September 1, 2015

Evaluation of Information Quality and Automation Visibility in Information Automation on the Flight Deck

Michael C. Dorneich, Iowa State University
Rachel Dudley, Iowa State University
William H. Rogers, Honeywell International
Emmanuel Letsu-Dake, Honeywell International

Available at: https://works.bepress.com/michael_dorneich/92/
Evaluation of Information Quality and Automation Visibility in Information Automation on the Flight Deck

Michael C. Dorneich1, Rachel Dudley2.
1Industrial and Manufacturing Systems Engineering, 2Human-Computer Interaction, Iowa State University

William Rogers, Emmanuel Letsu-Dake, Stephen D. Whitlow, Michael Dillard, Erik Nelson
Human-Centered Systems, Honeywell Laboratories

An empirical study was conducted to evaluate human factors issues related to the automation visibility and information quality of an information automation system. Information automation is responsible for the collection, processing, analysis, and presentation of information to the flight crew. Previous analysis had identified a set of characteristics particularly applicable to aircraft flight deck information automation and associated human factors issues. Air transport pilots interacted with an example information automation system in ways that allowed investigation of the information automation characteristics of automation visibility and information quality. The evaluation found that poor information quality appeared to be difficult for participants to detect, even when they were presented with the highest automation visibility level. In the times that they did not successfully compensate, participants tended to over-trust the automation, so when information was missing and they were under high workload, they chose the top plan suggested by the automation even though it was not the truly best plan. Trust in automation was reduced by low information quality, but compensated for by increased automation visibility.

INTRODUCTION

Beyond current operations, there is considerable research and development in new automation, procedures, and concepts to safely and efficiently handle an increasing demand for air travel. Next Generation Air Transportation System (NextGen) will utilize satellite-based navigation and interconnected database systems to guide and track air traffic more precisely than was previously feasible (FAA, 2013). NextGen operational concepts and technologies will dramatically affect both the types and amount of information available on flight decks (JPDO, 2009). This transformation will result in increasing automation to take advantage of the likely increase in the amount of available information (Landry, 2009). Information automation refers to automation devoted to the management and presentation of relevant information to flight crew members (Abbott, McKenney & Railback, 2013). Information automation processes and presents information to support a pilot’s tasks and awareness and does not include cases where automation decisions and actions directly affect the aircraft performance, flight path or systems.

The literature on human factors aspects of flight deck “automation” typically does not distinguish among different categories of automation (e.g. Tenney, Rogers, & Pew, 1998; Funk et al., 1999). However, there may be different human factors issues depending on the category of automation that is being considered. Work has focused on the implication of automated control of the aircraft. However, much of the automation currently being developed pertains to information support rather than control of the aircraft. Certain human factors issues and pilot errors which might be prevalent in interacting with information automation may be minimal for other types of automation. Conveying the right information at the right time to the flight crew and accepting input from them in a user-friendly manner is critical for safe operations.

Performance-based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team Flight Deck Automation Working Group has recommended that a stronger definition of information automation is needed, as well as defining terms associated with it (Abbott et al., 2013). Based on definitions of Billings (1991) and Parasuraman, Sheridan, & Wickens (2000), we developed a two-dimensional description of flight deck information automation that considered what is being managed (aircraft, systems, or information), and the stage of information processing involved (Dudley et al, 2014). A set of characteristics described a system’s operation or behavior to identify potential human factors issues (Dudley et al, 2014; Rogers et al, 2015).

An empirical study was conducted to evaluate human factors issues related the interaction between Automation Visibility and Information Quality. Results led to conclusions and considerations in the assessment of information systems.

INFORMATION AUTOMATION

Information Automation Characteristics

In broad terms, information automation is the programming logic that dictates what, when, and how information is presented to the flight crew. In previous work, nine characteristics, related to automation automation-human interaction, specific to information automation were identified (Dudley et al, 2014; see Table 1). The study focused on two characteristics: Information Quality and Automation Visibility.

<table>
<thead>
<tr>
<th>Table 1. Final set of information automation characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Characteristics</td>
</tr>
<tr>
<td>Functional Complexity</td>
</tr>
<tr>
<td>Information Quality</td>
</tr>
<tr>
<td>Adaptiveness</td>
</tr>
<tr>
<td>Level of integration</td>
</tr>
<tr>
<td>Degradation Behavior</td>
</tr>
</tbody>
</table>


Information Quality

Previous research in Information Quality originated in database administration and management of information systems (Reeves & Bednar, 1994; Wang & Strong, 1996; Myers, Kappelman, & Prybutok, 1997; Pipino, Lee, & Wang, 2002; Stvilia, Gasser, Twidale, & Smith, 2007; Batini, Cappiello, Francalanci, & Maurino, 2009). In this domain, there are several attributes that play an important part in the overall concept of information quality. Wang and Strong (1996) identified four properties of high quality data: 1) intrinsically good, 2) contextually appropriate for the task, 3) clearly represented, and 4) accessible to the data consumer. They further identified 15 dimensions within the four categories to capture the usefulness of information as a product, or commodity, to the consumers who seek it. In our empirical evaluation, we focused on the information quality dimensions of completeness and appropriate amounts of data.

The dimensions of information quality provided by Wang and Strong (1996) are relevant to pilots as consumers of flight deck information automation output. For example, automatic de-cluttering of a display aims to provide pilots with only the most relevant and timely information for a given situation (Billings & Woods, 1994). Information quality human factors issues include: erroneous input; incomplete, uncertain, or probabilistic input; bad assumptions or factors which are not accounted for (Wang and Strong, 1996); flawed or imprecise processing (e.g. due to a limited or constrained model) (Ockerman and Pritchett, 2002); variety of data sources (and potential for conflicting inputs); inaccurate or unreliable outputs (Rovira et al., 2007); and untimely information.

Information Automation Visibility

Information automation visibility refers to the ability of an automation system to provide adequate feedback about its current state, what information was used, and how the information was processed (Endsley, 1996; Whitlow, Dorneich, Funk, & Miller, 2002). This characteristic may also be referred to as opacity (e.g. Andre & Cutler, 1998; Bisantz, Marsiglio, & Munch, 2005). In order for automation to be visible, feedback must provide a view into the automation state in a manner which can be properly interpreted by the operator (Woods, 1996) and allows the operator to predict its behavior (Scerbo, 1996). This includes information about input sources and how the system is generating outputs. High automation visibility can enhance predictability by fostering the development of an accurate mental model and enable flightcrew to verify outputs. Poor automation visibility could result in a loss of situation awareness and an increase in workload (Sarter & Woods, 1994; 1995). However, the appropriate amount and timing of automation visibility (e.g., “explanatory” information) should be carefully evaluated because too much information presented at an inappropriate time (e.g., during time-critical tasks) could add workload and head down time (Deveans & Kewley, 2009; Degani, Barshi, & Shafto, 2013). Automation visibility attributes may contribute to information automation human factors issues including: poor predictability of behavior (Munaw, Sarter, & Wickens, 2001; Scerbo, 1996); hard-to-detect input or processing errors (Sarter & Woods, 1995, Sarter et al, 1997); difficult-to-assess information quality and verify outputs (Munaw et al., 2001, Sarter et al, 1997); and inadequate, inappropriately timed, or inappropriate amount of system feedback (e.g., modes) (Abbott et al., 2013; Kaber et al., 2001).

METHOD

An empirical study was conducted in a low fidelity simulator and an example information automation system.

Participants

Twelve air transport pilots from a cross section of regional and major airlines were recruited. All were male. Five were Captains and seven were First Officers. All had flown glass cockpits. Average age was 34 (range 24-56) and flight hours were about 7,000 hours (range: 2,000-14,000).

Task Application: Diversion Aid

Overview. The Diversion Aid was designed for possible NextGen decision support, when pilots might be called upon to take more responsibility for the decision of if, when, and where to divert an aircraft. The Diversion Aid integrates multiple information sources on the current state of flight, aircraft, maintenance, crew, and passenger schedules to display the implications of diversion decisions to pilots. The goal is for pilots to integrate the goals and priorities of interested airline operations stakeholders into the decision making process. This was based on a previously developed system (Dorneich et al., 2004) that uses a set of policy statements developed after conducting interviews with airline dispatchers, pilots, and various stakeholders.

Policies. A set of company policy statements was established to represent the operational priorities of all stakeholders affected by diversion decisions. These policies are used to assess the overall “goodness” of a diversion plan. Each policy was associated with cost points for each statement that is violated by a particular plan. For example, diverting a flight with an unaccompanied minor costs 10 points, while delaying a flight greater than 15 minutes costs 8 points. The policy statements are adapted from a list of policy statements developed by Dorneich et al, (2004) after conducting interviews with airline dispatchers, pilots, and various stakeholders. The goal of selecting a diversion option is to minimize the total cost incurred by the selected option: The lower the cost, the better the plan.

Displays. The Diversion Aid presented the original scheduled flight plan, followed by up to three diversion plans, based on minimizing the number of policy violations. Figure 2 depicts the graphical representations of the original plan with a ranked ordered list (best first) of diversion plan recommendations. The depiction of each individual policy violations (and their score) allows pilots to understand how the Diversion Aid ranked one plan ahead of another.

Tasks. A flight simulation was presented to the participants to help provide a sense of realism to the trials. The simulation began approximately ten minutes from top of
decent. After 60 to 90 seconds, the need for a diversion was announced and the participant was asked to make a recommendation within five minutes. The participant then started the Diversion Aid, reviewed its recommended plan(s), and decided whether to accept one of the plans or to reject its recommendation(s) if he or she could devise a better plan. The participant did not need to create a different plan. A help menu was available that displayed the set of policies

![Figure 1. Diversion Aid with the top three ranked options.](image)

**Independent Variables**

Table 2 describes the levels of Automation Visibility (low, medium, high) and Information Quality (low, high).

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>Levels</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information Quality</td>
<td>Low</td>
<td>Some relevant information not included in calculation of total diversion decision cost</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>All relevant information was included in the calculation of total diversion decision cost</td>
</tr>
<tr>
<td>Automation Visibility</td>
<td>Medium</td>
<td>Rank-ordered list of the three best options with no policy costs shown</td>
</tr>
<tr>
<td></td>
<td>High</td>
<td>Rank-ordered list of the three best options with policy costs shown</td>
</tr>
</tbody>
</table>

In the high Information Quality condition, the correct selection was always the top option on the display. In the low Information Quality condition, the correct selection was not the top selection, because the automation was missing information that resulted in incorrect scoring of the options. However, participants were given knowledge of this “missing” information during the pre-flight briefing. Thus in the low Information Quality conditions, the automation’s highest ranked plan was not actually the best plan – participants were expected to recognize that a different plan was better once they included the missing information (known to them) into their assessment. A third independent variable was originally included: Display Modality (text, graphic). The information content was identical between the two modes. However, since there were no statistically significant differences between the text and graphic displays, the data were collapsed over this variable.

**Dependent Variables**

*Decision Performance* was measured by the time to make a diversion decision, and correctness of the decision. Time to make a decision was the elapsed time from the start of the diversion task until participants made their diversion plan selection. The participant could accept one of the Diversion Aid plans or reject its recommendation(s). They could also reject all the plans shown if they felt that the options shown were flawed. In the medium/high Automation Visibility and low Information Quality trials, the best plan (correct selection) was listed below the automation’s highest ranked plan; in the low Automation Visibility condition, only one option shown by the automation, so if the participant recognized that there was missing information, he or she could choose to reject the plan. *Workload, Automation Awareness* and *Trust* were measured after each trial. See Table 3.

<table>
<thead>
<tr>
<th>Dep. Variable</th>
<th>Metric</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>Time to make a selection</td>
<td>once per trial</td>
</tr>
<tr>
<td>Performance</td>
<td>Selection of best plan</td>
<td>once per trial</td>
</tr>
<tr>
<td>Workload</td>
<td>NASA Task Load Index</td>
<td>once per trial</td>
</tr>
<tr>
<td>Automation Awareness</td>
<td>Understanding the basis of automation recommendations</td>
<td>once per trial</td>
</tr>
<tr>
<td>Trust</td>
<td>Survey question</td>
<td>once per trial</td>
</tr>
</tbody>
</table>

**Experimental Design and Procedure**

The experiment was a 2 (Information Quality) x 3 (Automation Visibility) within-subjects design. Scenario order was counterbalanced to avoid learning effects. Participants were trained on the Diversion Aid, tasks, and post-trial questionnaires. Participants practiced using the Diversion Aid in the simulator until they were able to make a diversion plan selection within a five minute time limit. Participants performed a total of six different diversion scenarios. After each diversion decision, they filled out the NASA-TLX workload scale and the post-trial questionnaire. After all six scenarios were completed, participants filled out a post-experiment questionnaire and were provided a short debrief.

**RESULTS**

A repeated measures analysis of variance was used. Results were considered significant if <= 0.05, and marginally significant if alpha > 0.05 and <= 0.1. All results charts depict means and standard error bars.

**Decision Performance**

*Plan Selection.* Participants correctly identified the best plan in 36% of the low Information Quality trials and 86% of
the high Information Quality trials. Information Quality was significant ($F_{(1,11)} = 32.98, p < 0.00013$). Automation Visibility was not significant.

**Time to make a Selection.** Automation Visibility was a significant factor ($F_{(2,22)} = 3.67, p < 0.042$) for this measure, with the low Automation Visibility condition being significantly faster ($t_{(22)} = 2.15, p < 0.043$) than the high Automation Visibility condition. Figure 1 shows the time to make a selection as a function of Automation Visibility.

**Workload**

**Total Workload.** Total workload showed no significant results for any of the dependent variables, nor their interaction.

**Mental Demand:** Automation Visibility was marginally significant ($F_{(2,22)} = 3.41, p < 0.051$). Low Automation Visibility level was significantly lower than the medium level ($t_{(22)} = 2.10, p < 0.047$; see Figure 2 right).

**Temporal Demand:** Automation Visibility was significant ($F_{(2,22)} = 4.56, p < 0.022$). The paired t-test showed that the low Automation Visibility level resulted in significantly lower temporal demand as compared to the medium level ($t_{(22)} = 2.38, p < 0.027$) and was also marginally significantly lower than the high Automation Visibility condition ($t_{(22)} = 1.92, p < 0.068$). Figure 2 (left) shows the results of the temporal demand as a function of Automation Visibility.

**Automation Awareness**

After every trial, participants were asked for their understanding of how the Diversion Aid arrived at its recommendations. Automation Visibility was significant factor for what? ($F_{(2,22)} = 5.08, p < 0.015$). Information Quality was marginally significant ($F_{(1,11)} = 4.11, p < 0.067$), and the interaction of these two independent variables was also marginally significant ($F_{(2,22)} = 2.66, p < 0.093$). The difference between medium and high Automation Visibility was significant ($t_{(22)} = 2.69, p < 0.013$) and the difference between the low and high Automation Visibility levels was marginally significant ($t_{(22)} = 1.69, p < 0.10$). See Figure 3.

**Figure 2. Selection time as a function of Automation Visibility.**

**Figure 3. Mental demand (left) and temporal demand (right).**

Although the low Automation Visibility level only provided one option, participants rated their understanding of its logic closer to that of high Automation Visibility level than medium level. Having only one option presented meant that participants only had to understand one plan, rather than having to understand three plans. With the costs included in the high Automation Visibility level, the details of the logic were much more readily available.

**Trust**

Trust was significantly affected by the level of automation visibility ($F_{(2,20)} = 4.18, p < 0.030$), information quality ($F_{(1,10)} = 6.26, p < 0.031$), as well as significant interaction between the two ($F_{(2,20)} = 4.15, p < 0.031$, respectively).

In the high information quality, trust in the Diversion Aid did not differ due to automation visibility. In the low information quality condition, however, trust was lowest when automation visibility was low and steadily increased as automation visibility increased (and participants were more able to understand what the automation was doing).

**DISCUSSION AND CONCLUSION**

Participants had a difficult time detecting when the automation was reasoning over incomplete information, despite the fact that the missing information was known to them. Participants only detected the incomplete information about a third of the time. Participants tended to over-trust the automation, so when information was missing and they were under high workload, they chose the top suggested plan even though it was not the truly best plan.

In our implementation of the Diversion Aid, trust in automation was reduced by low Information Quality, but was compensated for by increased Automation Visibility. In this study, high levels of trust in automation lead to failures to override sub-optimal Diversion Aid recommendations. This had a negative impact on decision performance when
ACKNOWLEDGEMENTS

This work was funded by the Federal Aviation Administration Human Factors Research and Engineering Division (contract #13-G-003, Iowa State University and Contract OTA DTFAWA-10-A-80031, Task 10, Honeywell). The opinions expressed herein are those of the authors and do not necessarily reflect the views of the FAA.

REFERENCES


