How Unexpected Events Produce An Escalation Of Cognitive And Coordinative Demands

David D. Woods
Emily S. Patterson
Institute for Ergonomics
The Ohio State University

In Stress Workload and Fatigue.
P. A. Hancock and P. Desmond (eds.)

Explaining the Clumsy Use of Technology

Each round of technological development promises to aid the people engaged in various fields of practice. After these promises result in the development of prototypes and fielded systems, those researchers who examine the reverberations of technology change have observed a mixed bag of effects, most quite different from the expectations of the technology advocates. Often the message practitioners send with their performance, their errors, and their adaptations is one of technology-induced complexity. In these cases, technological possibilities are used clumsily so that systems intended to serve the user turn out to add new burdens that congregate at the busiest times or during the most critical phases of the task (e.g., Woods, Johannesen, Cook, & Sarter, 1994, chapter 5, Woods & Watts, 1997).

Although this pattern has been well documented in a variety of areas such as cockpit automation (Sarter, Woods & Billings, 1997) and many principles for more effective human-machine and human-human cooperation have been developed (e.g., Norman, 1988), we have a gaping explanatory problem. There is a striking contrast between the persistence of the optimism of developers who before the fact expect each technological development to produce significant performance improvements and the new operational complexities that are observed after the fact. It seems difficult for all kinds of people in design teams to predict or anticipate operational complexities. Yet operational complexities are easy to see when the right scenarios are examined e.g., through using prototypes in appropriate scenarios or through incidents during practice.

Ultimately, we need to explain why this technology-induced complexity occurs so often when designers fully expect these systems to produce major benefits for the practitioners. There are many factors that could be invoked to explain this observation. Some may fall into hoary cliches about the need for human factors in the design process. Others may examine the pressures on development and developers. Here we explore one factor that contributes in part -- a fundamental
dynamic relationship between problem demands, cognitive and coordinated activities, and the artifacts intended to support practitioners.
The Escalation Principle

On the basis of observations of anomaly response in many supervisory control domains in both simulated and actual incidents, a pattern seemed to recur. When an anomaly occurred and people began to recognize various unexpected events, there was a process of escalation of cognitive and coordinated activities. During these periods of escalating demands, we observed the penalties associated with poor design of systems that had been intended to support practitioners.

Escalation Principle: The concept of escalation concerns a process – how situations move from canonical or textbook to non-routine to exceptional. In that process, escalation captures a relationship -- as problems cascade, they produce an escalation of cognitive and coordinative demands that bring out the penalties of poor support for work.

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 knowledge, monitoring, attentional control, information, and communication among team members (including human-machine communication) all tend to increase with the unusualness, tempo, and criticality of situations. If workload or other burdens associated with using a computer interface or with interacting with an autonomous or intelligent machine agent tend to be concentrated at these times, the workload occurs 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 or computerized device per se.

The concept of escalation captures a dynamic relationship between the cascade of effects that follows from an event and the demands for cognitive and collaborative work that escalate in response (Woods, 1994). An event triggers the evolution of multiple interrelated dynamics.

- There is a cascade of effects in the monitored process.
  A fault produces a time series of disturbances along lines of functional and physical coupling in the process (e.g., Abbott, 1990). These disturbances produce a cascade of multiple changes in the data available about the state of the underlying process, for example, the avalanche of alarms following a fault in process control applications (Reiersen, Marshall, & Baker, 1988).
• Demands for cognitive activity increase as the problem cascades. More knowledge potentially needs to be brought to bear. There is more to monitor. There is a changing set of data to integrate into a coherent assessment. Candidate hypotheses need to be generated and evaluated. Assessments may need to be revised as new data come in. Actions to protect the integrity and safety of systems need to be identified, carried out, and monitored for success. Existing plans need to be modified or new plans formulated to cope with the consequences of anomalies. Contingencies need to be considered in this process. All these multiple threads challenge control of attention and require practitioners to juggle more tasks.

• Demands for coordination increase as the problem cascades. As the cognitive activities escalate, the demand for coordination across people and across people and machines rises. Knowledge may reside in different people or different parts of the operational system. Specialized knowledge and expertise from other parties may need to be brought into the problem-solving process. Multiple parties may have to coordinate to implement activities aimed at gaining information to aid diagnosis or to protect the monitored process. The trouble in the underlying process requires informing and updating others – those whose scope of responsibility may be affected by the anomaly, those who may be able to support recovery, or those who may be affected by the consequences the anomaly could or does produce.

• The cascade and escalation is a dynamic process. A variety of complicating factors can occur, which move situations beyond canonical, textbook forms. The concept of escalation captures this movement from canonical to nonroutine to exceptional. The tempo of operations increases following the recognition of a triggering event and is synchronized by temporal landmarks that represent irreversible decision points.

The dynamics of escalation vary across situations. First, the cascade of effects may have different time courses. For example, an event may manifest itself immediately or may develop more slowly. Second, the nature of the responses by practitioners affects how the incident progresses – less appropriate or timely actions (or too quick a reaction in some cases) may sharpen difficulties, push the tempo in the future, or create new challenges. Different domains may have different escalation gradients depending on the kinds of complicating factors that occur, the rhythms of the process, and consequences that may follow from poor performance.

• Interactions with computer based support systems Interactions with computer based support systems occur in the context of these escalating demands on memory and attention, monitoring and assessment, communication and response.

In canonical (routinized or textbook) situations, technological systems seem to integrate smoothly into work practices, so smoothly that seemingly little cognitive work is required. However, cognitive work grows and patterns of distribution of this work over people and machines grow more complex as
situations cascade. Thus, the penalties for poor coordination between people and machines and for poor support for coordination across people emerge as the situation escalates demands for cognitive work.

The difficulties arise because interacting with the technological devices is a source of workload as well as a potential source of support. When interacting with devices or others through devices creates new workload burdens when practitioners are busiest, new attentional demands when practitioners are plagued by multiple voices competing for their attention, new sources of data when practitioners are overwhelmed by too many channels spewing out too much competing data, practitioners are placed in an untenable situation.

As active, responsible agents in the field of practice, practitioners adapt to cope with these bottlenecks in many ways – they eliminate or minimize communication and coordination with other agents, they tailor devices to reduce cognitive burdens, they adapt their strategies for carrying out tasks, they abandon some systems or modes when situations become more critical or higher tempo. Woods et al. (1994) devoted a chapter to examples of these workload bottlenecks and the ways that people tailor devices and work strategies to cope with this technology-induced complexity. Sarter et al. (1997) summarized this dynamic for cockpit automation. Cook and Woods (1996) captured this dynamic for a case of operating room information technology. Patterson and Woods (1997) described the strategies used in one organization, space shuttle mission control, for successfully coping with escalating demands following an anomaly. In this case, practitioners who are assigned on-call responsibility invest in building a prior understanding of the mission context before problems occur in order to be able to come into an escalating situation more effectively should an anomaly actually occur.

**An Example of Escalating Demands**

To illustrate the escalation principle, consider an anomaly that occurred during the ascent phase of a space shuttle mission (Watts, Woods, & Patterson, 1996; Watts-Perotti & Woods, 1997; for a more publicized case, examine the escalating demands on mission control during the Apollo 13 accident; e.g., Murray & Cox, 1989). As shown in Figure 1, an unexpected event produced an escalation of cognitive and coordinative demands and activities. In the figure, the escalating demands are grouped into three temporal units that roughly capture portions of the escalation process.

Several minutes into the ascent phase of the mission, one of the controllers responsible for monitoring the health and safety of the mechanical systems noticed an anomaly – an unexpected drop in hydraulic fluid in an auxiliary power unit (APU). The personnel monitoring immediately recognized that the symptoms indicated a hydraulic leak. Did this anomaly require an immediate abort of the ascent? In other words, how bad was the leak? Was it a threat to the safety of the mission? What were the relevant criteria (and who knew them, and where did they reside)? The mechanical systems controllers did a quick
calculation that indicated the leak rate was below the predetermined abort limit – the mission could proceed to orbit. The analysis of the event relative to an abort decision occurred very quickly, in part because the nature of the disturbance was clear and because of the potential consequences with an anomaly at this stage for the safety of the astronauts.

As the ascent continued, the figure points to a second collection of demands and activities that were intertwined and went on in parallel. The controllers for the affected system informed the Flight Director and the other members of the mission control team of the existence of the hydraulic leak and its severity. Because of the nature of the tools for supporting coordination across controllers (voice loops), this occurred in a very cognitively economical way for all concerned (see Watts et al., 1996). The team also had to plan how to respond to the anomaly before the transition from the ascent to the orbit phase was completed. As in all safety-critical systems, planning was aimed both at how to obtain more information to diagnose the problem as well as how to protect the affected systems. This planning required resolving conflicting goals of maximizing the safety of the systems as well as determining as confidently as possible the diagnosis of the anomaly. The team decided to alter the order in which the auxiliary power units (APUs) were shut down to obtain more diagnostic information. This change in the mission plan was then communicated to the astronauts.

A third group in the figure refers to demands and activities that occurred after the initial assessment, responses, and communications. Information about the assessments of the situation and the changed plans were available to other controllers who were or might be affected by these changes. This happened because other personnel could listen in on the voice loops and overhear the previous updates provided to the flight director and the astronauts about the hydraulic leak. After the changes in immediate plans were communicated to the astronauts, the controllers responsible for other subsystems affected by the leak and the engineers who designed the auxiliary power units contacted the mechanical systems controllers to gain further information. In this process, new issues arose, some were settled, but these issues sometimes needed to be revisited or reemerged.

For example, a series of meetings between the mechanical systems controllers and the engineering group was called. These meetings served an important role in the process to assess contingencies and to decide how to modify mission plans such as a planned docking with the MIR space station and entry. In addition, they provided opportunities to detect and correct errors in the assessment of the situation, to calibrate the assessments and expectations of differing groups, to anticipate more possible side effects of changing plans (see Watts-Perottti & Woods, 1997, for a complete analysis of these functions of cooperative work in this case).

Additional personnel were called in and integrated with others to help with the new workload demands and to provide specialized knowledge and expertise. In this process, the team expanded to include an impressive number of agents
acting in a variety of roles and teams all coordinating their efforts.

| Communicate with flight director, affected controllers, engineers, and management |
| Schedule meetings to discuss the impacts |
| Call in additional personnel and update them |
| Develop long-term response plan |
| Inform flight director of the leak |
| Plan how to protect system |
| Plan how to obtain more diagnostic information |
| Inform astronauts of leak and immediate changes to plan |
| Calculate hydraulic fluid loss |
| Recalculate hydraulic fluid loss |
| Diagnose as a leak |
| Justify diagnosis as a leak |
| Determine immediate abort is not needed |
| Debate how long shuttle should remain in orbit |

Fig. 1. Escalation of cognitive and coordinated work following an anomaly in space shuttle mission control.
Escalation Helps Explain Episodes in the Clumsy Use of Technology

The escalation principle helps to explain some recurrent phenomena in cognitive work. We will briefly refer to two recurrent patterns in the impact of technology on practitioners. One is clumsy automation where automation introduced to lower workload and free up resources actually creates new bottlenecks in higher tempo and more critical situations (Sarter, Woods & Billings, 1997). Another is how attempts to provide intelligent diagnostic systems with explanation capabilities have failed to make these artificial intelligence (AI) systems into team players (Malin et al., 1991).

Clumsy Automation

The escalation of problem demands helps explain a syndrome, which Wiener (1989) has termed “clumsy automation.” Clumsy automation is a form of poor coordination between human and machine in the control of dynamic processes where the benefits of the new technology accrue during workload troughs, and the costs or burdens imposed by the technology occur during periods of peak workload, high criticality, or high-tempo operations. Despite the fact that these systems are often justified on the grounds that they help offload work from harried practitioners, we find that they in fact create new additional tasks, force the user to adopt new cognitive strategies, require more knowledge and more communication at the very times when the practitioners are most in need of true assistance. This creates opportunities for new kinds of human error and new paths to system breakdown that did not exist in simpler systems (Woods & Sarter, in press).

We usually focus on the perceived benefits of new automated systems, and assume that introducing new automation leads to lower workload and frees up limited practitioner resources for other activities (Sarter et al., 1997). Our fascination with the possibilities afforded by automation often obscures the fact that new automated devices also create new burdens and complexities for the individuals and teams of practitioners responsible for operating, troubleshooting, and managing high-consequence systems. The demands may involve new or changed tasks such as device setup and initialization, configuration control, or operating sequences. Cognitive demands change as well, creating new interface management tasks, new attentional demands, the need to track automated device state and performance, new communication or coordination tasks, and new knowledge requirements. These demands represent new levels and types of operator workload.

The dynamics of these new demands are an important factor because in complex systems human activity ebbs and flows, with periods of lower activity and more self-paced tasks interspersed with busy, high-tempo, externally paced operations where task performance is more critical (Rochlin, La Porte, & Roberts, 1987). Although technology is often designed to shift workload or tasks from the human to the machine, the critical design feature for well-integrated cooperative
cognitive work between the automation and the human is not the overall or time-averaged task workload. Rather, it is how the new demands created by the new technology interact with low-workload and high-workload periods, how they affect the transition from canonical to more exceptional situations, and especially how they affect the practitioner’s ability to manage workload as situations escalate. It is these relationships that make the critical difference between clumsy and skillful use of the technological possibilities.

**Failure of Machine Explanation to Make AI Systems Team Players**

The concept of escalation helps us understand why efforts to add machine explanation to intelligent systems failed to support cooperative interactions with human practitioners. Typically, expert systems developed their own solution to the problem at hand. Potential users found it difficult to accept such recommendations without some information about how the AI system arrived at its conclusions. This led many to develop ways to represent knowledge in such systems so they could provide a description of how a system arrived at the diagnosis or solution (e.g., Chandrasekaran, Tanner & Josephson, 1989).

However they were generated and however they were represented, these explanations were provided at the end of some problem-solving activity after the intelligent system had arrived at a potential solution. As a result, they were one-shot, retrospective explanations for activity that had already occurred. The difficulties with explanations of this form generally went unnoticed. Effort was focused on building the explanation-generating mechanisms and knowledge representations. Development was directed toward contexts (or a simplified piece of a context was abstracted) where the underlying system was static and unchanging and where temporal relations were not significant. Even then a few noticed (e.g., Cawsey, 1992), in contrast to the assumptions of developers, when people engage in collaborative problem solving, they tend to provide information about the basis for their assessments as the problem-solving process unfolds to build a common ground for future coordination (e.g., Clark & Brennan, 1991; Johannesen, Cook & Woods, 1994).

Warnings about problems with one-shot, retrospective explanations were disregarded until AI diagnostic systems were applied to dynamic situations. After such prototypes or systems had to deal with beyond-textbook situations, escalation occurred. The explanation then occurred at a time when the practitioner was likely to be engaged in multiple activities as a consequence of the cascade of effects of the initial event and escalating cognitive demands to understand and react as the situation evolved. These activities included generating and evaluating hypotheses, dealing with a new event or the consequences of the fault(s), planning corrective actions or monitoring for the effects of interventions, attempting to differentiate the influences caused by faults and those caused by corrective actions, among others.
These kinds of expert systems did not act as cooperative agents. For example, the expert systems did not gauge the importance or length of their messages against the background context of competing cues for attention and the state of the practitioner’s ongoing activity. Thus, the system’s output could occur as a disruption to other ongoing lines of reasoning and monitoring (Woods, 1995).

In addition, the presence of the intelligent system created new demands on the human practitioner. The typical one-shot retrospective explanation was disconnected from other data and displays the practitioner was examining. This meant the practitioner had to integrate the intelligent systems assessment with other available data as an extra task. This new task required the practitioner to shift attention away from what was currently going on in the process, possibly resulting in missed events.

Overall, the one-shot, retrospective style of explanation easily broke down under the demands of escalation. Practitioners, rather than being supported by the new systems, found extra workload during high-tempo periods and a new source of data competing for their attention when they were already confronted with an avalanche of changing data. As a result, practitioners adapted. They simply ignored the intelligent system (e.g., Remington & Shafto, 1990, for one case; Malin et al., 1991 for the general pattern).

There are several ironies about this pattern of technology change and its surprising reverberations. First, it had happened before. The same experience had occurred in the early 1980s when nuclear power tried to automate fault diagnosis with non-AI techniques. The systems were unable to function autonomously and only exacerbated the data overload that operators confronted when a fault produced a cascade of disturbances (Woods, 1994). That attempt to automate diagnosis was abandoned, although the organizations involved and the larger research community failed to see the potential to learn about dynamic patterns in human-machine cooperation.

A second irony is that to make progress in supporting human performance, efforts have moved away from autonomous machine explanations and toward understanding cooperative work and the ways that cognitive activity is distributed (Hutchins, 1995). The developers had assumed that their intelligent system could function essentially autonomously (at least on the important components of the task) and would be correct for almost all situations. In other words, they designed a system that would take over most of the cognitive work. The idea that human-intelligent system interaction required significant and meaningful cooperative work adapted to the changing demands and tempo of situations was outside their limited understanding of the cognitive demands of actual fields of practice.

Implications
At the beginning of the chapter, we posed a question -- why is technology so often used clumsily, creating new complexities for already beleaguered practitioners?

The concept of escalation provides a partial explanation. In canonical cases the technology seems to integrate smoothly into the work practices. The practitioners are able to process information from machine agents. The additional workload of coordinating with a machine agent is easily managed. More static views of the work environment may be acceptable simplifications for textbook situations.

The penalties for poor design of supporting artifacts emerge only when unexpected situations dynamically escalate cognitive and coordination demands. In part, developers miss higher demand situations when design processes remain distant and disconnected from the actual demands of the field of practice. The current interest in field-oriented design techniques such as work analysis, cognitive task analysis, and ethnography reflects this state of affairs.

In part, developers misread and rationalize away the evidence of trouble created by their designs in some scenarios. This can occur because situations that escalate are relatively less frequent than canonical cases. Also, because practitioners adapt to escape from potential workload bottlenecks as criticality and tempo increase, the user hides the evidence that the system does not fit operational demands (Woods et al., 1994; Cook & Woods, 1996).

However, most important is that almost all design processes, including most human factors specialties, have missed the process of moving from canonical to exceptional that the concept of escalation captures. Supporting the escalation in cognitive and coordination activity as problems cascade is a critical design task (Patterson, Woods, Sarter, & Watts-Perotti, 1998). To cope with escalation as a fundamental characteristic of cognitive work, one needs to design:

- how more knowledge and expertise are integrated into an escalating situation,
- how more resources can be brought to bear to handle the multiple monitoring and attentional demands of escalating situations (Watts-Perotti & Woods, 1997),
- how to bring practitioners up to speed quickly when they are called in to support others (Patterson & Woods, 1997).

Many have noticed that scenario design is a critical activity for human-centered design processes (Carroll, 1997). Because escalation is fundamental to cognitive work, it specifies one target for scenario design. Field work techniques, such as building and analyzing corpuses of critical incidents, are needed to understand how situations move from textbook to nonroutine to exceptional in particular fields of practice, and particularly how this occurs after significant organizational or technological changes. Work is needed to identify general and specific complicating factors that shift situations beyond textbook plans (Roth & Mumaw, 1993).
The concept of escalation is not simply about problems, demands on cognition or on collaboration, or technological artifacts. Rather, it captures a dynamic interplay between all these factors. As a result, escalation illustrates a fundamental point distinguishing Cognitive Systems Engineering from other disciplines -- joint and distributed cognitive systems are the fundamental unit of analysis for progress on understanding and designing systems of people and technology at work (Woods & Roth, 1988; Hutchins, 1995; Woods, 1998). Escalation, in particular, and distributed cognitive systems, in general, are concerned with relationships between problem demands, cognitive and coordinated activity, and artifacts.

Acknowledgments

This work was supported by NASA Johnson Space Center (grant NAGW – 4560, Human Interaction Design for Anomaly Response Support, and grant NAG 9-786, Human Interaction Design for Cooperating Automation) with special thanks to Dr. Jane Malin and her colleagues at NASA Johnson. Additional support was provided by a National Science Foundation Graduate Fellowship. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.
References


Piaget's stages describe how cognitive development occurs in four discrete stages. Learn more about each stage of Piaget's influential theory. The theory outlines four distinct stages from birth through adolescence, focusing on how children acquire knowledge, reasoning, language, morals, and memory. Piaget's stages of development are: Stage.