

**FACTORS THAT MEDIATE FLIGHT PLAN MONITORING AND ERRORS IN PLAN REVISION:
PLANNING UNDER AUTOMATED AND HIGH WORKLOAD CONDITIONS**

Emily K. Muthard
Christopher D. Wickens
University of Illinois Aviation Human Factors Division
Savoy, Illinois

An experiment was conducted to explore the effects of automation and task loading on aviation plan monitoring and errors in plan revision. Pilots were asked to select one of two flight paths that traversed through hazardous airspace and to then monitor the safety of the path by seeking and reporting changes in dynamic traffic aircraft and weather systems. Following change detection, pilots were given the opportunity to revise their flight plan as a result of the changes. Attention guidance automation, which was reliable for plan selection, but failed to highlight a critical change that threatened safety after plan selection, was present on half of trials. Automation improved planning accuracy and confidence in high workload conditions. However, in nearly one third of trials, pilots failed to revise the flight plans as a result of a change, and were more likely to do so with imperfect automation in high workload.

INTRODUCTION

With the implementation of free flight, the responsibility of enroute flight planning is shifting from the air traffic controller to being jointly shared with the pilot. As such, the pilot should be adept in selecting, monitoring, and revising flight plans (Layton, Smith, & McCoy, 1993). These planning stages may be explained using an information processing model (Wickens & Hollands, 2000), in which the pilot must first filter and perceive environmental information and weigh the importance of these elements. This information is then integrated with that in memory stores, and an action is selected and executed. While plan construction and selection has been examined in both basic (Ward & Allport, 1997) and applied research (Xiao, Milgrim, & Doyle, 1997) domains, few studies have examined the underlying cognitive constructs of the plan monitoring and revision processes.

Though it may be argued that plan revision is simply a modified installment of an original plan, recent accident statistics have indicated that plan revision is a unique task, affected by somewhat different factors than plan selection. The National Transportation Safety Board, in investigating air carrier accident reports 1978 through 1990 found that two-thirds of decision errors could be attributed to pilots' unwillingness to revise established flight plans (NTSB, 1994). Plan revision errors, formally termed plan continuation errors (Orasanu, Martin, & Davison, 2001), are also a contributing cause to fatalities in aviation. For example, Goh and Wiegmann (2001a) examined general aviation accident reports from 1990 to 1997 and found that roughly 75% of accidents caused by revision errors were fatal, compared to only 18% for other types of general aviation accidents.

The need for plan revision is always characterized by a change, either in the pilot's goals or in the environmental conditions of the airspace. For example, deterioration in weather conditions along the pilot's flight path would dictate plan revision. Plan continuation errors result when the pilot fails to perceive the changing context of the airspace and subsequently consider alternate flight plans (McCoy & Mickunas, 2000). Though the failure to revise a plan is often attributed to overconfidence (Goh & Wiegmann, 2001b), lack of experience (Burian, Orasanu, & Hitt, 2000), or frequency of risk-taking behavior (Goh & Wiegmann, 2001c), recent investigations have found a link between plan continuation errors and situation awareness (Goh & Wiegmann, 2001b; Orasanu, Martin, & Davison, 2001). In agreement with this latter finding, the authors hypothesize that the failure to detect an environmental change that threatens flight safety may be an additional causal factor in plan continuation errors.

Change detection has been examined in basic visual tasks (Rensink, O'Regan, & Clark, 1997), automobile human factors (Pringle, Irwin, & Kramer, 2001), and aviation environments (Podczerwinski, Wickens, & Alexander, 2001). Together, the literature suggests that individuals are likely to miss more than half of the changes that occur (cf. Rensink, 2002). Such "change blindness" also occurs in skilled pilots. For example, Nikolic and Sarter (2001) investigated pilot performance in detecting mode changes in their flight management system and found that pilots missed over 10% of changes, and that this rate substantially increased with competition for visual attention. Though overall change detection is poor, performance is mediated by top-down information processing factors. In one study, detection of

relevant changes (i.e., changes that cause a potential conflict) in both traffic aircraft and weather systems was superior to that of irrelevant changes (Podczerwinski et al., 2001). Goh and Wiegmann (2001c), however, found that pilots consistently failed to detect severe deteriorations in weather, though these changes were directly relevant to the goal of safe flight. This finding may be explained by a common decision making phenomenon termed the confirmation bias (Wason, 1960). Under this bias, the pilot would seek cues that confirm the belief that the originally filed flight path was safe and ignore cues that refute the belief. Ambiguous cues, which hold no informational value to the task, are perceived to support the belief (Einhorn & Hogarth, 1978).

The confirmation bias may be linked with the tendency to commit a plan continuation error. Embedded within every plan continuation error is a change that *refutes* the pilot's belief that the flight path is safe. Under this bias, the pilot would be less likely to detect this change thereby failing to reassess the situation and the potential alternative flight paths. This link between the confirmation bias and plan continuation errors encompasses the first theme of the present study.

As we have shown thus far, planning is a highly demanding, complex task. Planning can be critically limited by the various sources and sheer amount of information in the environment (Xiao et al., 1997), the amount of information that can be held in working memory (Ward & Allport, 1997), the probabilistic nature of the hazards in the airspace, and poor situation awareness (Endsley, 2000). One solution to overcoming these concerns is the implementation of automation. Pilots have cited workload as one of the most important reasons for employing automation (Riley, Lyall, & Wiener, 1993), and attention guidance aids have been shown to assist performance in directing attention to the relevant elements of a scene (Mosier, Palmer, & Degani, 1992), and reducing workload and improving situation awareness (Wiener, 1985; Billings, 1991). Negative effects, such as cognitive tunneling, are also associated with imperfect automation (Wickens, 2000). Consequently, if an aid fails to highlight a relevant element of the visual scene, the pilot may fail to detect the important, but uncued element. Performance with imperfect automation comprises the second theme of the present paper.

The present experiment was conducted to assess the performance of pilots in selecting a safe flight plan and then in monitoring the plan with and without the aid of automation. In a low fidelity simulation,

participants selected one of two presented plans that traversed through hazardous airspace and then monitored the safety of the routes by detecting changes in traffic aircraft and weather systems. Change detection performance was assessed in detecting confirming, refuting, and ambiguous changes. Both workload and the presence of automation were manipulated.

In agreement with the literature presented, we first hypothesized that plan continuation errors were linked to the confirmation bias, which would be manifest in the inability to detect a refuting, safety-threatening change. Furthermore, we sought to show that automation, though beneficial during the stage of plan selection, would hinder plan monitoring and lead to a greater number of plan continuation errors when the aid failed to account for a safety-threatening change. Finally, we drew conclusions regarding the effects of workload on each of the planning stages (selection, monitoring, and revision).

METHODS

Procedure

Twenty-eight pilots from the University of Illinois Institute of Aviation, sixteen of whom were instrument certified, participated in the experiment. The pilots were asked to select the safer of the two presented paths, while considering the risk posed by terrain, traffic aircraft, and weather systems. The integrated hazard display, depicting ownship at the beginning of the trial is shown in Figure 1a.

After selecting a flight path, pilots were asked to monitor the safety of the plan by seeking and detecting changes in the altitude or trajectory of weather systems and traffic aircraft. Pilots were instructed to press a key when they detected a change and inform the experimenter of the item that changed and the nature of the change. In one fourth of trials, a change substantially threatened the safety of the flight path. At the midpoint of each trial, participants were asked to recommend the safer plan to a following aircraft that would be confronted with the same flight path choice as the participant, while taking into account the changes that had occurred up to that point. The progression of the trial is shown in Figure 1b.

Hazard Symbolology

The display depicted weather, terrain, and air traffic hazards. A risk value was calculated for each of the

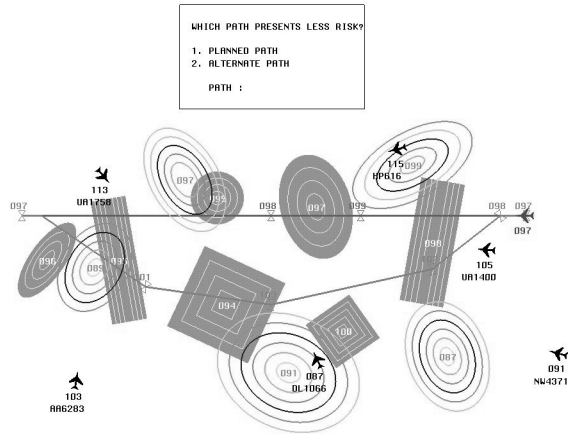


Figure 1a. Example of display at beginning of trial showing terrain (gray), weather (white ovals) and traffic aircraft hazards. The planned path runs in a straight line, and the alternate path branches to the south at the right of the display. Ownship is the aircraft located on the path to the right at the beginning of the trial.

hazards, based on the projected closest lateral and vertical distances of ownship to the hazard (see Muthard & Wickens, 2002). These risk values were then summed and weighted by a coefficient to ensure that all three types of hazard contributed equally to the overall integrated risk value. Risk algorithms were used to determine the safety of each of the two flight paths and to determine the impact of a change on flight safety.

Weather and Traffic Changes

Ten experimenter-induced changes in the altitude and trajectory of weather systems and traffic aircraft were implemented in each trial and were spaced randomly. Changes occurred every 25 to 75 s in the low workload condition and 17 to 50 s in the high workload condition. Each change was designed to increase, decrease, or have no noticeable effect on the safety of one of the two flight paths. To evaluate the presence of the confirmation bias, changes were classified as confirming, refuting, or ambiguous. A confirming change was defined as any change that

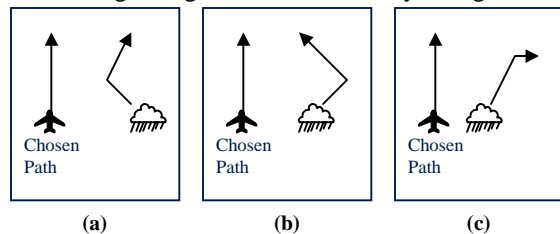


Figure 2. A change in weather heading, indicated by the bend in the arrow. (a) Change that confirms hypothesis that chosen path is safe. (b) Change that refutes hypothesis that chosen path is safe. (c) Change that provides ambiguous information regarding the safety of the chosen path.

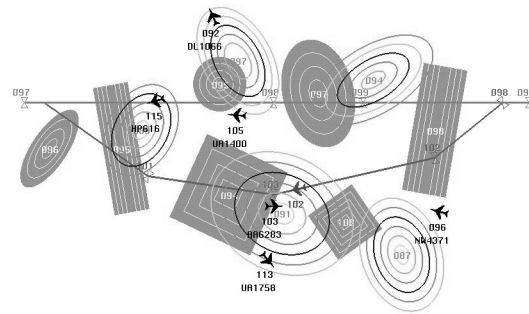


Figure 1b. Example of display near midpoint of trial in which pilot has selected the alternate path. Ownship is the aircraft located in the middle of the alternate path prior to the plan revision point of the trial.

increased the safety of the chosen flight path or decreased the safety of the unchosen path (i.e., that confirmed or reinforced the wisdom of the pilot's original flight path choice). A refuting change was defined as any alteration that decreased the safety of the pilot's chosen flight path or increased the safety of the unchosen path. Ambiguous changes had a minimal effect on the safety of either flight path. Graphic representations of these forms of changes are shown in Figure 2.

Plan Revision

One fourth of trials were designed to assess the frequency of plan continuation errors. During each of these trials, one of the changes occurring after the flight plan choice had been made, increased the risk of the initially safer flight path to a level that was then higher than that of the unchosen path. As a result of this change, the pilot should have optimally revised his or her flight plan at the midpoint of the trial when advising the following aircraft. On these trials, if the pilot failed to revise the plan, the behavior was classified as a plan continuation error.

Automation

Attention guidance automation, designed to support the plan selection task, was provided on one half of trials. Under this condition, the most hazardous one-third of elements was highlighted. The highlighting status of hazards remained constant throughout the entire trial (i.e., independent of whether the hazard changed), and was always 100% reliable for the plan selection task. The automation, however, failed in all of the trials described above in which a safety-

threatening change was implemented. In these trials, the automation failed to register the new risky status of the hazard and did not highlight the hazard, even though it changed and became essential to the safety of flight.

Experimental Design

The independent variables included workload (between-subjects) and the presence of attention guidance automation (within-subjects). Workload was manipulated in three ways. In the low workload condition, pilots were given 40 s to initially select a flight plan, the speed of the trial was ‘slow’ (i.e., the trial lasted 12 minutes), and no loading task was present. In the high workload condition, participants were only given 20 s to select a plan, the speed of the trial was increased by 50%, shortening the trial time to only 8 minutes, and pilots were also responsible for maintaining level pitch and roll in a secondary attitude tracking task.

RESULTS

Plan Selection

To assess the effects of workload and automation on plan selection performance, we examined plan selection accuracy, response time, and choice confidence in three ANOVAs. The main effect of workload was not significant for accuracy ($F_{(1, 27)} = 1.21, p > .10$), response time ($F_{(1, 26)} = 1.41, p > .10$), or confidence ($F_{(1, 27)} = 1.93, p > .10$). The automation main effect was also not significant for accuracy ($F_{(1, 27)} = 1.21, p > .10$), response time ($F_{(1, 26)} = .94, p > .10$), but was significant for the confidence dependent variable ($F_{(1, 27)} = 8.36, p = .01$), suggesting that pilots were more confident in plan selection with the automated aid than without.

Finally, the automation and workload interactions were not significant for accuracy ($F_{(1, 27)} = 1.21, p > .10$), response time ($F_{(1, 26)} = .02, p > .10$), or confidence ($F_{(1, 27)} = 1.37, p > .10$). Further examination, however, revealed important significant findings within the nonsignificant interactions for both accuracy and confidence.

When pilots were presented with an automated aid, both accuracy ($t_{(15)} = 1.94, p = .07$, marginally significant) and confidence ($t_{(15)} = 2.67, p = .02$, significant) were higher with automation than without the aid, though these differences were only observed for the high workload condition, as shown in Table 1.

Table 1. Plan selection accuracy and confidence was higher with automation, primarily when workload was high.

| | Workload | Control | Highlight |
|------------|----------|---------|-----------|
| Accuracy | Low | 65.4 % | 65.4 % |
| | High | 65.6 % | 78.1% |
| Confidence | Low | 4.7 | 4.9 |
| | High | 4.9 | 5.4 |

Plan Monitoring

In the plan monitoring phase, we wished to first examine the effect of highlighting on change detection and did so by comparing performance in detecting changes to hazards that were highlighted to those that were not highlighted as a function of workload. The analyses revealed that changes to highlighted hazards were detected more accurately ($F_{(1, 26)} = 27.72, p < .001$) and more quickly ($F_{(1, 22)} = 4.47, p = .05$) than changes to nonhighlighted elements. Additionally, a speed accuracy tradeoff was found for workload, such that changes in the low workload condition were detected 36.0% more accurately ($F_{(1, 26)} = 7.68, p = .01$), but 6.7 s more slowly ($F_{(1, 22)} = 8.88, p = .007$) than those in the high workload condition. The automation and workload interaction was not significant for accuracy ($F_{(1, 26)} = 1.75, p > .10$) or response time ($F_{(1, 22)} = .27, p > .10$).

Second, we assessed the presence of the confirmation bias in plan monitoring, by comparing detection performance for changes that confirmed or refuted the pilot’s hypothesis that the flight path was safe, and for changes that provided ambiguous information regarding the hypothesis. Results showed a significant main effect for the confirmation manipulation in both detection accuracy ($F_{(2, 54)} = 4.48, p = .02$) and response time ($F_{(2, 54)} = 6.30, p = .003$), showing that refuting changes ($M = 38.7\%$, $M = 15.7$ s) were detected more accurately and more quickly than confirming ($M = 33.7\%$, $M = 18.2$ s; $t_{(27)} = 1.71, p = .10, t_{(27)} = 1.66, p = .10$) and ambiguous changes ($M = 31.5\%$, $M = 22.4$ s; $t_{(27)} = 2.75, p = .01, t_{(27)} = 3.55, p = .001$). This finding does *not* support the confirmation bias described previously.

Plan Revision

The analyses conducted on the plan revision data were based on a limited sample size, because this data was only assessed for the one-fourth trials in which a critical change threatened flight safety. As such, conventional significance levels were not used to determine importance. Given the large differences in magnitude of the findings, and given that the power of the tests was limited because of small sample sizes, it remains important to stress the

potential meaningfulness of the following analyses, despite the lack of statistical significance. To examine the association between plan continuation errors and pilots' adeptness at plan monitoring, we looked first at the two trials for each pilot in which an experimenter-induced change substantially threatened the safety of the pilot's flight path. Results indicated that plan continuation errors were committed on 18 of 56, or nearly 32% of trials. We then examined the pilots' accuracy in detecting the safety-threatening change (whose overlooking defined the plan continuation error) and found that detection accuracy was lower on trials when a plan continuation error (33.3%) was committed than when it was not (48.3%), $X^2 = 1.29, p = .25$.

To test the hypothesis that imperfect attention guidance automation would result in a greater frequency of missed critical events, we assessed accuracy in detecting the safety threatening change in both the control and automation condition. Results revealed that a 60.9% of misses occurred with imperfect automation and only 39.1% of misses occurred in the control condition, $X^2 = 1.85, p = .17$.

Finally, we examined the frequency of plan continuation errors as a function of imperfect automation. Analysis revealed that, in the high workload condition only, a greater percentage of the trials with imperfect automation were comprised of plan continuation errors (7 of 16 or 43.7%) than in the baseline condition (3 of 16 or 18.7%), $X^2 = 2.33, p = .13$. In the low workload condition, plan continuation errors were committed more frequently in the control condition (5 of 12 or 41.7%) than with imperfect automation (3 of 12 or 25%), though this difference did not even approach significance, $X^2 = .72, p = .40$.

DISCUSSION

Two themes were researched in the present study. The first goal was to establish a link between the confirmation bias and plan continuation errors. Though overall results indicated that pilots were more sensitive to refuting than confirming changes, therefore *not* supporting the confirmation bias, the findings also showed that pilots often failed to detect a refuting, safety-threatening change, and were more likely to then commit a plan continuation error. These findings suggest that pilots generally do not exhibit the confirmation bias during plan monitoring, but a minority of pilots do so, and some of this tendency can be attributed to a lack of sensitivity to all refuting changes.

The second theme of the present paper addressed the influences of workload and automation on plan selection, monitoring, and revision. Results showed that attention guidance automation was beneficial in improving plan selection accuracy and confidence, especially in the high workload condition. This finding agreed with previous research, which has shown that automated aids are more likely to be employed under high workload (Riley et al., 1993), and are beneficial in assisting performance by directing attention to the relevant elements of a scene (Mosier, Palmer, & Degani, 1992), reducing workload, and improving situation awareness (Wiener, 1985; Billings, 1991).

Though automation assisted the stages of plan selection, the aid hindered performance at the plan monitoring stages. Change detection performance for hazards that were not highlighted by the aid deteriorated, and pilots were less likely to detect the safety-threatening change when the automation failed to highlight its new important status than when automation was not present. Consequently, results indicated that pilots in the high workload condition were also more likely to commit plan continuation errors with the imperfect automated aid. These findings collectively reflect a failure to properly monitor automated aids, to cognitively tunnel resources to primarily elements highlighted by the aid (Wickens, 2000; Yeh & Wickens, 200), and the resulting failure to revise the unsafe flight plan.

In conclusion, the findings show that when automation is imperfect, pilots are much less able to detect a change in an element of the airspace that threatens flight safety. Consequently, they are more likely to continue on routes that are no longer safe, thereby committing a plan continuation error. Given that automation can rarely account for the uncertainties in weather and traffic prediction, these findings represent a genuine concern for the application of these aids as planning tools.

ACKNOWLEDGMENTS

The authors wish to thank Sharon Yeakel for programming the experiment. Todd Kowalski and T.J. Hardy assisted in data collection. Funding for this research was provided through a grant from the NASA Ames Research Center (NASA NAG 2-1120) under technical monitor Dr. David Foyle.

REFERENCES

Billings, C. E. (1991). *Human-centered aircraft automation: A concept and guidelines* (NASA Technical

- Memorandum 103885). Moffet Field, CA: NASA Ames Research Center.
- Burian, B.K., Orasanu, J., & Hitt, J. (2000). Weather-related decision errors: Differences across flight types. *Proceedings of the IEA 2000 and HFES 2000 Congress* (pp. 22-25). Santa Monica, CA: Human Factors and Ergonomics Society.
- Einhorn, H.J., & Hogarth, R.M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, 70, 193-242.
- Endsley, M. (2000). Theoretical underpinnings of situation awareness: A critical review. In M.R. Endsley & D.J. Garland (Eds.), *Situation Awareness Analysis and Measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Goh, J., & Wiegmann, D.A. (2001a). Visual flight rules (VFR) flight into instrument meteorological conditions (IMC): A review of the accident data. *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University.
- Goh, J., & Wiegmann, D.A. (2001b). Visual flight rules flight into instrument meteorological conditions: An empirical investigation of the possible causes. *The International Journal of Aviation Psychology*, 11, 359-379.
- Goh, J., & Wiegmann, D.A. (2001c). An investigation of the factors that contribute to pilots' decisions to continue visual flight rules flight into adverse weather. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 26-29). Santa Monica, CA: Human Factors and Ergonomics Society.
- Koslowski, B., & Maqueda, M. (1993). What is the confirmation bias and when do people actually have it? *Merrill-Palmer Quarterly*, 39, 104-130.
- Layton, C.F., Smith, P.J., & McCoy, C.E. (1993). Design of a cooperative problem-solving system for en-route flight planning: An empirical evaluation. *Human Factors*, 36, 94-119.
- McCoy, C.E., & Mickunas, A. (2000). The role of context and progressive commitment in plan continuation error. *Proceedings of the IEA 2000 and HFES 2000 Congress* (pp. 26-29). Santa Monica, CA: Human Factors and Ergonomics Society.
- Mosier, K.L., Palmer, E.A., & Degani, A. (1992). Electronic checklists: Implications for decision making. *Proceedings of the Human Factors Society 36th Annual Meeting* (pp. 7-11). Santa Monica, CA: Human Factors Society.
- Muthard, E.K., & Wickens, C.D. (2002). *Factors that Mediate Flight Plan Monitoring and Errors in Plan Revision: An Examination of Planning under Automated Conditions* (Tech. Rep. AFHD-02-11/NASA-02-8). University of Illinois, Institute of Aviation: Savoy, IL.
- National Transportation Safety Board (NTSB) (1994). *Safety study: A review of flightcrew-involved major accidents of U. S. air carriers, 1978 through 1990* (NTSB/SS-94/01). Springfield, VA: National Technical Information Service.
- Nikolic, M.I., & Sarter, N.B. (2001). Peripheral visual feedback: A powerful means of supporting effective attention allocation in event-driven data-rich environments. *Human Factors*, 43, 30-38.
- Orasanu, J., Martin, L., & Davison, J. (2001). Cognitive and contextual factors in aviation accidents: Decision errors. In E. Salas & G. A. Klein (Eds.), *Linking Expertise and Naturalistic Decision Making* (pp. 209-225).
- Podczerwinski, E., Wickens, C.D., & Alexander, A.L. (2001). *Exploring the "Out of Sight, Out of Mind" Phenomenon in Dynamic Settings across Electronic Map Displays* (Tech. Rep. ARL-01-8/NASA-01-4). Savoy, IL: University of Illinois at Urbana-Champaign, Aviation Research Lab.
- Pringle, H.P., Irwin, D.E., & Kramer, A.F. (2001). The role of attentional breadth in perceptual change detection. *Psychonomic Bulletin and Review*, 8, 89-95.
- Rensink, R.A. (2002). Change detection. *Annual Review of Psychology*, 53, 245-277.
- Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8, 368-373.
- Riley, V., Lyall, B., & Wiener, E. (1993). *Analytic methods for flight-deck automation design and evaluation. Phase two report: Pilot use of automation* (Technical Report). Minneapolis, MN: Honeywell Technology Center.
- Ward, G., & Allport, A. (1997). Planning and problem-solving using the five-disc Tower of London task. *The Quarterly Journal of Experimental Psychology*, 50A, 49-78.
- Wason, P.C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.
- Weiner, E.L. (1985). Cockpit automation: In need of a philosophy. *Proceedings of the 1985 Behavioral Engineering Conference* (pp. 369-375). Warrendale, PA: Society of Automotive Engineers.
- Wickens, C.D. (2000). *Imperfect and unreliable automation and its implications for attention allocation, information access, and situation awareness* (Tech. Rep. ARL-00-10/NASA-00-2). University of Illinois, Savoy, IL: Aviation Research Lab.
- Wickens, C.D., & Hollands, J.G. (2000). *Engineering Psychology and Human Performance* (3rd Edition). Upper Saddle River, NJ: Prentice Hall.
- Xiao, Y., Milgrim, P., & Doyle, D.J. (1997). Planning behavior and its functional role in interactions with complex systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27, 313-324.
- Yeh, M., & Wickens, C.D. (2001). Display signaling in augmented reality: Effects of cue reliability and image realism on attention allocation and trust calibration. *Human Factors*, 43, 455-465.