ELSEVIER

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Int. J. Human-Computer Studies 58 (2003) 637–670

International Journal of
Human-Computer
Studies

www.elsevier.com/locate/ijhcs

Human performance and embedded intelligent technology in safety-critical systems

Martha Grabowski[a,b,*], Stephen D. Sanborn[c]

[a] *Information Systems Program, Le Moyne College, Syracuse, NY, USA*
[b] *Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute,*
CII 5213 110 8th Street, Troy, NY, USA
[c] *General Electric Global Research and Development Center, Schenectady, NY, USA*

Received 1 May 2002; accepted 13 January 2003

Abstract

Information technology continues to evolve rapidly. We see this particularly in the evolution of embedded intelligent systems—knowledge-based systems deployed in larger hosts with real-time response requirements, which provide real-time advice, guidance, information, recommendations and explanations to their users. These systems have recently been deployed in safety-critical large-scale systems, where humans and technology are jointly responsible for executing tasks, monitoring operations, and providing system safety. Thus, human interaction with intelligent technology in safety-critical systems has important implications. Those interactions can enhance or reduce system efficiency, enhance or compromise safety, and augment or negate the other benefits that technology provides. In this paper, we focus on interactions between human operators and embedded intelligent systems. We first consider the role of technology in safety-critical systems, and discuss studies of the impact of technology on human operators in such systems. We then describe embedded intelligent systems, and studies of their impacts on human operators. To illustrate these points, we consider the case of embedded intelligent technology introduction in one such setting, and the results of an empirical investigation of the impact of the technology on human performance in that system. We conclude with a discussion of the implications of the study and of the importance of understanding the impact of embedded intelligent technology on human operators in safety-critical systems.

---

*Corresponding author. Information Systems Program, Le Moyne College, Syracuse, NY, USA. Tel.: +1-315-445-4427; fax: +1-518-276-8227.

E-mail addresses: grabowsk@lemoyne.edu (M. Grabowski), sanborn@ge.com (S.D. Sanborn).

URL: http://web.lemoyne.edu/~grabowsk.

1071-5819/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S1071-5819(03)00036-3

1. Introduction

Information technology continues to evolve rapidly. We see this particularly in the evolution of intelligent systems over the past two decades. Recently, embedded intelligent systems—knowledge-based systems deployed in larger hosts with real-time response requirements—have been introduced in a variety of settings, providing real-time advice, guidance, information, recommendations and explanations to their users. These systems are essential to the functioning of many safety-critical large-scale systems, such as ship and space shuttle control systems (Heudin, 1991; Coenen et al., 1989), air traffic control systems (Perry, 1997), nuclear power plant control systems (Wong and Kalam, 1995), intelligent highway control systems (Dailey et al., 1993), flexible manufacturing systems (Ben-Arieh et al., 1988), patient monitoring systems in intensive care units (Leveson and Turner, 1993; Bogner, 1994) and military and defense systems (Rouse et al., 1990).

In safety-critical large-scale systems, humans and technology are jointly responsible for executing tasks, monitoring operations and providing system safety. Human interaction with technology in safety-critical systems therefore has important implications. Those interactions can enhance or reduce system efficiency, enhance or compromise safety, and augment or negate the other benefits that technology provides. The interactions can also have significant impact on human performance in such systems. These costs and benefits have been noted especially in the case of cockpit automation (Wiener, 1988), as well as space shuttle operations (Vaughan, 1996), nuclear power plants (Perrow, 1984; Sagan, 1993), and medical equipment and systems (Leveson and Turner, 1993; Institute of Medicine, 1999). Thus, understanding human–computer interactions and their implications in safety-critical systems is important.

In this paper, we focus particularly on interactions between human operators and one type of technology, embedded intelligent systems. We consider first the role of technology in safety-critical systems, and discuss studies of the impact of technology on human operators in such systems. We then describe embedded intelligent systems, and studies of their impacts on human operators. To illustrate these points, we consider the case of embedded intelligent technology introduction in one such setting, and the results of an empirical investigation of the impact of the technology on human performance in that system. We conclude with a discussion of the implications of the study and of the importance of understanding the impact of embedded intelligent technology on human operators in safety-critical systems.

2. Technology in safety-critical systems

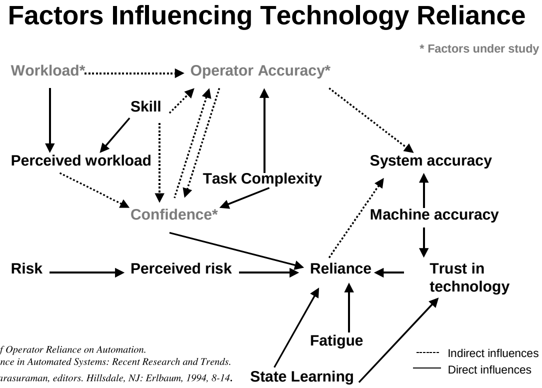
Safety-critical systems are comprised of human, technical, organizational and social elements, each important to the system's safety and reliability. Technology is often introduced into safety-critical systems to improve system performance, to remove error, and to increase safety. Technology can assist in problem recognition, identification of emergent failure, and in anticipating patterns that might lead to disaster (Sagan, 1993; Rochlin, 1997; Tenner, 1996). Technology can also improve hazard awareness, provide alerts, identify conflicts, eliminate routine actions that allow humans to concentrate on other tasks, and reduce unnecessary communication in congested situations. Thus, technology is often introduced into safety-critical systems in order to impact performance in the system and environment.

Technology impacts in safety-critical systems have often been studied, many times following a major incident or calamity. Both poorly designed technology and improper or pathological use of technology have been identified as contributing to major disasters (Rogers et al., 1992; Read, 1993; Vaughan, 1996), and the benefits of technology introduction have not always been realized. For instance, technology designed to reduce operator workload sometimes increases it (Bainbridge, 1983), and the introduction of technology can lead to a number of undesirable impacts on system operators—manual skill deterioration, alteration of workload patterns, poor monitoring, inappropriate responses to alarms and reductions in job satisfaction (Wiener and Curry, 1980; Vaughan, 1996). Problems can also arise from unanticipated interactions between technology, human operators and other systems in the environment (Tenner, 1996).

In some settings, entirely new human error forms can surface with the introduction of technology (Wiener, 1988), and "automation surprises" can puzzle the operator (Sarter and Woods, 1994). Technology that represents a considerable increase in complexity over previous systems can be one source of new error forms. Systems with high levels of complexity can increase the number of intervening subsystems between operators and the technology. This can have the effect of decreasing the direct control functions of the operator and increasing their "peripherilzation" (Norman, 1990). These problems have also been noted in studies addressing "clumsy automation" (Cook et al., 1990; Woods and Cook, 1991; Woods et al., 1991).

Trust in technology is an important factor influencing the impacts of technology in safety-critical systems. Trust is defined here as confidence or reliance upon the actions or information of another in an exchange (Ring and Van de Ven, 1994; Hosmer, 1995). Technology that is reliable, accurate and useful may nevertheless not be used if an operator believes that it is untrustworthy (Sheridan, 1988). Initially, users trust technology, and they expect a system to be accurate (Muir, 1988). However, trust is dynamic, and will change depending on user experience with a system (Lee and Moray, 1992). Users will weigh each experience with the technology differently, depending on the "risk" involved. The result is that trust will affect whether and how users use technology (Lee and Moray, 1994).

Riley (1994) suggests that reliance on technology is multiply determined, and varies over time (Fig. 1). A variety of factors can influence an operator in this regard: the system's accuracy and reliability; the complexity of the tasks being supported; the operator's workload, skills and abilities; operators' perceptions of risk in the system and their trust of technology; and the nature of the environment in which the technology is deployed. Users' confidence in their own skills and abilities can also determine trust in technology (Lee and Moray, 1994). Other contributing factors to reliance on technology may be the ease with which technology failures can be



Riley, V. A Theory of Operator Reliance on Automation. In Human Performance in Automated Systems: Recent Research and Trends. M. Mouloua & R. Parasuraman, editors. Hillsdale, NJ: Erlbaum, 1994, 8-14.

Fig. 1. Factors influencing technology reliance.

detected, and the ease with which the technology may be enabled and disabled (Lee and Moray, 1992).

Trust in technology is not always advantageous: in fact, some technology can be "over trusted," in the sense that operators may come to rely uncritically on it without recognizing its limitations or failing to monitor its inputs. Technology trust can lead operators to erroneous conclusions, to rely on single sources of information, and to fail to monitor displays and instruments, as cited in numerous accident reports (e.g. National Transportation Safety Board, 1973). The grounding of the cruise ship *Royal Majesty* provides an example of the dangers induced by over reliance or over trust in technology. In this case, crewmembers failed to monitor the ship's position using means independent of the electronic global positioning system (GPS). The result was the grounding of the vessel and lessons learned about the importance of a "trust but verify" attitude with respect to technology (National Transportation Safety Board, 1997).

Trusted systems, therefore, can pose a number of difficulties for operators in safety-critical systems—complacency, over reliance, over trust, lack of vigilance and error. In this research, we were able to measure several constructs that influence human reliance on technology, although we were not able to measure trust directly. We investigated the effects of one type of technology on operators in a safety-critical system. We describe that technology in the following section.

3. Embedded intelligent systems

Embedded systems are resident within a larger host and are constrained by the parameters, requirements and performance of that host. *Intelligent systems* exhibit cognitive processes that replicate, emulate or approximate human reasoning (Cohen, 1995). Intelligent systems are increasingly being embedded within larger host systems with demanding response requirements (Highland, 1994). Embedded intelligent systems are typically required to provide reasoning and decision support consistent with the host's decision support; provide support or notification when the host system slows or degrades; and provide decision support within a real-time response envelope (Paul et al., 1991; Chen et al., 1995). *Embedded intelligent systems*, therefore, provide advice, recommendations, conclusions or explanations in real time to their users—human operators, or their automated hosts. Operators and hosts rely on this guidance and advice for decision-making, for automated execution of tasks, for informing teams and individuals, and for enhancing productivity, efficiency, effectiveness and workflow.

There have been a number of studies investigating the impact of embedded intelligent technology on human performance. These studies focused on system validation, verification and software performance, asking the questions: Was the right system built (validation)? Was the system built right (verification)? And how did the system perform? (Florac and Carleton, 1999).

Validations of embedded intelligent systems have addressed whether the system provided reasoning and decision support consistent with the host's decision support; support or notification when the host system slows or degrades; or decision support within a real-time response envelope (Paul et al., 1991; Chen et al., 1995). For some systems, system predictability is more important than sheer speed, particularly for hard real-time scheduling in process-oriented embedded systems (Cullyer, 1991). Thus, evaluations of embedded intelligent systems have focused on the system's fault tolerance, reliability, predictability, transparency and viability (Shimeall and Leveson, 1991; Wen et al., 1997), as well as the system's modularity, robustness and architecture efficiency (Cohen, 1995), and the technology's response time and workload (Jaffe et al., 1991; Chandrasekaran et al., 1991).

How users use intelligent technology and how such technology impacts its users are also important considerations. Some studies have considered human performance with and without technology (Grabowski and Wallace, 1993; Entin and Serfaty, 1997), as well as a technology's impact on its users (Hogarth and Einhorn, 1992), its host (Leveson et al., 1991), and its environment (Rochlin, 1997). Evaluations of mappings between user mental models and intelligent systems architectures have also been performed, as have evaluations of the adequacy of intelligent systems' user interfaces (Adelman et al., 1993; Hall, 1988), and the impact of automated technology on users (Roth et al., 1987). Thus, the impacts of embedded intelligent systems have been studied, with differing results and metrics. However, little empirical work investigating the impacts of embedded intelligent technology on human performance in safety-critical systems has been undertaken. We present the results of such an investigation in the following section.

4. Example: human performance with embedded intelligent technology

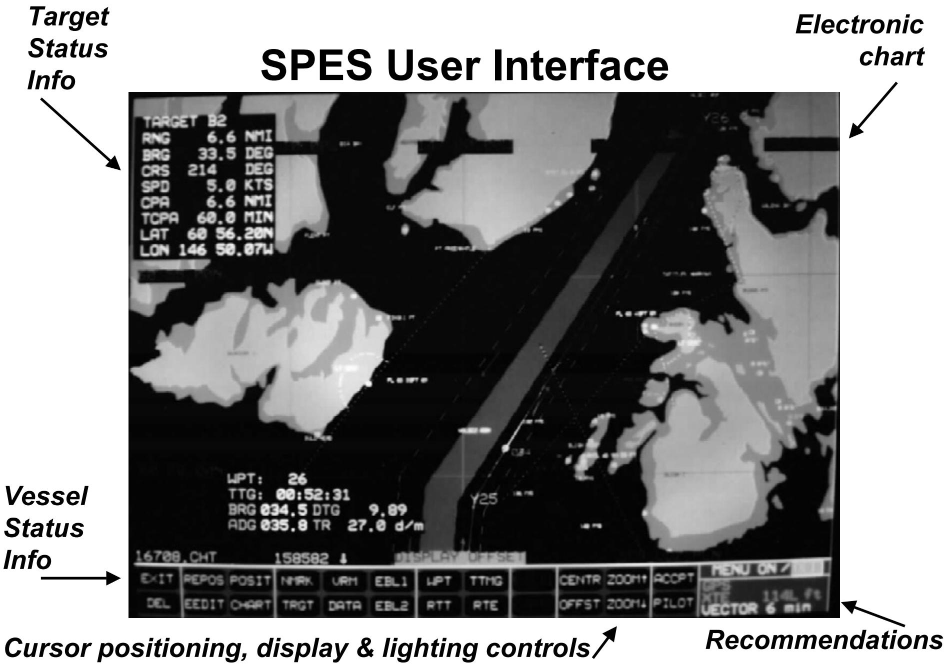
In this research, we considered the impact of embedded intelligent technology on human performance in one safety-critical system—marine transportation. Many studies of shipboard automation have been conducted over the past 30 years. These studies considered the potential for improved navigational safety in different settings, using a variety of treatments. However, few empirical studies considering the impact of an embedded intelligent system on human operators have been conducted. Shipboard automation studies have considered human performance during watch keeping and collision avoidance tasks, using such measures as vessel trackkeeping (Cook et al., 1981; Schuffel et al., 1989); closest point of approach (CPA) (Williams and Goldberg, 1982); and the frequency and magnitude of engine and rudder orders (Cooper et al., 1981). Safety of navigation in these studies was typically associated with close adherence to the vessel's intended track (i.e. small cross-track errors and CPAs), fewer rudder and engine orders, and with rudder and engine orders of small magnitude.

These shipboard automation studies also used mental workload measures. For instance, Schuffel et al. (1989) used an auditory Continuous Mental Task as a secondary mental task to infer mental workload during navigation. Fee et al. (1980) used a more complex auditory cue and response task for measuring a ship pilot's mental workload and Nieri (1980) used a two-tone auditory cued response task. It is significant that many of the studies cited were simulator-based evaluations. Given the hazardous nature of vessel operations, especially in confined waters, it is often difficult to conduct empirical technology impact studies in operational settings. Thus, earlier shipboard automation studies considered human performance measures as well as cognitive workload measures in their evaluations.

As a research vehicle for this study, we utilized an operational embedded intelligent system for ship navigation and piloting known as the Shipboard Piloting Expert System (SPES) (Grabowski and Sanborn, 2001). The SPES is an intelligent decision aid embedded within its host, a ship's navigation system. Following the *Exxon Valdez* oil spill in 1989, the SPES was developed by Rensselaer Polytechnic Institute for Exxon Shipping Company tank vessels in the trans Alaskan pipeline trade with support provided by the US Department of Transportation, Coast Guard and Maritime Administration; the National Oceanographic and Atmospheric Administration (NOAA); Exxon Shipping Company; and the Southwest Alaska Pilots Association. The SPES provides expert piloting knowledge to the ship's bridge watch team, and was embedded in its host aboard the 973-ft, 173 000 deadweight ton Exxon tankship, the *Exxon Benicia*, which operates between Valdez, Alaska and oil terminals on the West Coast of the United States.

The task to be supported by the technology was ship's piloting, a cognitively complex task comprised of three activities (trackkeeping, maneuvering and collision avoidance and the practice of good seamanship), utilizing three types of information (local knowledge, shiphandling knowledge and transit-specific knowledge) (Grabowski and Wallace, 1993; National Research Council, 1994).

The host for the embedded intelligent system, the *Exxon Benicia's* navigation system, was comprised of system sensors (i.e. depth sounders), radars, navigation sensors (i.e. radio direction finders, the GPS), and the ExxBridge integrated bridge system, a common display system for electronic chart information. The SPES was thus embedded in its host, and its information was visible to its users via the host's electronic bridge system, the ExxBridge (Fig. 2).



TARGET B2
RNG 6.6 NMI
BRG 33.5 DEG
CRS 214 DEG
SPD 5.0 KTS
CPA 6.6 NMI
TCPA 60.0 MIN
LAT 60 56.20N
LON 146 50.07W

WPT: 26
TTG: 00:52:31
BRG 034.5 DTG 9.89
ADG 035.8 TR 27.0 d/m
16700.CHT 158582

Fig. 2.

The SPES display provided a real-time plan view of the vessel's position in the waterway superimposed on an electronic chart of the waterway, along with the vessel's voyage plan, and graphic and text displays of shiphandling, navigation and maneuvering information (Fig. 3). Vessel status information was provided in the lower left-hand corner of the display, next to the cursor positioning, display and lighting, and configuration controls in the lower-right-hand corner of the display. Target status information was provided in the upper-left-hand corner of the display, and flashing colored alerts, alarms and recommendations were provided in the lower-right-hand corner of the display. The electronic chart consumed most of the screen display, which was consistent with user needs and expectations for graphical depictions of the vessel's transit through the waterway (National Research Council, 1994).

The SPES reasoned about the information from the ExxBridge in real-time; determined the implications of the information; generated alerts, alarms and advisories regarding potentially hazardous situations; and formulated recommendations about voyage plan alterations, courses to steer, times to turn and actions to follow. The SPES displayed its alerts, alarms and recommendations as graphical overlays on the vessel's electronic chart (e.g. as a flashing icon indicating collision danger with a target), as text output of recommendations and explanations, and as auditory signals associated with alerts and alarms. The ExxBridge display was the central navigational display on the ship's bridge, and was positioned next to the

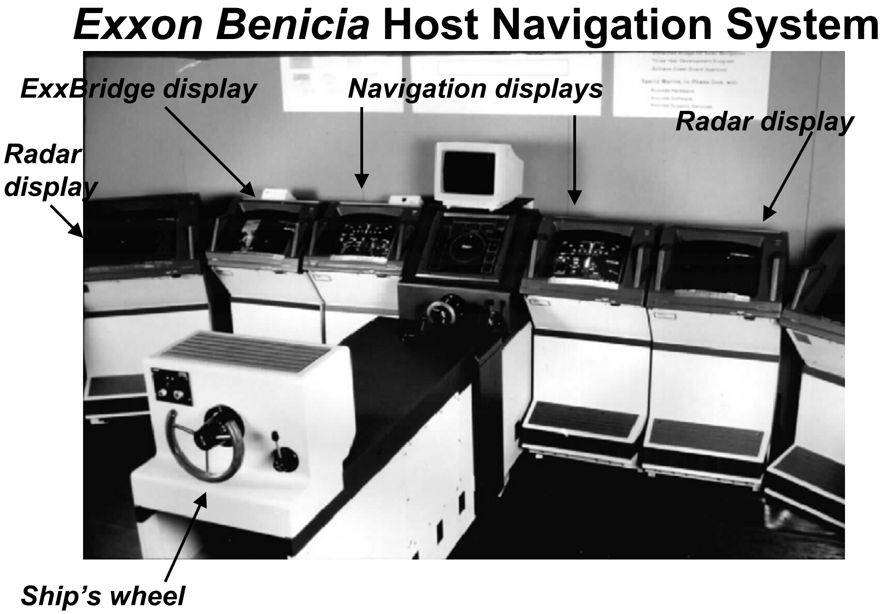


Fig. 3.

Table 1
Hypotheses

1	Operators using embedded intelligent technology will show better performance than operators not using the technology.
2	Operators using embedded intelligent technology will show less variance in performance than operators not using the technology.
3	Operators using embedded intelligent technology will consider more alternatives than operators not using the technology.
4	Operators using embedded intelligent technology will report lower workload than operators not using the technology.
5	Operators using embedded intelligent technology will report greater confidence and satisfaction than operators not using the technology.
6	Operators using embedded intelligent technology will show increased system usage, positively correlated to stress levels, compared to operators not using the technology.

ship's radars. All operators—masters, mates and pilots—had access to the ExxBridge display, as well as to print output and auditory signal messages that accompanied alerts and alarms.

4.1. Hypotheses and measures

The hypotheses considered are listed in Table 1, the operationalizations and measures used for each hypothesis are listed in Table 2. A fundamental question in

Table 2
Operationalizations and measurement of dependent variables by hypothesis

Hypothesis	Dependent variable	Variable operationalization	Data collection method
Hypothesis 1	Decision performance		
H1a	Trackkeeping	Mean cross-track error "Better" performance = smaller cross-track errors	Cross-track errors reported by automated data collection system on ship local area network (LAN)
H1b	Threat avoidance	Mean closest point of approach (CPA) (mean and minimum) "Better" performance = smaller CPA's	Mean CPA of radar targets designated and tracked by operators; collected by automatic data collection system on ship LAN
H1c	Threat maneuvering	Number of engine order commands "Better" performance = fewer commands	Audio recordings/coding of operator commands
H1d	Situation monitoring	Number of external (to vessel) communications "Better" performance = fewer communications	Audio recordings/coding of external communications
Hypothesis 2	Decision performance variance		
H2a	Trackkeeping variability	Variability of vessel and team performance measures described in previous section	
		Variability of Trackkeeping data above	Variance of cross-track errors
H2b	Threat avoidance variability	Variability of CPA data above	Variance of CPAs
Hypothesis 3	Number of alternatives considered		
		Number of maneuvering alternatives considered when a collision or grounding threat has been identified	Operator responses to maneuvering action formulation questions, reported in post-transit questionnaire
Hypothesis 4	Navigational workload		

Table 2 (*continued*)

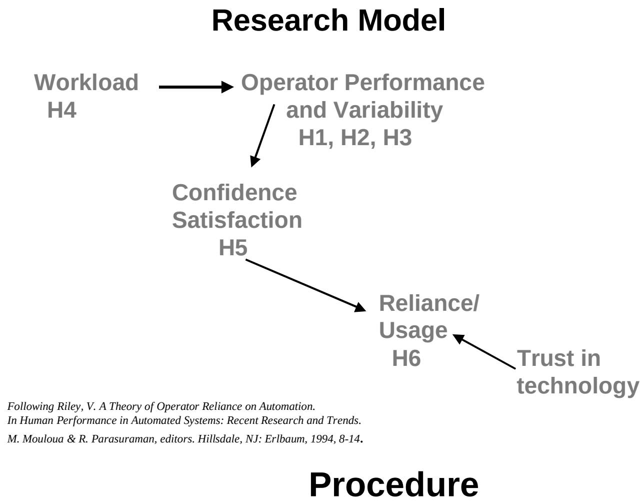
Hypothesis	Dependent variable	Variable operationalization	Data collection method
H4a	Navigational workload	Operators' experienced and perceived voyage stress	Operator responses to NASA Task Load Index (TLX) assessment of subjective workload
H4b	Role-based navigational workload		
Hypothesis 5	User decision confidence, satisfaction, familiarity		
H5a	Decision confidence	Self-reported operator confidence, satisfaction and familiarity	Operator responses to Likert scale questionnaire
H5b	Decision satisfaction		
H5c	Technology familiarity		
Hypothesis 6	System usage		
		Number of touch screen "touches"	Technology and host system usage recorded via host instrumentation

determining the impact of the embedded intelligent technology was how well the system supported the operators, and whether operator performance was enhanced when they used the system. Thus, Hypotheses 1–3 in Table 1 consider input–output measures: the performance of operators, the variability in their performance, and whether operators considered more alternatives in their decision-making. Hypotheses 4–6 considered other measures of human performance: whether operators reported lower workloads, greater confidence and satisfaction or showed increased system usage.

Our research model, illustrated in Fig. 4, follows Riley's (1994) constructs. Although we were interested in the impact of operator trust on system usage, we were not able to measure trust directly. Rather, we were able to measure other constructs that influence operator usage of and reliance on technology. The hypotheses that were tested are mapped to the research model constructs in Fig. 4.

5. Procedure

Prior to the study, the embedded intelligent technology (the SPES) was developed, integrated with the *Exxon Benicia's* navigation system, and installed aboard the *Exxon Benicia*, an Exxon tank ship which operates on an approximately 14 day cycle between Valdez, Alaska and oil terminals on the West Coast of California. Over a 2-year period, 91 subjects utilizing two technology treatments—one the host system alone, and the other, the host system with the embedded intelligent



Following Riley, V. A Theory of Operator Reliance on Automation.
In Human Performance in Automated Systems: Recent Research and Trends.
M. Mouloua & R. Parasuraman, editors. Hillsdale, NJ: Erlbaum, 1994, 8-14.

Fig. 4. Research model.

technology—were observed under different voyage conditions, and the contribution of the technology to its users was assessed.

The major shipboard environmental factors that were situation- or voyage-specific independent variables were operationalized into a single measure called "voyage stress." This is not a physiological factor, but rather an aggregate measure of the environmental situation in the shipboard setting. This measure, which has a history in the sociological, psychological and maritime literature, was estimated by recording the external visibility, vessel traffic, wind and current, along with the nature and degree of the confined waters or restricted maneuvering space encountered during the harbor transits (Williams and Goldberg, 1982; Kristiansen et al., 1989).

Low stress levels were represented by conditions such as clear visibility, no traffic, working propulsion and navigation equipment, no wind and no ice in the shipping lanes. Medium stress levels were represented by single occurrences of the conditions just described. Combinations of medium level stress conditions (e.g. high levels of traffic and a propulsion equipment failure) represented high-stress levels. High-stress situations could be encountered when vessels entered or transited the Valdez Narrows, a 0.5-mile wide restricted waterway characterized by tidal currents, speed restrictions, and traffic congestion, or when bad weather, heavy traffic and ice were encountered in other transit legs.

Each voyage of the *Exxon Benicia* was comprised of two transits of Prince William Sound, Alaska: an inbound transit from the Gulf of Alaska through Prince William Sound to the oil terminal in Valdez; and an outbound transit from Valdez, back through Prince William Sound, to the Gulf of Alaska. The transits were approximately 12 h in length. The data for this experiment were gathered during 16 voyages (i.e. 32 12-h transits) by the *Exxon Benicia* over a 2-year period. Each transit was subdivided into eight transit legs, consistent with pilot and ship master decompositions of the voyage and the problem space, with US Coast Guard approaches to licensing of ship's pilots, and with ship pilots' mental models of a voyage (Grabowski and Wallace, 1993; National Research Council, 1994). However, as seen in Table 5, the total number of transit leg observations is 247, rather than the expected 256 (32 transits × 8 transit legs), because during some voyages, data were not able to be collected because of heavy weather.

5.1. Subjects

The human subjects were members of the *Exxon Benicia*'s bridge watch teams, composed of an Exxon ship's master, an Exxon mate on watch and a pilot from the Southwest Alaska Pilot's Association. Members of the watch teams changed over the 2-year evaluation period, as shipboard officers (captains and mates) rotated on different cycles aboard the same vessel, and ship's pilots were assigned to vessels on various rotations. Although the members of the bridge watch teams aboard the *Exxon Benicia* changed over time, these effects were minimized because of the homogeneity of the subject pool (Grabowski and Wallace, 1993; National Research Council, 1994).

The Exxon subjects had completed training on the host system, had used the host system aboard the *Benicia* for a period of 1-year prior to the embedded intelligent technology installation, and had served as masters and mates aboard Exxon tank vessels for an average of 12 years. The Exxon personnel had also been familiarized with the embedded intelligent technology by a demonstration, video training, and simulator usage before participating in the evaluation. The Southwest Alaska pilots were familiarized with the host system during piloting transits over the year period prior to the technology installation, and were familiarized with the embedded intelligent technology after its installation.

5.2. Experimental design

The experiment was a 2 × 3 × 3 design: two technology treatments (the host system alone, and the embedded intelligent technology) were exposed to three types of subjects (a ship's master, mate on watch, and pilot), with three different levels of voyage stress encountered. Host system and technology transits were varied randomly throughout the 2-year assessment period, once subjects had been trained on both systems. Subjects were also exposed to different levels of voyage stress reflective of the environmental variables actually encountered during harbor transits (Tables 3 and 4).

Replications in the design cells depicted in Tables 3–5 were dependent on the voyage conditions encountered by operators over the 2-year evaluation period, which caused some variability in subject types across stress levels and technology treatments. For instance, very few subjects—pilots, masters or mates—experienced medium or high-voyage stress transits, as can be seen in Table 3. Of the 247 transit legs, only 39 were high stress, primarily as a result of high winds, restricted visibility

Table 3
Subjects exposed to voyage stress levels voyage stress level

	Low	Medium	High	Total
Pilots	12	2	0	14
Masters	25	3	1	29
Mates	38	8	2	48
Total	75	13	3	91

Table 4
Subjects exposed to technology treatments

	Host system	EIT technology	Total
Pilots	5	9	14
Masters	13	16	29
Mates	16	32	48
Total	34	57	91

EIT = embedded intelligent technology.

Table 5
Replication of transit leg experimental design cells voyage stress level

	Low	Medium	High	Total
Host	48	45	18	111
EIT Technology	62	53	21	136
Total	110	98	39	247

or traffic congestion (Table 5). Although regrettable from an experimental design standpoint, the low incidence of high-stress transits (e.g. high levels of traffic, high winds, ice in the channel, ship steering or propulsion failures, etc.) was not surprising, since after the *Exxon Valdez* oil spill, operators of Exxon vessels were encouraged to avoid medium and high-stress transits as those were the conditions that contributed to the *Valdez* grounding. In addition, it is also worth noting that not all operators were required to be on the bridge at the same time: ship's pilots are only aboard vessels for a portion of the inbound and outbound transits to Valdez, and masters may leave the bridge during the 12 h transit, which sometimes leaves the mate on watch as the sole operator on the bridge during portions of the transit. These differences account for some of the variability in subject types across technology treatments and stress conditions in Tables 3–5, and are challenges inherent in field evaluations of operational systems.

At the beginning of the voyage, operators were informed that the voyage was either an intelligent technology or a host system voyage. The ship's captain, who was informed by the researchers whether or not the intelligent technology was enabled for the voyage, informed the crew. The SPES could only be enabled for intelligent technology voyages and was disabled by the researchers for host system voyages, in keeping with the established research protocol.

5.3. Method

The hypotheses consider operator performance with embedded intelligent technology that was utilized for safe vessel navigation: the performance of the operators, the variability in their performance, and whether operators considered more alternatives, reported lower workloads, reported greater confidence and satisfaction, and showed increased system usage, compared to their experience with the host system. Hypotheses 1 and 2, which focused on operator performance and performance variability, analysed automatically collected data (within the host navigation system) about cross-track errors and closest points of approach (CPAs). In addition, Hirokawa and Pace's (1983) coding taxonomy was used to analyse audio recordings of operators' internal commands and external communications. The coding taxonomy used is provided in the appendix.

Hypothesis 3, which focused on the number of alternatives considered by the operators with and without the technology, utilized post-transit questionnaires about maneuvering alternatives formulated and considered during the transit. Hypothesis 4 focused on operator workload associated with the technology, using the NASA Task Load Index. Hypothesis 5 focused on operator familiarity with the technology, as well as operators' decision confidence and satisfaction when using the technology or its host, using Likert scale post-transit questionnaires. Hypothesis 6 assessed technology and host system usage as recorded by the host system instrumentation.

6. Analysis

Data for each hypothesis was collected for each voyage transit leg, for each treatment condition (host system, EIT technology) and voyage stress level encountered. Where appropriate, analysis of data by subject type was conducted. For Hypotheses 1a–c, 3, 5a and 5b, a multivariate test using Hotelling's T^2 test was used, as well as univariate tests of each individual measure. Where sample sizes between treatment conditions were not equal (N was not equal to M), the general linear model of the T^2 test was used, rather than the ANOVA test. The data for each of these hypotheses were found to be multivariate normal (α=0.001) with few outliers in terms of the treatment conditions.

The covariance matrices of the treatment vectors were compared to ensure that they were equal. Where they were not appropriate, measures were taken via the analysis tool. Multiple comparisons of the T^2 tests were made to compare with the univariate results in determining the tests of the null hypothesis. In the univariate tests, the variances of each treatment group were compared; where they were not equal, appropriate measures were taken via the analysis tool.

Analysis for Hypotheses 1d, 4 and 6 followed a similar pattern, although univariate *t*-tests were used to compare the variances of the treatment conditions and the analysis tools were used to compensate for unequal sample sizes between the treatment conditions. Analysis for Hypothesis 2 considered the variance of the trackkeeping and radar target CPA data of Hypothesis 1, and the analysis of variance and covariance was used to compare the host system and EIT technology treatment conditions.

7. Results

The hypotheses utilized in this study investigated operator performance with and reactions to embedded intelligent technology introduced into a safety-critical system. Operators reported using the embedded intelligent technology for navigational information, maneuvering and collision avoidance, and to obtain recommendations for the transit. Operators also reported that they relied on the system for advice, guidance, information, expertise and explanations. Operator technology use varied by voyage stress levels: operators utilized the technology significantly more than the host system alone in medium and high-stress situations (Table 11, Hypothesis 6, $p =$ **0.02**, **0.05**), a finding that indicates that operators used the technology's information and advice, even in high-stress situations.

Operator performance when using the EIT, however, was mixed. Operators using the EIT in low- and medium-stress conditions demonstrated significantly improved threat avoidance performance, compared to operators using the host system alone (Table 6, Hypothesis 1b, Mean Target CPA, $p =$ **0.03**, **0.06)**. Overall, operators demonstrated improved performance with the EIT primarily in low-stress conditions, demonstrating smaller cross-track errors, smaller CPAs, fewer engine commands and fewer external communications than operators using the host technology alone. This trend was also visible in operator variability measurements. Operator performance variability when using the EIT in low-stress conditions was significantly reduced: operators demonstrated significantly reduced trackkeeping variability (Table 7, Hypothesis 2a, $p =$ **0.09)** and significantly reduced threat avoidance variability (Table 7, Hypothesis 2b, $p =$ **0.00)**.

Operator performance in high-stress conditions, however, was not improved when using the EIT. Operator performance in medium- and high-stress conditions was improved when using the host system (Table 6, Hypothesis 1a, $p =$ **0.09**, **0.05)**, and during periods of high-stress, operators' trackkeeping variability (Table 7, Hypothesis 2a, $p =$ **0.00)** and threat avoidance variability (Table 7, Hypothesis 2b, $p =$ **0.00**) were significantly decreased with the host system alone. This may suggest that although operators used the technology, it did not provide a benefit to them. These findings are important, as it is during periods of high and medium stress that improved operator performance effects would be desirable, although those effects were not observed in this study. This result also provides important input to

Table 6
Operator performance results H1: decision performance

Measure/test/data source	Results				Finding	Accepted?
	Stress level	T^2	p	Relative magnitude		
H1a trackkeeping performance						
Mean cross track error (XTE)[a]	Low	0.74	0.82	H > EIT	EIT better in low stress condition	Rejected
Multivariate Hotelling's T^2 test	Medium	0.39	**0.09**	H < EIT	Host technology significantly better in medium and high stress conditions	Rejected
Ship's navigation data logs	High	0.30	**0.05**	H < EIT		Rejected
H1b threat avoidance performance						
Mean target CPA[b]	Low	0.03	**0.03**	H > EIT	EIT significantly better in low and medium stress	Partially supported
Multivariate T^2 test	Medium	0.02	**0.06**	H > EIT	Host better in high stress	Accepted
Ship's navigation data logs	High	0.23	0.14	H < EIT		Accepted
						Rejected
H1c threat maneuvering performance						
Maneuvering orders	Low	0.29	0.12	H > EIT	Reduced number of engine orders with EIT in low stress	Rejected
Multivariate T^2 test	Medium	0.74	0.70	H < EIT	Reduced number of engine orders with host in medium and high stress	Rejected

Audio and video recording/coding of engine order commands	High	0.35	0.81	H < EIT		Rejected
H1d situation monitoring performance						
External communications	Low		0.67	H > EIT	Reduced number of external communications with EIT in low stress	Rejected
Univariate T tests	Medium		**0.09**	H < EIT	Significantly reduced number of external communications with host technology in medium stress	Accepted
Audio recordings/coding of external communications	High		0.21	H < EIT	Reduced number of external communications with host technology in high stress	Rejected

---

[a] XTE = cross-track error, the deviation between the vessel's actual position and the intended track of the vessel, measured in meters or feet.

[b] **CPA** = closest point of approach, the closest distance another target will pass to own ship; measured in meters or feet.

Table 7
Operator performance results H2: variance in decision performance

Measure/test/data source	Results			Finding	Accepted?
	Stress level	Mean XTE variability	Relative magnitude		
H2a trackkeeping variability					
Mean cross track error variability	Low	**0.09**	H > EIT	Operator trackkeeping variability significantly reduced with EIT in low stress	Accepted
Mean cross track error variability	Medium	0.99		Operator trackkeeping performance not impacted in medium stress	Rejected
Covariance analysis	High	**0.00**	H < EIT	Operator trackkeeping performance variability significantly better with host alone in high stress	Rejected
Ship navigation data logs					
H2b threat avoidance variability					
Mean target CPA variability	Low	**0.00**	H > EIT	Operator threat avoidance variability significantly reduced with EIT in low stress	Accepted
Mean CPA variability	Medium	0.48	H > EIT	Operator threat avoidance variability significantly reduced with host in high stress	Rejected
Covariance analysis	High	**0.00**	H < EIT		Rejected
Ship navigation data logs					

technology designers and managers of safety-critical systems, who may introduce technology and perceive value in that technology, despite the fact that operator performance may not be enhanced (Table 8).

Significant role-related results with respect to operator workload were observed. Mates' workload was greatly impacted by the EIT: mates reported higher workload when using the EIT (Table 9, Hypothesis 4a), as well as higher workload than masters (Table 9, Hypothesis 4b, $p = \mathbf{0.09}$), significantly higher temporal workload than masters (Table 9, Hypothesis 4b, $p = \mathbf{0.00}$) and they required significantly more effort to use the EIT than did masters (Table 9, Hypothesis 4b, $p = \mathbf{0.01}$). Pilots, as well, reported significantly greater physical and temporal workloads than masters when using the technology (Table 9, Hypothesis 4b, $p = \mathbf{0.04}$, $\mathbf{0.05}$), and required significantly greater effort to use the EIT than did masters (Table 9, Hypothesis 4b, $p = \mathbf{0.01}$). These findings are consistent with the masters' greater familiarity with the technology, with masters' and pilots' cognitive demands, and with pilots' complaints about having to climb to the top of the bridge and down, hang onto a rope ladder in order to board and debark vessels, and then pilot the ship.

Role-related differences in operator decision confidence, satisfaction and familiarity were also observed. Pilots were more confident in and satisfied with information available from the host technology, regardless of whether the task at hand was voyage planning, situation assessment or assessment of maneuvering alternatives. Similarly, with the exception of confidence in voyage planning, mates also expressed greater confidence in and satisfaction with the host technology, rather than the EIT. These trends may reflect operator attitudes toward new technology. It would be interesting to have studied whether operator decision confidence and satisfaction changed over time, with greater technology use, but that was not measured in this study. Operator decision confidence and satisfaction appeared not to be related to technology familiarity: ship's masters were significantly more familiar with the technology than were pilots (Table 10, Hypothesis 5c, $p = \mathbf{0.02}$), but still expressed greater confidence and satisfaction with the host system, although the results are not significant. These role-related differences were important since different operators had different uses, and different expressions of utility, for the technology.

Operator use of the technology in medium- and high-stress conditions was significant (Table 11, Hypothesis 6, $p = \mathbf{0.02}$, $\mathbf{0.05}$), even though operator performance was not enhanced (Table 6, Hypothesis 1b, $p = \mathbf{0.03}$, $\mathbf{0.06}$). One master noted, for instance, that the embedded intelligent technology was helpful "when he had time to absorb all of its information," but less helpful when he needed to absorb critical information quickly. It is interesting that operators significantly utilized the technology in medium- and high-stress conditions, although they were not particularly satisfied with or confident in the information being provided by the EIT (Table 10, Hypotheses 5a and b). Moreover, operators persisted in using the technology in high-stress conditions when it no longer provided a benefit.

These results are interesting in the context of our research model, Fig. 5. Mates and pilots reported greater workload than masters when using the EIT, which may have contributed to mates' and pilots' greater satisfaction with and confidence in the

Table 8
Operator performance results, continued H3: consideration of maneuvering alternatives

Measure/test/data source	Results				Finding	Accepted?
	Operator type	T^2	# of alternatives	Relative magnitude		
H3 consideration of maneuvering alternatives						
Number of maneuvering alternatives considered	Masters	0.55	0.30	H < EIT	No significant results	Rejected
Multivariate Hotelling's T^2 test	Mates	0.61	0.35	H > EIT	Masters reported that they considered more alternatives with EIT	Rejected
Operator responses to post-transit questionnaires	Pilots	0.69	0.58	H > EIT	Mates and pilots reported that they considered more alternatives with host technology	Rejected

Table 9
Operator performance results, continued H4: navigational workload

Measure/test/data source	Results			Finding	Accepted?
	Stress level	*T* test of H_o	Relative magnitude		
H4a navigational workload					
Navigational workload	Low	0.79	$H < EIT$	No significant results	Rejected
Univariate *t* tests	Medium	0.99	$H = EIT$	Masters and pilots reported lower workload with EIT	Rejected
Operators' response to NASA Task Load Index (TLX) questionnaire	High	0.38	$H < EIT$	Mates reported higher workload with EIT	Rejected
	Operator type	*T* test of H_o	Relative magnitude		
	Masters	0.56	$H > EIT$		Rejected
	Mates	0.37	$H < EIT$		Rejected
	Pilots	0.55	$H = EIT$		Rejected
H4b role-based workload by workload type	**Masters vs. mates workload**				
	TLX categories	*T* test of H_o	Relative magnitude		
Navigational workload by role	Overall	**0.09**	Masters < Mates	Mates reported significantly higher workload overall, as well as significantly higher temporal workload, than masters	
Univariate *t* tests	Mental	0.73	Masters < Mates	Mates required significantly higher effort to utilize EIT than masters	
Operators' responses to NASA TLX	Physical	0.26	Masters < Mates		
	Temporal	**0.00**	Masters < Mates		
	Performance	0.49	Masters > Mates		
	Effort	**0.01**	Masters < Mates		
	Frustration	0.56	Masters > Mates		

Table 9 (*continued*)

Measure/test/data source	Results			Finding	Accepted?
	Stress level	T test of H_o	Relative magnitude		
	Masters vs. pilots workload				
	TLX categories	T test of H_o	Relative magnitude	Pilots reported significantly greater physical and temporal workload than masters	
	Overall	0.13	Pilots > Masters	Pilots required significantly more effort to utilize EIT than masters	
	Mental	0.69	Pilots > Masters		
	Physical	**0.04**	Pilots > Masters		
	Temporal	**0.05**	Pilots > Masters		
	Performance	0.41	Pilots < Masters		
	Effort	**0.01**	Pilots > Masters		
	Frustration	0.53	Pilots > Masters		

Table 10
Operator performance results, continued H5: decision-making confidence, satisfaction and familiarity

Measure/test/data source	Results				Finding	Accepted?
H5a decision confidence	Operator type	T^2	Voyage planning	Relative magnitude		
Decision confidence	Masters	0.99	0.32	H > EIT	No significant results	Rejected
Multivariate Hotelling's T^2 test	Mates	0.61	0.41	H > EIT	Pilots more confident with host, regardless of task	Rejected
Operators' responses to post-transit questionnaires	Pilots	0.83	0.69	H > EIT	With exception of situation assessment, mates also expressed more confidence with host rather than EIT	Rejected
	Operator type	T^2	Situation assessment	Relative magnitude		
	Masters	0.99	0.62	H > EIT		Rejected
	Mates	0.61	0.26	H < EIT		Rejected
	Pilots	0.83	1.00	H = EIT		Rejected
	Operator type	T^2	Maneuvering alternatives	Relative magnitude		
	Masters	0.99	0.58	H > EIT		Rejected
	Mates	0.61	0.30	H < EIT		Rejected
	Pilots	0.83	0.35	H > EIT		Rejected
H5b Decision satisfaction	Operator type	T^2	Voyage planning	Relative magnitude		
Decision satisfaction	Masters	0.87	0.78	H > EIT	No significant results	Rejected
Multivariate Hotelling's T^2 test	Mates	0.64	0.37	H < EIT	With the exception of mates for voyage planning and assessment of maneuvering alternatives, operators expressed greater satisfaction with host technology, rather than EIT	Rejected
Operators' responses to post-transit questionnaires	Pilots	0.89	0.61	H > EIT		Rejected

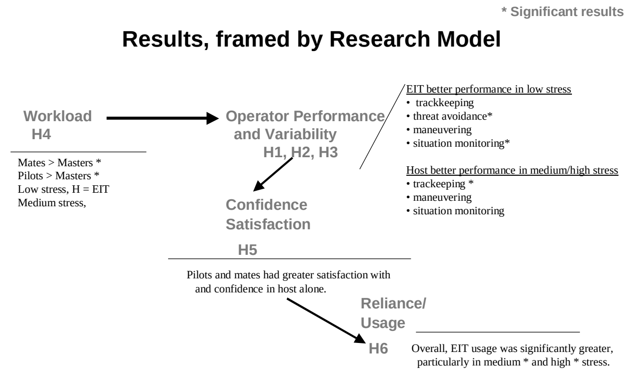
Table 10 (*continued*)

Measure/test/data source	Results				Finding	Accepted?
	Operator type	T^2	Situation assessment	Relative magnitude		
	Masters	0.87	0.60	H > EIT		Rejected
	Mates	0.64	0.54	H > EIT		Rejected
	Pilots	0.89	0.29	H > EIT		Rejected
	Operator type	T^2	Maneuvering alternatives	Relative magnitude		
	Masters	0.87	0.28	H > EIT		Rejected
	Mates	0.64	0.56	H < EIT		Rejected
	Pilots	0.89	0.29	H > EIT		Rejected
H5c Familiarity	Subjects	*T* test of H_o		Relative magnitude		
Familiarity with technology	Masters vs. Mates	0.15		Masters > Mates	Masters significantly more familiar with EIT than pilots.	Rejected
Univariate *T*-test	Pilots vs. Masters	0.02		Pilots < Masters		Accepted
Operators' responses to post-transit questionnaires	Pilots vs. Mates	0.30		Pilots < Mates		Rejected

Table 11
Operator performance results, continued H6: system usage

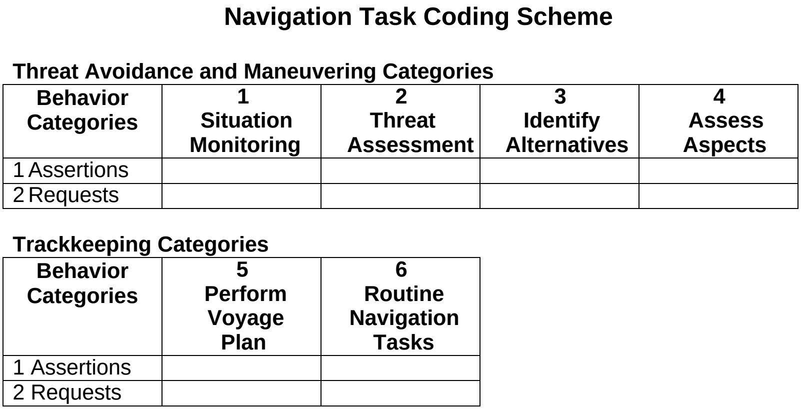
Measure/test/data source	Results			Finding	Accepted?
	Stress level	T test of H_o	Relative magnitude		
H6 system usage	Overall	**0.05**	H < EIT	Overall, EIT significantly greater usage	Accepted
“Touches” to touch screen	Low	0.63	H > EIT	EIT usage significantly greater usage in medium and high stress	Rejected
Univariate *t* tests	Medium	**0.02**	H < EIT		Accepted
Technology usage captured by host instrumentation	High	**0.05**	H < EIT		Accepted

Significant results are in bold.



Following Riley, V. A Theory of Operator Reliance on Automation.
In Human Performance in Automated Systems: Recent Research and Trends.
M. Mouloua & R. Parasuraman, editors. Hillsdale, NJ: Erlbaum, 1994, 8-14.

Fig. 5. Results framed by research model.



Navigation Task Coding Scheme

Threat Avoidance and Maneuvering Categories

Behavior Categories	1 Situation Monitoring	2 Threat Assessment	3 Identify Alternatives	4 Assess Aspects
1 Assertions				
2 Requests				

Trackkeeping Categories

Behavior Categories	5 Perform Voyage Plan	6 Routine Navigation Tasks
1 Assertions		
2 Requests		

Fig. 6. Navigation task coding scheme. Threat avoidance and maneuvering categories. Trackkeeping categories.

host technology, rather than the EIT. Similarly, operator workload increased when using the EIT in high-stress conditions, which may have contributed to operators' significantly improved performance in high-stress conditions when using the host system (Table 6, Hypothesis 1a, $p = \mathbf{0.05}$; Table 7, Hypothesis 2a, $p = \mathbf{0.00}$; Table 7, Hypothesis 2b, $p = \mathbf{0.00}$), but data were not collected to verify this supposition. Thus, in conditions when operator performance benefit would have been most desirable, those benefits were not observed.

Technology use was high, however, even though performance benefits were not observed in all conditions. Operator workload was generally higher with the technology, and despite greater operator confidence, satisfaction and familiarity with the host system. These are significant findings for operators, managers and regulators in safety-critical systems deploying technology to improve performance. The findings highlight the importance of empirical evaluations of technology before deployment, so as to manage technology introduction, use and training appropriately.

Since the nature of EIT use was not measured, it is difficult to say whether increased technology usage contributed to increased operator workload (e.g. operators engaged in more and frustrating searches for information, thus increasing workload) or whether increased operator workload contributed to usage (e.g. overloaded operators over utilized or inefficiently utilized the technology many times). It is also not clear whether the high technology use was associated with operator trust of the technology or not, as no direct trust measurements were made in this study. These are all interesting research questions that bear further investigation.

8. Methodology and limitations

Our methodology and results suffered from several limitations. For instance, we were not able to gather data on all transit legs because of heavy weather. This reduced the number of usable transit legs with data from an expected 256 to 247. We were also unable to gather much data about technology use and operator performance under high-stress conditions, primarily because of the limited number of high-stress transits undertaken by Exxon vessels following the grounding of the *Exxon Valdez*. Available data were collected, but statistically supportable conclusions about high-stress conditions cannot be made since the number of observations was small. In addition, the experiments conducted were not random. On a single voyage of two transits and eighy transit legs, the subjects were often the same, which creates dependencies in the observations. Additional work to analyse these dependencies and/or to conduct more random experiments would have enhanced the validity and generalizability of the study, but were not performed in this work.

It would also have been interesting to gather longitudinal data on individual subjects over the course of the study period, to see how operator opinions changed with technology utilization over time, but this data were not gathered.

An important input to technology impact studies is the boundary condition(s) within which the technology is useful. This includes, but is not limited to, determining such things as the environmental conditions under which the system is valid and the qualifications of operator required to appropriately utilize the system. This study took these things into account in several ways, including differentiating technology performance under conditions of low, medium and high stress. However, evaluation studies such as this one should explicitly point out the need to determine the constraints or boundary conditions of the technology, a point this study did not address.

9. Conclusions

With the proliferation of embedded intelligent technology in complex, safety-critical systems, the need for improved understandings of the impact of such technology on operator and system performance is clear. In this study, operators used the embedded intelligent technology in all stress conditions, even though improved operator performance was noted only in low-stress conditions. Operator use of the technology was significantly higher during periods of medium and high stress, as might be expected of technology being employed in safety-critical systems by operators who depend on the technology. However, although the operators relied on the technology, it did not enhance their performance in high-stress conditions. These findings are important for operators, whose performance enhancement is often the motivation for technology introduction and for technology designers and managers of safety-critical systems, who may be responsible for introducing technology in safety-critical systems.

Role-related differences between the operators in terms of their familiarity, confidence, satisfaction, and workload were noted, which underscore Riley's (1994) notion that the impacts of technology on operators are related to how familiar, confident and satisfied operators are with the technology, and with operator workload. Additional impacts observed in this study include operator performance and variability, which were favorably impacted by the EIT only in low-stress conditions.

Human operators of systems tend to be conservative in their work habits. New technology, when first introduced, tends to be looked at suspiciously and perhaps mistrusted. As experience is gained with new technology, however, and given that it works reliably and accurately, most operators will tend to like and come to trust the new device. In this study, the embedded intelligent technology was perceived to be both reliable and accurate, and technology usage, particularly during periods of medium and high stress, was high. This finding is important in light of the fact that operator performance did not improve significantly with use of the technology. In fact, operator performance when using the technology was only significantly enhanced during periods of low stress. This suggests that operator reliance and benefit were not calibrated, an important finding for operators and managers in safety-critical systems.

Human operators may also express leeriness of technology that they do not understand well. For instance, airline pilots have been found to have incomplete knowledge of the various modes and behaviors of their Flight Management Systems (Sarter and Woods, 1994). In this study, ship's pilots and mates were found to have less confidence and satisfaction in using the intelligent technology, compared to the ship's masters, who used the system most frequently. Time criticality and stress levels may also be related to an operator's tendency to use or trust technology. In this study, operators had a tendency to use the technology as a threat monitoring aid under conditions of low stress, a supposition supported by masters' reports of increased confidence and satisfaction with the intelligent technology in low-stress conditions.

Technology use in safety-critical systems, clearly, has important impacts on human and system performance. In safety-critical systems, both humans and technology are jointly responsible for executing tasks, monitoring safety and improving system performance. Understanding how and why operators utilize and rely on technology in safety-critical systems is thus an important step in enhancing safety and performance in safety-critical systems. The results in this study highlight the importance of such studies, and the need for empirical research investigating the impacts of technology on humans and systems in safety-critical settings.

Acknowledgements

The authors thank the US Department of Transportation, Maritime Administration (MARAD); the US Coast Guard; the National Oceanographic and Atmospheric Administration; Exxon Shipping Company; Sperry Marine, Inc.; and the Southwest Alaska Pilots Association, all of whom supported this research. The support and thoughtful guidance of the following individuals and organizations is gratefully acknowledged: Messrs. John Dumbleton and Paul Mentz of MARAD; Admiral Jim Card and Dr. Lee Alexander of the US Coast Guard; Messrs. Gus Elmer, Art Stevens and Bob Freeman from Exxon Shipping Company; Mr. Frank Iarossi from the American Bureau of Shipping; Messrs. Scott Bartlett and Tom King of Sperry Marine, Inc.; the Southwest Alaska Pilots Association, particularly Captains Jim Wright, Tony Joslyn and Richard Cochinos; and the masters and crew of the *Exxon Benicia* and other trans Alaskan Pipeline System (TAPS) trade tankers who participated in this research. The authors would also like to thank Professor Susan Wiedenbeck, the special issue editor; and three anonymous reviewers, who contributed greatly to this manuscript.

Appendix A. Navigation task coding scheme

Hirokawa's Function-Oriented Interaction Coding Scheme (FOICS) operationalizes communicative behaviors occurring in groups or teams that need to occur in order for the group to arrive as a satisfactory decision or solution (Hirokawa, 1982). The coding scheme specifically addresses interaction processes between members of problem-solving groups.

Hirokawa's coding approach is based on the recognition that while various possible communicative acts may be important to group performance; they may or may not be relevant to effective decision-making or problem solving since the group interaction process has both task and social dimensions. Hirokawa (1982) suggest that a two-level task-relevant coding scheme be utilized, incorporating *functional* categories (i.e. task achievement functions) and *behavior* categories (i.e. behaviors which mediate performance of the task function). The functional and behavior categories must be carefully selected for relevance to the group task being studied and to ensure that the resulting coding scheme is workable. The coding approach adopted for use in this experiment, adapted from Hirokawa's FOICS, is shown in Fig. 6.

In this research, the navigation task performed by the bridge watch team during transit of confined waters was the primary task. The essential task of the bridge watch team is to safely maneuver the vessel during the harbor transit while accomplishing their intended voyage plan. During a transit, the bridge watch team is responsible for attending to vessel navigation, maneuvering, threat avoidance and bridge maintenance activities as required by the transit and the practice of good seamanship (Cooper et al., 1981; Grabowski and Wallace, 1993; Schuffel et al., 1989; National Research Council, 1994).

A set of functional coding categories was formed based on the four task functions that vigilance action theory (Gouran and Hirokawa, 1983; Hirokawa and Scheerhorn, 1986; Janis, 1989) suggests must be performed if high-quality critical thinking is to be realized. These are problem analysis, identification of alternatives and thorough analysis of the positive and negative aspects of each alternative.

The equivalent task functions in maritime navigation are situation monitoring, threat assessment and avoidance action formulation (Cooper et al., 1981; Schuffel et al., 1989; National Research Council, 1994). Action formulation is comprised of identification of alternative avoidance maneuvers, and analysis of positive and negative aspects of each alternative. These functional categories effectively cover the need to make threat avoidance and maneuvering decisions during the vessel's transit. However, they do not address the concomitant execution of routine navigation tasks to carry out the vessel's intended voyage plan or performance of other necessary bridge and deck activities during the harbor transit.

Hirokawa (1982) utilized behavior categories that included *assertions* and *requests.* Assertions include introduction of new information: restatement, development, substantiation or modification of earlier information; agreement or disagreement with previous statements; or summarization or synthesis of previous information. Request behaviors include asks for ideas; asks for approval of statements; asks for clarification of previous statements; and asks for summary or synthesis of earlier portions of the discussion. Coury and Terranova (1991) and Terranova et al. (1991) utilized similar behavior categories in their studies of team coordination communications, although they did not consider the decision-making functional categories used by Hirokawa.

Application of the navigation task coding scheme depicted in Fig. 6 first required that the audio recordings collected aboard ship be transcribed and organized into a sequence of distinct statements by the bridge team members. Communication interaction analysis utilizes as a basis unit of analysis the "functional utterance"—that is "an uninterrupted utterance of a single group member which appears to perform a specific function within the group interaction process" (Fisher, 1970). Based on this unit of analysis, the distinction between two units of interaction is viewed as either (1) the interruption of one individual's contribution by the functional utterance of another group member or (2) the crossing of functions within the single contribution of a group member (Hirokawa, 1982).

The coding scheme depicted in Fig. 6 was then applied to analyse the group discussion data. Three coders were required to make two independent decisions: first, they had to identify the *function* that the utterance performs in the context of the particular watch team discussion. Second, the coders had to identify which utterance behavior was associated with the selected functional coding. Trained coders coded the group discussion using two single digit codes. The first digit indicated which of the threat avoidance and maneuvering (1–4) or trackkeeping (5 and 6) functional categories applied to the utterance. The second digit (1–2) was used to indicate the utterance's behavior type. The coding scheme was extended to indicate which watch team member (i.e. pilot, master or mate) was the source of the utterance. This enabled the communication patterns among the watch team members to be studied.

References

Adelman, L., Cohen, M.S., Bresnick, T.A., Chinnis, J.O., Laskey, K.B., 1993. Real time expert system interfaces, cognitive processes, and task performance: an empirical assessment. Human Factors 35 (2), 243–261.

Bainbridge, L., 1983. Ironies of automation. Automatica 19, 775–779.

Ben-Arieh, D.H., Moodie, D.L., Chu, C.D., 1988. Control methodology for FMS. IEEE Journal of Robotics and Automation 4 (1), 53–59.

Bogner, M.S. (Ed.), 1994. Human Error in Medicine. Lawrence Erlbaum Associates, Hillsdale, New Jersey.

Chandrasekaran, B., Bhatnagar, R., Sharma, D.D., 1991. Real-time disturbance control. Communications of the ACM 34 (8), 32–47.

Chen, I.-R., Bastani, F.B., Tsao, T.-W., 1995. On the reliability of AI planning software in real time applications. IEEE Transactions on Knowledge and Data Engineering 7 (1), 4–13.

Coenen, F.P., Smeaton, G.P., Bole, A.G., 1989. Knowledge-based collision avoidance. Journal of Navigation (UK) 41 (1), 107–116.

Cohen, P., 1995. Empirical Methods for Artificial Intelligence. MIT Press, Cambridge, MA.

Cook, R.C., Marino, K.L., Cooper, R.B., 1981. A Simulator Study of Deepwater Port Shiphandling and Navigation Problems in Poor Visibility (Report No. EA-80-U-099). Eclectech Associates Inc., North Stonington, CT.

Cook, R.I., Woods, D.D., Howie, M.B., 1990. The natural history of introducing new information technology into a high-risk environment. Proceedings of the 34th Annual Meeting of the Human Factors Society. Human Factors and Ergonomic Society, Orlando, FL: Washington, DC.

Cooper, R.B., Bertsche, W.R., McCue, G.J., 1981. Simulator Evaluation of Predictor Steering, Short Range Collision Avoidance and Navigation Displays (CAORF Report No. 30-7913-01). Kings Point, New York, National Maritime Research Center, Computer Aided Operations Research Facility (CAORF).

Coury, B.G., Terranova, M., 1991. Collaborative decision making in dynamic systems. Proceedings of the 35th Annual Meeting of the Human Factors Society, Baltimore, MD, pp. 944–948.

Cullyer, J., 1991. Safety-critical control systems. Computing and Control Engineering Journal 2 (5), 202–210.

Dailey, D.J., Haselkorn, M.P., Lin, P., 1993. Traffic information and management in a geographically distributed computing environment. Pacific Rim 3rd International Conference on Applications of Advanced Technologies in Transportation Engineering, Tokyo, Japan, July, pp. 159–165.

Entin, E.E., Serfaty, D., 1997. Sequential revision of belief: an application to complex decision-making situations. IEEE Transactions on Systems, Man and Cybernetics Part A: Systems and Humans 27 (3), 289–301.

Fee, J.J., Lascelle, W., Nieri, D., 1980. Concept Development and Simulation Test Plan for an Advanced Capabilities Vessel Traffic System (CAORF Report No. 43-7909-02). Kings Point, New York, National Maritime Research Center, Computer Aided Operations Research Facility (CAORF).

Fisher, B.A., 1970. Decision emergence: phases in group decision-making. Speech Monographs 37, 53–66.

Florac, W., Carleton, A., 1999. Measuring the Software Process. Addison-Wesley, New York.

Gouran, D.S., Hirokawa, R.Y., 1983. The role of communication in decision-making groups: a functional perspective. In: Maunder, M.S. (Ed.), Communications in Transition. Praeger Books, New York.

Grabowski, M.R., Sanborn, S.D., 2001. Evaluation of embedded intelligent real-time systems. Decision Sciences 32 (1), 95–123.

Grabowski, M.R., Wallace, W.A., 1993. An expert system for maritime pilots: its design and assessment using gaming. Management Science 3 (12), 1506–1520.

Hall, N.G., 1988. Diagnosing problems with the user interface for a strategic planning fuzzy DSS. IEEE Transactions on Systems, Man and Cybernetics 18 (4), 636–638.

Heudin, J.C., 1991. KOS: a knowledge-based operating system for real-time onboard applications of artificial intelligence. Proceedings of Euromicro '91. IEEE Computer Society Press, Silver Spring, MD, pp.138–144.

Highland, F., 1994. Embedded AI. IEEE Expert: Intelligent Systems and their Applications 2 (3), 18–20.

Hirokawa, R.Y., 1982. Group communication and problem-solving effectiveness 1: a critical review of inconsistent findings. Communication Quarterly 30 (2), 134–141.

Hirokawa, R.Y., Pace, R., 1983. A descriptive investigation of the possible communication-based reasons for effective and ineffective group decision-making. Communication Monographs 50, 363–380.

Hirokawa, R.Y., Scheerhorn, D.R., 1986. The role of communication in faulty group decision-making. In: Hirokawa, R.Y., Poole, M.S. (Eds.), Communication and Group Decision-Making. Sage Publishing, Beverly Hills, CA.

Hogarth, R.M., Einhorn, H.J., 1992. Order effects in belief updating: the belief-adjustment models. Cognitive Psychology 24, 1–55.

Hosmer, L.T., 1995. Trust: the connecting link between organizational theory and philosophical ethics. Academy of Management Review 20, 379–403.

Institute of Medicine, 1999. To Err is Human: Building a Safer Health System. National Academy Press, Washington, DC.

Jaffe, M.S., Leveson, N.G., Helmdahl, P.E.M., Melhart, B.E., 1991. Software requirements analysis for real time process control systems. IEEE Transactions on Software Engineering 17 (3), 241–258.

Janis, I.L., 1989. Crucial Decisions. The Free Press, New York.

Kristiansen, S., Rensvik, E., Mathiesen, L., 1989. Integrated total control of the bridge. Proceedings of the Society of Naval Architects and Marine Engineers' Annual Meeting, New York.

Lee, J.D., Moray, N., 1992. Trust, control strategies, and allocation of function in human–machine systems. Ergonomics 35, 1243–1270.

Lee, J.D., Moray, N., 1994. Trust, self confidence, and operators' adaptation to automation. International Journal of Human-Computer Studies 40, 153–184.

Leveson, N.G., Turner, C.S., 1993. An investigation of the therac-25 accidents. IEEE Computer 26 (7), 18–41.

Leveson, N.G., Cha, S.S., Shimeall, T.J., 1991. Safety verification of Ada Programs using Software Fault Trees. IEEE Software 13 (1), 48–59.

Muir, B.M., 1988. Trust between humans and machines, and the design of decision aids. In: Hollnagel, E., Mancini, G., Woods, D.D. (Eds.), Cognitive Engineering in Complex, Dynamic Worlds. Academic Press, London.

National Research Council, 1994. Minding the Helm, Marine Navigation and Piloting. National Academy Press, Washington, DC.

National Transportation Safety Board, 1973. Eastern Airlines L-1011, Miami, FL, 20 December 1972. Report No. NTSB-AAR-73-14, National Transportation Safety Board, Washington, DC.

National Transportation Safety Board, 1997. Marine Accident Report: Grounding of the Panamanian Passenger Ship Royal Majesty near Nantucket Island, MA, June 10, 1995. US Department of

Transportation, National Transportation Safety Board, Washington, DC, 12 March 1997. Also at http://www.ntsb.gov/Pressrel/031297.htm

Nieri, D., 1980. Communications Workload Investigation of Proposed Vessel Traffic Service in New York Harbor. CAORF Report No. 42-7909-01. National Maritime Research Center, Computer Aided Operations Research Facility, Kings Point, New York.

Norman, D.A., 1990. The 'problem' with automation: Inappropriate feedback and interaction, not 'over-automation'. Philosophical Transactions of the Royal Society of London B327, 585–593.

Paul, C.J., Acharya, A., Black, B., Strosnider, J.K., 1991. Reducing problem solving variance to improve predictability. Communications of the ACM 34 (3), 80–93.

Perrow, C., 1984. Normal Accidents: Living with High Risk Technologies. Basic Books, New York.

Perry, T.S., 1997. In search of the future of air traffic control. IEEE Spectrum 26 (4), 12–15.

Read, P.P., 1993. Ablaze: the Story of Chernobyl. Random House, New York.

Riley, V.A., 1994. Theory of operator reliance on automation. In: Mouloua, M., Parasuraman, R. (Eds.), Human Performance in Automated Systems: Recent Research and Trends. Erlbaum, Hillsdale, NJ.

Ring, P.S., Van de Ven, A.H., 1994. Developmental processes of cooperative interorganizational Relationships. Academy of Management Review 19, 90–118.

Rochlin, G., 1997. Trapped in the Net: The Unanticipated Consequences of Computerization. Princeton University Press, Princeton, NJ.

Rogers, W., Rogers, S., Gregston, J., 1992. Storm Center: the USS Vincennes and Iran Air Flight 655. Naval Institute Press, Annapolis, MD.

Roth, E.M., Bennett, K.B., Woods, D.D., 1987. Human interaction with an "Intelligent" machine. International Journal of Man-Machine Studies 27, 479–526.

Rouse, W.B., Geddes, N.D., Hammer, J.M., 1990. Computer-aided fighter pilots. IEEE Spectrum 27 (3), 38–41.

Sagan, S.D., 1993. The Limits of Safety: Organizations, Accidents, And Nuclear Weapons. Princeton University Press, Princeton, NJ.

Sarter, N., Woods, D.D., 1994. Pilot interaction with cockpit automation II: an experimental study of pilots' model and awareness of the flight management system. International Journal of Aviation Psychology 4, 1–28.

Schuffel, H.J., Boer, P.J., Van Breda, L., 1989. The ship's wheelhouse of the nineties: the navigation performance and mental workload of the officer of the watch. Journal of the Institute of Navigation 42 (1), 60–72.

Sheridan, T.B., 1988. Trustworthiness of command and control systems. In: Proceedings of the IFAC Conference on Man-Machine Systems, Oulu, Finland.

Shimeall, T.J., Leveson, N.G., 1991. An empirical comparison of software fault tolerance and fault elimination. IEEE Transactions on Software Engineering 17 (2), 173–182.

Terranova, M., Hartley, D., Coury, B.G., Hooper, K.N., 1991. Individual and Team Communication in a Dynamic Task. Proceedings of the 35th Annual Meeting of the Human Factors Society, pp. 954–958.

Tenner, E., 1996. Why Things Bite Back: Technology and the Revenge of Unintended Consequences. Alfred A. Knopf, New York.

Vaughan, D., 1996. The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA. University of Chicago Press, Chicago.

Wen, W., Callahan, J., Napolitano, M., 1997. Verifying stability of dynamic soft computing systems. In: Proceedings of the International Joint Conference on Artificial Intelligence Workshop on Validation, Verification, and Refinement of AI Systems, IJCAI, Nagoya, Japan, August 23–28.

Wiener, E.L., 1988. Cockpit automation. In: Wiener, E.L., Nagel, D.C. (Eds.), Human Factors in Aviation. Academic Press, San Diego.

Wiener, E.L., Curry, R.E., 1980. Flight deck automation: promises and problems. Ergonomics 23, 995–1011.

Williams, K., Goldberg, J. 1982. Standards of Watch Officer Performance: Validation of Situation Difficulty Measure Re: Collision Avoidance Problem Solving (CAORF Report No. 60-7908-01). Kings Point, National Maritime Research Center, Computer Aided Operations Research Facility (CAORF) New York.

Wong, S.K., Kalam, A., 1995. Development of a power protection system using an agent-based architecture. Proceedings of International Conference on Energy Management and Power Delivery 1, 433–438.

Woods, D.D., Cook, R.I., 1991. Nosocomial Automation: Technology-induced Complexity and Human Performance. Charlottesville, VA: In: Proceedings of the IEEE International Conference on Systems, Man & Cybernetics. IEEE Press, New York.

Woods, D.D., Potter, S.S., Johannesen, L., Holloway, M., 1991. Human Interaction with Intelligent Systems: Trends, Problems, and New Directions. Columbus, OH, The Ohio State University, CSEL Report 1991-001.