

Learning from Automation Surprises and “Going Sour” Accidents

David D. Woods¹ and Nadine B. Sarter²

¹Cognitive Systems Engineering Laboratory
Institute for Ergonomics
The Ohio State University

²Institute of Aviation - Aviation Research Laboratory
University of Illinois at Urbana-Champaign

In N. Sarter and R. Amalberti (Eds.)
Cognitive Engineering in the Aviation Domain,
Erlbaum, Hillsdale NJ, in press.

Introduction

Advances in technology and new levels of automation on commercial jet transports has had many effects. There have been positive effects from both an economic and a safety point of view. The technology changes on the flight deck also have had reverberating effects on many other aspects of the aviation system and different aspects of human performance. Operational experience, research investigations, incidents, and occasionally accidents have shown that new and sometimes surprising problems have arisen as well (Figure 1).

Insert figure 1 about here

Figure 1. The Reverberations of Technology Change on the Flightdeck for Human Performance

What are these problems with cockpit automation, and what should we learn from them?

- Do they represent over-automation or human error?
- Or instead perhaps there is a third possibility -- they represent coordination breakdowns between operators and the automation?

- Are the problems just a series of small independent glitches revealed by specific accidents or near misses?
- Do these glitches represent a few small areas where there are cracks to be patched in what is otherwise a record of outstanding designs and systems?
- Or do these problems provide us with evidence about deeper factors that we need to address if we are to maintain and improve aviation safety in a changing world?
- How do the reverberations of technology change on the flight deck provide insight into generic issues about developing human-centered technologies and systems (Winograd and Woods, 1997)?

Based on a series of investigations of pilot interaction with cockpit automation (Sarter and Woods, 1992; 1994; 1995; 1997a, 1997 b), supplemented by surveys, operational experience and incident data from other studies (e.g., Degani et al., 1995; Eldredge et al., 1991; Tenney et al., 1995; Wiener, 1989), we have found that the problems that surround crew interaction with automation are more than a series of individual glitches. These difficulties are symptoms that indicate deeper patterns and phenomena concerning human-machine cooperation and paths towards disaster. In addition, we find the same kinds of patterns behind results from studies of physician interaction with computer-based systems in critical care medicine (e.g., Moll van Charante et al., 1993; Obradovich and Woods, 1996; Cook and Woods, 1996). Many of the results and implications of this kind of research are synthesized and discussed in two comprehensive volumes, Billings (1996) and Woods et al. (1994).

This paper summarizes the pattern that has emerged from our research, related research, incident reports, and accident investigations. It uses this new understanding of why problems arise to point to new investment strategies that can help us deal with the perceived “human error” problem, make automation more of a team player, and maintain and improve safety.

The ability to step back and assess the implications of the research results was facilitated tremendously by our participation in a FAA team that examined the interface between flight crews and modern flight deck systems (Abbott et al., 1996). In this project, we were able to discuss the implications of observed difficulties with crew-automation coordination for investments to improve safety with a broad range of stakeholders in the aviation domain, including carrier organizations, line pilots, training managers, manufacturers, and industry groups. This effort helped us step back and assess the implications of the research for future investments to maintain and enhance aviation safety and safety in other related areas where new investments in automation are changing the roles of operational personnel.

Impact of Technology Change on Cognition and Collaboration

One way to recognize the pattern that underlies automation and human error is to listen to the voices we heard in our investigations. In these studies we interacted with many different operational people and organizations,

- directly in conversations about the impact of automation,
- through their judgments as expressed in surveys about cockpit automation,
- through their reported behavior in incidents that occurred on the line,
- through their performance in simulator studies that examined the coordination between crew and automated systems in specific flight contexts.

We will summarize the results of the multiple converging studies by adopting the point of view of different stakeholders and by expressing the research results and issues in their words. The statements are paraphrases of actual statements made to us in different contexts.

Automation Surprises: Coordination Breakdowns Between Crews and Automation

Pilots and instructors described and revealed the clumsiness and complexity of many modern cockpit systems. They described aspects of cockpit automation that were strong but sometimes silent and difficult to direct when time is short. We saw and heard how pilots face new challenges imposed by the tools that are supposed to serve them and provide “added functionality.” The users’ perspective on the current generation of automated systems is best expressed by the questions they pose in describing incidents (extended from Wiener, 1989):

- What is it doing now?
- What will do next?
- How did I get into this mode?
- Why did it do this?
- Stop interrupting me while I am busy.
- I know there is some way to get it to do what I want.
- How do I stop this machine from doing this?
- Unless you stare at it, changes can creep in.

These questions and statements illustrate why one observer of human-computer interaction defined the term *agent* as “A computer program whose user interface is so obscure that the user must think of it as a quirky, but powerful, person ...” (Lanir, 1995, p. 68).

Questions and statements like these point to automation surprises (Sarter, Woods and Billings, 1997), i.e., situations where crews are surprised by actions taken (or not taken) by the autoflight system. Automation surprises begin with miscommunication and misassessments between the automation and users which lead to a gap between the user’s understanding of what the automated systems are set up to do, what they are doing, and what they are going to do. The initial trigger for such a mismatch can arise from several

sources, for example, erroneous inputs such as mode errors or indirect mode changes where the system autonomously changes its status and behavior based on its interpretation of pilot inputs, its internal logic and sensed environmental conditions (Sarter and Woods, 1995; Sarter and Woods, 1997 a). The gap results in the crew being surprised later when the aircraft's behavior does not match the crew's expectations. This is where questions like, "Why won't it do what I want?" "How did I get into this mode?" arise.

It seems that the crew generally does not notice their misassessment from displays of data about the state or activities of the automated systems. The misassessment is detected, and thus the point of surprise is reached, in most cases based on observations of unexpected and sometimes undesirable aircraft behavior. Once the crew has detected the gap between expected and actual aircraft behavior, they can begin to respond to or recover from the situation. The problem is that this detection generally occurs when the aircraft behaves in an unexpected manner--flying past the top of descent point without initiating the descent, or flying through a target altitude without leveling off. If the detection of a problem is based on actual aircraft behavior, it may not leave a sufficient recovery interval before an undesired result occurs. Unfortunately, there have been accidents where the misunderstanding persisted too long to avoid disaster (cf., Billings, 1996).

The evidence shows strongly that the potential for automation surprises is greatest when three factors converge:

1. automated systems act on their own without immediately preceding directions from their human partner,
2. gaps in users' mental models of how their machine partners work in different situations, and
3. weak feedback about the activities and future behavior of the agent relative to the state of the world.

Automation surprises are one kind of breakdown in the coordination between crews and automated systems. Our investigations revealed a "funnel" of evidence about these kinds of coordination breakdowns. If we observe crews interacting with cockpit automation in full mission simulations, we find direct evidence of a variety of performance problems linked to the design of automation and to the training users receive. The problems observed are sometimes the result of "classic" human-computer interface design characteristics that lead to certain predictable forms of human error. If we look at operational experience we find that these coordination breakdowns and errors occur occasionally, but, in most cases, with no significant consequences. Unfortunately, we also have a small number of near misses or accidents where these same coordination breakdowns between crew and automation are a significant contributor to the sequence of events. In other words, there is a chain where:

- characteristics of the interface between automated systems and flight crews affect human performance in predictable and sometimes negative ways,

- there are precursor events where these performance problems occur but in innocuous circumstances or where the sequence of events is later redirected away from bad outcomes,
- occasionally, these problems occur in the context of more vulnerable circumstances, with other contributors present, and events spiral towards disaster.

The Going Sour Accident

These breakdowns in coordination between crew and automation create the potential for a particular kind of accident sequence -- the “going sour” accident (originally based on results from studying operating room incidents; Cook, Woods and McDonald, 1991). In this general class of accidents, an event occurs or a set of circumstances come together that appear to be minor and unproblematic, at least when viewed in isolation or from hindsight. This event triggers an evolving situation that is, in principle, possible to recover from. But through a series of commissions and omissions, misassessments and miscommunications, the human-automation team manages the situation into a serious and risky incident or even accident. In effect, the situation is managed into hazard.

Several recent accidents involving automation surprises show this signature. While they are classically referred to in aviation as controlled flight into terrain, some of these cases may better be described as managed flight into terrain since the automated systems are handling the aircraft and the flight crew is supervising the automation (for a brief overview of one vivid example of managed flight into terrain see Sarter et al., 1997).

The going sour scenario seems to be a side effect of complexity. Research and incident data raise the concern that new technology, when developed in a technology-driven rather human-centered way, is increasing the operational complexity and increasing the potential for the going sour signature (Billings, 1996).

After-the-fact, going sour incidents look mysterious and dreadful to outsiders who have complete knowledge of the actual state of affairs (Woods et al., 1994). Since the system is managed into hazard, in hindsight, it is easy to see opportunities to break the progression towards disaster. The benefits of hindsight allow reviewers to comment (Woods et al., 1994, chapter 6),

- “How could they have missed X, it was the critical piece of information?”
- “How could they have misunderstood Y, it is so logical to us?”
- “Why didn’t they understand that X would lead to Y, given the inputs, past instructions and internal logic of the system?”

In fact, one test for whether an incident is a going sour scenario is to ask whether reviewers, with the advantage of hindsight, make comments such as, “All of the necessary data was available, why was no one able to put it all together to see what it meant?”

The lesson learned from recent accidents involving breakdowns in the coordination between the automation and the flight crew is:

- the going sour scenario is an important general kind of accident category,
- there is a concern that this category represents a significant portion of the residual risks in aviation.

Only future data and events will reveal whether this is a growing part of the risk. Investments in turning cockpit automation into a team player and in training crews to better manage automated resources in a wide range of circumstances produce payoffs by guarding against this type of accident scenario.

Luckily, going sour accidents are relatively rare even in very complex systems. The going sour progression is usually blocked because of two factors:

- the expertise embodied in operational systems and personnel allows practitioners to avoid or stop the incident progression;
- the problems that can erode human expertise and trigger this kind of scenario are significant only when a collection of factors or exceptional circumstances come together.

Human Expertise and Technology-Induced Complexity

In our investigations we heard a great deal about how operators’ expertise usually compensates for the features of automation that contribute to coordination breakdowns. We heard about how training departments, line organizations, and individuals develop ways (through policies, procedures, team strategies, individual tactics and tricks) to get the job done successfully despite the clumsiness of some automated systems for some situations. Some of these are simply cautionary notes to pilots reminding them to “be careful, it can burn you.” Some are workarounds embodied in recipes. Some are strategies for teamwork. Many are ways to restrict the use of portions of the suite of automation in general or in particularly difficult situations. In other words, deficiencies in the design of the automation from a Human Factors point of view produce so few bad consequences because of human expertise and adaptation (Woods et al., 1994, chapter 5).

Overall, operational people and organizations tailor their behavior to manage the technology as a resource to get their job done, but there are limits on their ability to do this. Crew training is one of the primary tools for developing strategies and skills for managing automated systems as a set of resources (e.g., transition training as pilots move to a new glass cockpit aircraft and recurrent training). But there are many constraints that limit the amount and range of

training experiences pilots can receive. When we talked to training managers, we heard:

- “They’re building a system that takes more time than we have for training people.”
- “There is more to know--how it works, but especially how to work the system in different situations.”
- “The most important thing to learn is when to click it off.”
- “We need more chances to explore how it works and how to use it.”
- “Well, we don’t use those features or capabilities.”
- “We’ve handled that problem with a policy.”
- “We are forced to rely on recipe training much more than anyone likes.”
- “We teach them [a certain number of] basic modes in training, they learn the rest of the system on the line.”

Economic and competitive factors produce great pressure to reduce the training investment (e.g., shrink the training footprint or match a competitor’s training footprint). When there are improvements in training, these same forces lead people to take the benefit in productivity (the same level of proficiency in less time) rather than in quality (better training in the same time). People seem to believe that greater investments in automation promise lower expenditures on developing human expertise. However, the data consistently show that the impact of new levels and types of automation is new knowledge requirements for people in the system as their role changes to more of a manager and anomaly handler (e.g., Sarter et al., 1997). The goal of enhanced safety requires that we expand, not shrink, our investment in human expertise.

Complexity

The second reason why we see only a few accidents with the going sour signature is that breakdowns in coordination between human and automation are significant only when a collection of factors or exceptional circumstances come together. For example,

- human performance is eroded due to local factors (fatigue) or systemic factors (training and practice investments),
- crew coordination is weak,
- the flight circumstances are unusual and not well matched with training experiences,
- transfer of control between crew and automation is late or bumpy,
- small, seemingly recoverable erroneous actions occur, interact and add up.

Because there are always multiple contributors to a going sour incident and because these incidents evolve over time and a series of stages, it is easy to identify a host of places where a small change in human, team, or machine behavior could have re-directed the sequence away from any trouble. Focusing on any one of these points in isolation can lead to very local and manageable changes -- just shift the display slightly, modify a checklist, issue a

bulletin to remind crews of how X works in circumstance Y, reinforce a policy, add some remedial training.

While these changes may be constructive in small ways, they miss the larger lessons of this incident signature. When people and automation seem to mismanage a minor occurrence or non-routine situation into larger trouble, it is a symptom of overall system complexity. It is a symptom that all of the contributors to successful flight deck performance -- design, training, operational policies and procedures, certification -- need to be better coordinated.

The Escalation Principle

An underlying contributor to problems in human-automation coordination is the escalation principle (Woods et al., 1994). There is a fundamental relationship where the greater the trouble in the underlying process or the higher the tempo of operations, the greater the information processing activities required to cope with the trouble or pace of activities. For example, demands for monitoring, attentional control, information, and communication among team members (including human-machine communication) all tend to go up with the unusualness (situations at or beyond margins of normality or beyond textbook situations), tempo and criticality of situations. If workload or other burdens are associated with using a computer interface or with interacting with an autonomous or intelligent machine agent, these burdens tend to be concentrated at the very times when the practitioner can least afford new tasks, new memory demands, or diversions of his or her attention away from the job at hand to the interface per se. This is the essential trap of clumsy automation (Wiener, 1989)

Designer Reactions to Coordination Breakdowns: Erratic Human Behavior

Listen to how designers respond when they are confronted with evidence of a breakdown in the coordination between people and automation:

- The hardware/software system “performed as designed” (crashes of “trouble free” aircraft).
- “Erratic” human behavior (variations on this theme are “diabolic” human behavior; “brain burps,” that is, some quasi-random degradations in otherwise skillful human performance; irrational human behavior).
- The hardware/software system is “effective in general and logical to us, some other people just don’t understand it” (e.g., those who are too old, too computer phobic, or too set in their old ways).
- Those people or organizations or countries “have trouble with modern technology.”

- “We only provided what the customer asked for!” (or “we tried to talk them out of it, but we have to be customer-centered”).
- “I wanted to go further but ...” -- I was constrained by -- compatibility with the previous design, supplier’s standard designs, cost control, time pressure, regulations.
- Other parts of the industry “haven’t kept up” with the advanced capabilities of our systems (e.g., ATC does not accommodate the advanced capabilities and characteristics of the newer aircraft or ATC does not recognize what is difficult to do with highly automated aircraft under time pressure).

Some of these comments reflect real and serious pressures and constraints in the design world (e.g., design for multi-cultural users, economic pressures, very complex arrival and departure procedures).

Escaping from Attributions of Human Error versus Over-Automation

Overall, these kinds of comments from developers show how we remain locked into a mindset of thinking that technology and people are independent components --- either this electronic box failed or that human box failed. Too many reviewers and stakeholders, after-the-fact, attribute going sour incidents either to

- human error -- “clear misuse of automation ... contributed to crashes of trouble free aircraft” (La Burthe, 1997) or to
- over-automation -- “... statements made by ... Human Factors specialists against automation ‘per se’ ” (La Burthe, 1997).

This opposition is a profound misunderstanding of the factors that influence human performance. One commentator on human-computer interaction makes this point by defining the term interface as “an arbitrary line of demarcation set up in order to apportion the blame for malfunctions” (Kelly-Bootle, 1995, p. 101).

The primary lesson from careful analysis of incidents and disasters in a large number of industries is that going sour accidents represent a breakdown in coordination between people and technology (e.g., Norman, 1990). People cannot be thought about separately from the technological devices that are supposed to assist them. Technological artifacts can enhance human expertise or degrade it, “make us smart” or “make us dumb” (Norman, 1993).

The bottom line of recent research is that technology cannot be considered in isolation from the people who use and adapt it (e.g., Hutchins, 1995a). Automation and people have to coordinate as a joint system, a single team (Hutchins, 1995b; Billings, 1996). Breakdowns in this team’s coordination is an important path towards disaster. The real lessons of this type of scenario and the potential for constructive progress comes from developing better ways to coordinate the human and machine team -- human-centered design (Winograd and Woods, 1997).

Accident analyses suggest that breakdowns in human performance are a contributor to about 70 or 75% of aviation mishaps. Similar tabulations in other industries come up with about the same percentage. This should be interpreted as a motivation for paying increased attention to Human Factors. But some view these statistics superficially as an indication of a human error problem, and, as a result, they want to eliminate the human element, provide remedial training, or dictate all pilot action through expanded procedures.

However, research on the human contribution to safety and risk has found that “human error” is a symptom of deeper issues (Woods et al., 1994). To learn about these issues and constructively improve the system in which people function, these researchers have found that we need to go behind the label human error to identify and analyze the factors that influence human performance. In other words, there are organizational, training and design factors that influence human performance in predictable ways.

One simple and classic example of a kind of design induced error is the case of mode errors. Mode errors occur when an operator executes an intention in a way that would be appropriate if the device were in one configuration (one mode) when it is, in fact, in a different configuration. Note that mode errors are not simply just human error or a machine failure. Mode errors are a kind of human-machine system breakdown in that it takes both a user who loses track of the current system configuration, and a system that interprets user input differently depending on the current mode of operation (Sarter and Woods, 1995 a; Woods et al., 1994, chapter 5). The potential for mode error increases as a consequence of a proliferation of modes and interactions across modes without changes to improve the feedback about system state and activities. The resulting coupling, complexity and opacity of the automated system makes it difficult to train operators adequately for monitoring and managing these systems especially given resource limits for training. The result is gaps and misconceptions in users’ mental models of the automated system. In this example as in others, human, technological, and organizational factors interact, each affecting and being affected by the others.

Human Factors began and has always been concerned with the identification of design-induced error (ways in which things make us dumb) as one of its fundamental contributions to improved system design (e.g., Fitts, 1946; Fitts and Jones, 1947; Fitts, 1951 in the aviation domain). However, it is a profound misunderstanding of the research results to think that this implies a shift from – ‘the incident was caused by pilot error or operator error’ -- to – ‘the incident was caused by manager or designer error’. We make no progress if we trade pilot error for designer or manager error (Woods et al., 1994). There are always multiple contributors to failure each necessary but only jointly sufficient. Design and organizational factors often are a part of the set of contributors. But again the potential for progress comes from

understanding the factors that lead designers or managers inadvertently to shape human performance towards predictable forms of error through the clumsy use of technology or through inappropriate organizational pressures.

Strategies for Human-Centered Design

If diagnoses such as human error (be it operator, designer or manager) or over-automation are misleading and unproductive, then how do we make progress?

Human-Centered Design Is ...

A necessary first step is to adopt “human-centered” approaches to research and design (Billings, 1996). This perspective can be characterized in terms of three basic attributes: Human-centered design is problem-driven, activity-centered, and context-bound (Winograd and Woods, 1997).

1. Human-centered research and design is problem-driven.

A problem-driven approach begins with an investment in understanding and modeling the basis for error and expertise in that field of practice. What are the difficulties and challenges that can arise? How do people use artifacts to meet these demands? What is the nature of collaborative and coordinated activity across people in routine and exceptional situations?

2: Human-centered research and design is activity-centered.

In building and studying technologies for human use, researchers and designers often see the problem in terms of two separate systems (the human and the computer) with aspects of interaction between them. This focuses attention on the people or the technology in isolation, de-emphasizing the activity that brings them together. In human-centered design we try to make new technology sensitive to the constraints and pressures operating in the actual field of activity (Ehn, 1988; Flach and Dominguez, 1995).

New possibilities emerge when the focus of analysis shifts to the activities of people in a field of practice. These activities do or will involve interacting with computers in different ways, but the focus becomes the practitioner’s goals and activities in the underlying task domain. The question then becomes (a) how do computer-based and other artifacts shape the cognitive and coordinative activities of people in the pursuit of their goals and task context and (b) how do practitioners adapt artifacts so that they function as tools in that field of activity (Woods, in press).

3: Human-centered research and design is context-bound.

Human cognition, collaboration, and performance depend on context. A classic example is the representation effect -- a fundamental and much reproduced finding in Cognitive Science. How a problem is represented

influences the cognitive work needed to solve that problem, either improving or degrading performance (e.g., Zhang and Norman, 1994). In other words, the same problem from a formal description, when represented differently, can lead to different cognitive work and therefore different levels of performance. Another example is the data overload problem. At the heart of this problem is not so much the amount of data to be sifted through. Rather, this problem is hard because what data is informative depends on the context in which it appears. Even worse, the context consists of more than just the state of other related pieces of data; the context also includes the state of the problem solving process and the goals and expectations of the people acting in that situation.

Today, in most cases, new technology is developed based on its hypothesized impact on human cognition, collaboration, and performance (Winograd and Woods, 1997; Sarter et al., 1997; Woods, in press). Well-intentioned developers feel their work is human-centered because they are motivated thoughtful people, because they predict the new system will lead to improvements in performance, and because eventually they addressed the usability of the system developed. Despite such good intentions, development usually remains fundamentally technology-centered because developing the technology in itself is the primary activity around which all else is organized. The primary focus is pushing the technological frontier or creating the technological system, albeit a technology that seems to hold promise to influence human cognition, collaboration and activity. Eventually, interfaces are built which connect the technology to users. These interfaces typically undergo some usability testing and usability engineering to make the technology accessible to potential users. Knowledge of human-computer interaction and usability come into play, if at all, only at this later stage. However, there is a gap between designers' intentions to be user-centered and their actual technology-driven practice, which results in operational complexities like those on the automated flight deck. In other words, "the road to technology-centered systems is paved with user-centered intentions" (see Sarter et al., 1997).

Progress Depends on ...

At the broadest level, researchers have identified a few basic human-centered strategies that organizations can follow in an effort to increase the human contribution to safety:

- increase the system's tolerance to errors,
- avoid excess operational complexity,
- evaluate changes in technology and training in terms of their potential to create specific kinds of human error,
- increase skill at error detection by improving the observability of state, activities and intentions,

- invest in human expertise.

To improve the human contribution to safety several steps are needed. Design, operational, research, and regulatory organizations must all work together to adopt methods for error analysis and use them as part of design and certification. This creates a challenge to the Human Factors community - - to work with industry to turn research results into practical methods (valid but resource economical) that test for effective error tolerance and detection. The goal is to improve the ability to detect and eliminate design and other factors that create predictable errors.

Avoid Excess Operational Complexity

Avoiding excess operational complexity is a difficult issue because no single person or organization decides to make systems complex. But in the pursuit of local improvements or in trying to accommodate multiple customers, systems gradually get more and more complex as additional features, modes, and options accumulate. The cost center for this increase in complexity is the user who must try to manage all of these features, modes and options across a diversity of operational circumstances. Failures to manage this complexity are categorized as “human error.” But the source of the problem is not inside the person. The source is the accumulated complexity from an operational point of view. Trying to eliminate “erratic” behavior through remedial training will not change the basic vulnerabilities created by the complexity. Neither will banishing people associated with failures. Instead human error is a symptom of systemic factors. The solutions are system fixes that will involve coordination of multiple parties in the industry. This coordinated system approach must start with meaningful information about the factors that predictably affect human performance.

Mode simplification is illustrative of the need for change and the difficulties involved. Not all modes are used by all pilots or carriers due to variations in operations and preferences. Still they are all available and contribute to complexity. Not all modes are taught in transition training; only a set of “basic” modes is taught, and different carriers define different modes as “basic.” Which modes represent excess complexity and which are essential for safe and efficient operation? Another indication of the disarray in this area is that modes which achieve the same purpose have different names on different flight decks.

Making progress in simplifying requires coordination across an international, multi-party industry that is competitive in many ways but needs to be collaborative in others.

One place where mode simplification is of very great importance is the interaction across modes (indirect mode changes or mode reversions).

Indirect mode changes have been identified as a major factor in breakdowns in teamwork between pilots and automation. Simplifying these transitions and making transitions better fit pilot models is another very high priority area for improvement.

Error Detection through Improved Feedback

Research has shown that a very important aspect of high reliability human-machine systems is effective error detection. Error detection is improved by providing better feedback, especially feedback about the future behavior of the aircraft, its systems or the automation. In general, increasing complexity can be balanced with improved feedback. Improving feedback is a critical investment area for improving human performance and guarding against going sour scenarios. But where and how to invest in better feedback?

One area of need is improved feedback about the current and future behavior of the automated systems. As technological change increases machines' autonomy, authority and complexity, there is a concomitant need to increase observability through new forms of feedback emphasizing an integrated dynamic picture of the current situation, agent activities, and how these may evolve in the future. Increasing autonomy and authority of machine agents without an increase in observability leads to automation surprises. As discussed earlier, data on automation surprises has shown that crews generally do not detect their miscommunications with the automation from displays about the automated system's state, but rather only when aircraft behavior becomes sufficiently abnormal.

This result is symptomatic of low observability where observability is the technical term that refers to the cognitive work needed to extract meaning from available data. This term captures the relationship among data, observer and context of observation that is fundamental to effective feedback. Observability is distinct from data availability, which refers to the mere presence of data in some form in some location. For human perception, "it is not sufficient to have something in front of your eyes to see it" (O'Regan, 1992, p.475).

Observability refers to processes involved in extracting useful information. It results from the interplay between a human user knowing when to look for what information at what point in time and a system that structures data to support attentional guidance (see Rasmussen, 1985; Sarter, Woods and Billings, 1997). The critical test of observability is when the display suite helps practitioners notice more than what they were specifically looking for or expecting (Sarter and Woods, 1997 a).

One example of displays with very low observability on the current generation of flight decks is the flight mode annunciations on the primary

flight display. These crude indications of automation activities contribute to reported problems with tracking mode transitions. As one pilot mentioned to us, “changes can always sneak in unless you stare at it.” Simple injunctions for pilots to look closely at or call out changes in these indications generally are not effective ways to redirect attention in a changing environment. Minor tuning of the current mode annunciations is not very likely to provide any significant improvement in feedback. Researchers and industry need to cooperate to develop, test and adopt fundamentally new approaches to inform crews about automation activities.

The new concepts need to be:

- transition-oriented -- provide better feedback about events and transitions,
- future-oriented -- the current approach generally captures only the current configuration; the goal is to highlight operationally significant sequences and reveal what should happen next and when,
- pattern-based -- pilots should be able to scan at a glance and quickly pick up possible unexpected or abnormal conditions rather than have to read and integrate each individual piece of data to make an overall assessment.

For example, making vertical navigation modes more comprehensible and usable is likely to require some form of vertical profile display. The moving map display for horizontal navigation is a tremendous example of the desired target -- an integrated display that provides a big picture of the current situation and especially the future developments in a way that supports quick check reading and trouble detection. However, developing displays to support vertical navigation based on the above criteria is much more difficult because it is inherently a four dimensional problem. The industry as a whole needs to develop and test new display concepts to support pilot management of vertical navigation automation.

Going sour incidents and accidents provide evidence that improved feedback is needed. Despite the conflict with economic pressures, prudence demands that we begin to make progress on what is better feedback to support better error detection and recovery. To do this we need a collaborative process among manufacturers, carriers, regulators, and researchers to prototype, test in context, and adopt new innovations to aid awareness and monitoring. We need to move forward on this to ensure that, when the next window of opportunity opens up, we are ready to provide more observable and comprehensible automation.

How to Provide Better Feedback: Bumpy Transfer of Control

Let us look at one example of a coordination breakdown between crews and flight deck automation. Automation can compensate for trouble silently (Norman, 1990). Crews can remain unaware of the developing trouble until the automation nears the limits of its authority or capability to compensate.

The crew may take over too late or be unprepared to handle the disturbance once they take over, resulting in a bumpy transfer of control and significant control excursions. This general problem has been a part of several incident and accident scenarios. One example of this is asymmetric lift conditions caused by icing or engine trouble.

In contrast, in a well-coordinated human team, the active partner would comment on the unusual difficulty or increasing effort needed to keep the relevant parameters on target. Or, in an open environment, the supervisor could notice the extra work or effort exerted by his or her partner and ask about the difficulty, investigate the problem, or intervene to achieve overall safety goals.

How can we use the analogy to a well coordinated human team working in an open visible environment to guide how we can provide more effective feedback and better coordination between human and machine partners? For the set of feedback problems that arise when automation is working at the extreme ends of its envelope or authority, improved displays and warnings need to indicate:

- when the automation is having trouble handling the situation (e.g., turbulence);
- when the automation is taking extreme action or moving towards the extreme end of its range of authority;
- when agents are in competition for control of a flight surface.

This specifies a performance target. The design question is how to make the system smart enough to communicate this intelligently? How to define what are “extreme” regions of authority in a context sensitive way? When is an agent having trouble in performing a function, but not yet failing to perform? How and when does one effectively communicate moving towards a limit rather than just invoking a threshold crossing alarm?

From experience and research we know some constraints on the answers to these questions. Threshold crossing indications (simple alarms) are not smart enough -- thresholds are often set too late or too early. We need a more gradual escalation or staged shift in level or kind of feedback. An auditory warning that sounds whenever the automation is active (e.g., an auditory signal for trim-in-motion) may very well say too much. We want to indicate trouble in performing the function or extreme action to accomplish the function, not simply any action.

We know from experiences in other domains and with similar systems that certain errors can occur in designing feedback. These include:

- nuisance communication such as voice alerts that talk too much in the wrong situations,
- excessive false alarms,

- distracting indications when more serious tasks are being handled (e.g., a constant trim warning or a warning that comes on at a high noise level during a difficult situation -- “silence that thing!”).

In other words, misdesigned feedback can talk too much, too soon or it can be too silent, speaking up too little, too late as automation moves towards authority limits.

Should the feedback occur visually or through the auditory channel or through multiple indications? Should this be a separate new indication or integrated into existing displays? Should the indication be of very high perceptual salience; in other words, how strongly should the signal capture pilot attention? Working out these design decisions requires developing prototypes in terms of:

- perceptual salience relative to the larger context of other possible events and signals,
- along a temporal dimension (when to communicate relative to the priority of other issues or activities going on then),
- along a strength dimension (how much or how little to say and at what level of abstraction relative to ongoing activities)

and adjusting these attributes based on data on crew performance.

Developing effective feedback about automation activities requires thinking about the new signals or indications in the context of other possible signals and different kinds of situations. One cannot improve feedback or increase observability by adding a new indication or alarm to address each case one at a time as they arise. A piecemeal approach will generate more displays, more symbolic codings on displays, more sounds, more alarms. More data will be available, but this will not be effective feedback because it challenges the crew’s ability to focus on and digest what is relevant in a particular situation. Instead, we need to look at coherent sets and subsets of problems which all point to the need for improved feedback to devise an integrated solution.

Our analysis of this one example has identified the relevant human-machine performance targets, identified relevant scenarios for design and testing, set some bounds on effective solutions, identified some tradeoffs that must be balanced in design, and mentioned some of the factors that will need to be explored in detail through prototypes and user testing. The example illustrates the complexity of designing for observability.

Mechanisms to Manage Automated Resources

Giving users visibility into the machine agent's reasoning processes is only one side of the coin in making machine agents into team players. Without also giving the users the ability to direct the machine agent as a resource in their reasoning processes, the users are not in a significantly improved

position. They might be able to say what's wrong with the machine's solution, but remain powerless to influence it in any way other than through manual takeover. The computational power of machine agents provides a great potential advantage, i.e., to free users from much of the mundane legwork involved in working through large problems, thus allowing them to focus on more critical high-level decisions. However, in order to make use of this potential, the users need to be given the authority and capabilities to make those decisions. This means giving them control over the problem solution process.

A commonly proposed remedy for this is to allow users to interrupt the automated agent and take over the problem in its entirety in situations where users determine that the machine agent is not solving a problem adequately. Thus, the human is cast into the role of critiquing the machine, and the joint system operates in essentially two modes - fully automatic or fully manual. The system is a joint system only in the sense that either a human agent or a machine agent can be asked to deal with the problem, not in the more productive sense of the human and machine agents cooperating in the process of solving the problem. This method, which is like having the automated agent say "either you do it or I'll do it," has many obvious drawbacks. Either the machine does the entire job without benefiting from the practitioner's information and knowledge, and despite the brittleness of the machine agents; or the user takes over in the middle of a deteriorating or challenging situation without the support of cognitive tools. Previous work in several domains (space operations, electronic troubleshooting, aviation) and with different types of machine agents (expert systems, cockpit automation, flight path planning algorithms) has shown that this is a poor cooperative architecture. Instead, users need to be able to continue to work with the automated agents in a cooperative manner by taking control of the automated agents.

Using the machine agent as a resource may mean various things. In terms of observability, one of the main challenges is to determine what levels and modes of interaction will be meaningful to users. In some cases, the users may want to take very detailed control of some portion of a problem, specifying exactly what decisions are made and in what sequence, while in others the users may want only to make very general, high level corrections to the course of the solution in progress. Accommodating all of these possibilities is difficult and requires very careful iterative analysis of the interactions between user goals, situational factors, and the nature of the machine agent. However, this process is crucial if the joint system is to perform effectively in the broadest possible range of scenarios.

Enhancing Human Expertise

The last area for investment in the interest of improving the human contribution to safety is human expertise. It is ironic that the aviation industry seems to be reducing this investment at the very time when it points to human performance as a dominant contributor to accidents. This reflects one of the myths about the impact of automation on human performance is that, as investment in automation increases, less investment is needed in human expertise. In fact, many sources have shown how increased automation creates new and different knowledge and skill requirements.

In our investigations, we heard operational personnel say that the complexity of the automated flight deck means that pilots need new knowledge about how the different automated subsystems and modes function. We heard about investigations that show how the complexity of the automated flight deck makes it easy for pilots to develop oversimplified or erroneous mental models of the tangled web of automation modes and transition logics. We heard from training departments struggling to teach crews how to manage the automated systems as a resource in differing flight situations. Many sources offered incidents where pilots were having trouble getting a particular mode or level of automation to work successfully, where they persisted too long trying to get this mode of automation to carry out their intentions instead of switching to another means or a more direct means to accomplish their flight path management goals. For example, someone may ask, “Why didn’t you turn it off?” Response: “It didn’t do what it was supposed to, so I tried to get it to do what I had programmed it to do.” We heard how the new knowledge and skill demands are most relevant in relatively rare situations where different kinds of factors push events beyond the routine -- just those circumstances that are most vulnerable to going sour through a progression of misassessments and miscommunications. This increases the need to practice those kinds of situations.

For training managers and departments, the result is a great deal of training demands that must be fit into a small and shrinking training footprint. The combination of new roles, knowledge and skills as a result of new levels of automation with economic pressures creates a training double bind.

We heard about many tactics that have been developed to cope with this mismatch. For example, one tactic is to focus transition training on just a basic set of modes and leaving the remainder to be learned on the line. This can create the ironic situation that training focuses on those parts of managing automated systems that are the easiest to learn, while deferring the most complicated parts for individuals to learn later on their own. This tactic works:

- if the basics provide a coherent base that aids learning the more difficult parts or for coordinating the automation in more difficult circumstances,
- if there is an environment that encourages, supports and checks continued learning beyond minimum requirements.

Another tactic used to cope with this training double bind is to teach recipes. It is a time efficient tactic and helps prevent students from being overwhelmed by the complexity of the automated systems. Still, instructors and training managers acknowledge the limits of this approach and try to go beyond recipes as much as their time and resources limits allowed. All spoke of the need for pilots to practice what they have learned in realistic operational settings through line oriented simulation and line oriented flight training scenarios, although the scope of this training is limited by the economic and competitive forces squeezing training time. We saw evidence of an industry struggling to get better utilization of limited transition training time and limited recurrent checks.

As the training footprint shrinks, one response is to identify and focus in on the highest priority training needs. The US industry has increased freedom to do so under new programs with the FAA (the advanced qualification program or AQP). However, laudable as this is, it can inadvertently reduce resources for training and practice even further. Economic pressure means that the benefits of improvements will be taken in productivity (reaching the same goal faster) rather than in quality (more effective training). Trying to squeeze more yield from a shrinking investment in human expertise will not help prevent the kinds of incidents and accidents that we label human error after-the-fact.

Escaping from this double bind is essential. A first step is to recognize the limits of minimum requirements. Instead, we should produce a culture oriented towards continuous learning. Initial or transition training should produce an initial proficiency for managing the automated flight deck. This training should serve as the platform for mechanisms that support continued growth of expertise. An emphasis on continuous improvement beyond initial proficiency is needed because with highly automated systems we see an increase in knowledge requirements and the range of situations that pilots must be able to master. Developing accurate and useful mental models that can be applied effectively across a wide range of possible conditions depends on part-task or full mission practice in line-oriented situations.

The question then becomes how can we expand the opportunities to practice the management of automated resources across a wide variety of situations throughout a pilot's career? In many ways, the aviation industry is well prepared to adopt this approach. Pilots, in general, want to improve their knowledge and skills as evidenced by pilot-created guides to the automation that we noticed in several training centers. The industry already has invested heavily in line-oriented training. New training technology in the form of less expensive but high fidelity, part-task training devices is being utilized more.

Progress through Coordination

These comments also illustrate a general theme that emerges from research on Human Factors problems in industries with demands for very high levels of performance. Representatives of each segment of industry are under constraints and pressures (fit it all into this training footprint; minimize the changes from the previous flight deck, etc.). Each group knows that they are doing the best job possible given the constraints placed on them. So when evidence of glitches arises, it is natural that they look for solutions in other areas that contribute to flight deck performance, for example:

- trainers may advocate re-designing the system so that we can train people to use it within this limited resource window,
- designers may advocate efforts to get ATC to accommodate our automation's idiosyncrasies and capabilities,
- designers may encourage others to provide better training to enable people to cope with the large set of interconnected features designed as a result of multiple market demands,
- trainers may lobby for modified regulations so they are not forced to spend precious training time on items of lower priority for glass aircraft.

None of these solutions is wrong in detail -- all of these areas can be improved in isolation. But there is a deeper reading to these messages. This kind of circular reaction to evidence of glitches is symptomatic of a deeper need for coordination across areas that traditionally have functioned mostly autonomously -- training, design, operational procedures, certification. Each one of them, when considered alone, has improved a great deal, and this has created the generally extremely high safety levels in the aviation industry. However, the risk of failure exemplified by the going sour scenario involves the interaction or coupling between these individual areas.

In fact, increasing the level of automation increases the coupling between these areas. For example, many recognize how automation designers, in part, specify an operational philosophy. We have heard many people comment on the inadequacy of a 'throw-it-over-the-wall' linkage between design and training or between manufacturer and operator (and develop means to try to reduce this). There needs to be a closer integration of these multiple perspectives, in part, because of the advanced technology on new aircraft. This example of coupling can be extended to show how many other areas have become more inter-related with increased flight deck automation -- air traffic control (ATC) and advanced aircraft, safety and economics.

While improvements are still possible and desirable in each area as an isolated entity, progress in general demands the integration of multiple perspectives. In part, this is due to the fact that all parts of the system are under intense economic pressure. This means that training no longer has the room to make up for design deficiencies. Design for learnability becomes

another constraint on designers. ATC demands interact with the capabilities and the limits of managing advanced aircraft, yet ATC is a system undergoing change in the face of economic and performance pressures as well. A complex departure procedure may seem to increase throughput, at least on paper, but it may exact a price in terms of managing a clumsy team member -- the automation -- and erode safety margins to some degree. Coordination is needed precisely because change in any one part of the aviation system has significant effects for other parts of the system.

Conclusion

Overall, there are broad patterns behind the details of particular incidents and accidents. We need to better guard against the kind of incident where people and the automation seem to mismanage a minor occurrence or non-routine situation into larger trouble -- the going sour scenario. This scenario is a symptom of breakdowns in coordination between people and machines which in turn are a symptom of overall system complexity, at both operational and organizational levels (Woods, 1996).

Second, we can tame needed complexity

- through better feedback to operational personnel,
- through more practice at managing automated resources in a wide range of circumstances,
- by making the automation function as a team player,
- by creating “intuitive” automation designs that can be learned quickly, through better mechanisms to detect or predict where automation design will produce predictable kinds of human performance problems.

In general, we can act by first trying to limit the growth in complexity through checking for excess complexity, valuing simplicity of operation, increasing coordination between coupled areas.

Meeting the challenges of going sour scenarios in a coordinated manner is extremely difficult because any change will exact costs on the parties involved. Since the benefits are at a system level, it is easy for each party to claim that they should not pay the costs, but that some other part of the industry should. Since the aggregate safety level is very high (actuarial risk is low), it is easy to ignore the threat of the going sour scenario and argue that the status quo is sufficient. This is particularly easy because going sour incidents by definition involve several local contributing factors. Each case looks like a unique combination of events with the dominant common factor being human error. The certification and legal climate have produced a climate where change creates exposure to financial and competitive risk. This leads to minimum standards based on past practices (“you approved this before”; “it was safe enough before”) and progress crawls to a halt. Yet

progress, despite its pains, is exactly what is demanded if observed difficulties such as the going sour scenario are to be addressed. The question for regulators, manufacturers, and operators then is how to build the collaborative environment that can enable constructive forward movement. The goal of our research has been to point out specific areas for constructive continuing progress and more general directions that may help create a collaborative environment where progress is possible.

Acknowledgments

The research on which this assessment is based was sponsored by NASA Ames Research Center (under Cooperative Agreement NCC 2-592; Technical Monitors Dr. Everett Palmer and Dr. Kevin Corker) and NASA Langley Research Center (Grant ; Technical Monitor Dr. Kathy Abbott). We wish to thank the many pilots, instructors and designers who contributed to our specific investigations and who shared their views and experiences with us in many different ways. We are indebted to our fellow members on the FAA team that examined the interface between flight crews and modern flight deck systems -- the discussions and debates they sparked were critical to our reflections on the implications of the research results.

References

- Abbott, K., Slotte, S., Stimson, D., Bollin, E., Hecht, S., Imrich, T., Lalley, R., Lyddane, G., Thiel, G., Amalberti, R., Fabre, F., Newman, T., Pearson, R., Tigchelaar, H., Sarter, N., Helmreich, R., and Woods, D. (1996). *The Interface Between Flightcrews and Modern Flight Deck Systems*. Federal Aviation Administration : Washington D. C. June 18.
- Billings, C. E. (1996). *Aviation Automation: The Search For A Human-Centered Approach*. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- Cook, R. I. and Woods, D. D. (1996), Implications of automation surprises in aviation for the future of total intravenous anesthesia (TIVA). *Journal of Clinical Anesthesia*, 8:29s-37s.
- Cook, R. I., Woods, D. D. and McDonald, J.S. (1991). Human Performance in Anesthesia: A Corpus of Cases. Cognitive Systems Engineering Laboratory Report, prepared for Anesthesia Patient Safety Foundation, April 1991.
- Degani, A., Shafto, M., and Kirlik, A. (1995). Mode usage in automated cockpits: Some initial observations. In T. B. Sheridan (Ed.), *Proceeding of the International Federation of Automatic Control; Man-Machine Systems (IFAC-MMS) Conference*. Boston, MA: IFAC.

- Ehn, P. (1988). *Work-Oriented Design of Computer Artifacts*. Stockholm, Sweden: Arbetslivscentrum.
- Eldredge, D., Dodd, R.S., and Mangold, S.J. (1991). *A review and discussion of Flight Management System incidents reported to the Aviation Safety Reporting System*. (Battelle Report, prepared for the Department of Transportation). Columbus, OH: Volpe National Transportation Systems Center.
- Fitts, P. M. (1946). Psychological requirements in aviation equipment design. *Journal of Aviation Medicine*, 17(3), 270-275.
- Fitts, P. M. (1951). Engineering psychology and equipment design. In S. S. Stevens (ed.), *Handbook of Experimental Psychology*, 1287-1340. NY, Wiley.
- Fitts, P. M. and Jones, R. E. (1947). Analysis of factors contributing to 460 "pilot-error" experiences in operating aircraft controls. Memorandum Report TSEAA-694-12. Dayton, OH: Aero Medical Laboratory, Air Materiel Command.
- Flach, J. M. and Dominguez, C. O. (1995). Use-Centered Design: Integrating the User, Instrument, and Goal. *Ergonomics in Design*, July.
- Hutchins, E. (1995 a). How a cockpit remembers its speeds. *Cognitive Science*, 19, 265--288.
- Hutchins, E. (1995 b). *Cognition in the Wild*. Cambridge MA, MIT Press.
- Hutchins, E. (1996). The Integrated Mode Management Interface. Technical Report, Department of Cognitive Science, University of California, San Diego.
- Kelly-Bootle, S. (1995). *The Computer Contradictionary* (2nd edition). MIT Press, Cambridge MA.
- La Burthe, C. (1997). Human Factors perspective at Airbus Industrie. Presentation at International Conference on Aviation Safety and Security in the 21st Century. January 13-16, Washington DC.
- Lanir, J. (1995). Agents of alienation. *Interactions*, July, 66-72.
- Moll van Charante, E., Cook, R. I., Woods, D.D., Yue, L. and Howie, M.B. (1993). Human-computer interaction in context: Physician interaction with automated intravenous controllers in the heart room. In H.G.

Stassen, editor, *Analysis, Design and Evaluation of Man-Machine Systems 1992*, Pergamon Press, p. 263-274.

Norman, D. A. (1990). The 'problem' of automation: Inappropriate feedback and interaction, not 'over-automation.' *Philosophical Transactions of the Royal Society of London*, B 327:585--593.

Norman, D. A. (1993). *Things that Make us Smart*. Addison-Wesley: Reading, MA.

Obradovich, J. H. and Woods, D. D. (1996). Users as designers: How people cope with poor HCI design in computer-based medical devices. *Human Factors*, 38(4), 574-592.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46, 461-488.

Rasmussen, J. (1985). Trends in human reliability analysis. *Ergonomics*, 28 (8), 1185-1196.

Sarter, N. B. (1996). From Quantity to Quality, From Individual Pilot to Multiple Agents: Trends in Research on Cockpit Automation In R. Parasuraman and M. Mouloua, editors, *Automation and Human Performance*, Erlbaum.

Sarter, N. B. and Woods, D. D. (1992). Pilot Interaction with Cockpit Automation I: Operational Experiences with the Flight Management System. *International Journal of Aviation Psychology*, 2:303--321.

Sarter, N. B. and Woods, D. D. (1994). Pilot Interaction with Cockpit Automation II: An Experimental Study of Pilot's Model and Awareness of the Flight Management System. *International Journal of Aviation Psychology*, 4:1-28.

Sarter, N. B. and Woods, D. D. (1995). "How in the world did we get into that mode?" Mode error and awareness in supervisory control. *Human Factors*, 37: 5-19.

Sarter, N. B., Woods, D. D. and C. Billings, C. (1997). *Automation Surprises*. G. Salvendy, editor, *Handbook of Human Factors/Ergonomics*, second edition, Wiley, New York.

Sarter, N. B. and Woods, D. D. (1997 a). Teamplay with a Powerful and Independent Agent: A Corpus of Operational Experiences and Automation Surprises on the Airbus A-320. *Human Factors*, in press.

- Sarter, N. B. and Woods, D. D. (1997 b). Mode Errors of Omission and Commission: Observed Breakdowns in Pilot-Automation Coordination In A Full Mission Simulation Study. Manuscript submitted for publication.
- Tenney, Y.J., Rogers, W.H., and Pew, R.W. (1995). *Pilot Opinions on High Level Flight Deck Automation Issues: Toward the Development of a Design Philosophy* (NASA Contractor Report # 4669). Hampton, VA: NASA Langley Research Center.
- Wiener, E. L. (1989). *Human factors of advanced technology ("glass cockpit") transport aircraft*. (NASA Contractor Report No. 177528). Moffett Field, CA: NASA-Ames Research Center.
- Winograd, T. and Woods, D. D. (1997). Challenges for Human-Centered Design. In *Human-Centered Systems: Information, Interactivity, and Intelligence*. J. Flanagan, T. Huang, P. Jones, S. Kasif, (Eds.), National Science Foundation, Washington DC, July, 1997.
- Woods, D. D. (1996). Decomposing Automation: Apparent Simplicity, Real Complexity, In R. Parasuraman and M. Mouloua, editors, *Automation Technology and Human Performance*, Erlbaum.
- Woods, D. D. (1997). Designs are Hypotheses about How Artifacts Shape Cognition and Collaboration. *Ergonomics*, in press.
- Woods, D. D., Johannesen, L., Cook, R. I. and Sarter, N. B. *Behind Human Error: Cognitive Systems, Computers and Hindsight*. Crew Systems Ergonomic Information and Analysis Center, WPAFB, Dayton OH, 1994.
- Zhang, J. and Norman, D.A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, 18: 87--122