

Repeated Induction of Inattentional Blindness in a Simulated Aviation Environment

Kellie D. Kennedy
NASA Langley Research
Center

Chad L. Stephens
NASA Langley Research
Center

Ralph A. Williams
Analytical Mechanics
Associates, Inc.

Paul C. Schutte
U.S. Army

The study reported herein is a subset of a larger investigation on the role of automation in the context of the flight deck and used a fixed-based, human-in-the-loop simulator. This paper explored the relationship between automation and inattentional blindness (IB) occurrences in a repeated induction paradigm using two types of runway incursions. The critical stimuli for both runway incursions were directly relevant to primary task performance. Sixty non-pilot participants performed the final five minutes of a landing scenario twice in one of three automation conditions: full automation (FA), partial automation (PA), and no automation (NA). The first induction resulted in a 70% (42 of 60) detection failure rate with those in the PA condition significantly more likely to detect the incursion compared to the FA condition or the NA condition. The second induction yielded a 50% detection failure rate. Although detection improved (detection failure rates declined) in all conditions, those in the FA condition demonstrated the greatest improvement with doubled detection rates. The detection behavior in the first trial did not preclude a failed detection in the second induction. Group membership (IB vs. Detection) in the FA condition showed a greater improvement than those in the NA condition and rated the Mental Demand and Effort subscales of the NASA-TLX significantly higher for Time 2 compared Time 1. Participants in the FA condition used the experience of IB exposure to improve task performance whereas those in the NA condition did not, indicating the availability and reallocation of attentional resources in the FA condition. These findings support the role of engagement in operational attention detriment and the consideration of attentional failure causation to determine appropriate mitigation strategies.

INTRODUCTION

Inattentional blindness (IB) is a visual attention failure that can occur under periods of high and low workload (Cartwright-Finch & Lavie, 2007; Mack & Rock, 1998). IB occurs when observers fail to notice the presence of a clearly viewable but unexpected event when cognitive resources are diverted elsewhere. Errors and accidents attributed to IB have been identified across context and environment (Simons & Chabris, 1999). The information gathering conducted in the aviation environment is primarily visual and potentially safety-critical, thus requiring a better understanding of the natural tendencies and tolerances of the visual system.

A large human-in-the-loop experiment was conducted in three stages to evaluate IB in an aviation context. The first run explored if low workload conditions found in highly automated environments could produce an IB occurrence rates similar to those observed during high workload conditions (Kennedy, Stephens, Williams, & Schutte, 2014). The second run attempted a second IB induction with a highly task-relevant critical stimulus to explore IB rates and detection group memberships changes. This paper discusses the second run, specifically, the findings of participants who experienced the repeated induction of a second IB elicitation attempt.

Inattentional blindness. When objects in the world fail to reach conscious perception, individuals base their subsequent decisions on partial information. Simons (2000) noted observers must be aware of objects to make volitional changes in behavior. Failures to perceive visual information can occur when such information is relevant, detectable, and within the useful field of view (Mack & Rock, 1998; Simons & Chabris, 1999). Inattentional blindness (IB) is the failure of observers

to notice the presence of a clearly viewable but unexpected visual event when attentional resources are diverted elsewhere (Mack & Rock; Simons & Chabris, 1999). In this case, observers fail to notice a visual object or event; the object or event is clearly visible and detectable when observers look for it; and the failure to notice the object or event was not due to properties of the visual stimulus itself. The most damaging outcome of IB occurs when incomplete information leads to inaccurate representations of the external world that support incorrect decisions.

Task Load. High workload can result in stimulus detection failure (Cartwright-Finch & Lavie, 2007; Recarte, Perez, Conchillo, & Nunes, 2008). Furthermore, individuals engaged in visual search for specific stimuli often exhibit decreased detection of unexpected stimuli (Most & Astur, 2007). Researchers found that higher cognitive demands tend to reduce visual target detection and impair task performance (Reyes & Lee, 2008; Simons & Jensen, 2009; Strayer & Drews, 2007). In normal flight operations, pilots typically experience highest workload during take-off and landing (Wilson, 2002).

Automation is often the chosen solution for cognitively overloaded operators. However, low workload can also induce attentional failures. In a simulated driving task conducted by Kennedy and Bliss (2013), participants who reported low mental demand while following automated navigational directives were more likely to experience IB to a task-relevant critical stimulus than those participants who reported higher mental demand. In part one of this study, participants monitoring flight automation were just as likely to exhibit IB to a critical stimulus as those flying manually (Kennedy, Stephens, Williams, & Schutte 2014). IB occurrence rates

exhibited during automation-induced underload were equivalent to those exhibited during high workload, in line with the extended-U model of stress and performance (Hancock & Warm, 1989; Kennedy & Bliss, 2013).

The Malleable Attentional Resources Theory (MART) by Young and Stanton (2002) posited that the marked decrease in mental workload during periods of automation was due to attentional capacity varying directly with mental workload. Parasuraman and Manzey (2010) reported that maintaining appropriate attention is crucial for monitoring automated tasks. They found that without attention allocation, salient and critical information about an automated task could remain undetected despite eye fixations. They explained this automation-induced IB as automation complacency (AC). AC is defined as inferior detection of system malfunctions during periods of automation control as compared with those under manual control (Parasuraman & Manzey, 2010).

Task load, reliability, and system failure rate modulate AC (Bailey & Scerbo, 2007; Parasuraman & Manzey, 2010). The ideal circumstance for AC to occur is when an operator has a high, multiple-task load with a highly reliable automation system with infrequent and unexpected problems. The modern flight deck is just such an environment.

Importantly, Parasuraman and Manzey (2010) reported that unless an operator has experience specific to automation failures, AC cannot be overcome with experience and practice. The goal of this portion of the study was to investigate this finding using a repeated induction of IB with participants in low, moderate, and high automation conditions and explore the attentional resource allocation (Young & Stanton, 2002).

Current Study. This paper examines the relationship between levels of automation, workload, and IB occurrences for a task-relevant stimulus in a simulated flight task with repeated induction. This task utilized a simulated aircraft and three levels of flight control automation similar to autopilot, auto-throttle, and manual control. The measured outcome across automation conditions was IB occurrence for critical stimuli that were directly relevant and critical to primary task performance; specifically, two runway incursions.

Participants completed a simplified landing task twice each with a critical stimulus runway incursion. Following the first induction, all subjects were asked questions that clearly indicated the importance of attending to visual information during landing (Kennedy, Stephens, Williams, & Schutte, 2014). Next, subjects performed the second induction by performing the exact same landing task featuring a different critical stimulus runway incursion.

We predicted an overall reduction of IB occurrence rates across automation conditions in the repeated induction of IB (Parasuraman & Manzey, 2010; Young & Stanton, 2002). We predicted fewest IB occurrences when workload was moderate (Cartwright-Finch & Lavie, 2007; Kennedy & Bliss, 2013; Kennedy, Stephens, Williams, & Schutte, 2014). Of interest was the exploration of the IB group membership changes (IB/Detect) across automation conditions. We predicted that those in the low workload condition could improve attentional resource allocation to increase detection rates during landing whereas those in the manual condition could not (Parasuraman & Manzey, 2010; Young & Stanton, 2002).

METHOD

Experimental Design

The experimental task required non-pilot participants to twice perform the final five minutes of a simplified landing scenario. There were three automation conditions such that automation controlled all, some, or none of the aircraft operation similar to autopilot, auto-throttle, or manual control. The IB events occurred approximately 10 seconds before touchdown in both runs. In the first induction, the critical stimulus (a truck) began to move along a taxiway and intersected the active landing runway. The scenario ended just prior to touchdown, the simulation displays blanked, and the participant completed the post-experiment questionnaire. In the second induction, the critical stimulus (a plane) flew into ownship view in the airspace directly above the active landing runway. The simulation ended and the questionnaire followed as described. For a more detailed experimental description, please see Kennedy, Stephens, Williams, and Schutte, 2014.

Participants

Sixty participants (28 male, 31 female) completed this experiment and were compensated with \$50. Subject age range was 20-64 ($M=34.5$, $SD=13.3$). Subjects were required to be non-pilots, over the age of 20, have normal or corrected-to-normal vision and hearing.

Materials

Subjects signed an informed consent document and completed a background questionnaire to capture pertinent demographic information such as age, sex, abnormal vision or audition, and flight simulator experience. Subjects were given experimental instructions that described the flight simulator, the scenario, and the automation condition. Each subject performed three practice runs to achieve task proficiency. This study obtained Institutional Review Board approval at NASA.

Experimental Manipulation. Participants were randomly assigned to one of three automation condition for the entire experiment: full automation (FA), partial automation (PA), or no automation (NA). FA participants monitored the automation-controlled flight path and speed. PA participants manipulated flight path using a sidestick controller and monitored the speed. NA participants manipulated both the flight path using sidestick and speed using throttle. Participants with automation components were instructed to monitor the automation and report deviations. By design, no deviations existed and none were reported.

Task. The flight scenario consisted of daytime flight conditions with greater than 3 miles of visibility and light turbulence. The flight task required participants to perform the final five minutes of a simplified simulated landing scenario and pilot the aircraft down to Runway 29 at Louisville International Airport (SDF). The total run from starting point to touchdown point covered a distance of approximately 8 nmi

and lasted approximately 5 minutes. The specified airspeed was 180 knots until 2200 ft. then incorporated a speed reduction to 150 knots, which was to be maintained until touchdown. To avoid collision with critical stimuli, the simulation ended just prior to touchdown and the simulation displays blanked.

System Description. Participants conducted the task using a flight simulator that provided a highly simplified level of flight control fidelity to accommodate the non-pilot participants. The flight model used a twin turbo-prop commuter plane. The environment provided an out-the-window (OTW) view and a primary flight display (PFD) (Figure 1). The PFD displayed a repeated image of the OTW image and an instrumentation overlay with flight path marker with speed, altitude, and heading information (Figure 2).



Figure 1. Experimental apparatus.

Critical Stimulus. The FAA defines a runway incursion (RI) as any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle, or person on the protected area of a surface designed for the landing and take-off of aircraft (FAA, 2012). Seven vehicles were in the proximity of the landing runway: three non-moving, three moving, and one critical stimulus. The three moving vehicles were on the two taxiways adjacent to the active runway. All vehicles were in view for approximately 40 seconds and only the critical stimulus provided a conflict for any flight behaviors.

The first experimental run utilized a Category B, Vehicle Deviation runway incursion in the form of an orange and white box truck (FAA, 2012). The truck was positioned on an intersecting taxiway, entered the active landing runway, and presented a direct collision threat to the landing aircraft (Figure 2). The critical stimulus triggered approximately 10

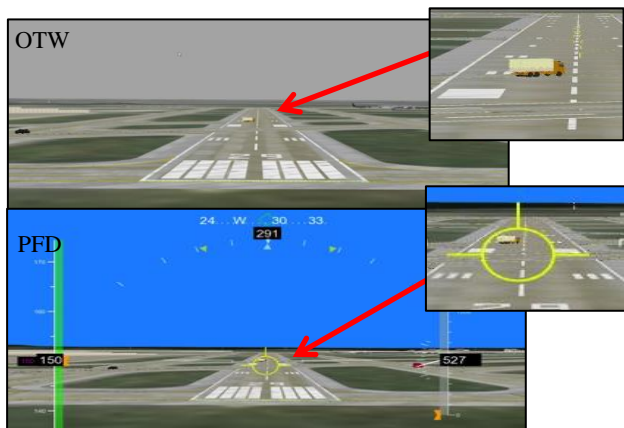


Figure 2. The OTW and PFD images showing the view of the first runway incursion (truck) critical stimulus.

seconds before the end of the scenario when the displays blanked, and the participant completed the post-experiment IB questionnaire.

The second experimental run (repeated induction attempt) utilized a Category B, Pilot Deviation RI in the form of a red and white general aviation plane. The plane flew an intersecting flight path, entered the active landing runway airspace, and presented a direct threat to the landing aircraft (Figure 3). The critical stimulus was in motion for approximately 8 seconds. The scenario ended just prior to touchdown, displays blanked, and the participant completed the post-experiment IB questionnaire.

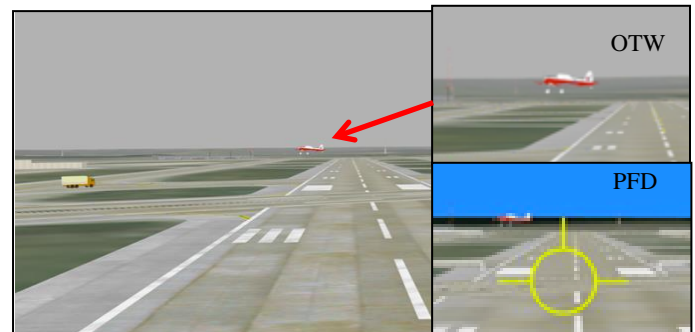


Figure 3. The OTW and PFD images showing the view of the second runway incursion (airplane) critical stimulus.

IB Questionnaire. After each scenario, participants completed an IB questionnaire about flight behaviors exhibited during the experiment. The conventional assessment of IB is the failure of a participant to consciously perceive the critical stimulus such that he or she is unable to report detection of the stimulus. Consistent with Mack and Rock's (1998) IB paradigm, the self-report post-experimental questionnaires prompted each participant to report detection of the critical stimulus. IB questions included "Did you see anything on or above the landing runway?" and "If so, please describe" for the participant to report detection. Participants who indicated that they did not detect the scenario specific critical stimulus were classified as exhibiting IB (Mack & Rock, 1998; Most & Astur, 2007). This technique indicates to the subject that there is something in the scene they did or did not detect which can influence the subject behavior in subsequent attempts to elicit IB. This prior exposure was intended and expected to influence the repeated induction.

NASA-TLX. Participants completed the NASA-Task Load Index (TLX) to provide a subjective rating of perceived workload. Task load was defined as the cost incurred by human operators to achieve a specific level of task performance (Hart & Staveland, 1988). The NASA-TLX includes six elements of workload: mental demand, physical demand, temporal demand, performance, effort, and frustration level. Both the overall and subscale score results were explored to investigate variations in task load for comparison with IB occurrences (Lee, Caven, Haake, & Brown, 2001; Nees & Walker, 2011).

Procedure

Subjects completed the Informed Consent Form and the background questionnaire. Next, subjects were randomly assigned to an automation condition and adjusted to the environment and equipment with three training runs. The three training runs used a scenario similar in operation to the test scenario but featured a Northerly approach to the SDF Runway 35L with no vehicles on or near the runway. Participants completed a post-training questionnaire and a NASA-TLX for the final practice run. Next, participants completed both experimental scenarios by piloting the simulated aircraft to Runway 29 with a Westerly approach. The screens blanked just prior to touchdown and participants completed the post-experiment IB questionnaires and NASA-TLX forms. Participants completed the remainder of the full experiment and were debriefed.

RESULTS

This paper examined the relationship of automation and IB occurrences for a task-relevant stimulus in a simulated flight task with repeated induction; Time 1 (T1), Time 2 (T2).

IB Across Induction. As predicted, the data revealed a decreased occurrence of IB across all conditions for T2 (50%, 30 of 60) as compared to T1 (70%, 42 of 60). The automation condition with the lowest IB occurrence rates remained PA followed by FA and NA (See Figure 4).

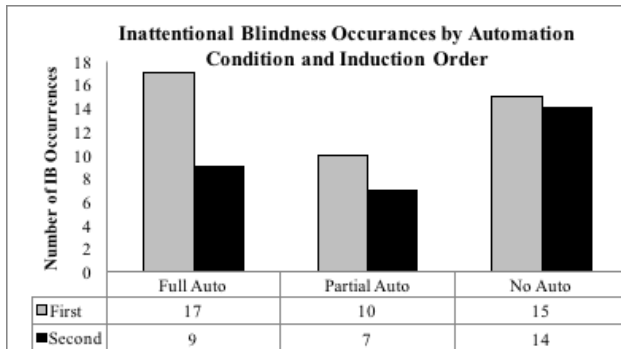


Figure 4. Number of IB occurrences grouped by automation condition for the first (truck) and second (plane) inductions.

From first induction to second induction, the subjects who changed detection group membership varied within automation conditions (See Figure 5). Also, as predicted, although PA had the best detection rates numerically, the FA condition was the most improved. The IB group membership change (IB or Detect) across automation conditions showed that the FA condition decreased IB by half (85% to 45%), then PA (50% to 35%) with NA as the least improved (75% to 70%). Approximately 70% of individuals remained in their original detection categories (IB to IB = 27, Detect to Detect = 15). While 25% moved from IB to Detect (15), only 5% moved from Detect to IB (3). Across automation conditions, FA had 8 participants (40%) move from IB to Detect while PA had 5 and NA had 2. PA had no members move from Detect to IB while PA had 2 and NA had 1.

TLX Across Inductions. In T1, the TLX served as a manipulation check of task difficulty across automation conditions. Kennedy, Stephens, Williams, and Schutte, 2014 confirmed that the overall NASA-TLX scores significantly

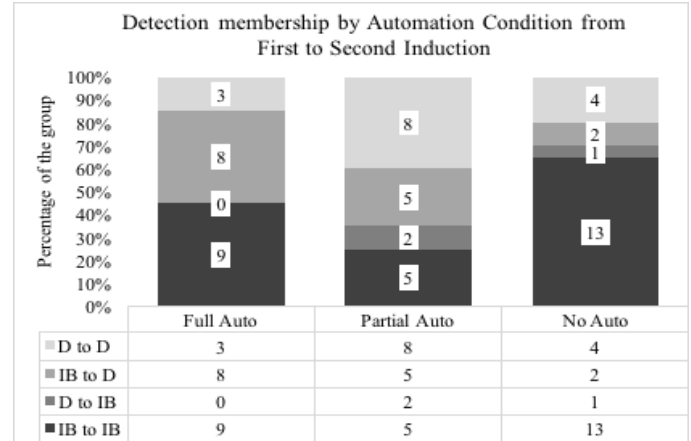


Figure 5. The change in detection membership (Detect or IB) from first to second induction including counts and grouped by automation condition.

differed between automation conditions and increased linearly from FA, to PA, to NA.

A repeated measures ANOVA was conducted to determine the change in the TLX ratings from T1 to T2. The data were normally distributed with homogeneity of variances. Two outliers were identified by boxplot but retained as plausible values. Unlike T1, the scores for T2 did not follow a linear path and did not vary significantly by automation condition. As shown in Figure 6, the Overall TLX scores from T1 and T2 decreased for PA and NA but increased for FA. There were statistically significant interactions between Automation Condition and IB-induction (T1 vs. T2) on the overall TLX ratings, $F(2, 57)=5.208$, $p<0.008$, partial $\eta^2 = 0.155$, and the two subscales of Mental Demand, $F(2, 57)=7.1$, $p<0.002$, partial $\eta^2 = 0.199$ and Effort $F(2, 57)=5.456$, $p<0.007$, partial $\eta^2 = 0.161$. There was also a statistically significant main effect in the Performance subscale between T1 and T2, $F(1, 58)=5.627$, $p<0.021$, partial $\eta^2 = 0.090$.

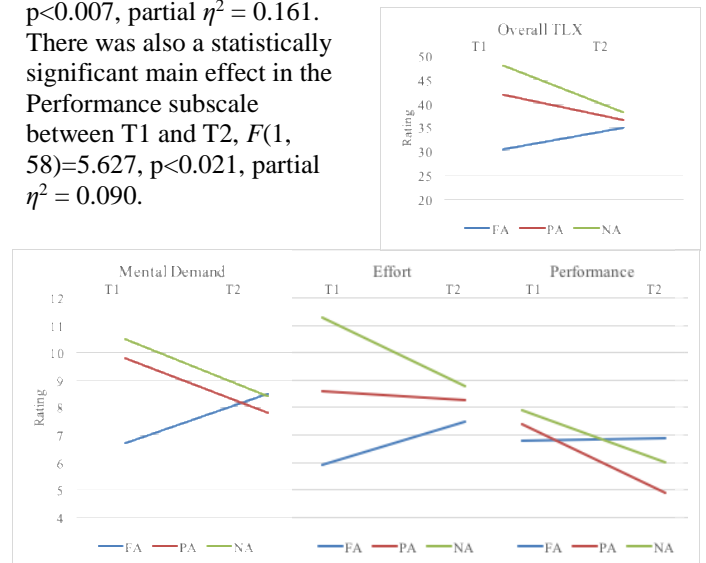


Figure 6. Time 1 and Time 2 TLX results showing the interaction effects for Overall TLX, Mental Demand, and Effort. Also showing the a significant main effect for Performance.

To further explore these differences, a one-way ANOVA was conducted on IB group membership and the difference in Mental Demand scores from T1 to T2. TLX Mental Demand scores significantly differed between IB group membership, $F(3, 56)=3.060$, $p<0.036$, partial $\eta^2 = 0.141$. Although

ANOVA is robust, there were several assumption violations; these included unequal sample sizes, small cell sample sizes, and outliers.

DISCUSSION

Sixty non-pilot participants performed two simplified landing scenarios in one of three automation conditions; FA, PA, and NA. Each landing scenario included a critical event object that was of high importance to task performance (two types of RIs). This study explored repeated induction of a visual attention failure event across workload conditions.

In the first IB attempt, the moderate workload condition (PA) performed the best with the low workload automation condition (FA) exhibiting high IB occurrences similar to those in the high workload condition (NA). The second IB attempt yielded reduced overall occurrences with the greatest improvement in the FA condition. This improvement coincided with a statistically significant increase in overall TLX, mental demand, and effort for the FA condition. These results suggest that the majority of participants in the FA condition recognized the first IB induction as a learning opportunity. With this experience of an unexpected event and questionnaire, these participants created and deployed an increased attentional awareness strategy which improved visual event detection, decreased IB, and increased workload.

While the workload levels for FA increased, the workload levels for the PA and NA decreased making the three automation levels no longer significantly different. All conditions moved towards a more moderate level of workload which is best suited for optimum performance as shown in the extended-U model of stress and performance (Hancock & Warm, 1989). Although all conditions improved, no condition achieved a 100% detection rate.

The capability of the cognitive system has a structurally-based upper limit. This is represented in the number of individuals that improved automation condition (FA=8, PA=5, NA=2). The relationship demonstrated between automation condition and repeated induction of IB could have potential impact in informing the appropriate use of automated systems as an error mitigation strategy. In particular, this research encourages consideration of the attention decrement and the attentional regulation behaviors required for successful task performance.

With this in mind, any mitigation strategy for a visual attention failure like IB should be tailored to the cause. IB caused by high workload requires solutions that reduce workload (e.g., task shedding or automation). In this condition, performance could not benefit significantly by previous exposure or “seeing the trick.” However, during periods of low workload while simply monitoring highly reliable automation, IB decrements can come from automation complacency. Unlike high workload, AC can be reduced through awareness (Cartwright-Finch & Lavie, 2007; Parasuraman & Manzey, 2010). The IB induced by low workload was reduced by revealing automation complacency and the tendency for the participants to become poor monitors. At the outset of the second induction, all participants knew they would be asked about features in the environment. Those

in the FA condition had the excess resources to use for attentional resource allocation during landing to improve performance whereas the high workload subjects did not. The NA condition did not allow an increased monitoring strategy because of the resources expended for manual control.

Further understanding the individual differences can help explain those who detected both times and those who failed to detect both times. Research has found that people with higher working memory capacity can successfully perform tasks of higher complexity (Hurt, Angell, & Perez, 2011). The flight deck can, at times, be a highly complex, overwhelming environment but it can also become monotonous to the point of disengagement, which is a condition rife with attentional failure opportunity (Lee et al., 2001).

Like high workload, low workload can induce a loss of attention. A pilot accustomed to highly reliable automation may experience automation-induced complacency and not recognize a time sensitive threat until the opportunity for intervention has passed. Although this study utilized a non-pilot sample, the change rate in critical stimulus detection warrants further testing with a pilot population using targeted scenarios to explore individual differences and attentional failures including the targeted recapture of attention.

In the National Airspace System of tomorrow, the role of automation in human error reduction continues to grow. Despite being increasingly rare, unpredictable events such as runway incursions and automation failure still require operator intervention. This line of research will help to identify situational and individual factors that increase occurrences of attention-related human error. By better understanding the cause of attentional events, we might hope to create mitigation strategies that ensure attention during maximized intervention opportunities without setting unrealistic expectations like constant vigilance.

REFERENCES

- Aviation Safety and Security Archives (ASSA). (2014). Aviation Safety and Security Archives. Retrieved from <http://prcsrc1.erau.edu/index.html>.
- Bailey, N. R. & Scerbo, M. W. (2007) Automation-induced complacency for monitoring highly reliable systems: the role of task complexity, system experience, and operator trust. *Theoretical Issues in Ergonomics Science*, 8:4, 321-348.
- Cartwright-Finch, U., & Lavie, N. (2007). The role of perceptual load in inattention blindness. *Cognition*, 102(3), 321-340.
- Federal Aviation Administration. (2012). *Runway Safety Report 2011-2012* (FAA Publication No. 2012-AJS-475). Retrieved from http://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/safety/media/2012-AJS-475-FY2011-Runway-Safety-Annual-Report.pdf
- Hancock, P.A., & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31, 519-537.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati, *Human mental workload* (pp. 139-183). Oxford, England: North-Holland.
- Kennedy, K.D. & Bliss, J.P. (2013). Inattention Blindness in a Simulated Driving Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (57)1, 1899-1903.
- Kennedy, K.D., Stephens, C.L., Williams, R.A., & Schutte, P. (2014). Automation and Inattention Blindness in a Simulated Flight Task. Presented at the 58th Annual Meeting of the Human Factors and Ergonomics Society, Chicago, IL.
- Lee, J. D., Caven, B., Haake, S., Brown, T. L. (2001). Speech-based interaction with in-vehicle computers: The effects of speech-based e-mail on drivers' attention to the roadway. *Human Factors*, 43, 631-640.
- Mack, A., & Rock, I. (1998). *Inattention blindness*. Cambridge, MA: The MIT Press.

- Most, S. B., & Astur, R. S. (2007). Feature-based attentional set as a cause of traffic accidents. *Visual Cognition*, 15(2), 125-132.
- Nees, M. A., & Walker, B. N. (2011). Auditory displays for in-vehicle technologies. *Reviews of Human Factors and Ergonomics*, 7(1), 58-99.
- Parasuraman, R. & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*. 52(3), 381-410.
- Recarte M. A., Perez E., Conchillo A., & Nunes, L. M. (2008). Mental workload and visual impairment: Differences between pupil, blink and subjective rating. *Spanish Journal of Psychology*, 11(2), 374-385.
- Reyes, M. L., & Lee, J. D. (2008). Effects of cognitive load presence and duration on driver eye movements and event detection performance. *Transportation Research Part F: Traffic Psychology and Behaviour*, 11(6), 391-402.
- Simons, D. J. (2000). Attentional capture and inattention blindness. *Trends in Cognitive Sciences*, 4(4), 147-155.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, 28(9), 1059-1074.
- Simons, D. J., & Jensen, M. S. (2009). The effects of individual differences and task difficulty on inattention blindness. *Psychonomic Bulletin & Review*, 16(2), 398-403.
- Strayer, D. L., & Drews, F. A. (2007). Cell-phone-induced driver distraction. *Current Directions in Psychological Science*, 16(3), 128-131.
- Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures, *The International Journal of Aviation Psychology*, 12(1), 3-18
- Young, M.S. & Stanton, N.A. (2002). Malleable attentional resources theory: a new explanation for the effects of mental underload on performance. *Human Factors*, 44(3), 365-375.