what the driver knows, is that by asking for identifying information or task-irrelevant details we might be asking the driver to report items crowded from awareness and may, as a result, greatly underestimate what they understand. When looking at the road, the driver still has a rough sense of where controls are in the cabin

but, particularly if they are visually similar (e.g., knobs, switches, toggles, or, increasingly, touchscreen buttons), they cannot identify and use them until they look at them to alleviate crowding. However, being unable to identify objects does not mean that all information from them is inaccessible or lost (Balas et al., 2009; Fischer & Whitney, 2011). In fact, peripheral information, when objects are crowded and unidentifiable, still provides enough information to tell you where to look (Harrison et al., 2013; Wolfe & Whitney, 2014). While looking at the road, you know that there are several buttons on your vehicle’s touchscreen and roughly where they are, but you cannot be sure of which one you need to touch until you make an eye movement toward it.

It is useful to model peripheral vision to gain intuitions about what information is available to the driver, and in this regard the field of human vision has made significant advances over the last decade. The texture tiling model (Balas et al., 2009; Rosenholtz, 2011, 2016; Rosenholtz, Huang, Ehinger et al., 2012; Rosenholtz, Huang, Raj et al., 2012) allows us to take an image, indicate a point of gaze (corresponding to where the driver is looking), and then visualize what information might survive based on our current understanding of processing in peripheral vision. Extensive behavioral experiments and modeling have demonstrated that the information that survives is often not perfectly detailed (Balas et al., 2009; Rosenholtz, Huang, Raj et al., 2012). The core of the texture tiling model of peripheral vision is the idea that peripheral input is coded in the brain as a set of summary statistics, an efficient way for the brain to summarize input over sizable regions of the visual field. This summarization may provide a solution to the problem of comparatively limited cortical real estate available for nonfoveal input. To accomplish this, the texture tiling model treats local peripheral regions as textures, rather than attempting to perfectly represent the entire scene. This process retains much, but not all of the detail of the original image, while sacrificing some spatial precision; critically, a great deal of useful information remains (Figure 3). Peripheral vision provides different, yet useful information across the full extent of

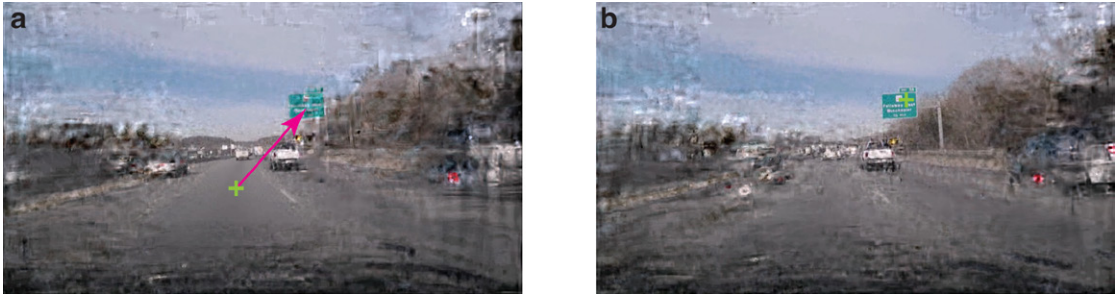


Figure 3. Peripheral vision in driving, as visualized here with the texture tiling model (Balas et al., 2009), is sufficient to guide search for specific information. The green cross in each panel indicates the fixation location. The magenta arrow in (a) indicates an eye movement planned to the sign, which remains visible but not readable in the periphery. Note how the visualization changes in (b) when the fixation location shifts. Sufficient information is available in (a) to enable an accurate eye movement to the sign.

the visual field, and understanding what drivers can and cannot learn from it is essential if we are to understand how drivers acquire information about the world and how the limitations of the visual system constrain and enable their ability to do so.

Critically, the visual system uses the information it extracts across the field of view, or the information available from peripheral vision, to enable rapid scene perception and understanding. Early work in this area demonstrated that scenes could be identified from brief presentations (Biederman, 1972; Biederman et al., 1974) potentially based on rapid extraction of global, rather than local, features from the scene (Navon, 1977). More recently, this has come to be known as the *gist* of the scene (Oliva & Torralba, 2006), since the observer gets the essence, rather than the full details. The gist of a road scene, for example, might be that it is a highway, not an urban scene, but does not consist merely of the scene category. Depending upon the scene, it might include whether or not a hazard was present, but perhaps not precise information about what vehicles are where, or what color each vehicle is. The gist of the road ahead of you is probably sufficient to begin to guide where you attend or look (Torralba et al., 2006) and may, in its own right, be enough for some tasks. Accordingly, asking drivers to report specific details of the scene that may not be operationally relevant (e.g., the color of a given vehicle) may underestimate the driver's ability to act (Endsley, 2019). In fact, information from

across the field of view likely supports many on-road tasks, including time-critical tasks like hazard perception, and supports drivers' information acquisition abilities.

Recent work in this space has investigated how long observers need to see a scene in order to determine how navigable it is, what kind of intersections lie ahead (e.g., a T-junction vs. a left turn; Ehinger & Rosenholtz, 2016; Greene & Oliva, 2009b), and how global properties of the scene are used to accomplish this (Greene & Oliva, 2009a). Observers can extract the information necessary to do these tasks with static images when the images are presented for 100 ms or less, faster than an eye movement, suggesting that observers (or drivers) do not need to look anywhere in particular to do so. In fact, we have shown that drivers watching video of real road scenes can detect hazards in road scenes they have never seen before with as little as 200 ms of video (Wolfe, Seppelt et al., 2020). How might this be possible? Certainly, drivers' expertise will play a key role, but experts' ability to benefit from their expertise necessarily relies upon the information they can extract from this single glance. Vision science studies of gist perception have found that observers can become better at extracting more comprehensive gist from static scenes with training (Greene et al., 2015), a finding that echoes work on driver expertise and the role of gist in driving (Pugeault & Bowden, 2015). Peripheral vision facilitates easy and rapid perception of the gist of a scene, providing global

information from across the field of view. However, this is only one facet of peripheral vision's contribution to the driver's representation of the world and their ability to acquire information; peripheral vision is also essential for planning shifts of attention and shifts of gaze, as we discuss next.

Visual Attention and Visual Search in Information Acquisition

While peripheral vision is informative, it cannot provide all the information we need to drive safely. How do drivers acquire detailed information about locations or objects in their environment, and how do they overcome the limitations of peripheral vision? Research on visual attention, particularly on visual search and attentional capture, suggests some mechanisms by which the visual system may accomplish the goal of selecting elements in the scene and acquiring more detailed information. However, before we can discuss how the driver uses attention to select elements from the scene and acquire details, we need to define what we mean by attention. To define it broadly, attention is a limited resource the visual system has for focusing on an object (say, the cyclist turning across your lane) or a location (like the area of road directly ahead of you, whether or not there is another car there). Broadly speaking, we attend to an object or location to acquire more detailed information about what we are attending to than is otherwise available. Attention is often likened to a spotlight, illuminating an object or location and, by inference, ignoring the rest (as in some interpretations of the useful field; Wolfe et al., 2017). However, the spotlight of attention can be separated from the current point of gaze (covert attention) and, in fact, we can have multiple spotlights in order to track multiple moving objects (Pylyshyn & Storm, 1988), for example, cars in other lanes near yours. Attention has also been thought of as a zoom lens, able to encompass multiple objects or a larger portion of the visual field (Eriksen & St. James, 1986), but without individuating the objects within it. Overall, attention is a flexible tool for selecting a subset of visual input in order to acquire more information, and this

leads us to two questions: what is attention good for, and how is it guided?

Considering how we use attention as a tool to acquire information takes us to work on attention and visual search, and first to Treisman's work on feature integration theory (Treisman & Gelade, 1980), which sought to explain why we perceive the world as detailed, yet we find it hard to find what we are looking for. If peripheral vision, for example, were to provide us with a perfectly detailed representation of the world, we should always be able to find what we are looking for easily, but we cannot. Treisman argued that the spotlight of attention was the key to understanding this problem, because attention was needed to know whether or not a given object was what you were looking for. The idea that attention is necessary for identification is the idea of attentional binding, where individual features (e.g., color, shape, and orientation) are bound into an object that we can identify. To take our example of the oblivious cyclist, you know something is there, you might even guess that the something is a cyclist, but you do not know for certain until you attend to it and bind the features into the cyclist you perceive. Treisman described attention as the glue with which we built objects, and this idea is sometimes erroneously understood as attention being required for any perception, rather than being required for accurate identification.

However, there is a bit of a wrinkle here: What can you perceive before attention binds a given group of features into an object? Feature integration theory implicitly argues that you have some awareness: You certainly know something is there, but you cannot identify it. To think of it another way, you are aware of the visual world beyond your current focus of attention, you are just not quite sure what everything is. The fact that you have some limited awareness of preattentive information is key: In order for attention to bind features into objects, features must be available to be bound. However, in feature integration theory, preattentive information is considered less than reliable; in subsequent work, Treisman discusses the idea of illusory conjunctions, where features away from what you are attending to are erroneously bound into plausible, but incorrect, mental percepts

without attention (Treisman & Schmidt, 1982). So, if you are attending to the cyclist, you may be unsure if the vehicle in the right lane is a red truck or a blue truck, although you probably know something is there. On the road, you may not need to know what color the truck was; you may not need a perfect representation of the environment, merely an adequate one.

There is a bit of a missing piece in the early versions of feature integration theory: How does the visual system know where to direct attention? What guides attention? Treisman's work does not address the question of where these features are in visual space, at least not directly. Guided search (Wolfe, 1994; Wolfe et al., 1989) addresses this problem, by asking "how does preattentive information guide search?" According to guided search, the visual system has feature maps, generated by a parallel process, which provide both the unbound features and some information about where they are in visual space, and, when you are sequentially searching for a target, attention is guided using these maps. In essence, you plan where to search based on where there are more of the features that potentially match your target and you work your way through these peaks on the feature map to complete a search task, thereby answering the question of how attention knows where to go. However, the information available preattentively is very simple: if you are looking for the red car, you might search through all red items in the scene, but guided search, in this early form, did not reflect your ability to understand the scene and would have you search serially through all the red items in the scene until you found it.

Both feature integration theory and guided search implicitly assume that you know what you are looking for in the scene, you just need to find it. What about elements in the scene that change and you did not know to look for beforehand? A common explanation for how we become aware of these elements comes from the literature on attentional capture, where a change in the world (like a cyclist abruptly steering into your lane) captures attention and forces you to become aware of it (c.f., Theeuwes, 1994). Intuitively, this is an appealing way to think about how drivers acquire the information they need, since it enables

them to become aware of these abrupt changes essentially immediately. However, this brings us to the same problem we just discussed in attentional guidance: If an object can capture attention, it is because there is information available to direct attention. The cyclist turning into your lane is certainly going to attract your attention, if only because their motion vector is orthogonal to yours, but you (probably) have some awareness of them before you can shift the focus of attention, and certainly before you can shift your gaze. The cyclist may still capture attention, but in order to do so, there was already information available to your visual system, which means that while attention might be captured, capture is not, in and of itself, a sufficient explanation for how drivers become aware of changes in the environment.

Lurking beneath the work on search and capture that we have discussed is the concept of preattentive information: in order to guide where you search, or for attention to know where to go when it is captured, the visual system has some preattentive information to operate on. However, ideas on what the visual system can do with information beyond the current focus of attention have changed dramatically in the last two decades. One body of evidence that pushed us beyond this view of preattentive information we have discussed is work on ensemble perception, our ability to look at a group of similar objects and take their average without needing to attend to each member of the group. This has been shown for properties ranging from size (Ariely, 2001) to orientation (Dakin & Watt, 1997) to facial emotion (Yamanashi Leib et al., 2014), and, more relevant for driving, the direction a group is walking (Sweeny et al., 2013). Since you can perceive the average of a group without attending, these results indicate that observers do not need to separately attend to each object in an array to understand them as a whole. Our ability to perceive the gist of a scene makes a similar argument as to how the visual system uses information more broadly and suggested that the accounts of attention and search we have discussed were incomplete.

This deeper understanding of what the visual system could accomplish without attention led to significant reworkings of both feature integration theory and guided search, respectively (Treisman, 2006; Wolfe, 2006). Treisman's 2006 version of

feature integration theory includes the idea of distributed attention, a diffuse ability to attend across the entire visual field without binding features into objects in order to encompass these results. However, while this change makes space for these results, it does not describe how we might use what we can perceive from the scene as a whole to guide attention and search, nor does it specify what tasks are possible with distributed attention. Other theories of search, including Wolfe's updates to guided search (Wolfe, 2006, 2012; Wolfe & Horowitz, 2017) and Rensink's work on theories of visual attention (Rensink, 2013), integrated these results, incorporating a more expansive idea of what the visual system is capable of before attention comes into play and arguing that these processes, like scene gist, can themselves guide search. These updated theories broadly recognize that the foundation upon which attention operates is informative in and of itself.

Building out from this idea, how might our knowledge of scenes impact search? If you are being a mindful driver and looking out for pedestrians (or the cyclist we have used as an example throughout), you are likely to look where you expect them to be (Torrallba et al., 2006). This idea has been developed in detail by Vö and colleagues as scene grammar and argues that we use our expertise in the structure of scenes to inform where we should look for a given target (Wolfe et al., 2011). Vö and colleagues describe scene-mediated guidance as being driven by semantic factors, the idea that some objects belong in certain scenes and not others along with syntactic factors, that objects ought to behave as we expect them to (Draschkow & Vö, 2017). As a result, objects which appear in entirely unexpected places will be harder to find, because they violate this expectation, suggesting that understanding how drivers acquire information also requires us to understand where they expect that information to be in the scene (Biederman et al., 1988; Castelhana & Heaven, 2011; Castelhana & Henderson, 2007).

To sum up, while drivers undoubtedly need to attend, focusing exclusively on what they attend to misses an important part of the puzzle. Indeed, by asking how attention is guided and how the information used for guidance is, itself, essential to drivers' ability to perceive the world, we can

better understand how drivers acquire the information they need. To summarize work on visual attention theory, the notion that preattentive information is only useful for attentional guidance has been supplanted by results showing how much we can achieve at a glance. In particular, work on scene perception and search suggests that attention and search is guided by this information and not just by the location of unbound features. This echoes what can be argued from a peripheral vision perspective, which is that we acquire detailed information from across the visual field using the information in peripheral vision (Ehinger & Rosenholtz, 2016), although this representation is not as accurate as we might wish and can include misbound features (Rosenholtz, Huang, Raj et al., 2012), similar to the illusory conjunctions we discussed (Treisman & Schmidt, 1982). These are inherently complementary ideas, because they both suggest that if we are interested in how the driver acquires information, we need to consider what information is available from across the visual field and recognize that drivers use this information, rather than only information from attended locations. Attention is a useful tool for augmenting the information the visual system has already gathered, similar to the idea of minimum required attention (Kircher & Ahlstrom, 2017), but selective attention is not responsible for drivers acquiring all the information they need. To take it one step further: Work on peripheral vision and visual attention together suggests that the information the driver acquires is often imperfect but adequate. Assuming that the driver's mental model of their environment needs to be perfect is at odds with their manifest ability to gather and represent information in the time available before they must act.

Eye Movements as a Tool for Information Acquisition

An additional piece of evidence for just how informative peripheral vision is, or how attending to one object or location cannot rob the visual system of all unattended information, comes when we consider how and why we move our eyes. Consider the cyclist turning across your lane of travel. To avoid hitting them, you are almost certain to make an eye movement in order

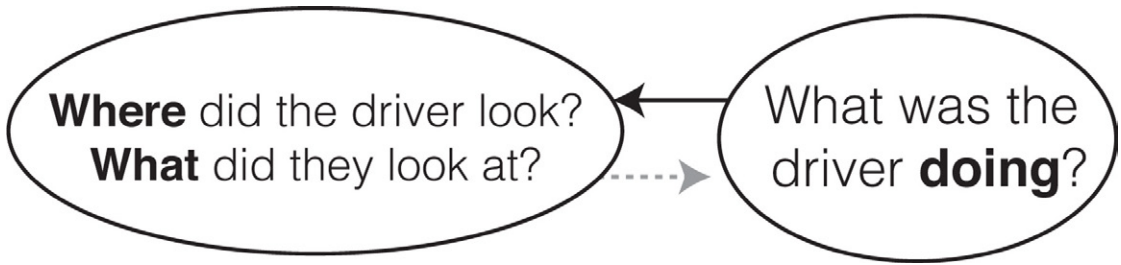


Figure 4. Understanding eye movements in information acquisition requires considering more than just where the driver looked. In order to understand how drivers acquire information, recorded gaze data must reflect where the driver looked, what they looked at, and what they were doing at the time. However, merely knowing where the driver looked and what they looked at may not be sufficient to indicate their task.

to better localize them in the environment, since knowing where they are is helpful when it comes to avoiding a collision. When we plan volitional eye movements (saccades), we attend to the target of the impending eye movement, gathering more information about the target or location before our eyes move. Presaccadic attention (Kowler et al., 1995) is essential for the accuracy of the eye movement (Kowler & Blaser, 1995) and even lets us better perceive objects we cannot otherwise identify before we look at them (Wolfe & Whitney, 2014). As we have discussed, eye movements are often studied to try to get at questions of how the driver acquires the information they need, but we need to understand how and why we make eye movements in order to successfully use them to probe questions of information acquisition.

If, for example, we are merely interested in whether the driver is looking at the road or not, as in *AttenD* (Kircher & Ahlstrom, 2009), there is little need to consider *why* the driver planned an eye movement from the phone in their lap back to the road. It is enough to know that they were looking away from the road, and we can surmise that they looked back to the road either because they detected something in their periphery (Wolfe, Sawyer et al., 2019) or simply got nervous about looking away. However, if the goal is to understand what the driver knows in order to model it, the problem becomes more complicated, since we need to ask *where* drivers looked, *what* they looked at, including what they could see in their periphery, and what task they were

doing that required them to make an eye movement. Critically, where the driver looked and what they looked at are inseparable; to make sense of the eye movement, we have to consider what the driver looked at specifically. To do this, we need to think about eye movements not just as shifts of gaze between large areas of interest, but as shifts of gaze between specific objects or locations in the driver's field of view (Figure 4). Therefore, we need to consider eye movement planning: in essence, what information is available from the periphery and how that explains why the eye movement was made to a given location.

The problem that arises is that these three questions, applied to driving, require accurate eye-tracking data coregistered with a recording of the driver's environment, as well as some way of determining the driver's task. In eye movement research, the idea of identifying an observer's task based on the gazepath and the stimulus is often credited to Yarbus. He showed observers a single image (Ilya Repin's *Unexpected Visitor*), asked them to perform various tasks, recorded different gazepaths for each task, removed the task labels from the gazepaths, and tried to determine what task the observer had been performing (Yarbus, 1967). This seems as if it would be fairly straightforward, but there is a problem here: Yarbus assumed that all observers doing the same task would, by and large, make the same set of eye movements. While observers will, often, look at the same areas in an image, particularly for locations within a face (c.f., Peterson et al., 2012), there are significant, stable individual differences

between observers' gazepaths even when they perform the same task (e.g., making a single saccade to a transient target, as in Kosovicheva & Whitney, 2017). These individual differences will significantly hamper efforts to determine exactly what a given individual was doing at a given point in time, since any analysis that tries to classify gazepaths by task has to contend with the added variance between individuals. This problem hampered later attempts to follow Yarbus' example (Greene et al., 2012) and has proven to be a stumbling block in using eye movements as a window into what drivers know about the world, as well as modeling their situation awareness based on their eye movements (c.f., Endsley, 2019).

There is some hope for broadly classifying task and behavior based on eye movements, although the classifications may be less precise than desired for characterizing information acquisition. We know that expert and novice drivers have very different scanpaths, both in hazard detection (Crundall, 2016) and more broadly (Mourant & Rockwell, 1972), while age is also a factor in how drivers use eye movements to acquire specific information (Sawyer et al., 2016) while gaze behavior is also influenced by the degree to which the driver trusts the vehicle (Sawyer, Seppalt et al., 2017) and their level of fatigue (Ji et al., 2004). For that matter, the features of the environment itself seem to have a profound impact on drivers' ability to predict immediate events (Wolfe, Fridman et al., 2019), which implies that gazepaths in different environments, for example, highway versus urban, are likely to be discriminable. Indeed, for some cases in which we might want to predict broad features of the driver or their environment, gazepaths are likely to be robust to individual differences, because these large-scale factors will swamp individual variability. However, more detailed classifications of driver state or task, like whether the driver started looking for their exit or, more granularly, what they are doing with the infotainment system, are likely to become progressively harder as individual differences account for more of the observed variance.

One way forward might be to take advantage of the recent availability of very large video-based datasets from instrumented vehicles (Dingus et al., 2006, Dingus et al., 2015;

Fridman et al., 2017; Owens et al., 2015), particularly when these datasets include gaze information, because with sufficient data, it may be possible to overcome the problem of individual variability. This can also be aided by new deep learning techniques for determining the actions occurring in a natural scene (Monfort et al., 2020), as this would reduce or eliminate the need to manually code the scene video. A combination of accurate gaze data, knowledge of scene contents, and some ability to determine the task a driver was performing at a given point in time promises to open a window to how the driver acquires the information they need and what they know about the environment at a given point in time. In particular, using these tools to ask where the driver might look next, based on an understanding of what information is available peripherally (and therefore available to eye movement planning), could enable far more robust driver state prediction and monitoring systems and increase road safety.

Conclusions and Future Implications

Having discussed our theory of information acquisition and its origins in vision science research, where can we go from here? Overall, our theory suggests new approaches to testing, modeling, and predicting what the driver knows, based on *how* they know it. Our approach has considerable overlap with the work of de Winter and colleagues (de Winter et al., 2019), since, like them, our goal is to use where drivers looked to inform our understanding of what they know, rather than relying on freeze-frame probes or other explicit tasks. Specifically, to test how drivers develop, maintain, lose, and restore their mental representation of the environment, we would focus on their ability to *use* the information they acquire to safely navigate, rather than their ability to report specific details of a given scene. Doing so will allow us to quantify this representation based on our estimate of driver performance and will facilitate predicting how it changes as drivers perform different tasks which change the information available to them.

To test how their representation of the environment is built up over time, we should develop experiments, probably in the laboratory at first, that ask observers to look at driving scenes and gain holistic understanding, rather than measuring their ability to report specific elements in the scene. For instance, we can probe how the acquisitive processes work here by asking observers to detect and evade road hazards, tasks that require the observer to develop an understanding of the scene as a whole (Wolfe, Seppelt et al., 2020). A key question in modeling these processes is how quickly, in terms of the required viewing duration, can drivers achieve a useful, if imperfect, degree of awareness. Critically, we have shown that the time required changes by a factor of 2 as a function of the task (220 ms to detect, but 400 ms to evade), emphasizing the task dependence of any measure used here. In addition, observers' ability to perform any task is likely a function of their expertise, and understanding the relationship between expertise and response is essential to be able to build useful models and requires testing across experience and age.

Drivers continually maintain their mental representation of the environment, but their ability to do so varies as a function of the information available to them and where they look. Because of the limits of peripheral vision, it is likely easier for drivers to do so if they look at the road, but what is the impact of looking away? Brake light detection and driver distraction studies suggest a way forward (Lamble et al., 1999; Summala et al., 1998; Wolfe, Sawyer et al., 2019; Yoshitsugu et al., 2000); by controlling where drivers or observers look, we can limit what information is available to them and determine the impact on their ability to maintain and update their awareness.

Additionally, drivers' mental representation is relatively robust to transient interruption like blinks and shifts of gaze (Matin & Pearce, 1965), and even to a total loss of visual input (Senders et al., 1967). This suggests that these representations are neither easily lost (Seppelt et al., 2017) nor develop slowly (Lu et al., 2017), simply because drivers must be able to quickly recover enough information

for action. Testing what it means for these representations to decay or be lost can be difficult, but laboratory experiments that ask observers to predict "what happens next" in driving scenes (Jackson et al., 2009; Wolfe, Fridman et al., 2019) might be one way to ask how the mental representation decays, particularly when combined with distracting tasks (although these representations are robust for experts, as reported by Endsley, 2019).

The robustness of these mental representations to some interruptions may be one reason that bridging the gap between the predictions of cognition-focused theories like situation awareness and performance of concurrent nondriving-related-activities has proved such a complex endeavor (de Winter et al., 2014). Indeed, while epidemiological approaches have had some success, attempting to use situation awareness to build process-focused indices, based on direct measurement, behavioral observation, performance assessment, and post hoc analysis (Endsley & Garland, 2000) have all struggled to adequately capture the impacts of these secondary tasks upon safety (as noted in Endsley, 2019). This is all the more frustrating because the automotive industry has a longstanding interest in these very questions (Endsley & Garland, 2000) and continues to actively explore ways forward within existing theoretical frameworks (i.e., Skrypchuk et al., 2019), even as an increasing number of patents describe what could be accomplished if driver knowledge could be well modeled.

The practical implications of information acquisition theory are significant; by giving traffic safety researchers a theory which explains how drivers acquire visual information, it enables more complete explanations of driver behavior. For example, novice and expert drivers move their eyes in distinctly different ways (Mourant & Rockwell, 1972), which may reflect expert's ability to interpret peripheral information in ways novices have not learned to yet. That, in turn, suggests that novice drivers need to make eye movements to perceive information that experts have learned to extract from peripheral vision. Along those lines, if we can better understand what information expert drivers are using from their

periphery, we can consider whether current training regimes teach drivers to use this information as well as they can and how we can train drivers to use their visual capabilities to the fullest extent possible.

Looking toward the future, particularly in the context of midlevel autonomy, understanding how drivers acquire information becomes ever more critical in L2 and L3 vehicles, since drivers are unlikely to focus on the road without operational need. Critically, developing a safe L2 or L3 system requires an implicit understanding of information acquisition, since safety will demand the vehicle account for the driver's capabilities and limitations. While information acquisition provides the upper bound for what the driver might know, such an upper bound is essential to developing safe autonomous and assistive systems, because these systems must reflect what the driver is actually capable of and how their capabilities change.

Consider the goal of assisting drivers by cuing them to things in the environment while they are distracted: understanding how drivers acquire information will let us develop better ways to augment their abilities and address their needs. Having a theory that helps us explore what the driver knows and how they know it is also useful in understanding the difference between what drivers know and what they do, helping us move beyond a focus on just where drivers look, to why they looked where they did, and what else they know about the situation. More broadly, in the context of understanding distraction in myriad contexts, understanding the interactions between peripheral vision, visual search, and eye movements can help to explain what distracts drivers, how the impacts are attenuated, and what we can do about them.

Information acquisition focuses on mechanisms and can provide necessary answers for developing models. Indeed, in conjunction with an understanding of how drivers use peripheral vision and plan eye movements, information acquisition provides a foundation toward predicting how drivers' representation of their environment changes based on where they look and what they are doing. By looking at this problem from a different perspective, future work will provide

both a more detailed understanding of what it means to be aware of the operating environment and a better understanding of how we use vision in the world, not just in the laboratory. Our approach complements work in situation awareness: going forward, we need both information acquisition's mechanistic understanding of how we acquire the information we need and the understanding of higher level cognition championed by Endsley (Endsley, 2015, Endsley, 2019). By fostering the exchange of ideas between vision science and human factors, we can better understand how the visual system works and how its capabilities and limitations govern what we can do on the road as drivers, and in the world as a whole.

ACKNOWLEDGMENTS


This work was supported by the Toyota Research Institute (TRI) through the CSAI-TRI Joint Partnership at MIT. However, this article solely reflects the opinions and conclusions of its authors. The authors also wish to thank the reviewers of this paper for their patience, and thank Anna Kosovicheva for many helpful discussions throughout the process.

KEY POINTS

- Information acquisition is a two-stage perceptual theory, beginning with information from across the visual field, forming the foundation for the driver's representation of the environment while guiding eye movements and attentional shifts as necessary to add detail.
- Information acquisition explains how drivers acquire and represent visual information, connecting applied problems in driver behavior and traffic safety to fundamental research findings in vision science.
- Information acquisition describes the mechanisms by which drivers acquire visual information and is a set of perceptual processes which precede the cognitive processes that give rise to situation awareness.
- Future work in driver behavior, including modeling driver knowledge and state, is enabled by the understanding that information acquisition provides, and the theoretical framework can inform visual perception research in both basic and applied domains.

- In addition to scientific knowledge, practitioner and engineering efficacy can be boosted by the application of our theory.

ORCID iD

Benjamin Wolfe  <https://orcid.org/0000-0001-9921-8795>

REFERENCES

- Anstis, S. (1998). Picturing peripheral acuity. *Perception*, 27, 817–825. <https://doi.org/10.1068/p270817>
- Anstis, S. M. (1974). Letter: A chart demonstrating variations in acuity with retinal position. *Vision Research*, 14, 589–592. [https://doi.org/10.1016/0042-6989\(74\)90049-2](https://doi.org/10.1016/0042-6989(74)90049-2)
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12, 157–162. <https://doi.org/10.1111/1467-9280.00327>
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9, 13–18. <https://doi.org/10.1167/9.12.13>
- Biederman, I. (1972). Perceiving real-world scenes. *Science*, 177, 77–80. <https://doi.org/10.1126/science.177.4043.77>
- Biederman, I., Blichle, T. W., Teitelbaum, R. C., & Klatsky, G. J. (1988). Object search in nonscene displays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 456–467. <https://doi.org/10.1037/0278-7393.14.3.456>
- Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology*, 103, 597–600. <https://doi.org/10.1037/h0037158>
- Blättler, C., Ferrari, V., Didierjean, A., & Marmèche, E. (2011). Representational momentum in aviation. *Journal of Experimental Psychology: Human Perception and Performance*, 37, 1569–1577.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226, 177–178. <https://doi.org/10.1038/226177a0>
- Braunagel, C., Geisler, D., Rosenstiel, W., & Kasneci, E. (2017). Online recognition of driver-activity based on visual Scanpath classification. *IEEE Intelligent Transportation Systems Magazine*, 9, 23–36. <https://doi.org/10.1109/ITS.2017.2743171>
- Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, 51, 1484–1525. <https://doi.org/10.1016/j.visres.2011.04.012>
- Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene GIST: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, 18, 890–896. <https://doi.org/10.3758/s13423-011-0107-8>
- Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 753–763. <https://doi.org/10.1037/0096-1523.33.4.753>
- Crundall, D. (2016). Hazard prediction discriminates between novice and experienced drivers. *Accident Analysis & Prevention*, 86, 47–58. <https://doi.org/10.1016/j.aap.2015.10.006>
- Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, 292, 497–523. <https://doi.org/10.1002/cne.902920402>
- Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, 37, 3181–3192. [https://doi.org/10.1016/S0042-6989\(97\)00133-8](https://doi.org/10.1016/S0042-6989(97)00133-8)
- de Winter, J. C. F., Eisma, Y. B., Cabral, C. D. D., Hancock, P. A., & Stanton, N. A. (2019). Situation awareness based on eye movements in relation to the task environment. *Cognition, Technology & Work*, 21, 99–111. <https://doi.org/10.1007/s10111-018-0527-6>
- de Winter, J. C. F., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transportation Research Part F: Traffic Psychology and Behaviour*, 27, 196–217. <https://doi.org/10.1016/j.trf.2014.06.016>
- Deubel, H. (2008). The time course of presaccadic attention shifts. *Psychological Research*, 72, 630–640. <https://doi.org/10.1007/s00426-008-0165-3>
- Dingus, T. A., Hankey, J. M., Antin, J. F., Lee, S. E., Eichelberger, L., Stulce, K. E., McGraw, D., & Stowe, L. (2015). Naturalistic driving study: Technical coordination and quality control (Report No. SHRP 2 Report S2-S06-RW-1).
- Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., Perez, M. A., Hankey, J., Ramsey, D. J., Gupta, S., Bucher, C., Doerzaph, Z., Jermeland, J., & Knipling, R. R. (2006). The 100-car naturalistic driving study. Phase II, results of the 100-car field experiment (Report No. DOT-HS-810-593): US Department of Transportation, National Highway Traffic Safety Administration.
- Draschkow, D., & Võ, M. L.-H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Scientific Reports*, 7, 16471. <https://doi.org/10.1038/s41598-017-16739-x>
- Ehinger, K. A., & Rosenholtz, R. (2016). A general account of peripheral encoding also predicts scene perception performance. *Journal of Vision*, 16, 13–13. <https://doi.org/10.1167/16.2.13>
- Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Proceedings of the Human Factors Society Annual Meeting*, 32, 97–101. <https://doi.org/10.1177/154193128803200221>
- Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32–64. <https://doi.org/10.1518/001872095779049543>
- Endsley, M. R. (2015). Situation awareness misconceptions and misunderstandings. *Journal of Cognitive Engineering and Decision Making*, 9, 4–32. <https://doi.org/10.1177/1555343415572631>
- Endsley, M. R. (2019). A systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of SAGAT and spam. *Human Factors*, 12, 001872081987537. <https://doi.org/10.1177/0018720819875376>
- Endsley, M. R., & Garland, D. J. (2000). *Situation awareness analysis and measurement*. CRC Press.
- Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics*, 40, 225–240. <https://doi.org/10.3758/BF03211502>
- Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, 106, 1389–1398. <https://doi.org/10.1152/jn.00904.2010>
- Fishman, G. A., Anderson, R. J., Stinson, L., & Haque, A. (1981). Driving performance of retinitis pigmentosa patients. *British Journal of Ophthalmology*, 65, 122–126. <https://doi.org/10.1136/bjo.65.2.122>
- Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 126–132. <https://doi.org/10.1037/0278-7393.10.1.126>
- Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., Terwilliger, J., Terwilliger, J., Ding, L., Seaman, S., Mehler, A., Sipperley, A., Pettinato, A., Seppelt, B., Angell, L., Mehler, B., & Reimer, B. (2017). MIT Autonomous Vehicle Technology Study: Large-Scale Deep 1049 Learning Based Analysis of Driver Behavior and Interaction with Automation. *arXiv preprint*, 1.
- Green, M. (2000). “How long does it take to stop?” Methodological analysis of driver perception-brake times. *Transportation Human Factors*, 2, 195–216. https://doi.org/10.1207/STHF0203_1
- Greene, M. R., Botros, A. P., Beck, D. M., & Fei-Fei, L. (2015). What you see is what you expect: Rapid scene understanding benefits from prior experience. *Attention, Perception, & Psychophysics*, 77, 1239–1251. <https://doi.org/10.3758/s13414-015-0859-8>
- Greene, M. R., Liu, T., & Wolfe, J. M. (2012). Reconsidering Yarbuss: A failure to predict observers’ task from eye movement patterns.

- Vision Research*, 62, 1–8. <https://doi.org/10.1016/j.visres.2012.03.019>
- Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology*, 58, 137–176. <https://doi.org/10.1016/j.cogpsych.2008.06.001>
- Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, 20, 464–472. <https://doi.org/10.1111/j.1467-9280.2009.02316.x>
- Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2013). Eye movement targets are released from visual crowding. *Journal of Neuroscience*, 33, 2927–2933. <https://doi.org/10.1523/JNEUROSCI.4172-12.2013>
- Higuchi, Y., Inoue, S., Hamada, H., & Kumada, T. (2019). Artificial optic flow guides visual attention in a driving scene. *Human Factors*, 137, 001872081984702.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36, 791–804. [https://doi.org/10.1016/s0896-6273\(02\)01091-7](https://doi.org/10.1016/s0896-6273(02)01091-7)
- Jackson, L., Chapman, P., & Crundall, D. (2009). What happens next? Predicting other road users' behaviour as a function of driving experience and processing time. *Ergonomics*, 52, 154–164. <https://doi.org/10.1080/00140130802030714>
- Ji, Q., Zhu, Z., & Lan, P. (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technology*, 53, 1052–1068. <https://doi.org/10.1109/TVT.2004.830974>
- Kircher, K., & Ahlstrom, C. (2009). *Issues related to the driver distraction detection algorithm attend* [Conference session]. Presented at the first International Conference on driver distraction and inattention, Gothenburg, Sweden.
- Kircher, K., & Ahlstrom, C. (2017). Minimum required attention: A human-centered approach to driver inattention. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 59, 471–484. <https://doi.org/10.1177/0018720816672756>
- Kosovicheva, A., & Whitney, D. (2017). Stable individual signatures in object localization. *Current Biology*, 27, R700–R701. <https://doi.org/10.1016/j.cub.2017.06.001>
- Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, 51, 1457–1483. <https://doi.org/10.1016/j.visres.2010.12.014>
- Kowler, E., Anderson, E., Doshier, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, 35, 1897–1916. [https://doi.org/10.1016/0042-6989\(94\)00279-U](https://doi.org/10.1016/0042-6989(94)00279-U)
- Kowler, E., & Blaser, E. (1995). The accuracy and precision of saccades to small and large targets. *Vision Research*, 35, 1741–1754. [https://doi.org/10.1016/0042-6989\(94\)00255-K](https://doi.org/10.1016/0042-6989(94)00255-K)
- Kristjánsson, Á., & Egeth, H. (2020). How feature integration theory integrated cognitive psychology, neurophysiology, and psychophysics. *Attention, Perception, & Psychophysics*, 82, 7–23. <https://doi.org/10.3758/s13414-019-01803-7>
- Lamble, D., Laakso, M., & Summala, H. (1999). Detection thresholds in car following situations and peripheral vision: Implications for positioning of visually demanding in-car displays. *Ergonomics*, 42, 807–815. <https://doi.org/10.1080/001401399185306>
- Lu, Z., Coster, X., & de Winter, J. (2017). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics*, 60, 293–304. <https://doi.org/10.1016/j.apergo.2016.12.003>
- Matin, L., & Pearce, D. G. (1965). Visual perception of direction for stimuli flashed during voluntary saccadic eye movements. *Science*, 148, 1485–1488. <https://doi.org/10.1126/science.148.3676.1485>
- McLeod, R. W., & Ross, H. E. (1983). Optic-flow and cognitive factors in time-to-collision estimates. *Perception*, 12, 417–423. <https://doi.org/10.1068/p120417>
- Monfort, M., Andonian, A., Zhou, B., Ramakrishnan, K., Bargal, S. A., Yan, T., Brown, L., Fan, Q., Gutfreund, D., Vondrick, C., & Oliva, A. (2020). Moments in time dataset: One million videos for event understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42, 1–1. <https://doi.org/10.1109/TPAMI.2019.2901464>
- Mourant, R. R., & Rockwell, T. H. (1972). Strategies of visual search by novice and experienced drivers. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 14, 325–335. <https://doi.org/10.1177/001872087201400405>
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9, 353–383. [https://doi.org/10.1016/0010-0285\(77\)90012-3](https://doi.org/10.1016/0010-0285(77)90012-3)
- Oliva, A., & Torralba, A. (2006). Building the GIST of a scene: The role of global image features in recognition. *Progress in Brain Research*, 155, 23–36. [https://doi.org/10.1016/S0079-6123\(06\)55002-2](https://doi.org/10.1016/S0079-6123(06)55002-2)
- Owens, J. M., Angell, L., Hankey, J. M., Foley, J., & Ebe, K. (2015). Creation of the naturalistic engagement in secondary tasks (nest) distracted driving dataset. *Journal of Safety Research*, 54, 33.e29–33.36. <https://doi.org/10.1016/j.jsr.2015.07.001>
- Owsley, C. (2011). Aging and vision. *Vision Research*, 51, 1610–1622. <https://doi.org/10.1016/j.visres.2010.10.020>
- Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*, 18, 445–451. <https://doi.org/10.1016/j.conb.2008.09.008>
- Peterson, M. F., & Eckstein, M. P., Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences*, 109, E3314–E3323. <https://doi.org/10.1073/pnas.1214269109>
- Pugeault, N., & Bowden, R. (2015). How much of driving is pre-attentive? *IEEE Transactions on Vehicular Technology*, 64, 5424–5438.
- Pylshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3, 179–197. <https://doi.org/10.1163/156856888X00122>
- Reimer, B. (2009). Impact of cognitive task complexity on drivers' Visual tunneling. *Transportation Research Record: Journal of the Transportation Research Board*, 2138, 13–19. <https://doi.org/10.3141/2138-03>
- Rensink, R. A. (2013). Perception and attention. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 97–116). Oxford University Press.
- Rosenholtz, R. (2011). What your visual system sees where you are not looking. In B. E. Rogowitz & T. N. Pappas (Eds.), *Proceedings of SPIE: Human Vision and Electronic Imaging* (Vol. 7865, pp. 7865–7910). International Society for Optics and Photonics.
- Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. *Annual Review of Vision Science*, 2, 437–457. <https://doi.org/10.1146/annurev-vision-082114-035733>
- Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology*, 3, 13. <https://doi.org/10.3389/fpsyg.2012.00013>
- Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, 12, 14. <https://doi.org/10.1167/12.4.14>
- Sawyer, B. D., Lee, J., Dobres, J., Mehler, B., Coughlin, J. F., & Reimer, B. (2016). Effects of a voice interface on mirror check decrements in older and younger multitasking drivers. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60, 16–20. <https://doi.org/10.1177/1541931213601004>
- Sawyer, B. D., Mehler, B., & Reimer, B. (2017). Toward an antiphony framework for dividing tasks into subAntiphony framework for dividing tasks into Subtasks. *9th Driving Assessment Conference*, 2017, 108–114.
- Sawyer, B. D., Seppalt, B., Mehler, B., & Reimer, B. (2017). Trust impacts driver glance strategy in multitasking. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61, 1441–1442. <https://doi.org/10.1177/1541931213601845>
- Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., & Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Research Record*, 14–33.
- Seppelt, B. D., Seaman, S., Lee, J., Angell, L. S., Mehler, B., & Reimer, B. (2017). Glass half-full: On-road glance metrics differentiate crashes from near-crashes in the 100-Car data. *Accident Analysis & Prevention*, 107, 48–62. <https://doi.org/10.1016/j.aap.2017.07.021>

- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28, 1059–1074. <https://doi.org/10.1068/p281059>
- Sivak, M. (1996). The information that drivers use: Is it indeed 90% visual? *Perception*, 25, 1081–1089. <https://doi.org/10.1068/p251081>
- Skrypchuk, L., Langdon, P., Sawyer, B. D., Mouzakitis, A., & Clarkson, P. J. (2019). Enabling multitasking by designing for situation awareness within the vehicle environment. *Theoretical Issues in Ergonomics Science*, 20, 105–128. <https://doi.org/10.1080/1463922X.2018.1485984>
- Spence, C., & Ho, C. (2015). Crossmodal attention: From the laboratory to the real world (and back again). In J. M. Fawcett, E. F. Risko, & A. Kingstone (Eds.), *The handbook of attention* (pp. 119–138). MIT Press.
- Summala, H., Lambale, D., & Laakso, M. (1998). Driving experience and perception of the lead CAR's braking when looking at in-car targets. *Accident Analysis & Prevention*, 30, 401–407. [https://doi.org/10.1016/S0001-4575\(98\)00005-0](https://doi.org/10.1016/S0001-4575(98)00005-0)
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 329–337. <https://doi.org/10.1037/a0028712>
- Theeuwes, J. (1994). Endogenous and exogenous control of visual selection. *Perception*, 23, 429–440. <https://doi.org/10.1068/p230429>
- Tootell, R. B., Silverman, M. S., Switkes, E., & De Valois, R. L. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, 218, 902–904. <https://doi.org/10.1126/science.7134981>
- Torralba, A., Oliva, A., Castelano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, 113, 766–786. <https://doi.org/10.1037/0033-295X.113.4.766>
- Traquair, H. M. (1927). *An introduction to clinical perimetry* (1st ed., pp. 4–10). Henry Kimpton.
- Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, 14, 411–443. <https://doi.org/10.1080/13506280500195250>
- Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107–141. [https://doi.org/10.1016/0010-0285\(82\)90006-8](https://doi.org/10.1016/0010-0285(82)90006-8)
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Vargas-Martín, F., & Peli, E. (2006). Eye movements of patients with tunnel vision while walking. *Investigative Ophthalmology & Visual Science*, 47, 5295–5302. <https://doi.org/10.1167/iovs.05-1043>
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, 15, 160–168. <https://doi.org/10.1016/j.tics.2011.02.005>
- Wolfe, B., Dobres, J., Rosenholtz, R., & Reimer, B. (2017). More than the useful field: Considering peripheral vision in driving. *Applied Ergonomics*, 65, 316–325. <https://doi.org/10.1016/j.apergo.2017.07.009>
- Wolfe, B., Fridman, L., Kosovicheva, A., Seppelt, B., Mehler, B., Reimer, B., & Rosenholtz, R. (2019). Predicting road scenes from brief views of driving video. *Journal of Vision*, 19, 8–14. <https://doi.org/10.1167/19.5.8>
- Wolfe, B., Sawyer, B. D., Kosovicheva, A., Reimer, B., & Rosenholtz, R. (2019). Detection of brake lights while distracted: Separating peripheral vision from cognitive load. *Attention, Perception, & Psychophysics*, 81, 2798–2813. <https://doi.org/10.3758/s13414-019-01795-4>
- Wolfe, B., Seppelt, B., Mehler, B., Reimer, B., & Rosenholtz, R. (2020). Rapid holistic perception and evasion of road hazards. *Journal of Experimental Psychology: General*, 149, 490–500. <https://doi.org/10.1037/xge0000665>
- Wolfe, B. A., & Whitney, D. (2014). Facilitating recognition of crowded faces with presaccadic attention. *Frontiers in Human Neuroscience*, 8, 103. <https://doi.org/10.3389/fnhum.2014.00103>
- Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, 1, 202–238. <https://doi.org/10.3758/BF03200774>
- Wolfe, J. M. (2006). Guided Search 4.0: Current progress with a model of visual search. In W. Grey (Ed.), *Integrated models of cognitive systems* (pp. 99–119). Oxford University Press.
- Wolfe, J. M. (2012, August 8). Approaches to visual search: Feature integration theory and guided search 2013.
- Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 419–433. <https://doi.org/10.1037//0096-1523.15.3.419>
- Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, 1, 0058. <https://doi.org/10.1038/s41562-017-0058>
- Wolfe, J. M., Võ, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, 15, 77–84. <https://doi.org/10.1016/j.tics.2010.12.001>
- Yamanashi Leib, A., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: A viewpoint-invariant mechanism to represent average crowd identity. *Journal of Vision*, 14, 26. <https://doi.org/10.1167/14.8.26>
- Yarbus, A. L. (1967). *Eye movements and vision*. Springer.
- Yoshitsugu, N., Ito, T., & Asoh, T. (2000). *JAMA's safety guideline on in-vehicle display systems (the study of monitor location of in-vehicle)* [Poster presentation]. Proceedings of the World Congress on Intelligent Systems, Turin, Italy.

Benjamin Wolfe is a postdoctoral associate at the Massachusetts Institute of Technology, affiliated with the Department of Brain and Cognitive Sciences and the Computer Science and Artificial Intelligence Laboratory, and will be an assistant professor at the University of Toronto Mississauga in the Psychology Department starting in 2021. He received his PhD in psychology from the University of California at Berkeley in 2015.

Ben D. Sawyer is an assistant professor at the University of Central Florida in the Department of Industrial Engineering and Management Systems. He received his PhD in human factors psychology from the University of Central Florida in 2015.

Ruth Rosenholtz is a principal research scientist at the Massachusetts Institute of Technology in the Department of Brain and Cognitive Sciences. She received her PhD in electrical engineering and computer science from the University of California at Berkeley in 1994.

Date received: March 22, 2019

Date accepted: June 9, 2020