FLIGHT CREW TASK MANAGEMENT
IN NON-NORMAL SITUATIONS

Paul C. Schutte
Anna C. Trujillo
Langley Research Center
National Aeronautics and Space Administration
Hampton, Virginia

Task management (TM) is always performed on the flight deck, although not always explicitly, consistently, or rigorously. Nowhere is TM as important as it is in dealing with non-normal situations. The objective of this study was to analyze pilot TM behavior for non-normal situations. Specifically, the study observed pilots' performance in a full workload environment in order to discern their TM strategies. This study identified four different TM prioritization and allocation strategies: 'Aviate-Navigate-Communicate-Manage Systems;' 'Perceived Severity;' 'Procedure Based;' and 'Event/Interrupt Driven.' Subjects used these strategies to manage their personal workload and to schedule monitoring and assessment of the situation. The 'Perceived Severity' strategy for personal workload management combined with the 'Aviate-Navigate-Communicate- Manage Systems' strategy for monitoring and assessing appeared to be the most effective (fewest errors and fastest response times) in responding to the novel system failure used in this study.

INTRODUCTION

There are four primary functions performed by the flight crew in civil transport aircraft to accomplish its mission. These functions are flight management, communications management, systems management, and task management (Abbott 1993). Flight management involves what is generally known as aviation (primary control of the aircraft) and navigation. Communications management is concerned with all communications occurring on the flight deck, including communications with air traffic control and with other crew members. Systems management occurs when the system that the flight crew is interacting with must be monitored or managed in performing a function.

"Task management is the first-level function of managing tasks and associated resources involved in conducting the mission. This is both a supervisory and a supporting function to the other three major flight deck functions. This function involves monitoring, scheduling, and allocating the tasks and task resources between and for each major function" (Abbott & Rogers 1993). Of the four functions, task management (TM) is perhaps one of the most underrated in that it has received the least amount of attention in flight deck design and airline training. TM is always performed on the flight deck, although not always explicitly, consistently, or rigorously. While some aids — such as procedures, checklists, and levels of alerts — do assist the crew in managing specific tasks and workload, the aids do not address the problem in a more general sense. For example, a procedure for shutting down an auxiliary power unit may not account for incoming air traffic control (ATC) calls or approach procedures. Likewise a standard approach procedure may not have a contingency for dealing with an oil system failure. It is the job of the flight crew to merge these procedures together in a safe and efficient manner. However,

little training is given on how to do so. Pilots often use rules of thumb — such as "first aviate, then navigate, then communicate, then administrate" — to prioritize tasks. These rules of thumb are helpful in many cases but are sometimes inappropriate for specific contexts, and are too general to help in others. Raby and Wickens (1994) concluded that "different pilots clearly differed in their scheduling" and that pilots prioritize tasks when faced with high workload situations.

Nowhere is TM as important as it is in dealing with non-normal situations. Here, the workload of the flight crew members is generally increased and they are required to respond to abnormalities that may affect several functions. For example, a systems failure may also affect the planned destination and aircraft handling (flight management) and may precipitate the need for communications. The crew often has to reschedule and reallocate a number of tasks and resources based on their assessment of the criticality of the problem. Non-normal situations require the flight crew to devote more time and attention to the function related to the cause of the non-normal condition (for example, systems management if an engine has failed, or flight management if a storm is encountered). However, it is rare that this function is the only one affected (Rogers, Schutte, Latorella, in press). TM has a large role to play in how the crew's attention is dispersed among the different functions.

OBJECTIVE

The objective of this study was to analyze pilot TM behavior in response to systems failures for which there was no specified procedure. Specifically, the study observed pilots' performance in a full workload environment to discern their TM strategies (if any existed).

METHOD

Subjects

Sixteen airline pilots from 4 major US airlines served as test subjects. Each of the subjects was rated in Extended Twin-engine Operations (ETOPS). The subjects were tasked with acting as Captain/Pilot-Not-Flying on an ETOPS oceanic flight. The over-water route was chosen because it forced the subject to stay aloft and make decisions as opposed to simply landing the aircraft at the first suitable airport. Also, ETOPS operations have specific rules which state that in the event of having only one operational primary system (e.g., engine or hydraulic) the aircraft must proceed immediately to the closest suitable alternate. A confederate first officer acted as Pilot-Flying. The confederate was fully capable in normal operations and was obedient to the Captain; however, he offered no help in decision making or situation assessment.

Apparatus

This study was conducted in a transport-class flight-deck simulator which was designed to provide full mission and full workload capability in a two-crew layout. The simulator's primary flight displays, navigation map displays, autopilot/flight director, and flight management system were designed to be analogous to current state-of-the-art transport aircraft such as the Boeing 747-400, the McDonnell Douglas MD11, and the Airbus A320. Radio communications with a confederate air traffic controller and confederate company dispatcher were provided.

Scenario

The scenario for this study was a large fuel leak near the engine. As in most modern aircraft, there is no appropriate checklist for a fuel leak. In this scenario, a large leak developed in the fuel line downstream of the fuel flow sensor but prior to the engine combustor. The location is important because it causes the fuel flow reading on the flight deck to appear normal for the throttle setting, while the thrust developed by the engine was abnormally low for the throttle setting. The fault indications — as seen by the subject — were an advance of the autothrottle (in order to maintain speed at reduced thrust), followed by a degradation in the engine thrust parameters (engine pressure ratio, engine rotational speeds, and exhaust gas temperature), followed by the inability of the aircraft to maintain altitude and speed. If nothing were done, the flight management system would alert the crew that there was a difference in the calculated fuel (based on fuel flow burn) and measured fuel in the tanks. The most a

ppropriate response for this failure was to shut down the engine in order to conserve fuel. With the engine shut down, the subject should command a descent, declare an emergency, and proceed to the nearest suitable alternate in accordance with ETOPS rules.

Measures

The subject's performance measures included factors such as response time, fault identification accuracy, and compensation accuracy. The subject's performance in non-systems tasks was also measured and included factors such as ETOPS compliance, route management, and fuel used.

The subjects were asked at regular intervals (approximately every 5 minutes) to state the tasks on which they were focusing. The subjects had three choices of response: systems (e.g., operations and status); spatial (e.g., altitude, aircraft position); or mission awareness (e.g., flight planning, fuel, weather). This measure was designed to provide some insight into their TM strategies. The subjects were instructed that they could postpone their response to the probes if they desired.

During the scenario, several "natural" conversational probes (e.g., "What happened back there?" "Could you give me any additional information that might help maintenance figure this out?") regarding the non-normal situation were made by the air traffic controller, the dispatcher, or the first officer. Audio-video recordings of the scenarios were made and transcribed to capture all of the utterances of the subjects.

The subject's performance was scored based on three criteria: correct diagnosis of the failure; correct system response to the failure; and correct mission response to the failure. For this scenario, the criteria were 'fuel leak,' 'engine shutdown,' and 'divert to alternate airport,' respectively.

The video tapes and transcripts were used to determine the management function that the subject was performing, and the TM strategy employed by the subject. The subject's utterances on the transcripts were scored according to a modified version of flight, communications, systems, and task management categories. TM strategies were attributed to each of the subjects subjectively by the authors based on review of the video tapes and the scoring data. A TM strategy was assigned based on the order and amount of time spent in each functional category and the subjects' TM statements about intent and priority. The TM strategies were compared with the subject's performance to see if there were any correspondences.

RESULTS

As mentioned earlier, there were no cautions or warnings or appropriate checklists for the fuel leak. Twelve of the subjects correctly identified the problem as a fuel leak. Nine of the sixteen subjects correctly shut down the engine, averting a low fuel situation. Of those nine, two did not divert to the closest alternate, in violation of ETOPS rules. Seven subjects did not shut down the engine which led to a low fuel situation (especially for the two subjects who did not divert). Table 1 shows a summary of the subjects' performance ordered by correctness. Subjects who did not shut down the engine received the lowest ratings since that left them in a fuel critical situation. These data were used to determine the effectiveness of the TM strategies derived below.

Two sets of data are the main emphasis of this paper. The first was their periodic assessment of their current focus based on the probes. The second was the transcripts and video tape of their performance. They were used to describe the subject's focus of attention during the scenarios.

The self assessment data showed that subjects were focusing their attention on systems an average of 46% of the time. This is not surprising, since the major events in the scenario were systems failures and the subjects were acting as pilot not flying, meaning they were charged with performing systems management duties. They stated that they were focusing on the mission an average of 36% of the time. Finally, only 18% of the their responses indicated that they were focusing on spatial awareness.

Correct System Response	Correct Mission Response	Correct Diagnosis	Number of Subjects
Yes	Yes	Yes	6
Yes	Yes	No	1
Yes	No	Yes	2
No	Yes	Yes	3
No	Yes	No	2
No	No	Yes	1
No	No	No	1

Table 1 - Summary of Subjects' Performance

For the second set of data, each of the subject's utterances were categorized into one of six functional categories. The categories were derived from Abbott (1993) and are described below. The flight management category was decomposed into aviate and navigate in accordance to Abbott's structure. After reviewing a few of the video tapes it became clear that an additional category was necessary for addressing the subject's questions about the simulation. These utterances were scored in a new category called, Flight Deck Systems.

Aviate - statements describing the Captain's manipulation or consideration of the motion of the aircraft; e.g., ascend, descend, turn left, turn right, pitch up, pitch down, roll.

Navigate - statements describing the Captain's manipulation or consideration of the aircraft route, change in destination, change in waypoints, point of equal time, etc. Also discussion of capabilities to complete the route (e.g., do we have enough fuel to make it to Honolulu?)

Communicate - since all utterances are a form of communication, this category had to be restricted to statements describing the initiation or termination of communications outside of the flight deck. The statements in-between the initiation and termination (i.e., the content of the communication) were scored using the appropriate management function.

System - any statements pertaining to the operation of airframe systems (i.e., engines, fuel, hydraulic, electrical, environmental).

Flight Deck Systems - any statements pertaining to the operation of flight deck instrumentation and automation that were peculiar to the simulator. This includes the FMS operation, display symbology, etc. Generally, these statements are in the form of questions about operation of the simulation and differences between the simulator and the aircraft that the subjects were used to flying.

Task - any statements pertaining to assigning duties, stating an intent or order of functions in the future, evaluation of resources, mission planning, etc.

Scoring the utterances in this manner provides some insight into how often such statements are made. Figure 1 shows the percentage of utterances for each category over all of the subjects. The standard deviations across subjects were no greater than 6.5%. The large percentage of Flight Deck Systems utterances was not surprising given that the subjects were unfamiliar with certain aspects of the aircraft simulator.
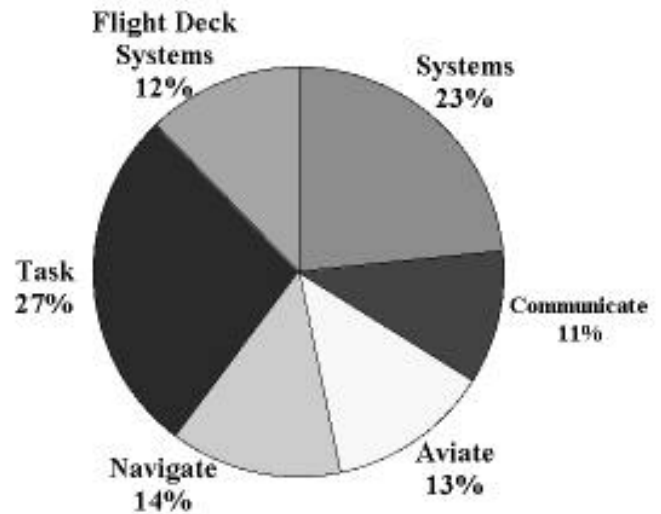

Figure 1. Categorization of Pilot Utterances

A review of time plots of the utterance categories showed the diversity of management functions that are addressed, acknowledged, or attended to over time. Figure 2 shows a sample plot of the utterance categories over time.
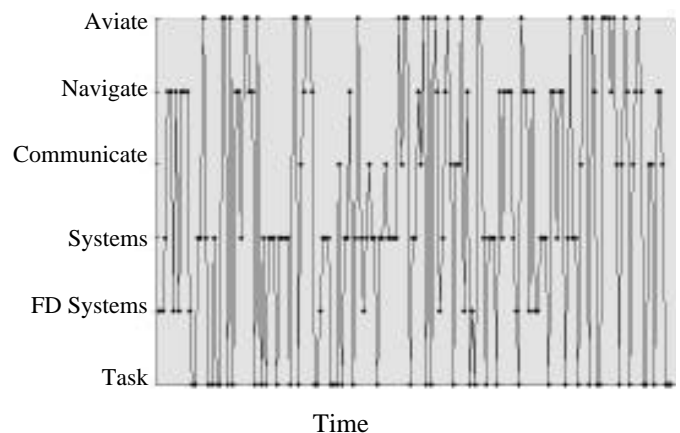

Figure 2 - Attention Transition Diagram

While it is tempting to use these plots to ascertain the subject's TM behavior, the plots alone do not provide sufficient data to do so. A single aviate utterance could be a deflection of an aviate task, a delegation of an aviate task, or a single step completion of an aviate task. An utterance could have been made to acknowledge a comment by the first officer or ATC, but not to attend to that comment.

In order to discern the TM strategy, careful reviews of the video tapes were required in conjunction with the scores. It is important to note that the strategies attributed to each of the subjects were assigned subjectively by the authors. A more rigorous approach in which independent raters score the utterances and categorize the strategies would be useful, but was beyond the scope of this project. Still, the results from this study bring out a number of interesting points. The authors recognized four different TM strategies that were employed by the subjects when faced with a non-normal system failure. The strategies represented how the subjects allocated resources and prioritized tasks. The first two strategies appeared to prioritize and allocate activities based on a goal-driven paradigm while the second two strategies appeared to be stimulus-driven. The four TM strategies follow.

Aviate-Navigate-Communicate-Manage Systems (ANCS) : subjects prioritize their attention with Aviate as the highest priority and Systems as the lowest priority. This prioritization occurs with little regard for the context of the situation.

Perceived Severity : subjects place highest priority on what they perceive to be the most threatening problem.

Procedure Based : subjects migrate towards tasks for which there are well defined procedures. These range from systems procedures to Federal Aviation Regulations (FARs).

Event/Interrupt (E/I) Driven : subjects' attention is given to a particular task based on an event or an interruption. They will typically continue pursuing that task until either the task is completed, the subject can do no more on the task, or until another event or interruption disrupts the task.

Subjects appeared to use the strategies for two distinct TM activities: monitoring and assessment of the situation; and personal workload management (i.e., where the subject's attention was focused). The monitoring activity comprises the 'assess current situation' and 'assess progress and status of active tasks' of Funk's (1991) Cockpit TM procedure. For example, one subject prioritized his personal workload based on Perceived Severity and monitored his situation based on ANCS strategy. When faced with the fuel leak, he first monitored to see that there were no Aviate, Navigate, or Communicate problems and that the First Officer had the airplane. Once the assessment was complete, he focused his attention on the most severe problem; i.e., the systems. Another subject, whose personal workload management was based on an ANCS strategy and monitoring was based on a Perceived Severity strategy, quickly assessed the systems problem to see if it was critical (e.g., engine fire).

Since it was not, he allocated most of his time to the aviate task, with the navigate and communicate tasks taking the lesser amounts of his time, respectively. When the other functional tasks were complete, he addressed the systems problem. Table 2 is a summary of the different TM strategies employed by the sixteen subjects ordered by their performance.

None of the subjects who had correct performance in all three measures (i.e., Correct Diagnosis / Correct System Response / Correct Mission Response) used the ANCS strategy or the Procedure Based strategy for their personal workload management. The seven subjects who did not shut down the engine (thus, placing them in a fuel critical situation) used the ANCS strategy for their personal workload management. ANCS was used for monitoring by five of the nine subjects who shut down the engine. There was no apparent correlation between the subject's airline and the strategies used.

SUB#	Personal Workload Management	Monitoring	Performance
S U B #	ANCS, Perceived Severity, Procedure Based, E/I Driven	ANCS, Perceived Severity, Procedure Based, E/I Driven	System Response/ Mission Response/ Diagnosis
1	E/I Driven	Procedure Based	Yes/Yes/Yes
2	E/I Driven	Perceived Severity	Yes/Yes/Yes
3	E/I Driven	ANCS	Yes/Yes/Yes
4	Perceived Severity	ANCS	Yes/Yes/Yes
5	Perceived Severity	ANCS	Yes/Yes/Yes
6	Perceived Severity	ANCS	Yes/Yes/Yes
7	ANCS	Procedure Based	Yes/Yes/No
8	Perceived Severity	ANCS	Yes/No/Yes
9	Procedure Based	Perceived Severity	Yes/No/Yes
10	ANCS	Procedure Based	No/Yes/Yes
11	ANCS	Perceived Severity	No/Yes/Yes
12	ANCS	E/I Driven	No/Yes/Yes
13	ANCS	Procedure Based	No/Yes/No
14	ANCS	Procedure Based	No/Yes/No
15	ANCS	Procedure Based	No/No/Yes
16	ANCS	Procedure Based	No/No/No

Table 2 - Summary of TM Strategy and Performance for the 16 Subjects

DISCUSSION

Perhaps the most important result is that TM appears to play a significant role in how the flight crew deals with an abnormality. The scoring of the utterances show that a large

percentage of the Captain/Pilot-Not-Flying's time in this experiment was spent in TM. However, the amount of training and procedures for TM is significantly less than for the other management categories. Thus, TM is largely dependent on individual differences between flight crews and personal style. This may be the only difference between an incident and an accident.

The strategy that appeared to be most useful in dealing with the systems failure presented in this experiment was to prioritize and allocate personal workload based on either Perceived Severity or E/I Driven strategy and to monitor based on an ANCS strategy. The subjects whose personal workload management was E/I Driven took longer to shutdown the engine (average 5320 seconds — standard deviation 510) than those based on Perceived Severity (average 3416 seconds — standard deviation 1297). Thus, the goal-driven combination of Perceived Severity/ANCS would appear to be the most useful for the role of Captain/Pilot-Not-Flying.

The success of the Perceived Severity/ANCS strategy combination is likely due to an appropriate emphasis being placed on the problems at hand and sufficient coverage from the strategy for monitoring (ANCS) to assure that all other flight deck tasks are assessed and performed. Subjects using the ANCS strategy as their monitoring seemed to cycle through the four categories to spot check for problems and to insure that all else was well.

Another reason that this strategy appears to work well is that the most severe problem (in this study, the fuel leak) is kept in the subjects' minds as they cycle through ANCS. Often, there is additional helpful information to be found by doing so. In the fuel leak scenario, navigation tasks such as position plots helped subjects identify excessive fuel usage prior to receiving the fuel disagree message from the flight management system. The aviate task showed that rudder was being deployed to compensate for the asymmetric thrust from the engines. Some subjects used this information to confirm that there was a real problem and not a sensor failure. Finally, the subjects used the communicate task with dispatch and with the flight attendants (e.g., "Could you go look out the window and see if there is any fluid streaming out of the engine?") to obtain additional information for dealing with the problem. Thus, the diversity of information provided by ANCS was beneficial in solving a problem in the systems.

Based on these results, it appears that using the ANCS for personal workload management was not helpful in dealing with the systems failure in this experiment. This is not surprising since systems is given the lowest priority in the ANCS strategy. It is possible that an aviate problem, such as traffic or weather, could be sufficiently handled by an ANCS strategy for personal workload management. However, it could be that an ANCS (personal workload) and Perceived Severity (monitor) strategy would lead the pilot to fixate on the aviate problem and miss a subsequent navigate or systems problem, whereas the reverse strategy could appropriately draw attention away momentarily to check on the status of those other areas. Controlled testing of these hypotheses would be required to definitively answer this question.

After reviewing the video tapes it became clear that interruptions played a significant part in the TM of the subjects — not only for those using E/I Driven strategies but for everyone. Some subjects had to work hard to stay true to their strategies in light of some very compelling interruptions. Not surprisingly, ATC was an ample supplier of interruptions. (The timing of ATC calls was not controlled.) However, there was another supplier from a less likely source. The confederate first officer was instructed to perform all the duties of first officer, flying the aircraft. He was well qualified for this and performed flawlessly according to the latest in training — especially in the aspect of Crew Resource Management (CRM). CRM dictates that a crew member clearly communicates his intent and actions to the rest of the crew. Since the first officer's task was an aviate task, these communications acted as distractions to the task that the subject was performing. Often, while scoring the utterances, it was seen that the subject would make an aviate comment, only in response to the first officer. In some cases, these comments actually disrupted the task that the subject was performing and led him to tend to the aviate task. Thus, CRM can potentially be misapplied if communications are not prioritized and managed.

An important methodological note is that providing a full-mission scenario where the subject (in this case, the pilot not flying) has access to other relevant personnel for realism (co-pilot, air traffic controller, dispatch) can produce natural conversations and "thinking out loud" that are a rich source of information for analysis of cognitive processes and decision making. Since the "think out loud" protocol is not explicitly instructed, it has the advantage of occurring unobtrusively. "Naturalistic" probes that require pilots to describe their thought processes and decisions can be successfully inserted into the scenarios as part of ATC or dispatch conversation without disruption to the flow of the experiment. The danger of providing realistic resources, such as dispatchers, is that some pilots may use them as resources and delegate responsibilities to them which, as in this study, were the very tasks that the subject was to perform.

In summary, task management is a significant and important part of the flight crew's duties. Improvements in the task management strategies used by pilots holds significant potential for the reduction of the negative consequences of pilot actions when responding to a system failure. More research needs to be performed to determine the most effective task management strategies. Those strategies should be considered for explicit incorporation into flight crew training.

ACKNOWLEDGMENTS

REFERENCES

Abbott, T. (1993). Functional categories for future flight deck designs (NASA Technical Memorandum TM-109005). Hampton, VA: NASA Langley Research Center.

Abbott, T. S., & Rogers, W. H. (1993). Functional categories for human-centered flight deck design. Proceedings of the 12th Digital Avionics Systems Conference. New York: AIAA/IEEE.

Funk, K. (1991). Cockpit task management: preliminary definitions, normative theory, error taxonomy, and design recommendations. The International Journal of Aviation Psychology, 1(4), 271-285.

Raby, M. & Wickens, C. D. (1994). Strategic workload management and decision biases in aviation. The International Journal of Aviation Psychology, 4(3), 211-240.

Rogers, W. H., Schutte, P. C., & Latorella, K. A. (in press). Fault management in aviation systems. In M. Mouloua & R. Parasuraman (Eds.), Automation and Human Performance: Theory and Applications. Hillsdale, NJ: Lawrence Erlbaum Associates.