Rapid Holistic Perception and Evasion of Road Hazards

Benjamin Wolfe[1], Bobbie Seppelt[2], Bruce Mehler[2], Bryan Reimer[2], Ruth Rosenholtz[1,3]

[1] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology
[2] AgeLab, Massachusetts Institute of Technology
[3] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

Corresponding Author: Benjamin Wolfe, bwolfe@mit.edu

Please cite this paper as

Wolfe, B., Seppelt, B. D., Mehler, B., Reimer, B., & Rosenholtz, R. (2019). Rapid Holistic Perception and Evasion of Road Hazards. Journal of Experimental Psychology: General. DOI:10.1037/xge0000665

Abstract: How quickly can a driver perceive a critical hazard on or near the road? Evidence from vision research suggests that static scene perception is fast and holistic, but does this apply in dynamic road environments? Understanding how quickly drivers can perceive hazards in moving scenes is essential because it improves driver safety now, and will enable autonomous vehicles to work safely with drivers in the future. This paper describes a new, publicly-available set of videos, the Road Hazard Stimuli, and a study assessing how quickly participants in the laboratory can detect and correctly respond to briefly presented hazards in them. We performed this laboratory experiment with a group of younger (20-25 years) and older (55-69 years) drivers, and found that while both groups only required brief views of the scene, older drivers required significantly longer to both detect (220 ms, younger; 403 ms, older) and correctly respond to hazards (388 ms younger; 605 ms older). Our results indicate that participants can perceive the scene and detect hazards holistically, without serially searching the scene, and can understand the scene and hazard sufficiently well to respond adequately with only slightly longer viewing durations.

Introduction

All drivers require visual information about the environment around them in order to drive safely (Schieber, Schlorholtz, & McCall, 2008; Sivak, 1996; Spence & Ho, 2008). For example, detecting road hazards, such as a moose walking into the road, requires vision. While drivers of traditional, manually-controlled vehicles must utilize this information to drive safely, their needs shift in autonomous vehicles. In particular, autonomous vehicles may request or require the driver to take over manual control of the vehicle (Gold, Damböck, Lorenz, & Bengler, 2013; Mackenzie & Harris, 2015; Samuel & Fisher, 2015; Samuel, Borowsky, Zilberstein, & Fisher, 2016) and these requests may either be planned takeovers (e.g., an approaching exit) or unexpected takeovers requiring a near-instant response. If autonomous vehicles are to be safe additions to the road, we must understand how quickly drivers can perceive their environment, and in particular, how quickly they can perceive and correctly respond to hazards. However, previous work on hazard perception has focused on drivers' need to *search* for hazards (c.f., (Crundall, 2016)). Given that search in road scenes is often thought to be a serial process (Underwood, Crundall, & Chapman, 2002), is it always necessary for hazard perception, or is some hazard perception holistic (Benda & Hoyos, 1983), and can drivers do it in a single glance? If a moose walks out of the woods towards the road, do you need to search for it, or do you notice it as soon as it exits the trees?

In planned takeovers, properly designed systems will ensure drivers have the time necessary to be fully aware of the roadway and the larger operating context prior to assuming control. However, in unanticipated handoff or takeover situations, it is likely that the driver will need to respond to an imminent hazard in the roadway. Note that there are some differences between the two; in an unanticipated handoff, the driver is forced to take control by the vehicle, and in a takeover situation, the driver chooses to of their own volition. However, both of these situations require the driver to rapidly understand their environment in order to take control. While drivers' ability to perceive hazards has been the focus of a considerable body of research (Brown & Groeger, 1988; McKenna & Crick, 1994; Pelz & Krupat, 1974), it is often understood that drivers must search for hazards in order to perceive them (Crundall & Underwood, 1998; Underwood, 2007; Underwood et al., 2002). In particular, much of this work operationalizes hazard detection by requiring the driver to look directly at the hazard, rather than asking if they can detect a change the driver believes to be hazardous. Such an operationalization lends itself to search-based explanations, because in order to look at the hazard, the driver must determine where the hazard is in the scene. More broadly, this view of hazard detection is often linked to results showing that expert drivers' eye movements cover more of the scene than novice drivers (Mourant & Rockwell, 1972). In turn, this is thought to reflect expert drivers searching for hazards and, implicitly, their need to attend to hazards to perceive them (Ranney, 1994). As a consequence, expert drivers, because they scan the scene more broadly than novice drivers, are likely to be better at detecting emerging hazards, such as moose walking into the road.

This view of driver information acquisition is often framed in terms of visual attention, particularly Treisman's Feature Integration Theory (Treisman & Gelade, 1980), which has been interpreted by human factors research as requiring the driver to attend to an object to have any awareness of it. If so, you would be unaware of the moose unless you attended to it, implicitly limiting awareness in and around the road. If attention is required for awareness, how, then, can you attend to something you are not aware of, and how impoverished is our awareness? One potential solution that researchers have suggested is for the moose to capture attention (Theeuwes, 1994). In Feature Integration Theory, basic features are available for basic processing without attention; the moose might move in a way that deviates from surrounding motion, leading it to "pop out", drawing attention, which allows it to be recognized as a moose and as a threat (Royden, Wolfe, & Klempen, 2001). Alternatively, the moose might capture attention precisely because it poses a threat, analogous to speeded detection of threatening objects (Blanchette, 2006; Subra, Muller, Fourgassie, Chauvin, & Alexopoulos, 2017; Williams, Palmer, Liddell, Le Song, & Gordon, 2006), which suggests that considerable processing of unattended stimuli might occur without attention or awareness. If, on the other hand, you know to be on the lookout for moose, and guide your attention (J. M. Wolfe, 1994) to detect them when they appear, this likely increases your awareness of wandering moose. While such a process may be beneficial if you know you need to search for wandering moose, the ability to guide attention is not unlimited (J. M. Wolfe & Horowitz, 2017), and the potential class of road hazards includes many possibilities beyond wandering moose. Overall, this notion that the driver needs to attend as a precondition for awareness leaves open the following question: does the driver need to search the scene to find hazards, become aware of them, and act, or can they detect hazards holistically, at a glance, and use this information to plan responses?

Within basic research on scene perception, there is strong evidence that visual perception does not rely solely on serial deployments of attention. Rather, research on getting the gist of a scene (Navon, 1977; Oliva & Torralba, 2006), the information available in a single glance, suggest that scenes are perceived holistically, with attention required only as necessary to refine details. This is in contrast to the view that one needs to attend to each object in the scene in order to perceive them (Treisman & Gelade, 1980), and then builds a scene piecewise from the attended objects (Mourant & Rockwell, 1972; Theeuwes, 1994). Results on scene gist prompted Treisman to rework Feature Integration Theory to include distributed attention (Treisman, 2006), which gathers information in parallel across the visual field. In this version of the theory, distributed attention enables perception of scene gist, which is then augmented by foveal attention driven in part by that gist information. Critically, the information extracted from the scene in 75 ms (Greene & Oliva, 2009a), the gist of the scene, is sufficient to classify the kind of scene. While classifying scenes as city or highway is fast, determining scene navigability (whether the path or road that is shown can be traversed) takes little additional time and can be accomplished with a viewing duration of 100 ms (Greene & Oliva, 2009b). Furthermore, there is significant evidence that gist perception can include the extraction of an abnormality signal; for example, the knowledge that something is wrong. Radiologists can correctly classify mammograms as containing abnormalities with less than 500 ms of viewing time (Evans, Georgian-Smith, Tambouret, Birdwell, & Wolfe, 2013; Evans, Haygood, Cooper, Culpan, & Wolfe, 2016), and there exists some evidence for similar abilities in hazard perception (Benda & Hoyos, 1983; Crundall, 2016). Critically, for studying hazard detection in driving, radiologists detect abnormality even though they often do not know the location of the lesion (Evans et al., 2013; 2016). If radiologists can detect abnormalities even if they cannot localize them, it suggests to us that drivers might be able to use a similar process for hazard detection, in contrast to theories which require them to localize hazards before they can be noticed. More broadly, these results in visual search suggests that we should be wary of laboratory tasks that require participants to name, identify, fixate, or otherwise localize a road hazard, since the driver might be sufficiently aware of it to respond, even if their localization is imprecise. Together, these results suggest that the visual system can very quickly extract information from across the visual field, that awareness is not limited to the current focus of attention, and that a driver might be able to detect an approaching moose very quickly indeed.

However, prior research on perceiving the gist of an image has exclusively used static images. Work on hazard perception has used video stimuli, but it has focused on the driver's need to search for hazards, operationalized as the participant looking directly at the hazard as an assumed precondition for awareness (Alberti, Shahar, & Crundall, 2014; Crundall, 2016; Crundall & Underwood, 1998; Crundall et al., 2012). However, Benda and Hoyos used static images in their hazard perception task, and found that drivers had little difficulty in immediately classifying and sorting static images by whether they contained a hazardous situation or not, suggesting a more holistic process (Benda & Hoyos, 1983); similar results have been reported recently by Huestegge and Bokler (2016). Other research in this area (Alberti et al., 2014) has used simulated environments, which may not represent the road environment accurately, and, as a consequence, behavioral responses to these simulated environments may not be representative of real-world behavior (Spence & Ho, 2015). In contrast, we ask how quickly drivers can detect and respond to hazards in moving scenes, without making them search for, fixate, and identify those hazards. Our approach bears some resemblance to the Hazard Perception Task developed by McKenna and Crick, but with two critical differences: first, they used a continuous-response hazard measure (drawing on the work of (Pelz & Krupat, 1974) and second, they focused on distinguishing between expert and novice drivers through response latency relative to events in the scene (McKenna & Crick, 1994). While this work is revealing, it illuminates the relative hazard detection criteria used by expert and novice drivers, rather than determining how long they would require to perceive and understand the road scene.

To facilitate this work, we developed a set of videos from real-world hazardous situations (the Road Hazard Stimuli, detailed in *Methods*). Participants view brief video clips, and either assess whether they holistically perceived a hazard or, in separate trials, what action they would take to evade that hazard. Given the evidence from classification of radiological images that abnormality information is available very quickly (with stimulus durations of 500 ms or less (Evans et al., 2016)), we posited that participants with driving experience might be able to extract a similar signal from videos of hazardous situations. Our aim was to probe the speed with which holistic perception and understanding of hazards in brief videos take place in a driving context, and more broadly to probe human ability to extract essential information from video of a dynamic real-world scene.

Materials and Methods

Participants

A total of 49 participants between the ages of 20 and 69 years old with one year or more of driving experience were recruited for this study from the MIT AgeLab's participant recruitment pool. Ten participants were excluded from the final analysis: three were excluded due to equipment failures during data collection, and an additional seven were excluded due to an inability to fit their data from one or more of the experimental conditions to a psychometric function (see *Analysis*). Six of the seven participants excluded for this reason generated data in the detection condition that could not be fit, meaning that their starting threshold in the hazard evasion condition did not reflect their individual performance on the hazard detection task, but rather used the default value (see *Task Conditions* for details of these tasks). As a consequence, their data in the hazard evasion task would not have been comparable to that of participants whose data could be fit on the hazard detection task, and their data was removed. All participants had normal or corrected-to-normal acuity, as assessed using the Federal Aviation Administration's test for near acuity (Form 8500-1), and the Snellen Eye Chart for distance acuity. Given the aging driving population and the ways in which visual perception changes with age (Owsley, 2011), we made a point of recruiting both older and younger drivers for this study. In pilot data, we estimated a within-subject main effect of video task across three levels (the cue-locked detection, response-locked detection, and evasion tasks, respectively) of Cohen's $f = 0.79$, corresponding to an approximately 165 ms average difference in thresholds across all pairs of conditions. Power calculations indicated that a minimum of 8 observers was required to detect this main effect at 95% power. All data reported were from a final set of 39 participants, with the younger participants ranging from 20-35 years old (19 total; 8 women and 11 men; mean age, 25.7 years, SD, 3.71 years) and the older participants ranging from 55-69 years old (20 total; 10 women and 10 men, mean age, 63.7 years, SD, 3.86 years). All participants provided written informed consent prior to participation as required by MIT's Committee on the Use of Humans as Experimental Subjects, in accordance with the Common Rule (45 CFR part 46) and were compensated $40 for their time.

Apparatus

Stimuli were presented using Matlab (Mathworks, Natick, MA) and Psychtoolbox-3 (Brainard, 1997; Pelli, 1997) on a 46” Sony Bravia HDTV (102 cm × 57 cm panel size; 1920 × 1080 pixel resolution and 60 Hz refresh rate) at an approximate viewing distance of 55-60 cm. The videos were shown on a gray background and subtended approximately 78^{o} horizontally and 44^{o} vertically at this viewing distance, where the road scene would subtend approximately the same visual angle it would for a driver. Head position was unconstrained, to approximate the driving experience, and the room was dimly illuminated, but not dark. To further increase realism and foster immersion (c.f. (Levy, Pashler, & Boer, 2006)), participants were provided with a wheel and pedal set (Apex Racing Wheel, connected to the stimulus computer over USB, and reporting as a gamepad device within Psychtoolbox), and used the pedals or wheel to make their responses (see *Procedure*).

Stimuli

The Road Hazard Stimuli set (available via the Open Science Framework at https://osf.io/uq6pc/) developed for this study comprises 503 8-second egocentric dashcam videos. The set includes 253 hazardous situation videos, which contain the events leading to a collision or near-collision event, broadly construed, and 250 non-hazardous situation videos. Videos were sourced from YouTube (in collaboration with the Moments project at MIT; (Monfort et al., 2019)) and were individually selected to avoid excessive in-frame text, hazard highlighting (e.g. added text or symbols to point out the hazard), and changes in frame rate. Videos were selected to include a wide variety of road environments (e.g., city streets, highway environments, rural roads), weather conditions, and forward-approach hazards. Critically, all hazards are visible from the camera position looking at the road ahead, although camera viewpoint varies from video to video. The primary goal in selecting videos for inclusion was to maximize the variability of hazards represented (e.g., uncontrolled objects, pedestrians, uncontrolled vehicles, loss of vehicle control), with the secondary consideration of varying the road environments and other conditions visible in the video. After downloading, videos were cropped to 8000 ms in duration for hazard and non-hazard videos, and the audio was removed for all videos. To control for environmental factors, when possible we extracted non-hazardous videos from epochs in the hazardous source video at least ten seconds prior to the emergence of the hazard (178 of 250 videos in set). The remaining 72 non-hazardous situation videos in the set were taken from videos which did not contain a hazard used in the final stimulus set. Critical timepoints in the hazardous situation videos were annotated as described below.

The 253 hazardous situation videos in the Road Hazard Stimuli set were annotated by three annotators (one experimenter plus two additional annotators who were naïve as to the goals of the study but trained to annotate driving behavior); any differences in the double-annotated data were mediated by the same experimenter who annotated the videos. For this study, we annotated two necessary timepoints (see Figure 1a): (1) the timepoint where there is the first visible deviation of the hazardous object from its normal state, in other words, the object has deviated from a non-threatening trajectory, and (2) the first point at which the driver's response is visible in the dashcam footage. The time of first visible deviation is the first time that the hazardous object can be seen to be moving in a way that is a cause for concern (e.g., a car starting to veer into the driver's lane; see event video 37); prior to this point, there is no visible indication that the object requires any more attention by the driver than any other object in the scene. A hazard did not need to physically enter the driver's lane of travel to be coded as the first visible deviation. In some videos, this point of first visible deviation corresponds to the first time that the object becomes visible in the footage (e.g., a deer running into the road; see event video 20). The first moment of driver response is when the driver slowed (braking), or began to swerve to the left or right to evade the hazard, as visible in the video; video from after that point in time included both the hazard and the driver's response, and was never seen by participants. Annotators also provided, based on the footage between these two timepoints, what they believed to be the ideal evasion response, based on the information they had from the video. This non-temporal annotation was limited to braking, swerving left or swerving right; these annotations accounted for 81.3%, 10.3%, and 8.3% of the hazardous situation videos, respectively. All annotators viewed the hazardous situation videos independently, and the experimenter assessed and moderated all annotations once this was completed to generate a final set of annotated time points. The mediated annotations are available for research use as part of the Road Hazard Stimuli.

Procedure

We measured the minimum video duration necessary to perceive hazard-related information, varying the task (see *Task Conditions*) and the time points that the video clips started from (see *Temporal Conditions*). Regardless of condition, each trial followed the same sequence of events (Figure 2). Participants were first shown a white noise luminance mask covering the same area on the display as the video, with a green cross (2° x 2°; line width: 0.4°) centered on the mask. Given the large size of the display, we presented the cross to orient participants to the center of the videos, but gave them no specific instructions on where to fixate. This was displayed for 250 ms, immediately followed by the video for that trial. The video duration on a given trial was set according to a staircase, with separate staircases for each combination of task and temporal condition (see *Staircase Control*). The video was followed by a second white noise mask and cross for 250 ms, after which time participants were free to respond as dictated by the task condition (hazard detection or hazard evasion). Note that there was no delay between the offset of the video and the onset of the mask to limit the amount of time participants could extract information from the video to the true duration. Responses made during either the video duration or post-stimulus mask were not recorded; responses were only logged following the post-stimulus mask. Following the response, the experiment advanced to the next trial after a 500 ms blank inter-trial interval.

Task Conditions

Participants completed each of the two task conditions in separate blocks of trials. The first block consisted of the hazard detection task (Figure 2a), in which participants reported whether they perceived a hazard in the video shown. They were not told what form the hazard might take, only that it might appear, and could include vehicles attempting to enter their lane of travel, objects falling on the road, or animals entering the road, as examples. Hazards were present in 50% of the trials, and participants were instructed to respond as quickly and as accurately as possible, using the accelerator pedal to indicate they had not perceived a hazardous situation, and the brake to indicate that they had perceived a hazard. Hazards are far more prevalent in this experiment than on the road, which allowed us to determine participants' viewing duration thresholds with a minimum of trials. Our approach bears some similarities to the Hazard Perception Task developed by McKenna and Crick (McKenna & Crick, 1994), but rather than using a continuous measure of perceived hazard, we used a two alternative forced choice paradigm, and determined the threshold display duration necessary to discriminate between hazard and no-hazard at 80% correct. Prior to completing the hazard detection task, participants did 20 practice trials to familiarize themselves with the pedal response and the timing of the experiment. In the practice trials, video duration was fixed at 750 ms, and participants were given visual feedback on their performance (the text “Correct” or “Incorrect” in green or red, respectively). After the practice trials, participants completed 200 trials without feedback (100 each for the cue-locked and response-locked conditions; see *Temporal Conditions*) with breaks every 50 trials. This took approximately 15 minutes for participants to complete.

Next, participants completed the hazard evasion task (Figure 2b), in which they were asked whether they would swerve left or right to evade the hazard shown in each new video. Since the majority of hazards in the stimulus set (81.3%) were coded as requiring a braking response, participants’ responses would have been predominantly braking if we had allowed for both braking and steering responses in the evasion task, and braking may be an acceptable response even for a number of videos in which the ideal response would be to steer. To avoid participants simply hitting the brakes on every trial, rather than truly judging every hazard, we exclusively used stimuli coded as requiring a steering response in the hazard evasion task. Hazards varied significantly, as in the hazard detection task, but had been coded by annotators as requiring a steering maneuver to evade, rather than braking. This represented approximately 19% of the hazardous situation videos in the Road Hazard Stimuli set. In the hazard evasion task, participants were not permitted to use the foot pedals, and were only permitted to respond with the wheel, having been told that they did not have the option to brake. Hazard evasion trials had a 100% hazard prevalence, equally split between hazards which required a left or right swerve. Again, this does not reflect the prevalence of abrupt steering responses in actual driving, but is necessary to determine the perceptual thresholds that were the focus of this study. Videos were never repeated between the detection and the evasion tasks. Participants first completed six practice trials with visual feedback, followed by 36 experimental trials with no feedback. The hazard evasion task took participants approximately 5 minutes to complete.

Temporal Conditions

In the *cue-locked* condition (used only with the hazard detection task; Figure 1b, upper), participants were shown 200 ms of video from immediately prior to the first visible deviation timepoint, to provide the gist of the scene prior to any visual indication of the hazard. This context duration was chosen to exceed the threshold required for accurate perception of static scene gist (Greene & Oliva, 2009b), since our stimuli were video, not still images. This was followed by a variable duration of video from after that timepoint. For example, if, in a given hazardous situation video, the first visible deviation was at 2500 ms into the video, and the staircased duration for that trial was 300 ms, the participant would have been shown a segment from the hazardous situation video running from 2300 to 2800 ms (200 ms of context, followed by 300 ms of hazardous situation). Also, based on pilot testing, participants were never shown more than 1000 ms of video in any one trial in the cue-locked condition, and never saw video past the point of driver response. This was ensured by randomly selecting each video (without replacement) from the set of videos that had a cue-to-response duration greater than or equal to the staircased duration value for that trial.

In the *response-locked* condition (used with both the hazard detection and hazard evasion tasks; Figure 1b, lower), participants were shown video ending at the point at which the driver of the vehicle began to respond, and beginning sometime after the first visible deviation. If, for example, the driver in a hazardous situation video had begun to respond at 4300 ms into the video, and the staircased duration value for a given trial was 450 ms, the participant would have been shown video from 3850-4300 ms in the 8000 ms video. Participants were never shown more than 1000 ms of video in the response-locked condition, and never saw video past the point of driver response. No neutral context video could be provided, because the changes in the scene are already occurring (see Figure 1b).

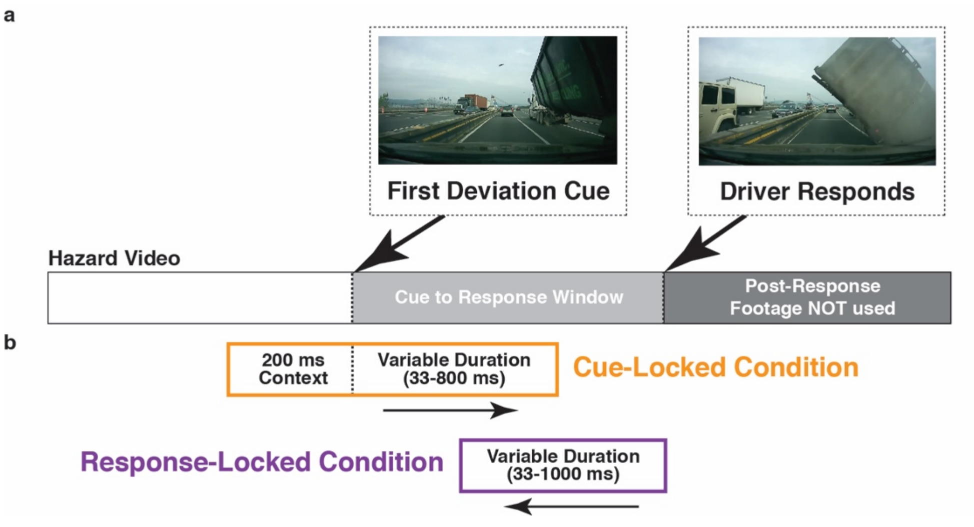

Figure 1: Illustration of Annotations and Temporal Conditions. (a) In the first panel ("First Deviation Cue") the trailer of the vehicle to the right has begun to tilt beyond the degree expected in normal driving, thus becoming the first visual indication of a potential hazard in the scene. In the second ("Driver Responds"), the trailer has tilted irrecoverably, and in the footage, the driver of the dashcam vehicle has begun to brake. The ideal action annotation exclusively used footage from between these timepoints for each video, and coded whether the annotator believed the hazard was best evaded by braking or turning to the left or right. (b) Visualizing where video for the cue-locked (on top, in orange) and the response-locked (bottom, in violet) was sourced from each hazardous situation video in the set, relative to annotated timepoints.

Staircase Control

For every trial in either the hazard detection or hazard evasion task, across cue-locked or response-locked conditions, stimulus duration was controlled by an independent three-down / one-up adaptive staircase, which held performance to approximately 80%. In the hazard detection task, there were two independent staircases, one for the cue-locked trials, and one for the response-locked trials, which were randomly interleaved. Both of these staircases started with the same initial stimulus duration (750 ms), but varied independently based on participant performance. Staircase step size was initially 167 ms, and decreased by 25% every 3 reversals, with a minimum possible value of 33 ms. In the hazard detection task, stimulus duration increased or decreased in response to incorrect or correct responses, respectively, using responses from all trials. In the hazard evasion task, the same staircase rule (three down, one up) was used, but the starting duration, rather than being fixed, was determined for each participant by taking the mean of all reversals in the response-locked hazard detection task, to start each participant at the duration they required to accurately detect hazards and reduce the total number of trials required in the evasion task. The hazard evasion task was only run with response-locked stimuli, because the correct response is only meaningful relative to the end of the stimulus window; in other words, a response that might be plausible with earlier information may prove to be a poor choice as the hazard evolves. Correct and incorrect responses for the hazard evasion task were determined relative to coding of the ideal response for the stimulus; for example, did the participant's response agree or disagree with the annotated ideal response.

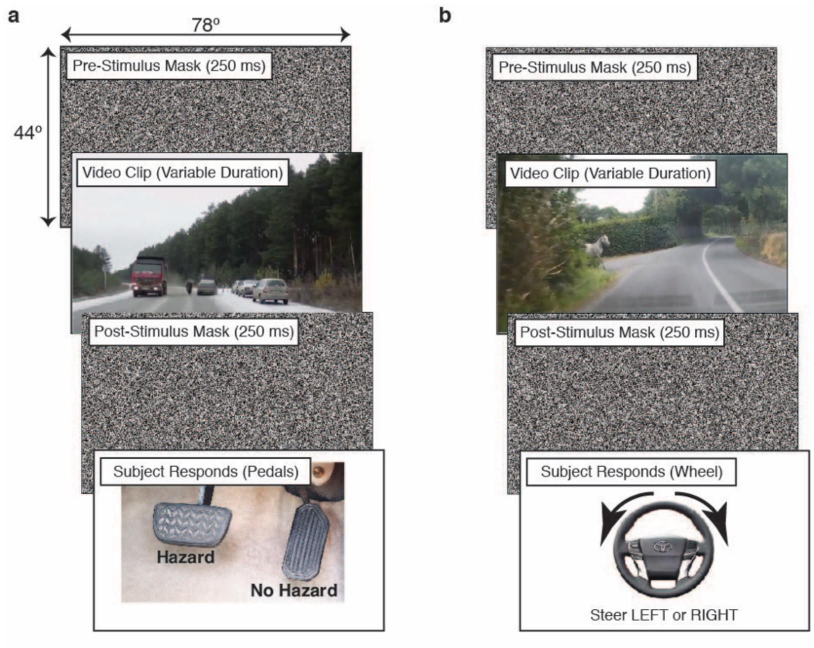

Figure 2: Stimulus sequence for (a) *hazard detection* and (b) *hazard evasion* tasks. Each trial began with a 250 ms pre-stimulus mask, followed by the video clip, followed by a 250 ms post-stimulus mask, followed by the participant's response (pedals in the hazard detection condition; wheel in the hazard evasion condition).

Analysis

Responses from each participant in each condition were fit to a two-parameter cumulative normal distribution (mean, μ, and standard deviation, σ) using maximum likelihood estimation in R (version 3.5.0), with chance and ceiling performance fixed at 50% and 100%, respectively. Seven participants had poor psychometric fits in at least one of the three conditions, with little or no relationship between stimulus duration and performance, and were removed from the analysis. These participants had a fitted linear slope of < .05 (i.e., an increase in accuracy of less than 5% per 1000 ms of video clip duration) and 80% thresholds outside the range of 0 - 1500 ms in at least one condition. Six of the seven participants whose data was excluded for this reason had poor fits in the hazard detection task, which meant they completed the hazard evasion task with the default starting value for the staircase, rather than one based on their performance in the detection task, and their data cannot be compared to other participants. The remaining participants' individual 80% performance thresholds were extracted from these fits and analyzed with a 3 (condition: detection task & cue-locked, detection task & response-locked, or evasion task & response-locked) x 2 (age) x 2 (gender) mixed-model ANOVA using the AFEX package

(0.20-a). Gender was included as a potential factor due to evidence for gender-based effects on driving tasks in older participants (Owsley & McGwin, 2010). Video condition was a within-participants factor, and age group and gender were between-participant factors. Values are reported using the Greenhouse-Geisser correction for sphericity. Reaction times were logged when the response was made (pedal depression for the hazard detection task, wheel turning for the hazard evasion task), and for reaction time analyses, reaction times in excess of 5000 ms were removed from the analysis (0.4% of trials removed) and were only calculated for correct trials. Because the difference in task between hazard detection (pedal response) and hazard evasion (wheel) precludes any comparison in reaction time across the two tasks, mean reaction times were analyzed with separate 2 (age) x 2 (gender) ANOVAs, one for the hazard detection task and one for the hazard evasion task.

Code Availability

All stimulus code, analysis code and anonymized data are available from Open Science Framework, at https://osf.io/cen28/.

Results

Fitted Thresholds

We observed a significant main effect of task between hazard detection (Figure 3) and hazard evasion (Figure 4), $F(1.82, 63.701) = 13.327$, $p < 0.0001$, $\eta_p^2 = 0.28$, with thresholds lower in the hazard detection task (younger participants, 220 ms, SD, 33 ms; older participants 403 ms, SD, 44 ms) compared to the hazard evasion task (younger participants, 388 ms, SD, 72 ms, older participants 605 ms, SD, 62 ms), indicating that longer viewing durations are needed for evasion than for detection. Using the Tukey method for pairwise comparisons, we observed significant differences between the cue-locked and response-locked conditions within the hazard detection task ($p < .0001$; see Figure 3), which is unsurprising since the cue-locked trials always had 200 ms of leading contextual video whereas the response-locked trials did not. We also observed a significant difference between the response-locked condition in the hazard detection task and the response-locked condition in the hazard evasion task ($p = 0.0005$). We also observed a significant main effect of age, $F(1,35) = 13.143$, $p = 0.0009$, $\eta_p^2 = 0.27$, with higher thresholds for older than younger participants. We did not observe a main effect of gender $F(1,35) = 0.31$, p

= 0.58, $\eta_p^2 = 0.008$. We observed no significant interactions between age and gender ($F(1,35) = 0.004$, $p = 0.95$, $\eta_p^2 = 0.0001$), age and task $F(1.82, 63.701) = 0.057$, $p = 0.93$, $\eta_p^2 = 0.001$ or gender and task $F(1.82, 63.701) = 1.38$, $p = 0.26$, $\eta_p^2 = 0.037$.

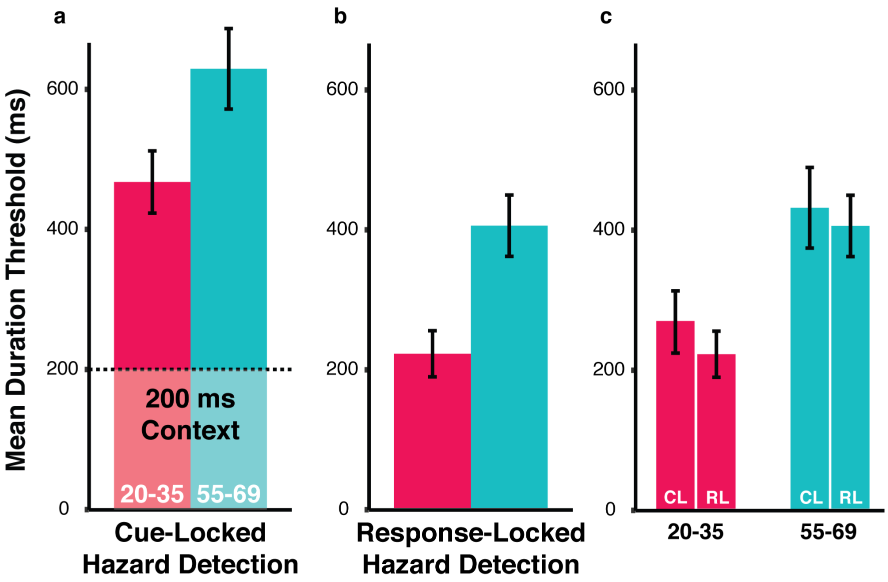

Figure 3: Hazard detection task, mean duration thresholds. (a) Mean thresholds for younger (crimson; left bars) and older (teal; right bars) participants in the cue-locked hazard detection task, showing a significant difference in mean threshold by age. (b) Mean thresholds for younger and older participants in the response-locked hazard detection task, showing a significant difference in mean threshold by age. (c) Thresholds from (a and b), labeled by condition, (CL for cue-locked, RL for response-locked) with the 200 ms of context removed from the mean threshold from the cue-locked condition. Notice the lack of a significant difference between the two conditions, indicating the relatively uniformly informative nature of the information within the cue to response window. Error bars are standard error of the mean.

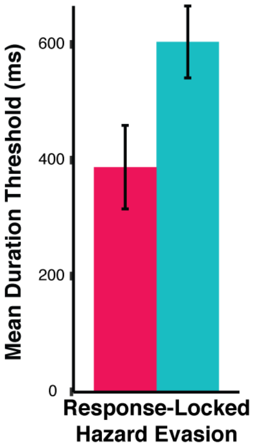

Figure 4: Hazard evasion task, mean duration thresholds. Mean thresholds for younger (crimson, left bar) and older (teal, right bar) participants are significantly higher than those in the hazard detection condition. Error bars are standard error of the mean.

Reaction Time

In the hazard detection task, we found no significant effect of age on correct reaction time relative to stimulus onset, $F(1,35) = 0.47$, $p = 0.50$, $\eta_p^2 = 0.013$ (reaction time for younger participants, 1210 ms, SD, 47 ms; for older participants, 1180 ms, SD, 50 ms; Figure 5a). We additionally found no significant effect of gender on reaction time in the hazard detection task, $F(1,35) = 0.15$, $p = 0.70$, $\eta_p^2 = 0.004$. In the hazard evasion task, we found the same pattern, with no effect of age on reaction time, $F(1,35) = 1.18$, $p = 0.28$, $\eta_p^2 = 0.032$ (reaction time for younger participants, 870 ms, SD, 84 ms; for older participants, 950 ms, SD, 50 ms; Figure 5b). We also saw no significant effect of gender on reaction time in the evasion task, $F(1,35) = 1.76$ $p = 0.19$, $\eta_p^2 = 0.05$.

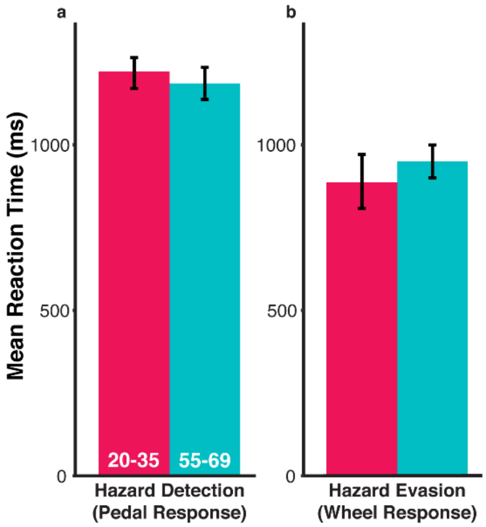

Figure 5: Mean reaction time for hazard detection (a) and hazard evasion (b), measured from stimulus onset. Mean reaction times for younger (crimson, left bar) and older (teal, right bar); mean reaction times are not significantly different within tasks. Note that response modalities changed between the hazard detection (a) and hazard evasion (b) tasks, and reaction times are, therefore, not comparable between tasks. Error bars are standard error of the mean.

Discussion

In the context of this experiment, drivers can detect hazards when presented with extremely brief video durations (220 ms for younger participants; 403 ms for older participants), which is not much longer than the display times necessary to perceive the gist of a static scene.

This suggests that a holistic process operates to detecting hazards in dynamic scenes, similar to previous results in holistic hazard detection (Benda & Hoyos, 1983; Huestegge & Böckler, 2016) and single-glance search of radiological images (Evans et al., 2016). Moreover, participants did not appear to benefit from prior contextual information in making this determination. This is in contrast to the hazard perception literature, which suggests that drivers need to understand the scene first before they could search it for likely hazards (Crundall, 2016). When we account for the contextual footage provided in the cue-locked condition (200 ms) we find no difference in the thresholds between the cue-locked and response-locked conditions. An alternate explanation for the results we observe would be that the greater informativeness of the response-locked video nearly-perfectly cancelled out the less-informative cue-locked videos with the addition of 200 ms of contextual video. While there may be subthreshold hazard cues before the annotated first deviation (a question we will address in future work), since these cues were not picked up in annotation suggests they are likely to be subtle and may be only minimally informative. Given this, it is likely that sufficient contextual information can be extracted simultaneously with the holistic hazard signal, rather than requiring prior context to notice the emergence of the hazard. In addition, mean reaction times (1180 - 1210 ms) in the hazard detection task are very similar to the mean brake reaction time reported by Green in a meta-analysis of on-road braking behavior in response to various events, who reports a mean brake reaction time of 1300 ms for unanticipated events (Green, 2000). This similarly suggests to us that drivers may respond on-road on a similar timeframe as we observed in the laboratory, although hazards on the road are far less prevalent than they were in this experiment.

Overall, our hazard detection task results and the brevity of participants' thresholds, indicate that drivers are able to accurately detect hazards without needing to search the scene. Notably, our results agree with prior work on holistic detection of road hazards (Benda & Hoyos, 1983), indicating that drivers can detect hazards without overtly searching for them. This is in contrast to accounts in hazard perception which assume that attention and overt shifts of gaze are preconditions for awareness of hazards (Alberti et al., 2014; Underwood et al., 2002; Underwood, Phelps, & Wright, 2005), and the idea that drivers must always search their environment (Mourant & Rockwell, 1972). It is important to note that while the fundamental capabilities of the human visual system can enable fast hazard detection in some circumstances, we in no way suggests that drivers do not need to scan the environment broadly to enable early hazard detection, since such expertise-driven scanning behavior can only benefit drivers' ability to detect hazards.

In the hazard evasion task, we find longer thresholds (388 ms for younger participants; 605 ms for older participants). Given the change in task, this is expected, because choosing to steer to the left or right encompasses a need both to accurately localize the hazard within the scene and to better understand the situation, for instance, the locations of other vehicles that might impact one's decision of which way to turn. This likely requires a more detailed understanding of the scene than simply detecting an abnormality signal (and is likely to be aided by one or more eye movements). In essence, being able to make a correct steering response to evade the hazard in a given video requires the participant to not only know that the hazard is present (as in the hazard detection task), but also to have some knowledge of where the hazard was, and where they might be able to steer to avoid it, relative to other objects and hazards in the scene. For example, if a moose is walking into the road from the right, and a vehicle is in the opposite lane, the driver must swerve to the right-hand shoulder to evade both the moose *and* the other vehicle. Of course, on the road a driver would have more information than what participants were provided in our study, both from their side mirrors and from multimodal sources, which might facilitate detecting such a hazard. The pattern in our results, with increased stimulus duration thresholds when the participants is acquiring information to plan a steering response rather than a detection response, is similar to results on takeovers in simulated driving (Gold et al., 2013). They found that drivers were faster to brake in abrupt handoff situations, and that when a handoff was initiated, they were slower to initiate a steering maneuver, suggesting a need for more information about the scene before they were comfortable doing so.

Critically, we found no interaction with age, suggesting that the shift from simple detection to gaining sufficient understanding to evade the hazard brings with it a relatively stable increase in duration thresholds of approximately 200 ms. The increased thresholds in the evasion task, as compared to the hazard detection task, suggest a critical difference between recognizing that a hazard is present in the environment, and having sufficient information to be able to act on that knowledge. In comparison, we find no difference in reaction time between our older and younger participants in either the detection or evasion tasks, a finding which may be attributable to older drivers' greater on-road experience, although our experimental design emphasized stimulus duration at the expense of reaction time measurement. Overall, however, perceptual thresholds on-road may exceed estimates from our experiment, as our participants were maximally attentive, and hazards are far more prevalent in our experiment than they are on the road, although this may be attenuated by the driver's multimodal sources of information (Spence & Ho, 2008).

Our work builds on a significant body of research with static real-world scenes, which has shown that participants can perceive the gist of a scene with brief presentations (Greene & Oliva, 2009b; Oliva & Torralba, 2006), suggesting that accounts of visual perception in driving which rely on serial attention to individual elements to comprehend a scene (Alberti et al., 2014; Crundall, 2016; Crundall, Underwood, & Chapman, 1999; 2002; Mourant & Rockwell, 1972; Underwood et al., 2002) may not adequately account for human capabilities. This work, to our knowledge, is the first to ask participants to rapidly perceive the events in a video of a real-world road scene, rather than providing a hazard embedded in a much longer video (Crundall, 2016). Unlike this previous work, however, our work focused on the stimulus duration our participants required to, respectively, detect and respond to imminent hazards, to determine how quickly drivers could acquire the necessary information for each task.

The implications of our results for our understanding driver behavior and capabilities are simple but profound: drivers can perceive aspects of their environments essentially at a glance, comprehending that hazards are present without needing to search them out, using the gist of the scene and detecting hazards holistically (Benda & Hoyos, 1983). This holistic detection of moving hazards is conceptually similar to radiologists' ability to holistically detect cancerous aberrations in briefly presented radiological images (Evans et al., 2013; 2016). In essence, drivers are likely detecting hazard cues that do not match the rest of the scene, which may often be atypical motion (e.g., the moose walking into the road on an orthogonal vector to the vehicle) or a deviation as comparatively subtle as another vehicle veering into one's lane. Detecting these deviations from the larger environment is, seemingly, sufficient to allow drivers to detect hazards, although the speed of the processes we observe suggest that drivers' representation of their environments will be imperfectly detailed. Far from this being a problem, it is likely a benefit because a driver will rarely need to know exactly what a hazard is, but knowing where it is and how it is moving is essential. However, this does not mean that drivers do not make eye movements or search for information that they need, merely that more information is available to them more quickly than accounts in driving research might suggest. Our results pose a significant challenge to accounts of driver behavior which assume the driver must actively search across the visual scene to be able to perceive that something is “wrong” or hazardous, and require the driver to attend to the hazard before they can be aware of it, much less respond to it.

While acquiring sufficient information for evasion requires a longer view of the road, our results also indicate that this additional time is far less than might be supposed based on work on non-emergency handoffs (Samuel et al., 2016), which suggests that drivers will require several seconds to view the road prior to reassuming control. The speed with which our participants can understand the scenes they were shown may be accounted for by their level of attention to their task, and drivers are known to respond to emergency situations similarly quickly (c.f., (Lee, McGehee, Brown, & Reyes, 2002). However, while the thresholds we report are brief, the window in our stimuli between the annotated deviation cue and the first visible response are also brief (1200 ms on average; similar to the duration reported by (Green, 2000) in a meta-analysis of brake reaction time to unanticipated hazards), suggesting that drivers can notice, understand and respond to a hazard on this timescale. Critically, the driver only has a limited window in which to acquire the information they need, understand that information and respond to the perceived hazard. Given this, the driver may anticipate where potential hazards may occur (as suggested by results which show expert drivers have different patterns of eye movements than novice drivers, and that these patterns correspond to changes the driver may need to know about in the scene (Alberti et al., 2014; Crundall, 2016)), and this knowledge of where to look will certainly aid them in perceiving the scene (to say nothing of sources of information beyond their view of the road ahead). However, this process must take place very quickly to begin developing the degree of awareness necessary to respond to changes in the environment in traditional driving, because there is no time for a slow process on the road.

However, our results should be considered in their context; that is, the fact that we used a pair of laboratory-based tasks with hazard prevalences that exceed those of any conceivable road environment. Our participants were maximally attentive and undistracted, and fully expected to be shown a variety of hazardous situations, even if they had no specific knowledge of what the hazards might be or where they might appear in any given scene. More critically, the effect of prevalence on search performance is well-known (J. M. Wolfe, Horowitz, & Kenner, 2005), and one might expect our participants to have missed hazards more frequently had they been rare. However, prevalence effects in the laboratory can be far weaker in more critical tasks, such as asking radiologists to look for abnormalities in medical images (Gallas et al., 2019; Gur et al., 2003). For that matter, expertise is a significant factor in hazard detection (Underwood, Ngai, & Underwood, 2013), which may aid hazard detection and evasion planning. Furthermore, drivers have multisensory information to draw upon (Spence & Ho, 2008) and do not have to rely on merely a single view of the road presented for a few hundred milliseconds to detect hazards. That said, the agreement we see between the reaction time reported by Green and our own suggests that, these caveats aside, we have been able to probe the perceptual process which underlies hazard detection. Future work will need to investigate whether our results hold up under low-prevalence conditions, and whether they do translate to actual driver behavior.

Thinking towards how they might translate to the road, a case in which the speed of detection and response is particularly critical is in the case of unanticipated takeovers in autonomous vehicles. In these cases, it may not be feasible to give the driver enough time to fully perceive that a hazard is present; drivers must, if at all possible, be given enough time to localize the hazard and, if at all possible, to act accordingly. The takeover problem is compounded by the difference in thresholds we observed as a function of age; while we observed no difference in reaction time, older drivers required longer to integrate information (Owsley, 2011; Owsley & McGwin, 2010), as shown by the longer thresholds we observed. As a result, it is unlikely that they would be able to perceive environments as quickly as younger drivers, a fact which should be accounted for when timing handoff events of all types in autonomous vehicles and that urges a cautious approach to developing this technology. Our results show that hazards in real-world road scenes can, under certain conditions, be perceived and acted upon quickly, suggesting that drivers can acquire some of the information necessary for these tasks using the gist of the scene. This requires input from across the visual field (B. Wolfe, Dobres, Rosenholtz, & Reimer, 2017). Models of driver behavior should account for this ability, and for the speed with which the visual system acquires information, although it is also necessary to consider the disruption to the driver's state and representation of the world in unanticipated handoffs. Future applied research on this question may need to consider what kinds of takeover events exist on the road, and the implications for our results on drivers' ability to reassert control across the lifespan. In scene perception more generally, future work should further characterize what information humans can acquire on this brief timescale, and how that information, together with the task, directs eye movements to gather further information from the scene.

Acknowledgements

The authors wish to acknowledge funding support to BW and RR from the CSAIL-TRI Joint Fellowship, and support to the project overall from the AHEAD Research Consortium and the MIT AgeLab. The authors also particularly wish to thank Anna Kosovicheva for many helpful discussions. Appreciation is also extended to Alea Mehler, Anthony Pettinato, Ryn Flaherty and Luca Russo at MIT AgeLab for their help with data collection.

References

Alberti, C. F., Shahar, A., & Crundall, D. (2014). Are experienced drivers more likely than novice drivers to benefit from driving simulations with a wide field of view? *Transportation Research Part F: Traffic Psychology and Behaviour*, *27(A)*, 124–132. http://doi.org/10.1016/j.trf.2014.09.011

Benda, von, H., & Hoyos, C. G. (1983). Estimating hazards in traffic situations. *Accident; Analysis and Prevention*, *15*(1), 1–9. http://doi.org/10.1016/0001-4575(83)90002-7

Blanchette, I. (2006). Snakes, spiders, guns, and syringes: How specific are evolutionary constraints on the detection of threatening stimuli? *Quarterly Journal of Experimental Psychology*, *59*(8), 1484–1504. http://doi.org/10.1080/02724980543000204

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436.

Brown, I. D., & Groeger, J. A. (1988). Risk perception and decision taking during the transition between novice and experienced driver status. *Ergonomics*, *31*(4), 585–597. http://doi.org/10.1080/00140138808966701

Crundall, D. (2016). Hazard prediction discriminates between novice and experienced drivers. *Accident; Analysis and Prevention*, *86*, 47–58. http://doi.org/10.1016/j.aap.2015.10.006

Crundall, D. E., & Underwood, G. (1998). Effects of experience and processing demands on visual information acquisition in drivers. *Ergonomics*, *41*(4), 448–458. http://doi.org/10.1080/001401398186937

Crundall, D., Chapman, P., Trawley, S., Collins, L., Van Loon, E., Ben Andrews, & Underwood, G. (2012). Some hazards are more attractive than others: Drivers of varying experience respond differently to different types of hazard. *Accident; Analysis and Prevention*, *45*, 600–609. http://doi.org/10.1016/j.aap.2011.09.049

Crundall, D., Underwood, G., & Chapman, P. (1999). Driving experience and the functional field of view. *Perception*, *28*(9), 1075–1087.

Crundall, D., Underwood, G., & Chapman, P. (2002). Attending to the peripheral world while driving. *Applied Cognitive Psychology*, *16*(4), 459–475. http://doi.org/10.1002/acp.806

Evans, K. K., Georgian-Smith, D., Tambouret, R., Birdwell, R. L., & Wolfe, J. M. (2013). The gist of the abnormal: Above-chance medical decision making in the blink of an eye.

Psychonomic Bulletin and Review, *20*(6), 1170–1175. http://doi.org/10.3758/s13423-013-0459-3

Evans, K. K., Haygood, T. M., Cooper, J., Culpan, A.-M., & Wolfe, J. M. (2016). A half-second glimpse often lets radiologists identify breast cancer cases even when viewing the mammogram of the opposite breast. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(37), 10292–10297. http://doi.org/10.1073/pnas.1606187113

Gallas, B. D., Chen, W., Cole, E., Ochs, R., Petrick, N., Pisano, E. D., et al. (2019). Impact of prevalence and case distribution in lab-based diagnostic imaging studies. *Journal of Medical Imaging*, *6*(1), 1–10. http://doi.org/10.1117/1.JMI.6.1.015501

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1938–1942. http://doi.org/10.1177/1541931213571433

Green, M. (2000). "How long does it take to stop?" Methodological analysis of driver perception-brake times. *Transportation Human Factors*, *2*(3), 195–216. http://doi.org/10.1207/sthf0203_1

Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cognitive Psychology*, *58*(2), 137–176. http://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*(4), 464–472. http://doi.org/10.1111/j.1467-9280.2009.02316.x

Gur, D., Rockette, H. E., Armfield, D. R., Blachar, A., Bogan, J. K., Brancatelli, G., et al. (2003). Prevalence effect in a laboratory environment. *Radiology*, *228*(1), 10–14. http://doi.org/10.1148/radiol.2281020709

Huestegge, L., & Böckler, A. (2016). Out of the corner of the driver's eye: Peripheral processing of hazards in static traffic scenes. *Journal of Vision*, *16*(2), 11–15. http://doi.org/10.1167/16.2.11

Lee, J. D., McGehee, D. V., Brown, T. L., & Reyes, M. L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors*, *44*(2), 314–334. http://doi.org/10.1518/0018720024497844

Levy, J., Pashler, H., & Boer, E. (2006). Central interference in driving: Is there any stopping the psychological refractory period? *Psychological Science*, *17*(3), 228–235. http://doi.org/10.2307/40064523

Mackenzie, A. K., & Harris, J. M. (2015). Eye movements and hazard perception in active and passive driving. *Visual Cognition*, *23*(6), 736–757. http://doi.org/10.1080/13506285.2015.1079583

McKenna, F. P., & Crick, J. L. (1994). *Hazard Perception in Drivers: A Methodology for Testing and Training* (No. 313). *TRL Contractor Report*. Crowthorne, Berkshire.

Monfort, M., Andonian, A., Zhou, B., Ramakrishnan, K., Bargal, S. A., Yan, Y., et al. (2019). Moments in Time Dataset: one million videos for event understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1. http://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*, *14*(4), 325–335.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*(3), 353–383.

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. http://doi.org/10.1016/S0079-6123(06)55002-2

Owsley, C. (2011). Aging and vision. *Vision Research*, *51*(13), 1610–1622. http://doi.org/10.1016/j.visres.2010.10.020

Owsley, C., & McGwin, G. (2010). Vision and driving. *Vision Research*, *50*(23), 2348–2361. http://doi.org/10.1016/j.visres.2010.05.021

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pelz, D. C., & Krupat, E. (1974). Caution profile and driving record of undergraduate males. *Accident; Analysis and Prevention*, *6*(1), 45–58. http://doi.org/10.1016/0001-4575(74)90015-3

Ranney, T. A. (1994). Models of driving behavior: A review of their evolution. *Accident; Analysis and Prevention*, *26*(6), 733–750. http://doi.org/10.1016/0001-4575(94)90051-5

Royden, C. S., Wolfe, J. M., & Klempen, N. (2001). Visual search asymmetries in motion and optic flow fields. *Perception and Psychophysics*, *63*(3), 436–444. http://doi.org/10.3758/BF03194410

Samuel, S., & Fisher, D. L. (2015). Evaluation of the minimum forward roadway glance duration. *Transportation Research Record: Journal of the Transportation Research Board*, *2518*, 9–17. http://doi.org/10.3141/2518-02

Samuel, S., Borowsky, A., Zilberstein, S., & Fisher, D. L. (2016). Minimum time to situation awareness in scenarios involving transfer of control from an automated driving suite. *Transportation Research Record: Journal of the Transportation Research Board*, *2602*, 115–120. http://doi.org/10.3141/2602-14

Schieber, F., Schlorholtz, B., & McCall, R. (2008). Visual requirements of vehicular guidance. In C. Castro (Ed.), *Human Factors of Visual and Cognitive Performance in Driving* (1st ed., pp. 31–50). Boca Raton, FL: Human Factors of Visual and Cognitive Performance in Driving. http://doi.org/10.1201/9781420055337.ch2

Sivak, M. (1996). The information that drivers use: Is it indeed 90% visual? *Perception*, *25*(9), 1081–1089. http://doi.org/10.1068/p251081

Spence, C., & Ho, C. (2008). Crossmodal information processing in driving. In C. Castro (Ed.), *Human Factors of Visual and Cognitive Performance in Driving* (pp. 187–200). Boca Raton, FL: CRC Press. http://doi.org/10.1201/9781420055337.ch10

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). Cambridge, MA: MIT Press.

Subra, B., Muller, D., Fourgassie, L., Chauvin, A., & Alexopoulos, T. (2017). Of guns and snakes: testing a modern threat superiority effect. *Cognition and Emotion*, *32*(1), 81–91. http://doi.org/10.1080/02699931.2017.1284044

Theeuwes, J. (1994). Endogenous and exogenous control of visual selection. *Perception*, *23*(4), 429–440. http://doi.org/10.1068/p230429

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4-8), 411–443. http://doi.org/10.1080/13506280500195250

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. http://doi.org/10.1016/0010-0285(80)90005-5

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. *Ergonomics*, *50*(8), 1235–1249. http://doi.org/10.1080/00140130701318707

Underwood, G., Crundall, D., & Chapman, P. (2002). Selective searching while driving: the role of experience in hazard detection and general surveillance. *Ergonomics*, *45*(1), 1–12. http://doi.org/10.1080/00140130110110610

Underwood, G., Ngai, A., & Underwood, J. (2013). Driving experience and situation awareness in hazard detection. *Safety Science*, *56*, 29–35. http://doi.org/10.1016/j.ssci.2012.05.025

Underwood, G., Phelps, N., & Wright, C. (2005). Eye fixation scanpaths of younger and older drivers in a hazard perception task. *Ophthalmic and Physiological Optics*, *25*(4), 346–356. http://doi.org/10.1111/j.1475-1313.2005.00290.x

Williams, L. M., Palmer, D., Liddell, B. J., Le Song, & Gordon, E. (2006). The 'when' and 'where' of perceiving signals of threat versus non-threat. *NeuroImage*, *31*(1), 458–467. http://doi.org/10.1016/j.neuroimage.2005.12.009

Wolfe, B., Dobres, J., Rosenholtz, R., & Reimer, B. (2017). More than the useful field: considering peripheral vision in driving. *Applied Ergonomics*, *65*, 316–325. http://doi.org/10.1016/j.apergo.2017.07.009

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin and Review*, *1*(2), 202–238. http://doi.org/10.3758/BF03200774

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 0058. http://doi.org/10.1038/s41562-017-0058

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*(7041), 439–440. http://doi.org/10.1038/435439a