NASA Technical Memorandum 4547

Effects of Historical and Predictive Information on Ability of Transport Pilot To Predict an Alert

Anna C. Trujillo
Langley Research Center • Hampton, Virginia

National Aeronautics and Space Administration
Langley Research Center • Hampton, Virginia 23681-0001

March 1994

Abstract

In the aviation community, the early detection of a possible subsystem problem developing during a flight is potentially useful for increasing the safety of the flight because the extra time may allow the flight crew more options for dealing with a failure. Commercial airlines are currently using twin-engine aircraft for extended transport operations over water, and the early detection of a possible problem might increase the flight crew's options for safely landing the aircraft. One method for decreasing the severity of a developing problem is to predict the behavior of the problem so that appropriate corrective actions can be taken. To investigate the pilots' ability to predict long-term events, a computer workstation experiment was conducted in which 18 airline pilots predicted the alert time (the time to an alert) using 3 different dial displays and 3 different parameter-behavior complexity levels. The three dial displays were as follows: (1) standard (resembling current aircraft round dial presentations); (2) history (indicating the current value plus the value of the parameter 5 sec in the past); and (3) predictive (indicating the current value plus the value of the parameter 5 sec into the future). The time profiles describing the behavior of the parameter consisted of constant rate-of-change profiles, decelerating profiles, and accelerating-then-decelerating profiles. Although the pilots indicated that they preferred the near-term predictive dial, the objective data did not support its use. The objective data did show that the time profiles had the most significant effect on performance in estimating the time to an alert.

Introduction

In the aviation community, the early detection of a possible subsystem problem developing during a flight is potentially useful for increasing the safety of the flight because the extra time may allow the flight crew more options for dealing with a failure. An Aviation Safety Reporting System (ASRS) (ref. 1) database search revealed a significant number of incidents involving slowly developing consequences from failures. These failures included leaks in the fuel, oil, hydraulic, and vacuum subsystems and engine flameouts. Furthermore, in some accidents investigated by the National Transportation Safety Board (NTSB), fault consequences occurred well before a subsystem parameter entered an alert range. One example is the Eastern Airlines flight 855 accident (ref. 2), whose root cause was an oil leak due to missing O-rings in the engines. In that accident, after the number 2 engine had failed and been shut down because of a low-oil alert, the oil quantities of the number 1 and number 3 engines decreased for 15 to 20 minutes before the low-oil alerts occurred for those engines. By that time, it was too late to avert a near-catastrophic failure of the engine subsystem. If the crew had noticed the problem earlier, they possibly could have saved the affected systems for landing.

Also, a rapidly developing area in commercial aviation that presents additional motivation for detecting a possible problem early is the use of twin-engine aircraft for extended transport operations over water, known as ETOPS (extended transport operations). ETOPS-rated aircraft are allowed to be as far as 90 minutes away from the nearest suitable airport. If the consequences of a fault can be minimized in this situation, then the effect of the fault on the flight may also be minimized. Thus, an earlier recognition of a possible problem may decrease the severity of a failure and thus increase the safety of the flight.

One method for enhancing the recognition of a developing problem is to present information to the pilot on the predicted behavior of the system. This information could also allow for an earlier indication of the severity and urgency of a problem, as compared with the case in which the first symptom is a caution or warning alert. Currently, pilots must make predictions based on "raw" information; that is, they must calculate how quickly a parameter indicator is increasing or decreasing, whether it is accelerating or decelerating, and how far the indicator must travel to reach the alert threshold. Then, they must decide if this information signals an existing or potential problem, how much time is available to deal with it, and how urgent the problem is. Unfortunately, Wickens (ref. 3) states that a conservative bias is present in any prediction. This would result in underestimating the time to an alert, which would affect the criticality of attending to the problem.

Aids designed to improve the pilot's ability to make these predictions could show a near-term historical value of the parameter or could compute and display a near-term predictive value of the parameter. A history of the parameter value is exact because the actual past values are known, but this requires the pilot to calculate future values from past parameter behavior. However, if historical information proved to be as beneficial as predictive information, then displaying historical information to the pilot would be preferred because of the easier computational task. Unfortunately, evidence shows that humans have some difficulty in applying historical values in making predictions. For example, when estimating the next point in a time series from a static display, Van Heusden (ref. 4) found that when fewer historical data points were displayed, subjects forgot the essential information given in the preceding points that were no longer visible. This forgetfulness resulted in errors in estimating the next point in the time series, and thus these errors contributed to an overestimated velocity and an underestimated acceleration. Spenkelink (ref. 5) also found that historical information hindered a subject's ability to detect an oncoming abnormality in a dynamic situation, and he concluded that the historical information had an inhibiting effect.

On the other hand, providing predictive values will more directly aid the pilot in determining how much time remains until an alert occurs, but these values may be less accurate depending on the forecast time. Therefore, in order to test both historical and predictive information in an aviation-type task, the workstation study described in this paper evaluated pilot information aids for predicting the alert time (the time to an alert).

Objectives

The main objective of this research effort was to examine how presenting near-term historical or predictive information affected the pilot's ability to make a long-term prediction of when an alert would occur. Thus, the primary factor studied was the type of information provided rather than its format. The historical or predictive information presented was near term, that is, 5 sec into the past or future. All alerts that the pilot had to predict occurred in the long term, that is, an order of magnitude greater than the near-term historical or predictive information provided. Besides determining whether this information aided the pilot, this study began to delineate the effects of various factors on the pilot's ability to judge the time to an alert.

A secondary objective of this effort was to evaluate subjectively how intuitive the display designs were. Although the focus was on information content instead of format, obtaining some indication that the format chosen was reasonable was also desirable.

To address these objectives, a controlled experiment was conducted by using a computer workstation. A description is given of the independent variables chosen as well as the rationale for examining them in this context.

The four independent factors studied were (1) dial type, (2) scenario level of complexity, (3) display viewing time, and (4) direction of parameter movement. Each factor is described below.

Dial Type

The three types of dial displays evaluated were current values (standard), current values plus historical information (history), and current values plus predictive information (predictive). All displays depicted round dials because pilots were most familiar with this format. The displays used were intended to be generic and thus did not depict any particular subsystem gauge with which a pilot may have been familiar. This prevented the pilot from associating the behavior and the design of the dial with a specific subsystem. For all dials, the green normal range was 40 to 175 units, the amber caution range was 175 to 200 units, and the red warning range was 0 to 40 units. (See fig. 1.) Thus, the total range of the dial was 0 to 200 units, encompassing 220° of a circular arc. The digital readout of the value was always green in color because the value was always in the normal range during this experiment.

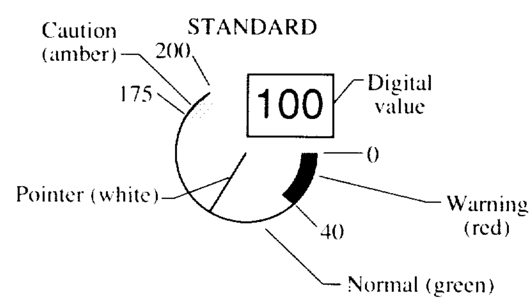


Figure 1. Standard dial.

The standard dial was labeled "STANDARD" above the dial. (See fig. 1.) The history dial (shown

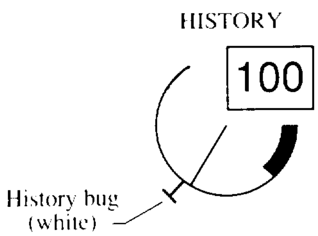


Figure 2. History dial.

in fig. 2), was similar to the standard dial, but outside the arc was a white *T* called the history bug and above the dial was the word "HISTORY." This display bug showed the dial value 5 sec in the past. The predictive dial (shown in fig. 3), which added a different piece of information to the standard dial, had a white diamond-shaped bug called the predictive bug which showed the value 5 sec into the future. Above the dial was the word "PREDICTIVE." For this experiment, the predictive dial was ideal in that the actual parameter value in 5 sec was exactly as the predictive bug indicated, although pilots were not told this.

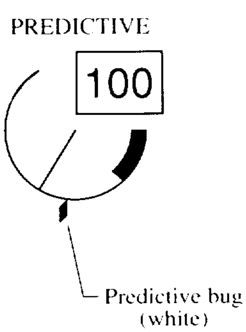


Figure 3. Predictive dial.

The different shaped bugs and the dial title added salient cues about which display the pilot was currently using. The history and predictive dials looked similar, and confusion between the two would have arisen if these cues had not been present.

Scenario Level of Complexity

The second factor examined was the different ways that the parameter behaved. This factor was accomplished by using time profiles of varying difficulties, or levels of complexity. Each profile followed one of three levels of complexity: simple, medium, or difficult. Simple parameter behavior had a constant rate of change of the parameter value. Medium profiles decelerated throughout the profile, and difficult profiles first accelerated and then decelerated. These three levels of complexity were employed for several reasons. First, failures may have unique manifestations that the pilot probably would not know *a priori*. Second, a pilot's ability to estimate the time when the value would reach an alert range would probably depend on the level of complexity of the parameter behavior. Finally, for constant rates of change of parameter values, the history and predictive dials would look identical except for the relative position of the bug, which would trail the value for the history dial or lead the value for the predictive dial.

Because the simple-level profiles had a constant rate of change (fig. 4), the distance between the bug and the actual value did not change. Thus, the time to an alert was a simple extrapolation of the distance between the history or predictive bugs and the actual value divided into the distance between the actual value and the beginning of the alert range. This value then had to be multiplied by 5 sec (the lag/lead time of the bug) to get the time to an alert.

Medium-level time profiles followed the square root of time. (See fig. 4.) Constants were set so that the profiles were always decelerating.

The difficult-level profiles first accelerated and then decelerated. Figure 4 shows the general profile for increasing trials. For these trials, the deceleration began at least 2 sec before the pilot had to estimate the time to an alert so as to ensure that the time profile did not purposely mislead the pilot about its deceleration.

The three profiles had several aspects in common. During the trial, the dial pointer did not change direction because the viewing time was assumed to be insufficient for the pilot to factor in directional changes. The increasing profiles stopped at 125 units, and the decreasing profiles stopped at 90 units. In both cases, the value was 50 units from an alert range at the end of a trial. The trials were designed so that the predictive bug was never in an alert range at the end of a trial. This forced the pilot to extrapolate the time to an alert from the information available, and it did not give an unfair advantage to the predictive dial. At the beginning of each trial, neither the bugs nor the actual value started in an alert area. Thus, pilots did not confuse the alert range for which they were estimating the time to an alert.

If each scenario could continue uninterrupted after reaching 90 or 125 units, all parameter values would reach a caution or a warning range 20 to 80 sec later. The pilots were not told this. Furthermore, the response choices were between 10 and 120 sec so that the pilots were not biased to choose between 20

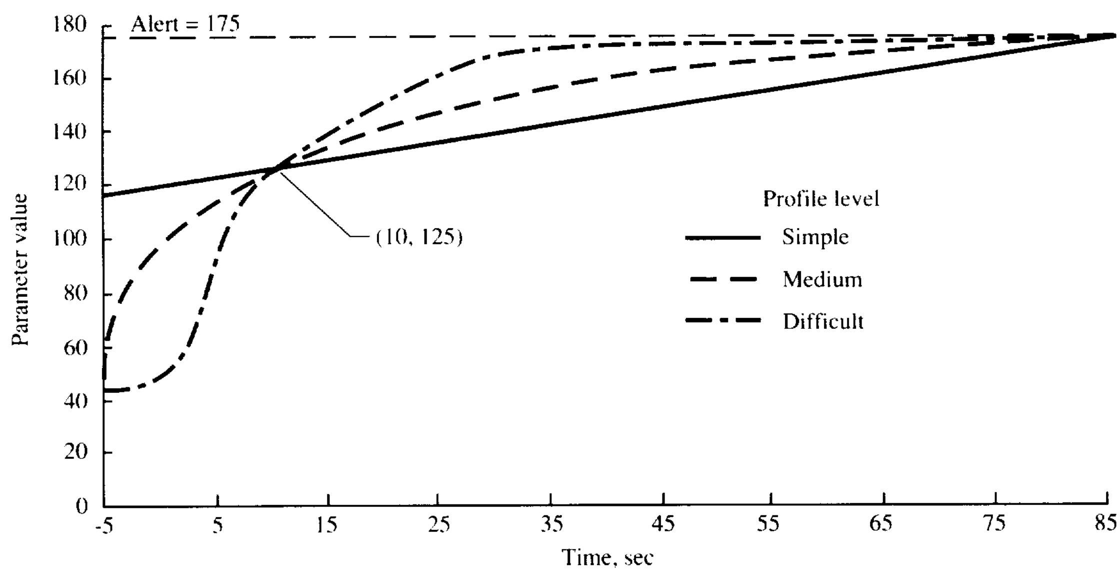


Figure 4. Complexity levels of scenario.

and 80 sec. No alerts occurred during the dynamic presentation.

Display Viewing Time

The third factor was the amount of time during which the pilot could study the dial (5 or 10 sec) before having to estimate the time to an alert. The two display viewing times were chosen to find their influence on the pilot's ability to estimate the time to an alert. They were also representative of the time that a pilot might normally view an instrument for monitoring purposes.

Direction of Movement

The fourth and last factor was the direction of parameter movement. Half the scenarios had increasing parameter values, and the other half had decreasing parameter values.

Experiment Design

Subjects

Eighteen male active-airline pilots used the displays described above. The pilots averaged 7000 hours of flight over 13 years of flight experience, with half of those years being commercial experience. The maximum number of hours that a pilot had was 16 000 and the minimum was 3000. The average age was 38, with the oldest being 59 and the youngest being 29.

Test Design

The test design of the experiment was a four factor ($3 \times 3 \times 2 \times 2$), within-subject repeated-measures design. As described above, the four independent factors were (1) the dial type (standard, history, or predictive); (2) the scenario level of complexity for the parameter behavior (simple, medium, or difficult); (3) the display viewing time (5 or 10 sec); and (4) the direction of parameter movement (increasing or decreasing). The dial types were grouped, whereas the three scenario levels of complexity, the two display viewing times, and the two directions of movement were randomized for each display type. Trials for each dial type were conducted consecutively. Because the display types were blocked, each pilot saw one of six dial sequences. All possible permutations of the three dial types were seen equally among the pilots. The experiment consisted of 24 data trials per dial type with a total of 72 trials per pilot. This resulted in two trials for each combination of the four independent factors. Furthermore, the profiles were blocked, that is, one set for the introduction, another for the demonstration trials, one for the practice trial, and the last set for the data collection trials.

Dependent Measures

The three dependent measures collected were (1) the accuracy of predicting when an alert would occur, (2) the time required to make that prediction, and (3) the subjective rankings of the various display

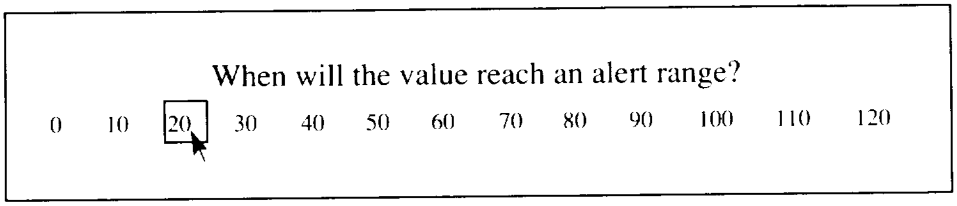


Figure 5. Question screen.

factors. The computer recorded the pilots' predictions and response times. Subjective data, collected mainly through a questionnaire, explicitly solicited pilots' likes and dislikes concerning the information.

Hypotheses

In considering the four independent factors and objectives of this study, the following were hypothesized. For the main factor of dial type, pilots would make predictions with explicitly displayed predictive information more quickly and accurately, but historical information would not be as beneficial (as Van Heusden (ref. 4) and Spenkelink (ref. 5) found). However, having the information would be better than having no information at all. The dial sequence, an artifact of the experiment design, should not have an effect on predicting the time to an alert. Regarding the three time profiles, pilots would be the most accurate with constant rate-of-change time profiles and would have the most difficulty with time profiles that have accelerating and then decelerating portions because of conservative biases in prediction. For constant rate-of-change trials, no difference should occur between displaying historical and predictive values. In considering the display viewing time, pilots would be more accurate with the longer display viewing time because they would have more time and information on which to base their prediction. Lastly, the direction of parameter movement should not affect predicting the time to an alert.

Procedure

First, a pilot received written instructions describing the experiment and a full description of each display. In general, he was told that for the data trials, a computer workstation would display a dial for 5 or 10 sec. After the dial animation, a question would replace the dial on the screen. He would answer the question by using the "mouse" to choose one of the possible answers.

Next, the pilot saw six demonstration trials that included the three scenario complexity levels. The pilots were not told about the different parameter behavior complexities. At the end of each demonstration trial, the pilot was told the amount of time needed for the parameter to reach the appropriate alert range, to the nearest 10 sec. This time was the answer sought from the pilot during the data trials.

Before the data collection trials, a medium-level practice trial was run in which the procedure was similar to the data collection trials described below. The only difference from a data trial was that after the pilot estimated the time to an alert, the next screen displayed the correct answer. The demonstration scenarios and the practice scenarios provided feedback on the length of time needed for the parameter to reach an alert range. No feedback was given during the data trials.

After the practice trial, the data collection trials began for that dial. Before each trial started, a screen reminded the pilot of the display type that he would see and the length of time that it would appear on the screen. This minimized any startle effects at the beginning and end of each trial. Following the pilot's push of the mouse button, the dial animation began 1 sec after the dial appeared. After 5 or 10 sec, the question that the pilot needed to answer replaced the dial, and he chose the answer with the mouse. For each data trial, the question that the pilot had to answer as quickly and as accurately as possible was, When will the value reach an alert range? (See fig. 5.) Pilots were not instructed on how to trade speed for accuracy. With the mouse, the pilot chose the estimated time to an alert, from that point in time, to the nearest 10 sec. The 10-sec intervals, which forced all pilots to use the same interval in predicting, helped to control one between-subject difference.

Once the pilot chose an answer, the computer recorded his response to the question and the time that he took to answer the question. Then, the next introductory screen appeared. When the pilot finished the 24 trials for a particular dial type, the next dial type in the sequence repeated the above procedure.

At the end of the experiment, the pilot filled out a questionnaire ranking the information given on the

Table 1. Significant Objective Results

[N.S. indicates data that are not significant]

Effector	Difference in 10-sec intervals		Absolute difference in 10-sec intervals		Time to choose answer, sec	
	Mean	σ	Mean	σ	Mean	σ
Dial type:						
Standard	N.S.	N.S.	1.8	1.5	N.S.	N.S.
History	N.S.	N.S.	2.1	1.7	N.S.	N.S.
Predictive	N.S.	N.S.	2.0	1.7	N.S.	N.S.
Complexity:						
Simple	0.8	1.7	1.3	1.2	N.S.	N.S.
Medium	−0.5	2.4	1.9	1.5	N.S.	N.S.
Difficult	−2.2	2.3	2.6	1.8	N.S.	N.S.
Viewing time:						
5 sec	N.S.	N.S.	N.S.	N.S.	10.00	8.6
10 sec	N.S.	N.S.	N.S.	N.S.	8.60	8.0
Direction:						
Decreasing	N.S.	N.S.	1.9	1.5	N.S.	N.S.
Increasing	N.S.	N.S.	2.0	1.7	N.S.	N.S.

displays. (See the appendix.) Other questions on the usefulness of this added information were also included.

Data Analysis

The main objective of this experiment was to examine how presenting near-term historical or predictive information affected the pilot's ability to make a long-term prediction of when an alert would occur. Thus, the difference was calculated between the pilot's estimate of the time required for the value to reach an alert and the actual time required to reach an alert, rounded to the nearest 10 sec. The actual time that the dial took to reach an alert was rounded to the nearest 10 sec because the pilots could answer only in increments of 10 sec. Both this difference and the absolute value of the difference were analyzed. The second dependent measure analyzed was the time required for the pilot to choose his answer. The objective data were analyzed by using the general linear models (GLM) procedure in the SAS Institute's SAS/STAT statistical computer program. (See ref. 6, pp. 549–640.) Also analyzed with the GLM package were the data indicating the differences in predictions and response times among the varying complexities of parameter behavior, the amount of time that the pilot could study the dial, and the direction of parameter movement. The Newman-Keuls posttest (ref. 7, pp. 346–351) was used to analyze multiple pairs of means for significant effects ($p \leq 0.05$) if the combinations involved were less than eight in number; otherwise, further postanalysis involved the Tukey HSD (honestly significant difference) method (ref. 7, pp. 352 and 353) because it controlled the "family-wise" error rate better when making all pairwise comparisons among several group means.

For the secondary objective of evaluating subjectively the intuitiveness of the display designs, the data consisted primarily of answers to the questionnaire administered at the end of the test. For ranking data, −3 was assigned to the lowest rating and +3 was assigned to the highest rating. The rankings were analyzed by the SAS/STAT nonparametric analysis of variance (NPAR1WAY) on ranks (ref. 6, pp. 713–726) and the SAS/STAT GLM procedure. Frequencies and averages were presented for the subjective data for factors that were significant ($p \leq 0.05$). Comments made by pilots during the test were also recorded and reported.

Results and Discussion

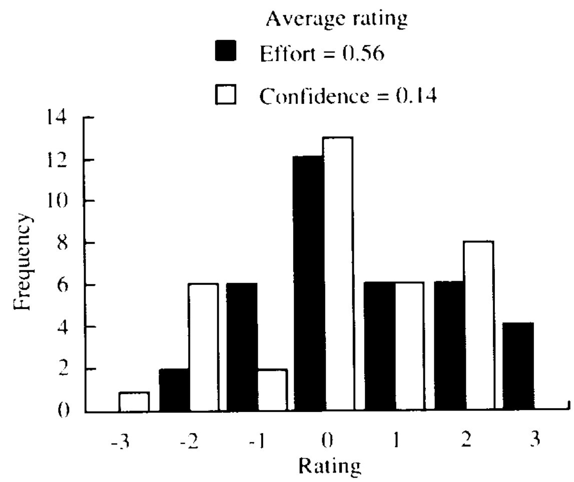
Dial Type

The hypothesis was made that the pilots would make their predictions of the time to an alert more quickly and accurately when using the predictive dial than when using the standard or history dials. Although no significant main effects were found with respect to response time, a significant effect of dial type ($F_{2,11} = 5.39$, $p \leq 0.03$) was found for the absolute value of the accuracy of their predictions, but it accounted for less than 1 percent of the variation. Further analysis showed that even though the history and predictive dials were not significantly different from each other, both dials produced larger errors than the standard dial. (See table 1.) On the other hand, from the subjective questionnaire, the pilots had more confidence in their predictions for the history and predictive dials then for the standard dial ($F_{1,68} = 7.04$, $p \leq 0.01$).

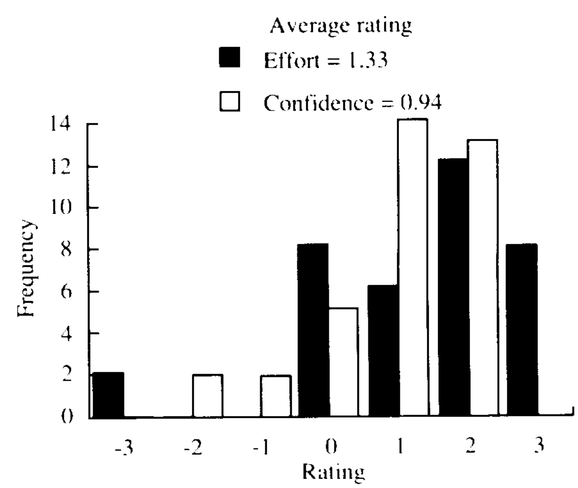
The dial type in the confidence question data accounted for 9 percent of the total variation. The predictive dial had the highest confidence rating. (See fig. 6.) Furthermore, when compared with the effort required to estimate the time to an alert for the standard dial, the predictive dial was rated as requiring the least effort, and the history dial was rated as requiring less effort than the standard dial ($F_{1,68} = 5.12$, $p \leq 0.03$). (See fig. 6.) The dial type accounted for 7 perent of the variation in the effort question data. Thus, the pilots thought the added information increased their accuracy in estimating the time to an alert, but the pilots' familiarity with the standard dial may have overshadowed the perceived benefits of the added information, or perhaps the history and predictive dials added some unforeseen complexity that degraded prediction performance.

A significant dial-by-sequence interaction ($F_{10,22} = 3.37$, $p \leq 0.01$) for the time required for the pilots to predict when an alert would occur was also found. (See fig. 7.) This interaction accounted for, at most, 15 percent of the total variation, which was not surprising in that the pilot took the shortest time in choosing the time to an alert for the last dial seen but the longest time for the first dial seen in the sequence. This can be partially attributed to learning effects, including learning effects involved in using the mouse.

Dial display design. The subjective questionnaire queried pilots about some aspects of the dial



(a) History dial.



(b) Predictive dial.

Figure 6. Average effort and confidence subjective ratings compared against standard dial.

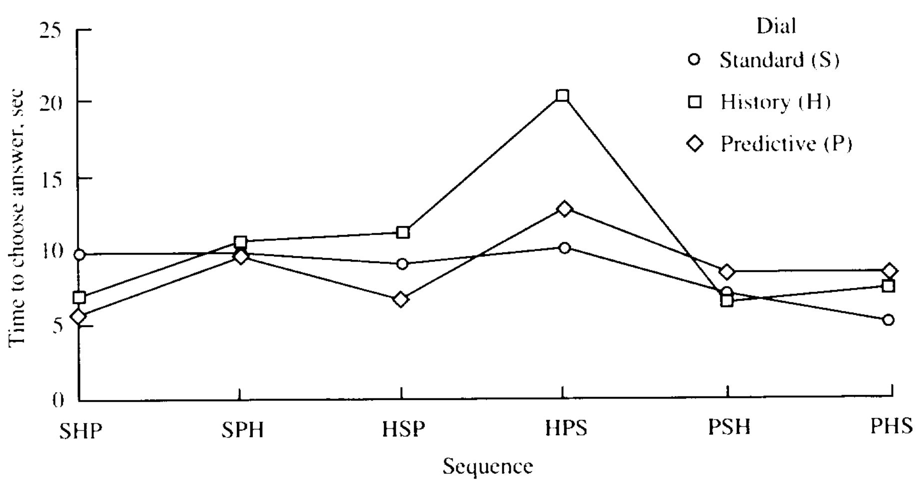
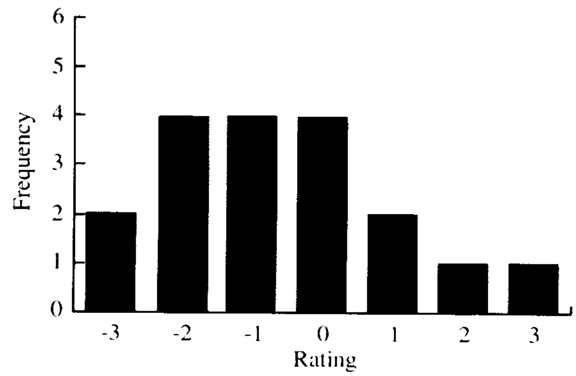
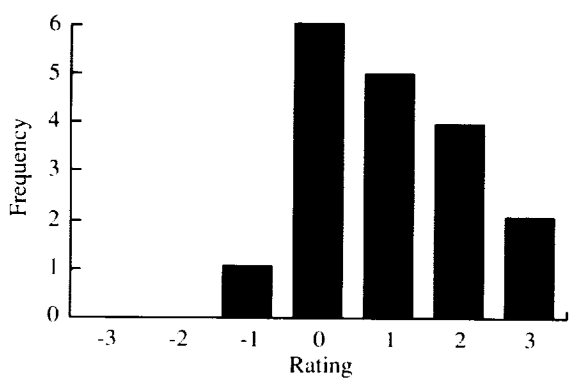


Figure 7. Dial-by-sequence interaction.

display design. Most of the comments pertained to the history and predictive bugs. The adequacy of the lag/lead times of the history and predictive dial bugs showed significant differences ($F_{1,34} = 11.63$, $p \leq 0.01$), which accounted for 25 percent of the total variation. The pilots thought that the predictive bug lead time of 5 sec was slightly greater than adequate, whereas the history bug lag time of 5 sec was less than adequate. (See fig. 8.)



(a) History dial. Average rating = −0.61.

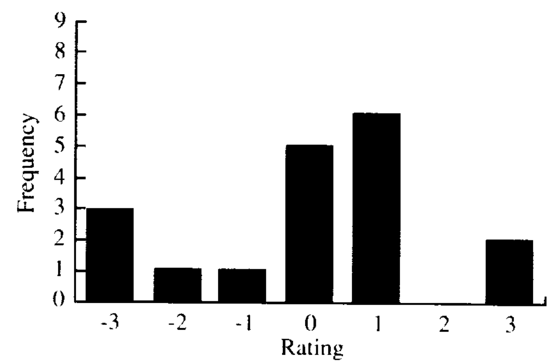


(b) Predictive dial. Average rating = 1.00.

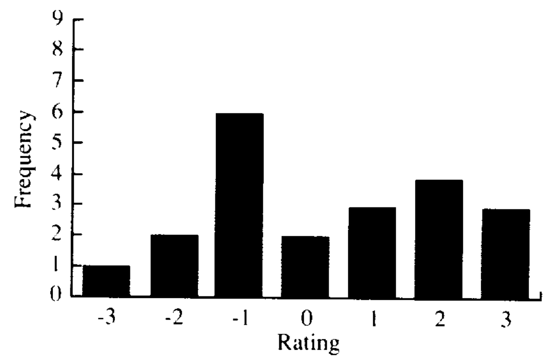
Figure 8. Subjective ratings of lag/lead time of bug.

Most of the comments about the bug lag/lead time concentrated on the predictive bug. Even though pilots rated the predictive bug lead time as adequate, most said they would have preferred that the bug have a longer lead time, with the average being approximately 10 sec. In considering the predictive bug lead time, one pilot mentioned that any lead time would be helpful, but another remarked, "The farther into the future the better."

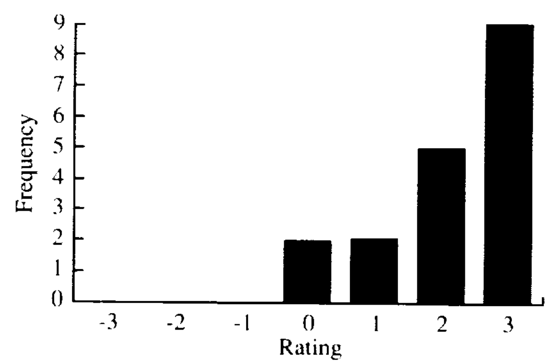
The overall ratings of the dials were significant ($F_{2,51} = 8.62$, $p \leq 0.01$), accounting for 25 percent of the total variation. As expected, the predictive dial



(a) Standard dial. Average rating = 0.



(b) History dial. Average rating = 0.56.



(c) Predictive dial. Average rating = 2.17.

Figure 9. Overall dial ratings.

rated the highest, whereas the history and standard dials had similar lower ratings. (See fig. 9.)

A few pilots gave reasons for disliking a dial and thus rating it low. Two pilots, who commented on the history dial, said that it was of no use in calculating what will happen and that the bug was distracting. One pilot rated the standard dial high because the bugs were too distracting. Two pilots did not like the predictive dial because they were not sure if they could trust it. These were the only pilots who had concerns about the accuracy of the predictive bug, even though no mention of the accuracy of the prediction algorithm was provided. On the other hand, one pilot rated the predictive dial high because, according to him, it provided what pilots want to know.

Written explanations to some of the questions provided insight into how some pilots would use the information. Concerning the history dial, one pilot wanted it for confirmation, whereas another liked it because it was useful for catching up on the behavior of a subsystem. The comments regarding how they would use the predictive dial dealt mainly with having an advanced warning to an alert. One pilot did mention that he would use it to try to keep that subsystem out of the alert ranges. No other comments were made regarding active behavior toward subsystem management.

Other comments pertained to subsystems that would benefit from this information. Thirteen pilots wanted this information for engine instruments, and the majority felt that this was the place where it would be the most beneficial. Other areas in which pilots would like this information are systems involving quantity, pressure, and temperature, as well as airspeed and altitude indicators.

Pilots' methods of determining when an alert would occur. Most pilots were unable to verbalize their methods of determining when an alert would occur for the standard dial. However, when asked how they estimated the time to an alert with the history or predictive dials, most could provide a method. The majority said that they attempted to estimate the distance between the bug and the value at the end of the trial. Next, they tried to calculate how many times that distance divided into the distance to the alert range, which was approximately 50 units away. That number was then multiplied by 5 sec (the bug lag/lead time) to get the approximate time to the alert. They then added more time to account for the deceleration of the dial. The pilots' methods of estimating the time to an alert for the history and predictive dials suggest that the bugs required more processing, thus moving the pilots from knowledge-based behavior to skill- or rule-based behavior (ref. 8).

The pilots' inability to verbalize their method of determining when an alert would occur for the standard dial contributed to the lower overall rating of the standard dial. This may have also affected the pilots' confidence ratings of the dials. The confidence ratings for the history and predictive dials were above neutral (0) when compared with the standard dial. Thus, the pilots may have had less confidence in their estimate for the time to an alert with the standard dial.

Scenario Level of Complexity

The author also hypothesized that the level of profile complexity would affect both the speed and accuracy of the pilots' responses. Although no significant main effects were found with respect to response time, significant main effects for accuracy were discovered. These effects accounted for approximately 38 percent of the variation in the difference ($F_{2,11} = 190.59$, $p \leq 0.01$) between the actual and predicted alert times, and for approximately 17 percent of the variation for the absolute difference ($F_{2,11} = 33.72$, $p \leq 0.01$). As seen in table 1 when looking at the difference, the pilots overestimated the time to an alert for the trials with a simple complexity level and underestimated the time to an alert for the medium and difficult trials. Therefore, pilots underestimated the constant rate of change of the parameter value for the simple parameter behavior, and they appeared to underestimate the deceleration of the medium and difficult parameter behavior profiles, thus supporting the conservative bias in prediction. Unexpectedly, analyzing the difference showed that the smallest errors occurred for the medium complexity level, but the absolute value of the difference may be a more accurate measure because errors cannot cancel one another. When considering the absolute value of the difference, simple behavior caused the smallest errors and difficult behavior caused the greatest errors, a result that was expected because humans have some difficulty in estimating deceleration.

Although not asked directly in the questionnaire, 5 out of 18 pilots did mention the differences in parameter behavior complexities. Only three pilots made direct comments that the scenarios did not all follow the same general behavior. Two pilots mentioned that estimating the time to an alert was easier in the trials with constant or nearly constant rate of change than in the trials that rapidly decelerated. Overall, most pilots felt that all scenarios had approximately the same difficulty level; hence, the effort and confidence of prediction remained constant within the dial.

Calculation of time to alert assuming constant rate of change. Because the pilots mentioned that their prediction method used a constant rate of change plus an extra time factor to account for the deceleration for the history and predictive dials, it was interesting to explore whether the extra time factor differed for dial type and scenario level of complexity. If the rate of change were constant, the amount of time for the value to reach an alert range was estimated from the rate of change (the distance between the bug and the actual value at the end of the 5- or 10-sec viewing time divided by 5 sec), that is, the lag/lead time of the bug. The time to an alert was then estimated by dividing the rate of change into 50 units (the distance to the alert range at the end of the viewing time) and rounding that time to the nearest 10 sec. This time was subtracted from the pilot's estimate of the time to an alert to get an error difference used in analysis.

Results of assuming constant rate of change. In the analysis of this error difference, the complexity level of the parameter behavior was a significant factor ($F_{2,11} = 23.26$, $p \leq 0.01$) accounting for 6 percent of the total variation. (See table 2.) Further analysis found that the parameter behavior complexities varied from each other significantly. If the rate of change were constant for all cases, pilots overestimated the time to an alert. Because the pilots had larger errors for the medium and difficult levels of parameter behavior, the pilots were apparently attempting to account for the deceleration in the actual scenarios, but they were not accounting for it adequately, as seen in the accuracy data mentioned above. The difficult scenarios had the most time added to their estimates, probably due to the acceleration at the beginning of the scenario accentuating the deceleration at the end. Thus, although most of the pilots did not directly comment on the different parameter behaviors, they did seem to notice some difference between the scenarios in that they added more time at the end of their calculations for the medium and difficult levels of parameter behavior.

Table 2. Significant Results If Velocity Were Constant

Complexity	Difference in 10-sec intervals	
	Mean	σ
Simple	0.7	1.7
Medium	1.2	1.6
Difficult	2.0	1.7

Display Viewing Time

The third main experimental hypothesis was that a longer viewing time would allow the pilots to be more accurate in their predictions. This effect of display viewing time was not detected, although a significant effect was discovered for the time required to choose an answer ($F_{1,12} = 13.20$, $p \leq 0.01$). The viewing time accounted for only approximately 1 percent of the variation in the dependent measure. As might be expected, the longer that the pilots could watch the dial, the less time they took in choosing an answer. (See table 2.)

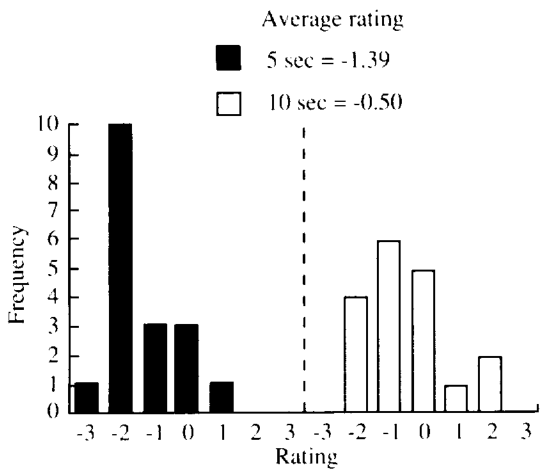
In the subjective data, the adequacy of the display viewing time for the different display types had two significant factors, the dial type ($F_{2,102} = 15.74$, $p \leq 0.01$) and the display viewing time ($F_{1,102} = 16.07$, $p \leq 0.01$), which accounted for approximately 21 percent and 11 percent of the total variation, respectively. As illustrated in figure 10, the predictive dial had the highest rating, and the 10-sec viewing time had a higher rating than the 5-sec viewing time for all dials.

Most pilots commented that they wanted to observe the dial for at least 10 sec, with the average being around 15 sec; therefore, it became interesting to see if their ratings supported their comments. Thus, the viewing times were extrapolated from the subjective ratings of the display viewing time. To achieve a rating of 3, the pilots would supposedly need to view the history and predictive dials for approximately 18 sec and 19 sec, respectively. Therefore, the pilots' comments regarding the desire to view the dial for 15 sec were corroborated by their ratings. Notice that increasing the predictive bug lead time to 10 sec and increasing the viewing time to 15 sec is near the earliest time to an alert in this experiment. Furthermore, if the standard dial viewing time is extrapolated to achieve a rating of 3, pilots would supposedly need to see the dial for nearly 30 sec. Notice that 30 sec is the earliest time to an alert for the trials.

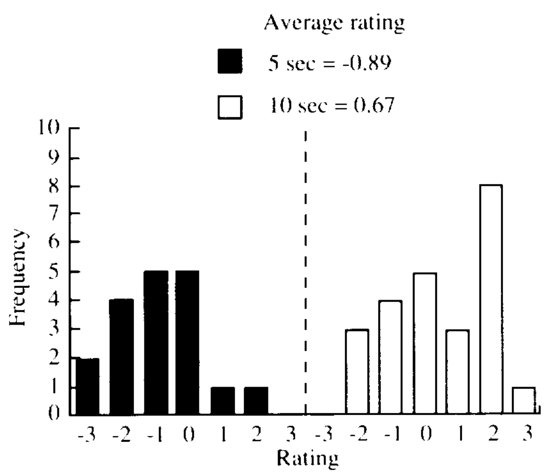
Two pilots did not care how long the dial was shown because they were going to take action only when the value reached an alert range, and thus they were not concerned over what happened before the alert. Two pilots wanted to be told directly when the value would reach an alert range because they felt that watching the dial and estimating the time to an alert would lead to a fixation on that dial. As a result, some pilots wanted to know when an alert was going to occur, whereas others wanted to know the information only if required actions were associated with it.

Direction of Movement

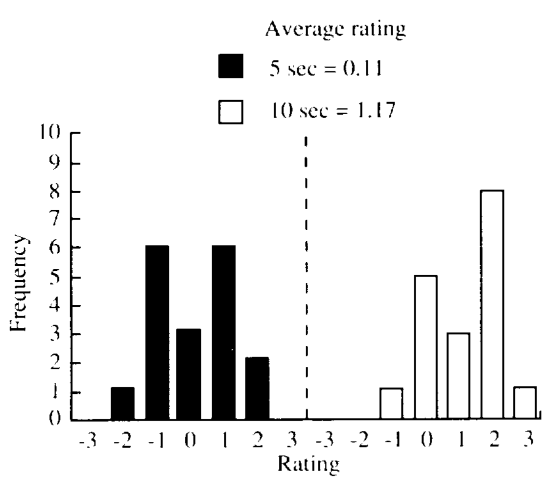
Unexpectedly, the direction of parameter movement was a significant factor ($F_{1,12} = 6.55$, $p \leq 0.03$),



(a) Standard dial.



(b) History dial.



(c) Predictive dial.

Figure 10. Subjective ratings of average display viewing time.

but it accounted for less than 1 percent of the total variation. (See table 2.) Comments involving the ranking questions showed a minimal effect on the effort rating because of the direction of movement of the value. Thirteen of the 18 pilots perceived no difference in their effort between trials when the value was increasing and trials when the value was decreasing.

Concluding Remarks

Although the pilots said that they preferred the near-term predictive information, the objective data showed no performance advantage in using it for estimating the alert time (the time to an alert). Even though a small positive effect occurred because of dial sequencing, which was attributed to learning effects, the standard dial led to smaller absolute prediction errors. Comments made by the pilots suggest that with the new information, many were busy trying to calculate the time to an alert, whereas with the standard dial, predicting was more of a perceptual process. Because minimal explicit mental calculations were made for the standard dial, pilots were better able to estimate the time to an alert. However, the lack of a conscious method used on the standard dial to calculate the time to an alert led to poorer ratings for that dial, even though pilots performed better with it. The history and predictive bugs may have also been a distraction, or perhaps pilots simply did better with the standard dial because they were familiar with it.

Presentation variables also influenced the effectiveness of the historical and predictive information. For instance, the longer the pilots watched the dial, the quicker they could estimate the time to an alert. Furthermore, the direction of movement may have influenced the pilots' perceived speed changes in the value of the parameter. Also, several pilots mentioned that the lag/lead times of the history and predictive bugs were too short. Many would have preferred a longer lag/lead time. Lastly, pilots' comments suggest that the use of the bugs led them to predict the time to an alert primarily on the rate of change of parameter value as judged by the distance between the bug and the actual value. This may not have occurred if a different format for the information had been used. Thus, the confidences in and the preferences for a particular format do not guarantee the effective use of that format, as seen in the objective results not supporting the hypothesized benefit of this form of presenting the near-term historical and predictive information.

As hypothesized, the level of complexity of parameter behavior was a significant factor, but the dial type did not affect the pilots' ability to predict the time to an alert for any of the scenario complexities. Pilots were unable to compensate completely for the differences in the behavior. For the medium and difficult parameter behaviors, pilots considered the decelerating trend in predicting the time to an alert, but the time that they added to their estimate was not sufficient to fully overcome their underestimation of the rate of change of parameter value. As a result, both deceleration and rate of change were underestimated, thus supporting a general conservative bias in prediction.

NASA Langley Research Center
Hampton, VA 23681-0001
January 6, 1994

Appendix

Subjective Evaluation

For each of the following questions, please either write out your answer or mark the block that best describes your answer. The blocks in between the extremes and the middle of each scale indicate not as much. Do not mark on the block dividers. If you run out of room for the written answers, feel free to use the back of a sheet.

Definitions: much more effort - much more mental effort required
about the same - neither particularly difficult nor easy
much less effort - much less mental effort required

very unsure - not very confident
about the same - neither particularly sure nor unsure
very sure - very confident

very inadequate - not enough to accomplish task
adequate - just enough to accomplish task
very adequate - more than enough to accomplish task

As you probably remember, the trials were of different lengths. Half the trials only had the dial on the screen for 5 seconds while the other half had the dial on the screen for 10 seconds. In the following questions

the 5 second trial = the trials where the dial was on the screen for 5 seconds

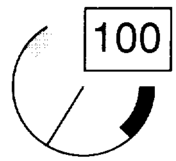
and

the 10 second trial = the trials where the dial was on the screen for 10 seconds

The following page reviews the dials you have just seen.

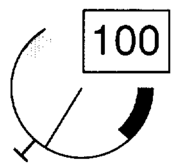
The standard dial refers to the dial with no extra information pictured.

STANDARD



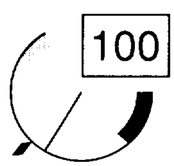
The history dial refers to the dial with the T outside the dial, which displayed the parameter's value 5 seconds ago.

HISTORY



The predictive dial refers to the dial with the filled-in diamond, which showed what the parameter's value will be in 5 seconds.

PREDICTIVE



1. Compared to the standard dial, how much effort was needed during the 5 second trials to determine when the value would reach a caution or warning region

   i) with the history dial?

much more effort			about the same			much less effort

   ii) with the predictive dial?

much more effort			about the same			much less effort

   iii) During the 5 second trials, were there any differences in your effort to predict the time to an alert between trials with increasing values or trials with decreasing values? If yes, describe. ______________________________

   iv) During the 5 second trials, were there any differences in your effort to predict the time to an alert among the trials? If yes, describe. ______________________________

2. Compared to the standard dial, how much effort was needed during the 10 second trials to determine when the value would reach a caution or warning region

i) with the history dial?

much more effort	about the same	much less effort

ii) with the predictive dial?

much more effort	about the same	much less effort

iii) During the 10 second trials, were there any differences in your effort to predict the time to an alert between trials with increasing values or trials with decreasing values? If yes, describe. ______

iv) During the 10 second trials, were there any differences in your effort to predict the time to an alert among the trials? If yes, describe. ______

3. Compared to the standard dial, how sure were you during the 5 second trials of your decision of when a value would reach a caution or warning region

i) with the history dial?

very unsure			about the same			very sure

ii) with the predictive dial?

very unsure			about the same			very sure

iii) During the 5 second scenarios, were there any differences in how sure you were of your prediction time to an alert between trials with increasing values or trials with decreasing values? If yes, describe. ______________________________

____

____

____

____

____

____

____

____

iv) During the 5 second scenarios, were there any differences in how sure you were of your prediction time to an alert among the trials? If yes, describe. ______________

____

____

____

____

____

____

____

____

4. Compared to the standard dial, how sure were you during the 10 second trials of your decision of when a value would reach a caution or warning region

i) with the history dial?

very unsure		about the same		very sure

ii) with the predictive dial?

very unsure		about the same		very sure

iii) During the 10 second scenarios, were there any differences in how sure you were of your prediction time to an alert between trials with increasing values or trials with decreasing values? If yes, describe. ____________________

iv) During the 10 second scenarios, were there any differences in how sure you were of your prediction time to an alert among the trials? If yes, describe. ____________________

5. How adequate was the 5 second viewing for determining when an alert was going to be reached

i) with the standard dial?

very inadequate	adequate	very adequate

ii) with the history dial?

very inadequate	adequate	very adequate

iii) with the predictive dial?

very inadequate	adequate	very adequate

6. How adequate was the 10 second viewing for determining when an alert was going to be reached

i) with the standard dial?

very inadequate	adequate	very adequate

ii) with the history dial?

very inadequate	adequate	very adequate

iii) with the predictive dial?

very inadequate	adequate	very adequate

7. How much time would you like to see the dial for determining when an alert would be reached and why? ____________________

8. How adequate was the 5 second look back time for the bug which displayed the previous value for determining when an alert was going to be reached?

very inadequate			adequate			very adequate

9. How adequate was the 5 second look ahead time for the bug which displayed a future value for determining when an alert was going to be reached?

very inadequate			adequate			very adequate

10. How much time backward and forward would you like the history and predictive bugs to show and why? ____________________

11. On the scale below, please rate the displays. You may put more than one display type in a box. Please look at the example below before making your choices.

Example:

Displays:									
a									
b					c		ab		
c	least liked								most liked

Displays:										
standard										
history										
predictive	least liked									most liked

12. Why did you choose the above order? ______________________________

____

____

____

____

____

____

13. How could the displays you liked the most be improved further? ____________

____

____

____

____

____

____

14. How would you use the history and predictive information? ______________

____

____

____

____

____

____

15. Which instruments would you like to have the display for and why?

16. Please record any other comments, suggestions, or criticisms you may have about any of the display types, the scenarios, or the way the experiment was conducted?

References

1. Battelle. *Aviation Safety Reporting System.* Search Request Number 1929, Aircraft Subsystems Failures, Mountain View, California, 1990.

2. *Aircraft Accident Report Eastern Airlines, Inc., Lockheed L-1011, N334EA, Miami International Airport, Miami, Florida, May 5, 1983.* NTSB-AAR-84-04, National Transportation Safety Board, Mar. 1984. (Available from NTIS as PB84 910 404.)

3. Wickens, Christopher D.: *Engineering Psychology and Human Performance.* Scott, Foresman & Co., 1984.

4. Van Heusden, Arnold R.: Human Prediction of Third-Order Autoregressive Time Series. *IEEE Trans. Syst., Man & Cybern.*, vol. SMC-10, no. 1, Jan. 1980, pp. 38–42.

5. Spenkelink, G. P. J.: Aiding the Operator's Anticipatory Behaviour: The Design of Process State Information. *Appl. Ergon.*, vol. 21, no. 3, Sept. 1990, pp. 199–206.

6. *SAS/STAT® User's Guide, Release 6.03 ed.* SAS Institute Inc., 1988.

7. Howell, David C.: *Statistical Methods for Psychology, Second ed.* Duxbury Press, 1987.

8. Hudlicka, E.; Corker, K.; Schudy, R.; and Baron, S.: *Flight Crew Aiding for Recovery From Subsystem Failures.* NASA CR-181905, Jan. 1990.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 0704-0188
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.			
1. AGENCY USE ONLY(Leave blank)	2. REPORT DATE March 1994	3. REPORT TYPE AND DATES COVERED Technical Memorandum	
4. TITLE AND SUBTITLE Effects of Historical and Predictive Information on Ability of Transport Pilot To Predict an Alert			5. FUNDING NUMBERS WU 505-64-13
6. AUTHOR(S) Anna C. Trujillo			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) NASA Langley Research Center Hampton, VA 23681-0001			8. PERFORMING ORGANIZATION REPORT NUMBER L-17305
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001			10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA TM-4547
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified–Unlimited Subject Category 06			12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

In the aviation community, the early detection of a possible subsystem problem developing during a flight is potentially useful for increasing the safety of the flight because the extra time may allow the flight crew more options for dealing with a failure. Commercial airlines are currently using twin-engine aircraft for extended transport operations over water, and the early detection of a possible problem might increase the flight crew's options for safely landing the aircraft. One method for decreasing the severity of a developing problem is to predict the behavior of the problem so that appropriate corrective actions can be taken. To investigate the pilots' ability to predict long-term events, a computer workstation experiment was conducted in which 18 airline pilots predicted the alert time (the time to an alert) using 3 different dial displays and 3 different parameter-behavior complexity levels. The three dial displays were as follows: (1) standard (resembling current aircraft round dial presentations); (2) history (indicating the current value plus the value of the parameter 5 sec in the past); and (3) predictive (indicating the current value plus the value of the parameter 5 sec into the future). The time profiles describing the behavior of the parameter consisted of constant rate-of-change profiles, decelerating profiles, and accelerating-then-decelerating profiles. Although the pilots indicated that they preferred the near-term predictive dial, the objective data did not support its use. The objective data did show that the time profiles had the most significant effect on performance in estimating the time to an alert.

14. SUBJECT TERMS			15. NUMBER OF PAGES
Predictive information; Historical information; Parameter display; Predicting alerts; Time to an alert			24
			16. PRICE CODE A03
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT