# Beyond Automation Surprises: A Simulator Study of Disturbance Management on Highly Automated Flight Decks

Mark I. Nikolic

Nadine B. Sarter

# BEYOND AUTOMATION SURPRISES: A SIMULATOR STUDY OF DISTURBANCE MANAGEMENT ON HIGHLY AUTOMATED FLIGHT DECKS

**Mark I. Nikolic**
The Ohio State University
Department of Industrial and Systems Engineering
Columbus, OH

**Nadine B. Sarter**
The University of Michigan
Center for Ergonomics
Ann Arbor, MI

Error prevention alone will never be sufficient for improving safety in complex high-risk systems, such as aviation. This approach needs to be combined with better support for error and disturbance management which, in turn, requires an improved understanding of current strategies for coping with errors and the resulting disturbances to the flight. The present research has sought systematic empirical evidence to expand our understanding of the disturbance management process on modern flight decks. A simulator study was conducted with twelve B747-400 airline pilots in order to examine (the effectiveness of) their strategies for diagnosing and recovering from disturbances, and the impact of current automation design on these processes. Pilots flew a one-hour scenario (with a confederate copilot) which contained challenging events that probed pilots' knowledge of, and proficiency in, using the autoflight system. A process tracing methodology was used to analyze and identify patterns in strategies across pilots. Overall, pilots completed the scenario successfully but varied considerably in how they coped with disturbances to their flight path. Our results show that aspects of feedback design delayed the detection, and thus escalated the severity, of a disturbance. Diagnostic episodes were rare due to pilots' knowledge gaps as well as time-criticality. Our findings can inform the development of design and training solutions to observed difficulties with error and disturbance management in a variety of domains.

## Introduction

Human error is cited as the cause, or a contributing factor, in the majority of aviation incidents and accidents (e.g., Boeing, 1994). Yet the very low accident rate in this domain illustrates that aviation exhibits a strong degree of error resilience thanks to pilots' successful management of their errors and associated disturbances to the flight. In highly complex, dynamic, and event-driven domains, such as aviation, operators often need to manage consequences of breakdowns in human-machine performance that interact, cascade and escalate over time while continuing to maintain the ongoing process (such as flying the airplane). This activity can be characterized as disturbance management, since, from a practitioner's perspective, the potentially negative system effects of an error are more relevant than the error per se.

Disturbance management refers to the activity of diagnosing the underlying source(s) of a disturbance (i.e., a deviation from a desired state) in parallel with coping with the disturbance itself by maintaining the integrity and goals (i.e., efficiency, safety) of an underlying dynamic process (Woods, 1988). In the aviation domain, for example, a pilot needs to diagnose the source (for example, an erroneous input to the FMS) of an observed disturbance (such as a deviation from the flight path) and cope with the disturbance (by bringing the airplane back on course) while maintaining the integrity of the underlying process (i.e., while continuing to fly the airplane). While disturbance management is usually discussed in the context of system faults, the same activities tend to be involved in handling the consequences of breakdowns in the interaction between humans, machines, and the complex dynamic environment in which they collaborate. We will therefore use the term "disturbance management" to refer to pilots' efforts to cope with the effects of automation-related erroneous actions and assessments.

Despite the importance of disturbance management for system safety, few studies have examined its components in real-world dynamic environments (for some examples, see Klinect et al., 1999; Woods, 1984). The majority of work in this area has focused on error detection, leaving unanswered questions about the other stages of disturbance management (i.e., diagnosis and recovery). Earlier studies suggest that diagnosis does not necessarily occur or precede recovery during dynamic disturbance management (Kanse and Schaaf, 2001). Furthermore, an examination of how technological tools shape disturbance management seems to be missing from most earlier efforts.

.In the context of pilot-automation interaction and performance breakdowns on modern flight decks, our goal was to determine whether, and under what circumstances, pilots attempt to diagnose before they respond, and to what extent diagnosis is required for successful disturbance handling. Another objective was to examine the range of recovery strategies used, especially when they are influenced by the design of flight deck automation.

## Methods

As the final step in a research program that included jump-seat observations, a flight instructor survey, and an incident database analysis, a high-fidelity simulator study was conducted with type-rated airline pilots in order to examine error and disturbance management in a semi-controlled full-mission flight simulation context.

*Participants*

Pilot volunteers were recruited from two major U.S. carriers and one airplane manufacturer. Twelve type-rated Boeing 747-400 pilots (11 current, 1 recently retired; mean hours on type = 3837.75, SD = 2478) participated in the study and were paid $100 for their involvement.

*Simulator*

The simulation was conducted on a fixed-base 747-400 flight simulator. The 747-400 is a highly automated four-engine long-haul passenger aircraft. The simulator was equipped with fully functional displays and control interfaces. An Evans & Sutherland ESIG 3350 image generation system rendered a panoramic out-of-window visual scene which covered 45° horizontally and 34° vertically for each pilot.

*Procedure*

After briefing the flight with the experimenter and reviewing all flight-related paperwork, the participating pilot joined the confederate pilot in the simulator. The confederate knew the purpose of the study, occupied the right (co-pilot) seat, and helped ensure that scenario events occurred as designed. The confederate pilot was instructed not to be overly proactive in helping participating pilots detect their errors. However, he was instructed to intervene (by directing the participant's attention) if the detection delay jeopardized the experimenter's likelihood to observe a recovery. The confederate was also asked to elicit pilots' reasoning about problems by asking relevant questions to expose the pilot's intentions and reasoning. Interactive air traffic control was provided by the experimenter/observer to help ensure the proper evolution of the scenario by issuing planned and improvised clearances. After reviewing the planned route and the current state of the aircraft, the scenario began in-flight with the aircraft level at 9000 feet, during the initial climb-out phase. The scenario ended once the aircraft landed at Los Angeles and came to a complete stop on the runway. The pilot then remained in the simulator cab and was debriefed by the experimenter for another 30-60 minutes.

*Scenario*

All participants flew the same one-hour daytime scenario from San Francisco to Los Angeles in the role of pilot-in-command. Weather throughout the scenario was clear with minimal winds. Based on data gathered from our earlier survey, observations, and consultations with domain experts, several scenario events were designed that created a high probability of observing automation-related disturbances by placing heavy knowledge and attentional demands on pilots resulting in the potential for breakdowns in human-machine communication and coordination. Since errors and disturbances were not introduced through experimenter-induced system failures or unrealistic clearances, they were not necessarily observed for each pilot on each event.

*Selected Scenario Events*

Because of space limitations, this paper will present results from two of the events that were used in the scenario.

*LNAV Capture.* After crossing PESCA, ATC instructed the aircraft to continue on a 140 degree heading instead of turning left to continue on the flight plan. As a result, the aircraft will not physically cross the next two waypoints that are programmed into the Flight Management Computer (FMC) and are kept in the route. Thus, if the route is not reprogrammed by the pilot, the autopilot will attempt to return to these waypoints and result in unwanted aircraft behavior when the pilot attempts to rejoin the course by activating the LNAV mode.

*VNAV ALT Mode.* In order to begin an automated descent, the autoflight system must be in the 'VNAV PTH' mode. However, in our scenario, the automation was likely to enter the 'VNAV ALT' mode due to cruise altitude changes given by air traffic control. If the pilot does not actively change the mode back to ‘VNAV PTH’ (typically, by changing the cruise altitude in the CDU interface of the Flight Management System and then pushing the altitude knob), the aircraft will not descend as expected at the top-of-descent (TOD) point, and may potentially miss an altitude target. This event could elicit a mode error due to either incomplete system knowledge or a monitoring breakdown, and could have resulted in an altitude violation if it was not detected and corrected in a timely manner.

*Data Collection and Analysis*

Multi-angle video and audio recordings were made to assist in recreating verbal and behavioral protocols. This information was supplemented by an observer, who sat directly behind the pilots in the simulator cab and noted pilot responses to events. Upon completion of the scenario, the participating pilot was debriefed by the experimenter in order to review and clarify any ambiguities about his scenario performance and to probe participants’ knowledge of the automated flight system. These sources of data were combined to form a coherent process trace (Woods, 1993) of participant behavior which can be compared across participants as well as to canonical or “standard” recovery paths for each event.

## Results

All twelve pilots completed the scenario for this study “successfully” in the sense that they all made a safe landing. However, every pilot struggled at some point with handling events during the simulated flight, and every scenario unfolded in a unique way because pilots used a variety of strategies for managing events and recovering from disturbances. When possible, a canonical solution path was defined by a subject matter expert for the event. This path represented the most efficient but not necessarily the only correct or successful sequence of pilot actions for the event. It provides a single frame of reference from which to compare performance across pilots.

*LNAV Capture*

This event examined how pilots recovered their original course after an air traffic control clearance caused them to bypass two of the waypoints on the original route. As a consequence, the FMS continued to consider them as “active” (i.e., as valid targets) since the airplane never came close enough for them to be removed by the automation’s logic. As a result, pilots who re-activated LNAV to resume the course without first modifying the route in the CDU caused the airplane to turn off-course to a 090 heading instead of the 070 heading as instructed by ATC.

All pilots eventually managed to recover their course, although minor deviations occurred for two pilots (see Figure 1). After receiving the ATC clearance to intercept their normal course via a 070 heading, all pilots made the initial turn using the HDG SEL mode (a lateral mode at a low level of automation). The recovery processes from that point on fall into two categories. One group of five pilots reprogrammed the route prior to engaging the LNAV mode (represented on the lower half of Figure 1). A second group of seven pilots activated the LNAV mode without updating the original route in the FMS. In general, the group that reprogrammed the route first was more aware of the current state and logic of the automation. These pilots recognized that the FMS route contained “stale” information that was no longer applicable to the new context.

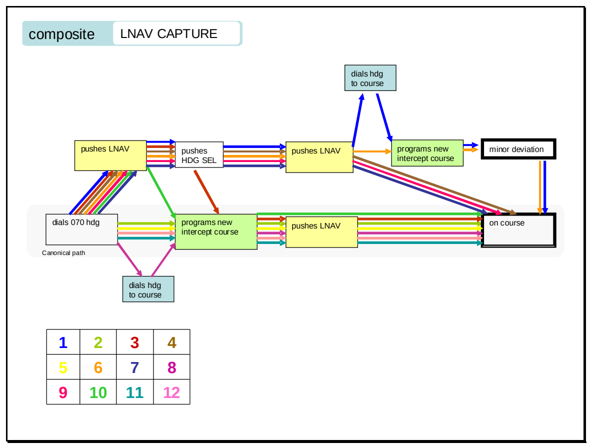


**Figure 1.** *Composite of abstracted solution paths for LNAV event. Each line represents one pilot and is color-coded for all pilots.*

For the second group of pilots, it was not immediately obvious that there was a problem because the incorrect FMS route produced aircraft behavior that was initially consistent with pilot expectations. Since the FMS believed that the floating waypoint (a turn to a 090 degree heading) was the current target, activation of the LNAV mode resulted in a turn in the expected direction (left) but not to the assigned heading of 070 degrees. This confirming cue initially masked the problem and led some pilots to assume the aircraft was on the correct course.

Six of the seven pilots in the second group (top portion of Figure 1) recovered the correct heading by reverting back to the HDG SEL mode, after detecting either the unexpectedly rapid engagement (or “capture”) of the LNAV mode or the subsequently incorrect heading of 090 degrees that was commanded by the autopilot. Of the six pilots that recovered with HDG SEL, one of them detected the active waypoint mismatch at this point and reprogrammed the route, while five of them reattempted to engage the LNAV mode, again, without reprogramming the route. This repeat strategy worked for three pilots but it worked by chance, since enough time and distance had elapsed for the FMS to automatically advance to the next waypoint and thus for the route to be corrected.

None of the seven pilots who prematurely engaged the LNAV mode was able to explain the cause of the unexpected behavior prior to beginning recovery actions. The debriefing confirmed that the seven pilots who did not understand the observed LNAV behavior were either unaware of which waypoint was active during the event and/or were generally unfamiliar with floating waypoints and their effect on arming the LNAV mode after a deviation. One pilot believed that the unexpected LNAV behavior was a "malfunction."

*VNAV ALT Mode*

The canonical path for handling this event involves two steps: 1) entering the new cruise altitude into the FMS, and 2) pushing the altitude knob to make the FMS accept the new value (Figure 2). Completing these actions results in the activation of the VNAV PTH mode, which is necessary to achieve the desired descent profile. Otherwise, the automation remains in the VNAV ALT mode.

The event was "successful" in the sense that the VNAV ALT mode became active during cruise in ten of the twelve cases. In the other two cases, the pilots (4 and 10) proactively reprogrammed the FMS prior to reaching the new cruise altitude and went directly to the VNAV PTH mode. Three of the ten 'VNAV ALT' pilots (1, 2, and 3) successfully returned to the VNAV PTH mode by completing the canonical path.

The solution path for Pilot 3 is an example of a pilot who recovered the VNAV PTH mode from the VNAV ALT mode, though using an extraneous sequence of actions in addition to the canonical path . This strategy was described later by the pilot as "pushing buttons until it worked" and "resetting" the system, but also reflected incomplete knowledge of how to deal with this problem efficiently. Although the pilot did not understand why he was in the wrong mode, he knew that it was incorrect, and worked to resolve that discrepancy.

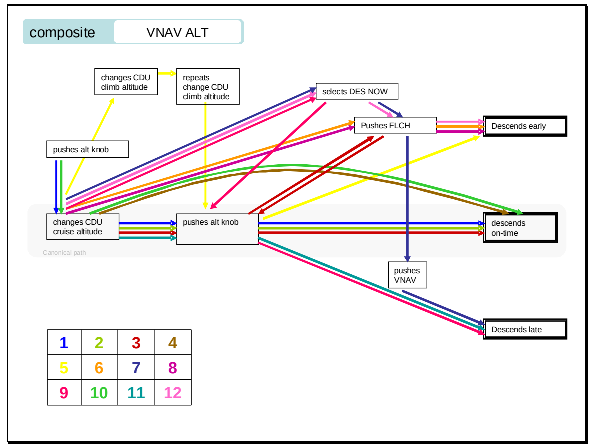


**Figure 2.** *Composite of abstracted solution paths for VNAV ALT event. Each line represents one pilot and is color-coded for all pilots.*

The other seven pilots remained in the incorrect mode (VNAV ALT) for a majority of the cruise phase. Note that there were no observable consequences of being in the VNAV ALT mode during this phase, since the aircraft was flying at a level altitude. Unwanted consequences would only appear when the aircraft reached the TOD point, approximately 20 minutes later, and would fail to begin the descent, creating the potential for the aircraft to miss programmed altitude restrictions. In the debriefing, all of these pilots were found to have gaps in their knowledge related to the functioning of the VNAV PTH and VNAV ALT modes. Interestingly, four of these seven pilots (5, 6, 8, and 12) avoided the consequence of the incorrect automation setting – the failure to descend automatically at the TOD – by deciding to descend earlier than the TOD point. In other words, the gaps in their mental model of the VNAV mode were either masked or worked-around by their early descent strategy, which they stated was based on the desire to alleviate workload during the descent. In contrast, three pilots (7, 9, and 11) remained in VNAV ALT at the TOD, and the aircraft did not descend as they had intended.

Of the three pilots (7, 9, and 11) who did not initiate an early descent, two recognized quickly that the aircraft had not started to descend and recovered by engaging the FLCH mode. One pilot (Pilot 11) was distracted with arrival preparations for almost 10 minutes after passing the TOD point, and recovered late by engaging the FLCH mode. During the event, none of these three pilots were able to explain why the aircraft did not descend as expected, suggesting an incomplete understanding of the automation that was later confirmed during the debriefing.

## Discussion

All pilots completed the scenario "successfully" in the sense that they managed to complete the flight and land safely. At the same time, all participants experienced at least one disturbance during the course of the scenario. Note that these disturbances did not result from system faults. Rather, potentially unproblematic events were "managed" into disturbances from which pilots then had to recover.

One important goal of the current study was to explore the need for, and the effectiveness of, diagnosing errors and disturbances in the context of dynamic event-driven systems. In the present study, pilots rarely attempted to diagnose the source of a disturbance, except in two unsuccessful cases (two different pilots during two different events) in which pilots remained fixated on an incorrect diagnosis. This finding is in agreement with earlier findings from other dynamic domains where the absence of diagnostic activities was explained by time pressure and the need for immediate recovery to avoid negative consequences (Kanse and Schaaf, 2001; Kontogiannis, 1999; Reason, 1990). While time pressure and the immediate need to recover from disturbances (i.e., in cases of impending or actual deviations from assigned routes or altitudes) may have precluded diagnosis in many cases, it was also absent from contexts that were not time critical (i.e., the majority of the cruise phase in the VNAV ALT event). This may, in part, be explained by considerable knowledge gaps in pilot mental models of the automation which were observed in earlier research (Sarter and Woods, 2000; Mumaw et al., 2000) and confirmed in this study. For example, nine of 12 pilots in this study were found to have incomplete or inaccurate knowledge of the vertical navigation (VNAV) submodes of the FMS. These misconceptions – which were sometimes masked by serendipitous pilot actions that produced apparently seamless performance - likely contributed to problems with detecting, diagnosing and recovering from disturbances, and in some cases, even exacerbated the existing disturbance.

While the absence of diagnostic activities did not result in catastrophic outcomes, it may have affected the success and efficiency of recovery. In most cases, pilots used generic recovery strategies (repeating actions or resetting the automation) or engaged in trial-and-error behavior, rather than developing and implementing a problem-specific solution. In most cases, these generic recovery strategies, and also the observed tendency to use high levels of automation to manage disturbances (contrary to what is typically prescribed by training), were not successful and instead led to a delay in recovery, which further exacerbated the disturbance.

After detecting the disturbance in the LNAV Capture Event, pilots commonly resorted first to a "quick fix" by reverting to a lower-level mode (HDG SEL) in order to immediately correct the heading. This choice was likely prompted by the urgency of this disturbance which, over time, was producing an escalating divergence between the required and actual course. The use of such quick-fixes has been observed by other authors (Kontogiannis, 1999; Kanse and Schaaf, 2001) in process control domains. In those cases, they served to stabilize a situation in order to allow for an analysis of the problem and/or more thorough corrective actions. In contrast, five pilots in our scenario followed the "quick fix" with just a generic repetition of the LNAV engagement, without any further analysis or modification of the automation's instructions.

The repetition strategy - observed primarily in the LNAV event - seemed to be based on pilots' erroneous belief that the original action was appropriate but that the automation, for some reason, did not accept the pilot's input or execute the command as intended. This example illustrates that coincidentally successful strategies can lead to erroneous beliefs which can become incorporated into a pilot's mental model of the system. As a result, pilots may develop misrepresentations of functional system architecture that can lead to miscalibration of their system knowledge.

The resetting strategy - observed for 2 pilots during the VNAV ALT event - appears to be a type of workaround that did not require deeper system knowledge of how the disturbance occurred or how to avoid it in the future. Interestingly, both repetition and resetting strategies were observed by Plat and Amalberti (2000) in a simulator study of pilot responses to experimenter-induced software "bugs" or malfunctions in the behavior of the flight deck automation. This suggests that some pilots in our scenario treated disturbances as if they were discrete malfunctions which were unavoidable (i.e., not attributable to their actions) and required only generic fixes that did not require accurate or detailed system knowledge. However, these strategies can be brittle in that they may work in some contexts, but may not be effective in others, especially in unforgiving environments.

In addition, our findings indicate that disturbance management was not always well-supported by the available feedback to pilots. In the case of the VNAV ALT event, pilots were unable to visualize the implications of the active mode for the descent since there is no predictive vertical profile display. Instead, pilots are shown only a symbol and adjacent alphanumeric label ("T/D") representing the top-of-descent point on the map display. Aside from an alphanumeric mode annunciation on the PFD (i.e., "VNAV ALT") they receive no salient indication on the map display of whether the top-of-descent will be honored by the system. As a result, the current feedback may contribute to delays in detecting the error, which in turn, allow the disturbance in the aircraft's profile to escalate.

## Conclusion

Error prevention alone will never be sufficient for improving safety in complex high-risk systems. Rather, a deeper understanding is needed of how human operators cope with the consequence of inevitable errors and thus the disturbances to the processes they monitor and control. The problems of inadequate feedback of autonomous system changes have been widely discussed (Sarter and Woods, 1995; Wiener, 1989) and have also been observed in the current study. However, these problems have often been discussed in the context of detecting the existence of an erroneous setting (e.g., "mode awareness" and "automation surprises"). Observations of pilot performance in the present study have shown that current automation design not only delays detection, but is too ambiguous for diagnosis, and does not support operators in recovering from disturbances in the most optimal way. Continued efforts in this area will inform the design of cognitive tools that effectively support this process.

## Acknowledgements

The authors would like to thank Jack Bard, Randy Mumaw, Jack Butcher, Steve Kimball, and the airlines and pilots that participated in the study. This work was funded by a grant from the Federal Aviation Administration (Grant # **DTFA 96-G-043;** Technical monitor: Dr. Tom McCloy).

## References

Boeing Commercial Airplane Group. (1994). *Statistical summary of commercial jet aircraft accidents: Worldwide operations 1959-1993* (Boeing Airplane Safety Engineering B-210B). Seattle, WA: Boeing.

Kanse, L. and Schaaf, T. (2001). Recovery from Failures in the Chemical Process Industry. *International Journal of Cognitive Ergonomics*, *5*(3), 199-211.

Klinect, J.R., Wilhelm, J.A., & Helmreich, R.L. (1999) Threat and error management: Data from line operations safety audits. In Jensen, R. S. (Ed.), *Proceedings of the Tenth International Symposium on Aviation Psychology*, Columbus, OH: Ohio State University.

Kontogiannis, T. (1999). User strategies in recovering from errors in man–machine systems. *Safety Science, 32*, 49-68.

Mumaw, R.J., Sarter, N., Wickens, C., Kimball, S.,Nikolic, M., Marsh, R., Xu, W., & Xu, X. (2000). *Analysis of pilots' monitoring and performance on highly automated flight decks*. (Final Project Report: NASA Ames Contract NAS2-99074). Seattle, WA: Boeing Commercial Aviation.

Plat, M., & Amalberti R. (2000). Experimental crew training to deal with automation surprises. In N.B. Sarter and R. Amalberti (Eds.), *Cognitive Engineering in the Aviation Domain* (pp. 287-308). Hillsdale, NJ: LEA.

Reason, J. (1990). *Human Error*. New York: Cambridge University Press.

Sarter, N.B. and Woods, D.D. (1995). "How in the world did we ever get into that mode?" Mode Error and Awareness in Supervisory Control. *Human Factors, 37(1),* 5-19.

Sarter, N.B. and Woods, D.D. (2000). Teamplay with a Powerful and Independent Agent: A Full-Mission Simulation Study. *Human Factors*, *42*(3), 390-402.

Wiener, E.L. (1989). Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft. *NASA Contractor Report 177528*. Moffett Field, CA: NASA Ames Research Center.

Woods, D.D. (1984). Some results on operator performance in emergency events. In D. Whitfield, editor, *Ergonomic Problems in Process Operations*, Inst. Chem. Eng. Symp. Ser. 90, 21-31.

Woods, D.D. (1988). Coping with complexity: The psychology of human behavior in complex systems. In: Goodstein, L. P., Andersen, H. B., and Olsen, S. E. (Eds.), *Tasks, Errors, and Mental Models*. New York: Taylor & Francis.

Woods, D.D. (1993). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In G. Kln, J. Orasanu, & R. Calderwood (Eds.), *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing Corporation.

Woods, D.D. (1994). Cognitive Demands and Activities in Dynamic Fault Management: Abduction and Disturbance Management. In N. Stanton, editor, *Human Factors of Alarm Design*, Taylor & Francis, London, 1994.

Woods, D.D., Johannesen, L.J., Cook, R.I., & Sarter, N.B. (1994). *Behind human error: Cognitive systems, computers, and hindsight.* WPAFB: CSERIAC.