

CHANGE DETECTION AFTER PRELIMINARY FLIGHT DECISIONS: LINKING PLANNING ERRORS TO BIASES IN PLAN MONITORING

Emily K. Muthard & Christopher D. Wickens
University of Illinois, Aviation Research Lab
Savoy, Illinois

The present study investigated a link between plan continuation errors and plan monitoring. Pilots were asked to execute a flight plan that traversed through hazardous airspace and then monitor the success of the plan by seeking and detecting changes in the airspace that could affect the safety of the plan. Following change detection, pilots had the opportunity to revise these plans. In nearly one-third of trials, pilots failed to revise flight plans, thereby committing a plan continuation error, and were more likely to do so when plan monitoring was inadequate. Overall, more than half of changes went undetected, though detection response times were improved when changes were relevant to the flight planning task or when traffic aircraft were changed rather than weather systems. Findings imply that plan monitoring is less than perfect, which may be a substantial contributing factor to plan continuation errors.

An evolution toward more distributed air-ground decision making has correspondingly increased the responsibilities of the pilot. Following the execution of a flight plan, the pilot must effectively monitor the airspace for changes that can impact the safety of the route and revise the preliminary flight plans if necessary. In an examination of general aviation accidents from 1990 to 1997, Goh and Wiegmann (2001a) found that roughly 75% of accidents caused by revision errors were fatal, compared to only 18% of other types of general aviation accidents. The failure to revise a flight plan despite emerging evidence that suggests it is no longer safe has been formally termed a plan continuation error (Orasanu, Martin, & Davison, 2001). In such situations, an element of the airspace that was of limited significance becomes central to the adequacy of the planning process, though the pilot does not detect this change or correctly assess its significance (McCoy & Mickunas, 2000; Orasanu et al., 2001). Nearly two-thirds of decision errors can be classified as plan continuation errors (NTSB, 1994), a finding that clearly shows the importance of examining these errors in greater detail.

While the failure to revise a plan is often attributed to overconfidence (Goh & Wiegmann, 2001b; Wilson & Fallshore, 2001), lack of experience (Burian, Orasanu, & Hitt, 2000), or frequency of risk-taking behavior (Goh & Wiegmann, 2001c), recent examinations have also indicated a link between plan continuation errors and poor situation awareness (Goh & Wiegmann, 2001b; Orasanu, Martin, & Davison, 2001). The current authors propose that revision errors should also be investigated by examining pilots' adeptness at monitoring established plans and consequently updating their situation assessment, if necessary, on the basis of changing environmental evidence. Plan monitoring, though essential in understanding the underlying causes of plan continuation errors, remains a topic that has not received adequate attention from the research community.

Successfully monitoring the airspace after plan execution is dependent upon the pilot's ability to detect changes in traffic aircraft or weather systems that may impact the safety of the flight plan. Research examining change detection performance, however, has revealed that observers are likely

to miss more than half of the changes that occur (cf. Rensink, 2002; Simons, 2000). Problems in change detection are also present with skilled pilots in realistic simulation studies. Nikolic and Sarter (2001) investigated pilot performance in detecting mode changes and found that pilots missed over 10% of changes in FMS modes, and this rate substantially increased with competition for visual attention. Furthermore, research examining the detection of unexpected experimenter-induced changes to the mode revealed that only one of thirty-two pilots detected the mode change (Mumaw, Sarter, & Wickens, 2001). Though change detection overall is poor, performance is improved when the change is salient (Pringle, Irwin, Kramer, & Atchley, 2001; Scholl, 2000) or occurs within or near focused attention (Rensink, O'Regan, & Clark, 1997; Thomas & Wickens, 2000), as is the case with most visual attention tasks.

Change detection performance is also mediated by the expectancies, beliefs, or goals of the pilot at the time of the change. Performance is superior in detecting alterations in items that are relevant to the task at hand, when compared to irrelevant changes (Podczerwinski, Wickens, & Alexander, 2001; Pringle et al., 2001). Though relevant changes usually improve performance, Goh and Wiegmann (2001c) found that some pilots did not consistently detect severe changes in weather, though these changes were directly relevant to safely navigating the aircraft. Such detection failures may be related to the pilot's belief that their flight plan was safe, and consequently pilots ignored evidence that refuted this belief. This behavior describes a common decision making phenomenon called the confirmation bias (Wason, 1960), in which observers seek information that confirms their beliefs and ignore that which refutes their beliefs. Under this bias, ambiguous cues, which hold no informational value to the task, are frequently used to support the established hypothesis (Einhorn & Hogarth, 1978; Wason & Johnson-Laird, 1972). If the confirmation bias were present in flight planning tasks, the pilot would likely ignore changes that hinder the safety of his or her plan, and correspondingly fail to recognize the threat that such changes pose, while seeking changes that confirm the safety of the chosen flight path.

The present study sought to assess the performance of pilots in monitoring his or her flight plan. In a low fidelity simulation, participants selected one of two potential flight plans that traversed through hazardous airspace. Following the selection of a flight plan, pilots were asked to monitor the airspace for changes in traffic aircraft and weather systems. Change detection performance was assessed as a means of measuring the effects of the physical characteristics of changes (salience, location with respect to visual attention) and the pilots' beliefs, expectancies, and goals (relevance, confirmation bias) on plan monitoring.

We also sought to establish a relationship between the ability to detect those changes that pose a serious threat to the risk of a flight path and the likelihood of committing a plan continuation error.

METHODS

Twelve pilots from the University of Illinois Institute of Aviation, ten of whom were instrument certified, were asked to select one of two flight plans that traversed through airspace while considering the risk posed by terrain, weather systems, and traffic aircraft. A representation of the display at the beginning of the trial is shown in Figure 1a.

Pilots were asked to view the dynamic display and were given up to forty seconds to select one of the two presented flight paths. Participants then monitored the airspace for changes in the altitude or heading of weather systems or traffic aircraft that could impact the safety of their chosen flight path. Pilots were instructed to press a key when they detected a change in the airspace and inform the experimenter of the item that changed and nature of the change. At the midpoint and end of each trial, participants were given the opportunity to revise their chosen flight path. We framed these revision actions by asking pilots to recommend the less hazardous path, while taking into account the changes that had occurred up to

that point, to an aircraft that would be entering the airspace and confronted with the same planning task as the participants. The progression of the trial can be seen in Figure 1b. Pilots participated in one practice trial and twelve experimental trials, and the total experiment lasted nearly two and a half hours.

Hazard Symbology

Hazards encountered during the flight consisted of weather systems, terrain, and possible conflict traffic. A risk value, based on the projected closest lateral and vertical distances of ownship to the hazard, was obtained for each hazard. These risk values were summed and weighted by a coefficient to ensure that all forms of hazards contributed equally to the total risk value. Risk algorithms were designed to measure the level of risk for a given path, and were also used to measure the impact of a hazard change upon the safety of a flight path.

Changes in Weather and Traffic

To assess the effects of both physical change characteristics and the expectancies and beliefs of the individual on plan monitoring, changes to the heading and altitude of both traffic aircraft and weather systems were implemented, as shown in Figure 1b. These changes occurred every 25 to 75 seconds, with ten changes occurring within each trial. Each change was designed to increase, decrease, or have no noticeable effect on the overall risk of a flight path. Changes were also varied to affect both the planned and alternate flight paths. These manipulations allowed for the classification of these changes in five manners.

Hazard. Changes were categorized by the type of hazard that was changed and included alterations to weather systems or traffic aircraft.

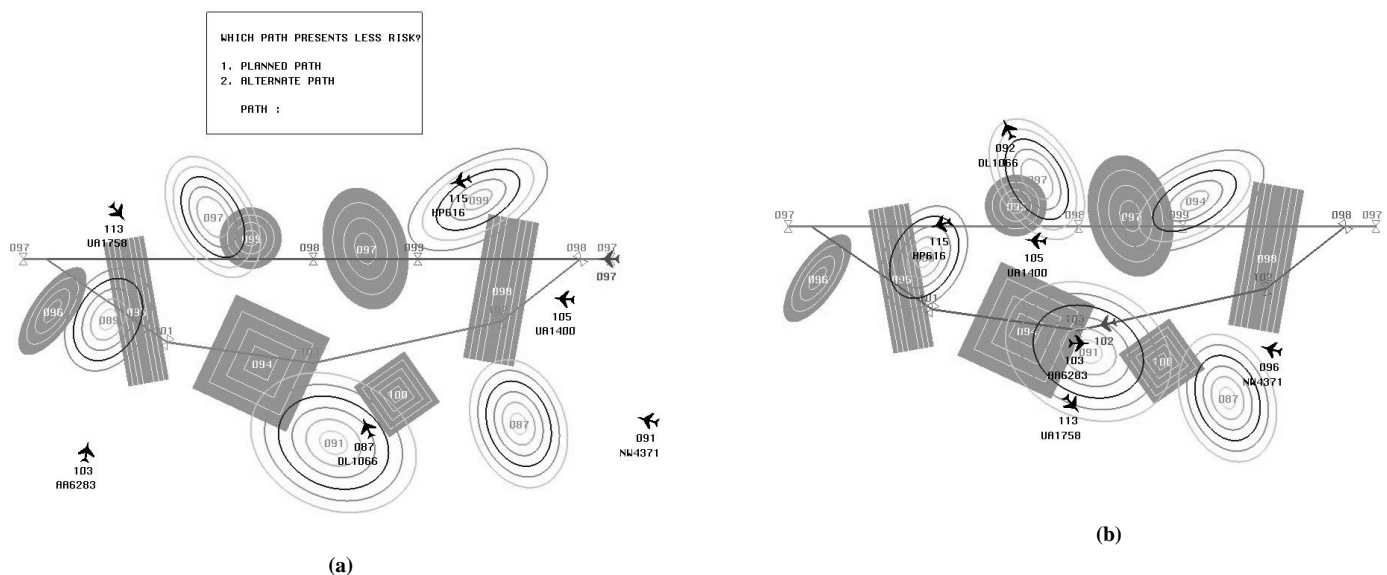


Figure 1. Example of display showing terrain (gray), weather (white ovals), and traffic aircraft hazards. The planned path runs in a straight line, and the alternate path branches to the south at the right of the display. Ownship is the aircraft located on the path at (a), to the right at the beginning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

Parameter. Changes were classified by the form of the change. Changes within this category included heading (spatial) changes and altitude changes (changes in the digital data tag which indicated a traffic aircraft's altitude or a weather system's top).

Location. Changes were classified to affect one of the two flight paths, and included changes to the pilot's chosen flight path or the unchosen flight path. Neutral changes were also included in this categorization and did not affect either flight path to a noticeable degree.

Relevance. Changes that increased or decreased the risk of any one of the two flight paths was characterized as relevant to the flight planning and monitoring task. Those changes that did not affect the risk of either of the two paths were termed irrelevant. Irrelevant changes were no different than neutral changes described previously.

Confirmation. In order to assess the role of the confirmation bias in plan monitoring, confirming, refuting, and neutral changes were included. A confirming change was defined as any change that decreased the risk of the pilot's chosen flight path or increased the risk 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 increased the risk of the pilot's chosen flight path or decreased the risk of the unchosen path. Neutral changes had no effect on the safety of either flight path.

Plan Continuation Errors

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

RESULTS

We were interested in establishing a relationship between plan continuation errors and pilots' adeptness at plan monitoring. To examine this association, we looked first at the three trials for each pilot in which an experimenter-induced change substantially increased the risk of the pilot's chosen flight path. Results indicated that plan continuation errors were committed on 11 of 36, or nearly 31%, of trials. We then examined the pilots' accuracy in detecting the particular change that posed a threat to the safety of the flight path and found that accuracy (45.5%) was substantially lower on trials when a plan continuation error was committed than when it was not (60.7%). Though this difference was not statistically significant ($t(18) = 0.83, p > 0.15$), the statistical power of this test was low (0.24) due to the limited sample size in measuring this rare event. Because the results indicate a link between detection of the change that threatened flight safety and the likelihood to commit a plan continuation error,

it is important to explore the factors that mediate change detection in plan monitoring.

Change detection response time and accuracy were measured to assess the pilots' performance in plan monitoring. These variables were analyzed in each of the five categories, including the type of hazard that was changed (weather or aircraft), type of parameter than was changed (altitude or heading), the location of the change (chosen, unchosen, or neutral), the relevance of the change (relevant or irrelevant) and the relationship of the change relative to the pilot's prior beliefs (confirming, refuting, or neutral).

Change detection accuracy ($M = 47.6\%$) was not significantly influenced by any of the change classifications. Additionally, weather heading changes, which were detected at rates reaching only 3%, were eliminated from the analysis to reduce unwanted variance. The speed of weather movement was simply too slow for any of the participants to notice the heading changes.

Mean change detection latencies were 18.5 secs. A repeated measures ANOVA was conducted for each of the five change variables, and the mean detection response times for each is shown in Table 1. Analysis revealed a significant main effect for changes categorized by hazard ($F(2, 20) = 22.61, p = 0.001$), relevance ($F(1, 10) = 7.93, p = 0.02$), location ($F(2, 18) = 8.30, p = 0.003$), and confirmation ($F(2, 18) = 8.30, p = 0.003$). Detection response time was not significantly influenced by parameter (altitude, heading), $F(2, 12) = 0.25, p > 0.15$.

Planned comparisons were then conducted on each of the dimensions for which we found significant main effects. These comparisons revealed that changes to traffic aircraft were detected 4.7 seconds faster than to weather systems, $t(10) = 4.59, p = 0.001$, and relevant changes were detected 3.7 seconds faster than irrelevant changes, $t(10) = 2.76, p = 0.02$. Recall that irrelevant changes were no different than neutral changes in the location and confirmation categorizations. The effect for relevance was replicated in the location and confirmation categorizations. Changes that affected both the

Table 1. Mean change detection response time for each change category and type. Note that neutral changes in location and confirmation categories are the same by definition, and therefore have the same reported means.

Change Classification	Change Type	Mean Detection RT (sec.)
<i>Hazard</i>	Plane	16.9
	Weather	21.6
<i>Parameter</i>	Heading	19.0
	Altitude	18.3
<i>Location</i>	Chosen Path	16.1
	Unchosen Path	18.3
	Neutral	20.8
<i>Relevance</i>	Relevant	17.1
	Irrelevant	20.8
<i>Confirmation</i>	Confirming	18.3
	Refuting	15.9
	Neutral	20.8

chosen and unchosen path were detected significantly faster than neutral changes ($t(11) = 2.56, p = 0.03$ and $t(11) = 1.93, p = 0.08$, respectively) and both confirming and refuting changes were detected significantly faster than neutral changes ($t(11) = 1.96, p = 0.08$ and $t(11) = 4.05, p = 0.002$, respectively). In comparing detection response times for changes to the chosen and unchosen flight path, no significant difference was found ($t(11) = 1.42, p > 0.15$). In a similar manner, confirming and refuting changes were not found to significantly differ ($t(11) = 1.36, p > 0.15$), though an examination of the response time means shows that refuting changes were detected 2 seconds faster than confirming changes. This finding does *not* support the confirmation bias described previously.

DISCUSSION

The goal of the present study was to propose and investigate the link between the prevalence of plan continuation errors and the ability of the pilot to accurately monitor the chosen flight plan and to then determine how physical features of changing elements or pilot expectancies, beliefs, and goals affect the accuracy of the monitoring task.

Pilots were found to commit plan continuation errors on nearly one-third of the trials in which we initiated a change to threaten the safety of their flight path. This finding is consistent with the prevalence and severity of errors found in the accident statistics discussed previously (NTSB, 1994).

Previous research has shown that plan continuation errors result, in part, from poor situation awareness (Goh & Wiegmann, 2001b). This finding was replicated in the present study, which showed that pilots were not proficient at detecting the severe changes in hazards that threaten the safety of their flight path, a task that corresponds directly to Endsley's (1988) perception stage of situation awareness. Results revealed that pilots were 15% less likely to commit a plan continuation error when they correctly detected the change that threatened the safety of their flight path than if they did not, showing that poor situation awareness was a cause in committing such errors. These results pointed to the hypothesized link between the stages of plan monitoring and plan revision and further exploration was conducted to determine the factors that mediate successful plan monitoring and change detection.

Change detection rates reached levels of nearly 50%, showing that over half of the changes went undetected and suggesting that plan monitoring performance leaves much to be desired. Though such rates seem low, they are similar to those found in change detection research (cf. Rensink, 2002; Simons, 2000). Performance improved, however, with changes along several dimensions.

The success of plan monitoring performance was determined by both the physical characteristics of the change and the expectancies and beliefs of the pilot. Changes to traffic aircraft were detected nearly 5 seconds faster than to weather systems, suggesting that pilots were visually sampling aircraft hazards more frequently than weather. This finding is consistent with visual sampling theory (Senders, 1967;

Wickens, Xu, Helleberg, & Marsh, 2001) in that traffic aircraft moved faster than weather. Performance in detecting changes to heading and altitude did not significantly differ, though performance was especially poor in detecting changes to the heading of weather systems, suggesting that monitoring performance may be particularly vulnerable in these cases.

Pilot beliefs, goals, and expectancies were also found to significantly affect change detection performance. Changes that were relevant to the pilot's flight planning task were detected nearly four seconds faster than irrelevant changes, replicating previous research on change detection in applied contexts (Podczerwinski et al., 2001; Pringle et al., 2001).

The final change categorization examined the presence of the confirmation bias in plan monitoring. Theories of the confirmation bias would predict that the pilot would seek (and therefore be more likely to detect) changes that confirmed the greater safety of the chosen flight path, either by increasing the safety of this path or decreasing the safety of the unchosen path. In contrast, we did not find support for the confirmation bias and indeed found a trend that refuting changes were detected more quickly than confirming changes.

How do we resolve the findings that many pilots showed the plan continuation error syndrome reflective of the confirmation bias (31% of the trials), with the fact that the change detection data indicated equal, if not greater, vigilance for disconfirming, rather than confirming changes, a pattern that is at odds with the results of several studies on the confirmation bias? Indeed these two trends in the current data are not irreconcilable. The plan continuation error was only demonstrated on one-third of the trials in which it could be directly measured, and was therefore not a behavior pattern that was reflective of the general population. In contrast, the statistical analyses of change detection, which by definition reflects the 'typical' behavior of the population, showed that the typical population was *not* less sensitive to refuting changes. Thus the data suggest that, while the pilot population in general may not fall prey to the confirmation bias, a substantial minority do so (a minority reflected in the accident statistics) (Wickens, 2001), and at least some of this tendency can be attributed to a lack of sensitivity to all environmental changes, including both confirming and disconfirming ones.

Conclusions

The results from this study suggest that the inability to adequately monitor flight plans is a contributing cause to plan continuation errors. Visual attention was particularly driven to information that was relevant to the flight planning task, suggesting that pilots can successfully distinguish between relevant and irrelevant information. The general population of pilots did not succumb to the confirmation bias during plan monitoring as might have been suspected, although evidence suggested that in some individual cases, pilots did show lowered sensitivity to refuting information. Less encouraging, however, is the finding that pilots failed to detect nearly half of the changes that occurred within the airspace, particularly when the alterations were to weather systems. Such poor detection rates may be worsened under competition for visual

resources or automation-induced visual tunneling, potentially leading to an increase in the frequency of plan continuation errors. These issues will be investigated in future studies.

ACKNOWLEDGMENTS

The authors wish to thank Sharon Yeakel for her programming skills and two anonymous reviewers for their comments. Todd Kowalski assisted in data collection. Funding for this research was through a grant from the NASA Ames Research Center (NASA NAG 2-1120), under technical monitor Dr. David Foyle.

REFERENCES

- 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*, 22-25.
- Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, 70, 193-242.
- Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the 32nd Annual Human Factors Society Meeting* (pp. 97-101). Santa Monica, CA: Human Factors Society.
- 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: 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. (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*, 26-29.
- McCoy, C. E., & Mickunas, A. (2000). The role of context and progressive commitment in plan continuation error. *Proceedings of the IEA 2000/HFES 2000 Congress*, 26-29.
- Mumaw, R. J., Sarter, N. D., & Wickens, C. D. (2001). Analysis of pilots' monitoring and performance on an automated flight deck. In *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: The Ohio State University
- 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). Mahwah, NJ: Lawrence Erlbaum.
- 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. L., Irwin, D. E., Kramer, A. F., & Atchley, P. (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.
- Scholl, B. J. (2000). Attenuated change blindness for exogenously attended items in a flicker paradigm. *Visual Cognition*, 7, 377-396.
- Senders, J. W. (1967). On the distribution of attention in a dynamic environment. *Acta Psychologica*, 27, 349-354.
- Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, 7, 1-15.
- Thomas, L. C., & Wickens, C. D. (2000). *Effects of display frames of reference on spatial judgments and change detection* (Technical Report ARL-00-14/FED-LAB-00-4). Savoy, IL: University of Illinois, Aviation Research Lab.
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 12, 129-140.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. London: Batsford.
- Wickens, C. D. (2001). Attention to safety and the psychology of surprise. Keynote address at *The 11th Annual International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.
- Wickens, C. D., Xu, X., Helleberg, J., & Marsh, R. (2001). Pilot visual workload and task management in freeflight: A model of visual scanning. In *Proceedings of the 11th Annual International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.
- Wilson, D. R., & Fallshore, M. (2001). Optimistic and ability biases in pilots' decisions and perceptions of risk regarding VFR into IMC. *Proceedings of the 11th Annual International Symposium on Aviation Psychology*. Columbus, OH: Ohio State University.