

# Evaluating Hazard Detection Effects of In-Vehicle Safety Automation

Daniel Quinn  
Clemson University  
Clemson, South Carolina  
dbquinn@clemson.edu

Cazembe Kennedy  
Clemson University  
Clemson, South Carolina  
cazembk@clemson.edu

Austin Guy  
Clemson University  
Clemson, South Carolina  
alguy@clemson.edu

Gary Vestal  
Clemson University  
Clemson, South Carolina  
gvestal@clemson.edu

## ABSTRACT

Myriad factors increase drivers' workload and decrease their ability to identify hazards. Drivers' hazard perception under high workload conditions can benefit from the assistance of automated hazard detection systems. However, automation is inherently imperfect and system failures biases (false-alarm-prone vs. miss-prone) may affect the performance benefits of these systems. The current study explored how systems with different failure biases affect drivers' detection and recognition of on-road hazards. We hypothesized that automation false alarms better support driver hazard recognition than automation misses. Our results found no significant differences in detection or recognition times between the failure bias conditions. Lessons learned and directions for future research are discussed.

## CCS CONCEPTS

### ADDITIONAL KEYWORDS AND PHRASES

Automation, Hazard Detection, Eye Tracking, Visual Attention, Workload

## 1 INTRODUCTION

Recent safety recommendations by the National Highway Safety Administration (NHTSA, 2016) highlight the importance of visual attention on avoiding bicycle-motorist collisions. For example, risk of collision may decrease if drivers yield to and pass on-road bicycles the same as they would other vehicles (NHTSA, 2016). Before a driver can react to the presence of a potential on-road hazard (e.g., bicyclist), he or she must visually detect the object, then recognize it as a hazard. However, myriad distractions can occur under typical driving conditions (i.e., navigation system use; Peters & Peters, 2001) that increase drivers' cognitive workload and inhibit their ability to recognize visual targets (Recarte & Nunes, 2003).

The effect of workload on attention allocation is evidenced by changes in a variety of eye behaviors. For example, high workload may result in increased blink frequency, saccade speeds (Savage, Potter, & Tatler, 2013), and pupil diameter (Tsai, Viirre, Strychacz, Chase & Jung, 2007). Importantly, workload increases can produce visual tunneling, which refers to a reduction of the overall size of the attentive visual field (Rentanen and Goldberg, 1999; Tsai, Viirre, Strychacz, Chase & Jung, 2007). Visual tunneling primarily affects the vertical axis

of the visual field, with additional reduction occurring on the horizontal axis (Rentanen and Goldberg, 1999, Tsai, Viirre, Strychacz, Chase & Jung, 2007; Savage, Potter, & Tatler, 2013). One study determined that workload can limit the overall visual field by as much as 14%, thereby potentially reducing drivers' abilities to recognize hazards in their peripheral vision (Rentanen and Goldberg, 1999). Gaze fixation is also related to successful hazard detection. Fixations on visual targets tend to be longer for those that are task relevant (e.g., potential hazards while driving; Velichkovsky, Rothert, Kopf, Dornhoefer, & Joos, 2002) and shorter when simultaneous tasks are attended (Tsai, Viirre, Strychacz, Chase & Jung, 2007). Therefore, decreasing the harmful impacts of cognitive workload by removing distractions is one way to improve drivers' ability to detect and recognize visual targets. However, a more practical solution may be to support drivers' hazard detection abilities through the use of target detection automation.

## 2 BACKGROUND

Automation can improve users' abilities to detect visual targets under conditions of high workload (e.g., Parasuraman & Riley, 1997). Automation used in vision enhancement systems (VESs) can help drivers identify hazards during low speed and low visibility conditions (Tsimhoni & Green, 2002). One type of VES projects infrared or thermal imagery of the roadway to a secondary display (e.g., a real-time infrared video image of the roadway projected on a heads-up display). However, these types of VESs replicate the full roadway and divide drivers' visual attention between the road and the display. Therefore, they are unsafe at high speed and increase driver workload (Rumar, 2002). Other VES displays are highly specific and display only task-relevant information in the visual field. For example, Caird, Horrey, and Edwards (2001) determined that automation which detects high-value targets (e.g., bicyclists) can effectively enhance drivers' target recognition and response timing. However, automation is inherently imperfect, and it is important to consider how system failures affect the drivers' perceptions and behaviors.

People are less likely to trust and use automation that is unreliable than that which is reliable (Lee & See, 2004). When systems are unreliable, two types of automation failures, false alarms and misses, affect how drivers rely on systems and comply with system alarms (Wickens & Dixon, 2007). These failures result from system sensitivity to signals in the environment (signal detection theory; see Nevin, 1969) Highly sensitive systems are more likely to alert drivers to hazards that are not present (e.g. false-alarm-prone). Minimally

sensitive systems are more likely to fail to detect hazards that are present (e.g., miss-prone). Robust literature on these failure types have demonstrated that false alarms tend to primarily affect compliance and misses tend to primarily affect reliance (Dixon, Wickens, McCarley, 2007). However, some studies have found the contrary. For example, Schwarz and Fastenmeier (2017) found increased compliance rates in false-alarm-prone automation. Their study utilized a simulated driving task wherein participants were exposed to automation hazard detector false alarms. The authors determined that driver braking times improved regardless of the presence of the false alarms; this finding did not support their hypotheses. A possible explanation of their findings is task workload. Users are more likely to comply with automation when experiencing high workload (McBride, Rogers, & Fisk, 2011). Thus, high workload may dampen the effects of false alarms on compliance in hazard detection tasks. Driver workload may be an important determinant of the sensitivity threshold for automated hazard detection systems.

The current study investigated a high-workload task to determine whether false-alarm-prone systems more successfully support hazard perception than miss-prone systems. We hypothesized that hazard detection and recognition is better supported by automation false alarms than misses.

### 3 METHODOLOGY

#### 3.1 Participants

Twelve undergraduate students, ages 18-28 ( $M = 23.24$ ,  $SD = 2.8$ ), were recruited through in-class announcements. Participants volunteered their efforts for this study.

#### 3.2 Experimental Design

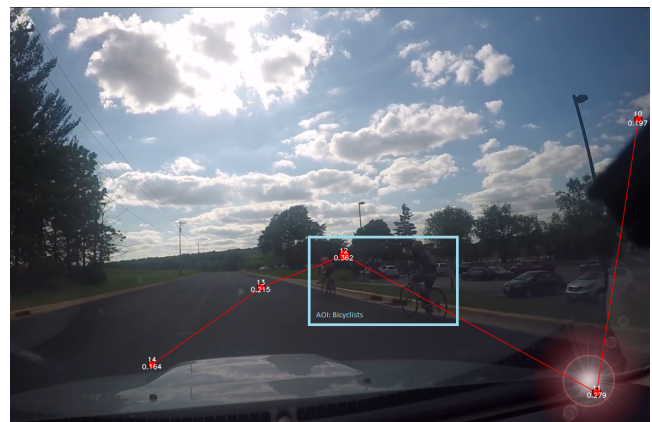
This study used a 2 (automation type: miss-prone, false-alarm-prone) x 2 (automation presence: present, not present) mixed-factors design. Automation type was the between-subjects factor. Therefore, participants were exposed to automation that is false-alarm-prone or miss-prone, but not both.

Participants experienced two experimental blocks composed of 30 target-detection trials (60 total). Each block contained either automation or no automation. A-B counterbalancing ensured that order presentation of the two blocks did not confound the effects of automation on task performance. Participants in the false-alarm-prone automation condition only experienced false alarm failures and participants in the miss-prone automation condition only experienced miss failures. Prior research identified that 70% automation reliability is the cutoff point at which individuals perceive automation as reliable (Wickens & Dixon, 2007); therefore, this study used automation that was 80% reliable (e.g., 6 failures during 30 trials) so that participants would be inclined to use the system. To increase participant workload and provide a mask between stimuli, a math problem was displayed at the center of a blank screen between each two-roadway image. An experimenter recorded participants' verbal responses to the questionnaires.

**Performance measures.** Target detection and recognition time were our primary variables of interest for this study. To

measure target detection time, areas of interest (AOI; see figure 1) were placed around each bicyclist in the roadway images and an eye tracker measured the average time to first gaze fixation on each AOI. To measure target recognition time, we instructed participants to left click the mouse only when they were confident a bicyclist was present in the roadway scenes. The average time-to-click when bicyclists were actually present in the scene was then used to measure target recognition time.

**Subjective measures.** Subjective trust in the automation, subjective reliance on the automation, and cognitive workload are known to affect human use of automation. Thus, these questions were measured after each block using a 1-7 Likert scale (1 = not at all, 7 = extremely). To measure trust and reliance, we used a questionnaire adapted from Lee and See (2004), and to measure workload, we used the NASA TLX (Hart & Staveland, 1988).



**Figure 1.** An example AOI positioned over bicyclists in a roadway scene.

#### 3.3 Materials

Researchers captured 60 screenshots from footage recorded by a high definition (1920 x 1080) video recorder. 30 of these images included bicyclists and 30 did not. An automated hazard warning light was simulated by presenting a red starburst in the bottom of each screenshot. Thus, when the simulated automation detected a bicyclist in the scene, the warning light was shown. Example stimuli are shown in figures 2 and 3.



**Figure 2.** An example false alarm (automation detects a bicyclist that is not present)



**Figure 3.** An example miss (automation fails to detect a bicyclist that is present)

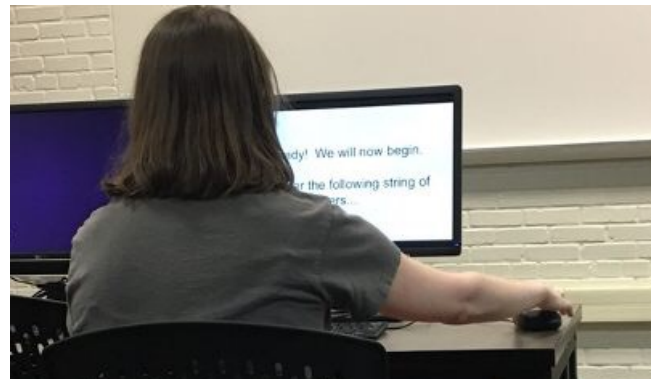
### 3.4 Apparatus

This experiment used a Gazepoint GP3 eye tracker to measure eye gaze. This system processed pupil/corneal reflection (PCR) with a sampling rate of 60 Hertz, an accuracy of 1-degree visual angle, and a Savitzky-Golay filter. Participants sat at a viewing distance 25.59 inches in front of a 22-inch, 1680 x 1050 resolution Dell LED backlit LCD display.

### 3.5 Procedure

First, researchers directed participants to their seat in front of a computer and eye tracker (see figure 4). Participants provided their informed consent and demographic information (e.g., age and gender). We calibrated the eye tracker to the participants' eye gaze and then provided the task instructions. We informed participants that the tasks may be difficult and that they should complete each task to the best of their ability. When participants were given the aid of automation, they received instructions that the system is highly reliable but imperfect. They then saw an example of the automation performing correctly and failing so that they would better be able to recognize errors during the task. Participants only witness automation failures for the condition to which they were assigned.

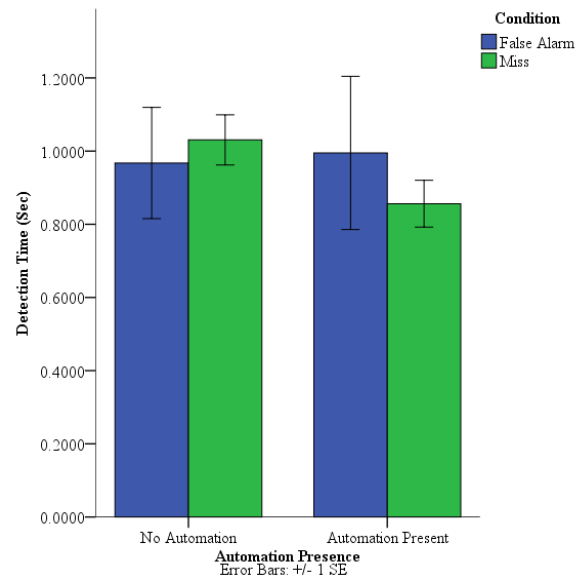
Participants completed several simple tasks during the experiment to increase their workload. Participants first memorized a string of five letters presented at the beginning of each block. Next, participants verbally answered simple arithmetic problems (e.g.  $5 + 3$ ,  $8 - 2$ ) presented for a short time on screen before each roadway image is shown. In addition to increasing workload, this task also served to mask consecutive roadway images. Next, participants identified bicyclists from roadway images. Participants were instructed to click the left mouse button when they were completely confident that they saw a bicyclist. The roadway images were each shown for five seconds. After going through both blocks of trials and questionnaires, the participants were thanked for their time and dismissed.



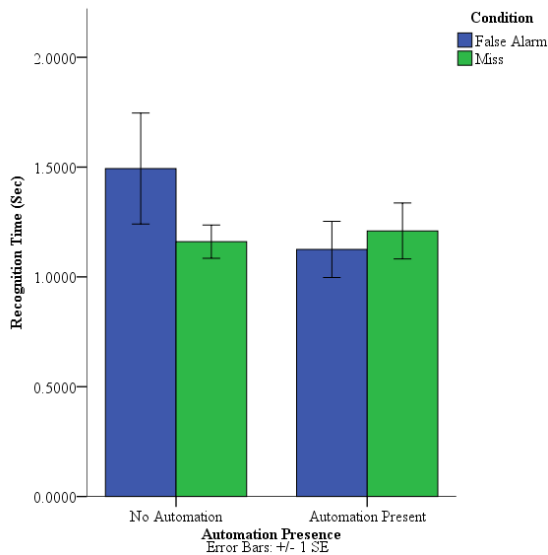
**Figure 4.** Participant receiving instructions prior to beginning task.

## 4 RESULTS

A 2 (automation type: miss-prone, false-alarm-prone) x 2 (automation presence: present, not present) repeated measures analysis of variance (ANOVA) was conducted for bicyclist detection time (see Figure 5) and recognition time (see Figure 6). Results revealed no significant main effect of automation presence,  $F(1,10) = 1.22$ ,  $p > .05$ , and no significant interaction of automation presence and failure bias,  $F(1,10) = 3.04$ ,  $p > .05$ . These findings suggest that detection and recognition time were not influenced by either the presence of automation or failure bias.



**Figure 5.** Mean bicyclist detection times (in seconds) for automation false alarm and miss conditions, clustered by automation presence.



**Figure 6.** Mean bicyclist recognition times (in seconds) for automation false alarm and miss conditions, clustered by automation presence.

**Subjective Measures.** A separate 2 (automation type: miss-prone, false-alarm-prone) x 2 (automation presence: present, not present) repeated measures ANOVA was conducted for workload. Consistent with prior literature, mean workload was lower in the automation present condition ( $M = 16.67, SD = 5.24$ ) than in the no automation condition ( $M = 21.33, SD = 5.16$ ),  $F(1,10) = 8.95, p < .05$ . No significant interaction of automation presence and failure bias was detected for workload,  $F(1,10) = .129, p > .05$ . One-way ANOVAs revealed no significant differences between the automation conditions for either trust,  $F(1,11) = .940, p > .05$ , or reliance,  $F(1,11) = .094, p > .05$ .

## 5 DISCUSSION

We hypothesized that participants would have faster hazard detection and recognition times within the false-alarm prone condition compared to the miss prone condition. The results of our data did not show significant differences in performance between the automation conditions. While the data did not support the hypothesis, we see various factors that could explain this finding. One possible explanation is that the automation was not salient enough to effectively aid hazard detection and recognition. Future iterations of this research should maximize automation salience. It also may be possible that participants were inadequately trained prior to beginning the task (e.g., were confused about what the automation looked like). Our workload measures indicated relatively low scale scores, suggesting that the task did not contain enough workload to consider the experimental tasks *high* workload. Most participants had no difficulty answering the math problems while performing the search task, and even did well with remembering the string. This is counter to the secondary task performance we might expect under high workload conditions. Future studies should consider using a longer string (7 characters instead of 5) and consider either using more complex math problems or simultaneously displaying the math problems during the visual search task (e.g., by using the auditory modality). This would require

them to switch between two simultaneously occurring tasks instead of switching between consecutive tasks.

## 6 LIMITATIONS

This experiment has a few limitations that would be remedied if the work were replicated. One particularly notable limitation is the small sample size. Our sample size resulted in low power in our statistical analyses, thereby impairing our ability to detect mean differences. Another limitation encountered was controlling the environment of the experiment. This being a project for class, researchers had to sometimes run the study with participants while other people were in the room running studies of their own. This extra distraction may have added statistical noise and could have the potential to have affected participants' ability to focus on the experimental tasks. Future studies should attempt to isolate the participants by accessing the experimental room when there was no one occupying it or otherwise reserving the space. Another possible confound is participant understanding of the stimuli. Researchers found when analyzing the data that participants seemed to not always agree on what a "bicyclist" was considered. Some of the stimuli showed people walking with bicycles but not actually riding on the bicycles. For these trials, some participants did not identify them as "bicyclists" whereas others did. To remedy this, the researchers in the future would make clear in the instructions what constituted a target, or use less ambiguous stimuli so that the distinction would not have to be made. A final limitation to consider is that participants were told that their task was to detect hazards on the road. This does not place participants in a natural environment of driving, where hazards can appear at any second. The knowledge that their primary task was hazard detection might have led participants to default to searching for bicyclists, and minimally use or ignore altogether the automated aid that was provided.

## 7 CONCLUSIONS

This experiment set out to examine different failure biases within an automated hazard identification system to see how they might affect performance. Although the experiment conducted was not able to find significant results, the underlying work lays down a foundation to for a future study to be designed that could yield significant and useful results. Using an eye-tracking device and software, the researchers were able to observe participants performing a search task for hazards on a simulated road. The design decision of having high cognitive workload to simulate the different activities occurring while driving was found to not be effective enough to give significant data, but still has a basis in research to be useful with a tweaked experimental design. Future work would include running the experiment again at a larger scale, improving the quality and clarity of the stimuli, increasing the cognitive workload, working in a more controlled environment, and increasing the saliency of the automation provided.

## REFERENCES

- [1] Caird, J., Horrey, W., & Edwards, C. (2001). Effects of conformal and nonconformal vision enhancement systems on older-driver performance. *Transportation Research Record: Journal of the Transportation Research Board*, (1759), 38-45.
- [2] Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses?. *Human factors*, 49(4), 564-572.
- [3] Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, 130(2), 169.
- [4] Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, Space, and Environmental Medicine*, 78(5), B176-B185.
- [5] Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80.
- [6] McBride, S. E., Rogers, W. A., & Fisk, A. D. (2011). Understanding the effect of workload on automation use for younger and older adults. *Human factors*, 53(6), 672-686.
- [7] Meyer, J., Wiczorek, R., & Günzler, T. (2014). Measures of reliance and compliance in aided visual scanning. *Human factors*, 56(5), 840-849.
- [8] National Center for Statistics and Analysis. (2016, May). Bicyclists and other cyclists: 2014 data. (Traffic Safety Facts. Report No. DOT HS 812 282). Washington, DC: National Highway Traffic Safety Administration.
- [9] Nevin, J. A. (1969). Signal Detection Theory and Operant Behavior: A Review of David M. Green and John A. Swets' Signal Detection Theory and Psychophysics. *Journal of the Experimental Analysis of Behavior*, 12(3), 475-480.
- [10] Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors*, 39(2), 230-253.
- [11] Peters, G. A., & Peters, B. J. (2001). The distracted driver. *The Journal of the Royal Society for the Promotion of Health*, 121(1), 23-28.
- [12] Rantanen, E. M., & Goldberg, J. H. (1999). The effect of mental workload on the visual field size and shape. *Ergonomics*, 42(6), 816-834.
- [13] Rumar, K. (2002). Night vision enhancement systems: What should they do and what more do we need to know? The University of Michigan. *Transportation Research Institute*. UMTRI-2002-12.
- [14] Rusch, M. L., Schall, M. C., Gavin, P., Lee, J. D., Dawson, J. D., Vecera, S., & Rizzo, M. (2013). Directing driver attention with augmented reality cues. *Transportation Research Part F: Traffic Psychology and Behaviour*, 16, 127-137.
- [15] Savage, S. W., Potter, D. D., & Tatler, B. W. (2013). Does preoccupation impair hazard perception? A simultaneous EEG and eye tracking study. *Transportation Research Part F: Traffic Psychology and Behaviour*, 17, 52-62.
- [16] Schwarz, F., & Fastenmeier, W. (2017). Augmented reality warnings in vehicles: Effects of modality and specificity on effectiveness. *Accident Analysis & Prevention*, 101, 55-66.
- [17] Velichkovsky, B. M., Rothert, A., Kopf, M., Dornhöfer, S. M., & Joos, M. (2002). Towards an express-diagnostics for level of processing and hazard perception. *Transportation Research Part F: Traffic Psychology and Behaviour*, 5(2), 145-156.
- [18] Yamada, K., & Kuchar, J. K. (2006). Preliminary study of behavioral and safety effects of driver dependence on a warning system in a driving simulator. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 36(3), 602-610.
- [19] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.