

2008

Enhancing decision making by implementing likelihood alarm technology in integrated displays

Rylan M. Clark
University of Idaho

Ernesto A. Bustamente
University of Idaho

Follow this and additional works at: <https://scholar.utc.edu/mps>



Part of the [Psychology Commons](#)

Recommended Citation

Clark, Rylan M. and Bustamente, Ernesto A. (2008) "Enhancing decision making by implementing likelihood alarm technology in integrated displays," *Modern Psychological Studies*: Vol. 14 : No. 1 , Article 5.

Available at: <https://scholar.utc.edu/mps/vol14/iss1/5>

This articles is brought to you for free and open access by the Journals, Magazines, and Newsletters at UTC Scholar. It has been accepted for inclusion in Modern Psychological Studies by an authorized editor of UTC Scholar. For more information, please contact scholar@utc.edu.

Rylan M. Clark &
Ernesto A. Bustamante

University of Idaho

Enhancing Decision Making by Implementing Likelihood Alarm Technology in Integrated Displays

Complex environments with automated systems, such as aircraft cockpits and nuclear control rooms, require critical decisions to be made about human intervention. Human monitors operating in these roles must interact with copious amounts of information. Decision support tools within integrated displays, especially alarms, aid people in monitoring these systems by capturing their attention to focus on possibly dangerous conditions. Once signaled, monitors choose whether they wish to acknowledge the alarm and search for more process status, or ignore it. This study investigates the impact of likelihood alarm technology versus traditional binary alarms on decision making accuracy and response bias in this acknowledgement phase using a two-stage Signal Detection Model. Participants performed two low-fidelity, twenty-min flight missions consisting of dual primary tasks, compensatory tracking and fuel management, and a secondary engine-monitoring task. Probability of engine malfunctions (10%, 90%) and type of alarm system (Binary vs. Likelihood) were manipulated for each participant. It was hypothesized that the probability of engine malfunctions (P), and likelihood alarm technology (LAT) would interact with decision making accuracy. Additionally, a main effect of P on decision making bias was expected. Results showed that LAT significantly increased accuracy, especially under low P conditions, but had no effect on response bias. The results of this study support prior literature's findings on the superiority of LAT over binary alarms in complex tasks characterized by high workload, translating to better monitoring performance for many practical applications.

Technology has granted the ability to allocate functions, traditionally carried out by humans, to automated systems. Nearly every facet of our infrastructure is impacted by automation, as its supplements to human performance are well established (Sheridan, 2002; Sheridan & Parasuraman, 2006). For examples of implementation, consider nuclear power, ground transportation, air traffic control, aviation, and collision-warning systems (Parasuraman, Hancock, & Olofinboba, 1997), as well as medical diagnosis (Li, Lin, & Chang, 2004), luggage screening (Drury,

Ghylin, & Holness, 2006), and cockpit display monitoring (Bailey & Scerbo, 2005).

Automation and Human Performance

Automation harnesses computational power to perform tasks otherwise apportioned to human operators, with the intent of a higher performing human-machine system. This is especially true in complex environments characterized by high levels of workload. Humans and computers have unique strengths for specific tasks that, once identified, can be properly merged into a smoothly functioning team. For instance, the employment of alarms in

complex task scenarios has augmented human performance as well as safety (Sorkin, Kantowitz, & Kantowitz, 1988). Nevertheless, as the field of human factors demonstrates, there is always room for improvement, and concerns over traditional binary alarm systems (BAS) have rallied support for new technology. Likelihood alarm technology (LAT) provides differing alarm signals based on the predicted validity of there being an actual problem. One main concern with BAS is unreliability. LAT offers superior decision support because human monitors tend to match their response rates to the indicated probability of a real problem existing. This study provides evidence for the superiority of LAT, in terms of decision making, over traditional BAS.

A definite benefit of automated systems is executing specific tasks in a more accurate and reliable manner than sole human operators. However, it is important to understand that automation must be applied properly, considering both its own and the human operator's limitations. Empirical research shows that humans tend to trust overly-reliable alarm systems too much, becoming complacent to the extent that performance degrades (Parasuraman & Riley, 1997; Wickens and Dixon, 2007). Wickens and Dixon also remind us that when diagnostic automation is too imperfect and unreliable (their analysis concludes .70), it can literally be useless, or even result in worse than baseline human performance. Judicious function allocation among humans and automation is required to profit from the merits of automated systems. Sorkin and Woods (1985) also advocate the importance of designing automated monitoring systems around total human-machine performance instead of just automated subsystem performance.

System Operators to System Monitors

Along with procuring traditional operator functions, automation has qualitatively transformed the role of people in complex tasks from operators to monitors (Parasuraman & Riley, 1997). Technological development has changed the human role to chiefly that of supervisory control (Sheridan & Hennessy, 1984; Woods, 1982), whose main function is to monitor a number of automated displays. Automated factories using extensive robotics, the flight decks of commercial aircrafts,

and power plant control rooms are examples described by Sorkin and Woods (1985) of decision-making environments in which human operators act an alerted-monitor role. Xiao and Seagull (1999) assert that monitoring is not a passive task but rather the act of maintaining continual situational awareness. In these settings, an automated monitor subsystem assists human operators by executing preprogrammed decisions about system conditions. These decisions are made based on inputs, such as incoming data and process state, which are weighed against expectations of what constitutes normal and abnormal operating conditions (Woods, 1995). This shift of humans' roles into a supervisory monitoring state necessitates dynamic fault management.

Dynamic Fault Management

The above examples of environments where humans interact with and rely on automation are permeated with what is called dynamic fault management. As the name denotes, dynamic fault management represents a situation where a complex system's conditions are being monitored to identify and address problems that arise. In complex environments, such as the control room of a nuclear power plant, there are safety-critical tasks that demand more resources than a lone human operator can provide. It is important to monitor the conditions of a nuclear reactor for more than just safety reasons, as millions of people are relying on it for power, and the monetary costs of compromising system integrity are extreme. Humans are infamous for conducting poor dynamic fault management. This occurs for a variety of reasons, many being related to the problem of sustaining vigilance while monitoring some type of display. Decision support tools (DSTs) were created to combat this inadequacy. In addition, legal concerns and technologies' rapid development compel system designers to incorporate alarm systems into dynamic fault management situations. Alarm systems are a type of DST that are discussed below.

Decision Support Tools

Previously active system users, human operators' now-common role of vigilant attention tasks in complex environments is made even more difficult by high levels of workload. Consequently, DSTs

were created to aid human monitors for these types of environments. Alarm systems are a major form of DST. According to Xiao and Seagull (1999), they are affixed to displays and devices on human-machine interfaces (out of direct consideration to human performance limitations) to report process status. Woods (1995) argues the alarm role in dynamic fault management settings as that of functioning to create an attentional capture effect on alerted human monitors. The usefulness of alarm systems lies in their plausible ability to enhance human performance in numerous complex tasks (Bustamante & Bliss, 2004; Gupta, Bisantz, & Singh, 2002; Sorkin, Kantowitz, & Kantowitz, 1988). Especially in complex systems, human operators may become overwhelmed with status information that taxes memory, attention, and decision-making capacities (Fallon, Bustamante, Ely, & Bliss, 2005). In this study, two types of DST's we compared, BAS and LAS, are discussed after the succeeding section. We used the EICAS engine monitoring alarm system, which is currently being used in commercial aviation, and manipulated it as either a binary or likelihood system with 10% or 90% probabilities of engine malfunctions.

Positive Predictive Value of Alarm Systems

Alarm signal 'reliability' is a common term for positive predictive value (PPV). PPV is defined as the conditional probability that given an alarm, a problem actually exists (Bustamante & Bliss, 2005). A phenomenon known as 'probability matching' occurs when people approximately match their response rate to a given alarms' PPV (Bliss, Gilson, & Deaton, 1995). For instance, a human monitor that knows a particular alarm system's PPV is 80% will tend to acknowledge about four out of five alarm signals. Similarly, if he or she knows an alarm system's PPV to be 20%, then roughly one in five signals will get acknowledged. Evidence for altered acknowledgement frequencies, motivated by PPV, has been documented by researchers (Bliss & Dunn, 2000; Bliss, Gilson, & Deaton, 1995; Getty, Swets, Picket, & Gonthier, (1995). PPV is dependent on three factors: accuracy, threshold, and the probability of problematic conditions. Problematic condition probabilities (or base rates) cannot be manipulated by alarm designers, but thresholds and

accuracy can. Accuracy is now a minor issue owing to sophisticated fault diagnosis algorithms and sensor technologies that are so highly developed as to operate nearly perfectly. Threshold is the predetermined limit of these sensors and algorithms that, when exceeded by system-condition cues, trigger an alarm's signal. To ensure that most dangerous system conditions are detected, alarm designers are forced to set thresholds at low levels in a so called "fail-safe approach." This means that given an actual problem, there is a very high probability that the alarm will transmit it to the human monitor. However, given an alarm, the probability of an actual problem is quite low, because a whole range of system conditions may exist falling above the threshold but not necessarily indicating a real problem. Probability of problematic conditions, PPV's third determining factor, is self explanatory. Also known as the 'base rate' of dangerous conditions, this factor represents the likelihood of dangerous conditions actually occurring. For instance, take an overly sensitive car alarm that activates nearly every day. Most of its alerts are false alarms because the probability, or base rate, of dangerous conditions may only be 5%. That is, only five percent of the time will there actually be an attempted break in—which is what the car owner purchased the alarm for. It is not difficult to identify the conundrum of setting thresholds too conservatively in systems with low base rates; false alarms (FAs) are frequent. A high number of FAs impart numerous undesirable effects on the human monitor. Stressed factory workers may easily learn to ignore various systems known to generate many FAs, figuring they are probably not indicating a true problem among more pressing issues. PPV is often low for alarm systems due to low base rates in most settings, and preprogrammed sensitive thresholds. The result of alarms' undesirable ratio of FAs to true alarms (TAs) is explained in more detail later.

Binary Alarm Technology

Sorkin & Woods (1985) parsimoniously explain BAT: automated monitors display messages to the system operator when a preset threshold value is surpassed by measured conditions. Engineers have traditionally utilized this alerting approach, which has now come to be known as the binary alarm

tradition. BAT produces only one type of alarm signal, regardless of the underlying conditions. When that preset threshold (determined by system designers) is violated, an alarm is activated. This is a problem when considering how most systems emit a high amount of false alarms. Out-of-the-loop human monitors encounter much trouble trying to distinguish true from false alarms. In fact, Sorkin and Woods (1985) go on to cite that BAT monitors often have problems identifying, prioritizing, and reacting to novel situations with this type of alarm (Banks & Boone, 1981; Cooper, 1977; Kragt & Bonten, 1983). Decision making under circumstances of complexity and stress becomes further confounded when considering monitors' out of the loop role.

Likelihood Alarm Technology

As opposed to BAT, LAT emits various signals indicating PPVs that depend on the extent of sensitivity-threshold violation. Sorkin, Kantowitz, & Kantowitz (1988) explain LAT as an automated monitoring system that encodes dangerous event likelihood into an alerting signal for the human operator. In their study of LAT displays, results showed improved primary and secondary task performance as well as decision making impacts without compromising attentional load. In a previous study, Sorkin & Woods (1985) contend that if an automated subsystem could provide multiple criteria indicating the conservatism of threshold violation, human monitors could re-evaluate resource allocation strategies. Support exists for the contention that people change their responsiveness depending on differing alarm system outputs (Meyer & Ballas, 1997; Robinson & Sorkin, 1985), which is the basis for LAT development. This is useful in situations characterized by high workload, where several tasks must be monitored and limited attention must be allocated based on critical decisions. Summarily, LAT is valuable because it empowers human monitors with PPV knowledge that they can use to respond more often to true alarms compared to false alarms (Bustamante & Bliss, 2005; Bustamante, Fallon, & Bliss, 2005).

Display Integration and Out-of-the-Loop Performance

Again, it is imperative to apportion automation's roles accordingly. Improper function delegation

within complex tasks, such as an aircraft cockpit, may cause more mayhem than no automation at all. Two issues of concern here are reliance, explained earlier in concurrence with PPV, and compliance. Compliance is the degree to which monitors follow alarm advisories (which are typically unreliable given low PPV). Given low base rates and consequent PPV found in most task settings, display integration has become popular. Display integration is the inclusion of many displays into one or few, based on the proximity compatibility principle. This principle, as explained by Wickens & Carswell (1995), declares that displays relevant to a similar mental operation or task should be located close together. Although display integration has many merits, operator compliance becomes an issue because it is difficult for the human monitor to know what is actually happening within the system. This is known as out-of-the-loop performance; where the human monitor/operator is one step removed from the system with an alarm conveying no information beyond one alarm signal. An overly compliant, or completely untrusting, monitor cannot be expected to make premium decisions under complex, high workload situations where the only information available is that an alarm threshold has been violated. Performance may suffer when users reduce compliance and begin to distrust alarm technology (Parasuraman & Riley, 1997).

Alarms and the "Cry-Wolf Effect"

To improve alarm effectiveness, consider the following. Since highly developed fault diagnosis algorithms and sensor technologies solve the accuracy factor of alarms, and base rates of problematic conditions are out of human's control, let's analyze threshold sensitivity. Traditional binary alarm displays emit one signal, regardless of how extreme thresholds have been violated, and minimally indicative of actual urgency. A large proportion of BATs are unreliable because they emit so many false alarms (Getty, Swets, Pickett, & Gonthier, 1995). As Bustamante, Fallon, & Bliss (2005) point out, a common user response to this unreliability is distrust in the alarm- known as the "cry-wolf effect" (Breznitz, 1983). Obviously, human monitors working in vital roles would benefit from trustworthy alarm technology. Monitors falling

victim to the cry wolf effect may ignore true alarms, risking system integrity and even people's lives. Bustamante, Fallon, & Bliss (2005) go on to propose that one way to empower users in better distinguishing FAs from TAs is an alerting display that presents the likely validity of each alarm signal. LAT accomplishes this vision through urgency mapping (Edworthy & Adams, 1996), providing monitors with an indication of how hazardous system conditions may be based upon the degree of sensitivity-threshold violation. Decision making accuracy and response bias were calculated based on the *ab* Signal Detection Theory developed by Bustamante, A.E. (under review), which is specially adapted to these measures.

Goal of this research

This study sought to analyze the effects of LAT on human monitors' decision making during use of aviation displays in a complex, multi-task, flight simulation. Participants' decisions were analyzed in terms of response bias and accuracy using a two stage decision making model introduced by Bustamante, Bliss, and Newlin (2008). Dangerous condition probabilities, alarm reliability, and LAT versus BAT were manipulated in an effort to examine how much LAT impacts decision making.

Hypothesis

It was hypothesized that the probability of engine malfunction (*P*), and LAT would interact with decision making accuracy. Additionally, a main effect of *P* on decision making bias was expected. Manipulating *P* by default changes the EICAS's PPV, so participant response rates were expected to parallel reliability levels, a tendency first noted by Dorfman (1969), and later by Bliss, Gilson, & Deaton (1995). A main effect of *P* on bias, which is the tendency of participants to affirmatively respond, was expected based on prior research by Bustamante & Bliss (2005) that found workload to have main effects on overall and true alarm response rates.

Method

Experimental Design

A 2 x 2 between-groups design was used for this study. The reliability of the alarm system was

manipulated at two extremes of the continuum (i.e., 10% and 90%) following a similar methodology as Bustamante, Bliss, and Anderson (2007). The use of LAT was manipulated following Bustamante (2005)'s methodology comparing two types of alarm systems (i.e., a BAS and a LAS).

Participants

A power analysis revealed that approximately 40 participants would be necessary to obtain statistically significant effects at a .05 alpha level, assuming a power of .80 and a medium effect size for each factor (Cohen, 1988). Therefore, 40 (20 females, 20 males) undergraduate and graduate students participated in this study. Participants ranged from 18 to 30 years of age ($M = 21.90$, $SD = 3.28$). They all had normal or corrected-to-normal vision and hearing. Participants were compensated with one and half research credits as a form of incentive to participate in this study. In addition to this, a \$25.00 gift card was awarded to the participant with the best performance to motivate participants to perform at their maximum level.

Materials and Apparatus

This study took place in a laboratory with an average ambient noise level of 45 dB(A). As part of this experiment, participants completed a simulated roundtrip flight. Each simulated flight leg lasted 20 min. To complete each simulated flight leg, participants had to perform two main flight tasks in addition to a secondary engine-monitoring task.

Primary Flight Tasks. The primary flight tasks were simulated using the Multi-Attribute Task Battery (MATB) designed by Comstock and Arnegard (1992) and consisted of a dual-axis compensatory-tracking task and a resource-management task (see Figure 4).

Compensatory-Tracking Task. The main purpose of this task was to simulate the primary function that pilots need to perform to fly an airplane, which is to maintain level flight. Participants were tasked with keeping a circle that randomly fluctuated along the vertical and horizontal axes as close to the centre as they could.

Resource-Management Task. The main purpose of this task was to simulate another important function that pilots need to perform as they fly an airplane, which is to make sure that they have an

optimal level of fuel in their tanks. Participants were required to keep an optimal level of fuel on the two main tanks, while preventing any of the secondary tanks from being depleted.

Secondary Engine-Monitoring Task. The main purpose of this task was to simulate a crucial secondary function that pilots need to perform to maintain flight safety, which is to ensure that they have at least one fully functioning engine at all times. Participants performed this task with the aid of a simulated Engine Indicating and Crew Alerting System (EICAS), which varied in its degree of reliability (i.e., 10% or 90%), which was manipulated through changes in the probability of engine malfunctions, and the use of LAT (i.e., BAS or LAS), which was implemented following the same methodology as the one used by Bustamante (2005). Participants were tasked with either acknowledging or ignoring alarms emitted by the EICAS. Each alarm was composed by a visual stimulus (see Figure 5) as well as an auditory stimulus, which was presented to participants at 55 dB(A) through a set of sound-attenuated headphones. In case they decided to acknowledge a particular alarm, they gained access to additional system-status information (see Figure 6) to help them make a corrective action when necessary.

Procedure

Participants came to the laboratory individually. First, the experimenter greeted them and provided them with the informed consent form. Second, the experimenter asked participants to silence or turn off their cellular phones if they had one. Third, the experimenter assigned each participant an identification number. Fourth, the experimenter asked participants if they had any questions regarding the nature of the study. If participants decided to participate, the experimenter asked them to sign and date the informed consent form. Fifth, the experimenter asked participants to complete the background and contact information form. Sixth, the experimenter showed participants the instructions for completing the study and asked them to read them carefully. Seventh, the experimenter instructed participants to place the set of stereo headphones on their heads and adjust them to fit comfortably. Eighth, the experimenter showed participants the

familiarization screen and instructed them about how to use the graphical user interface of the program. Ninth, the experimenter showed participants how to navigate through the program to the first session and explained all the information displayed on the screen. Last, the experimenter answered any final questions participants had regarding the completion of the study.

Results

Alarm Response Accuracy

Table 1 shows the observed means and standard deviations of participants' alarm response accuracy across all four conditions.

A 2 x 2 between-groups ANOVA was conducted to examine the effects of the probability of engine malfunctions (.10, .90) and the type of alarm system (BAS, LAS) on participants' alarm response accuracy. Results showed a statistically significant interaction effect, $F(1, 36) = 34.17, p < .01$, partial $\eta^2 = .49$, observed power = .1.00. Results also showed statistically significant main effects for both factors (see Table 2).

As shown in Figure 1, the use of the LAS significantly increased participants' alarm response accuracy only when the probability of engine malfunctions was .10.

Alarm Response Bias

Table 3 shows the observed means and standard deviations of participants' alarm response bias across all four conditions.

A 2 x 2 between-groups ANOVA was conducted to examine the effects of the probability of engine malfunctions (.10, .90) and the type of alarm system (BAS, LAS) on participants' alarm response bias. Results showed a statistically significant interaction effect, $F(1, 36) = .06, p < .01$, partial $\eta^2 = .00$, observed power = .1.00. There were no other statistically significant effects (see Table 4).

As shown in Figure 2, response bias did not vary with the type of alarm system, but did depend on the probability of engine malfunction.

Discussion

Results demonstrate that LAT improved decision making given a low, realistic base rate of dangerous conditions. Many applied settings are characterized by low base rates, which just happen to be ideal for realizing the benefits of LAT owing to its particular merits during high workload. This is especially true in free flight aviation cockpits, the applied aim of this study. Furthermore, LAT had no effect on decision making bias but a significant positive effect on accuracy during high operator workload. Using the DM model as an assessment tool, participants acknowledged more true alarms and ignored more false alarms when assisted by LAT. According to Xiao & Seagull (1999), this is desirable in that high rates of false alarm acknowledgments (and low PPVs) have been empirically shown to decrease performance of human-machine systems (Sorkin & Woods, 1985; Lawless, 1994; Breznitz, 1984; Bliss, Gilson, & Deaton, 1995; Bliss, Dunn, & Fuller, 1995).

Likelihood alarm technology has workload-minimizing and efficiency-maximizing benefits too. These findings also have more implications for minimal attentional costs of LAT compared to BAT, because workload and attention are similar constructs (Kantowitz & Casper, 1988). Consider that if an alarm inflicts a high enough mental workload, performance decreases on other parts of the human-machine system too (Kantowitz & Casper). This is where LAT really excels, because it provides the monitor with a PPV that permits him/her to prioritize tasks and decide whether to acknowledge the alarm in an effort to save cognitive resources, or respond and take possible action. Using the a-b Signal Detection Theory Model in this study as a more robust framework than traditional SDT theory allowed accuracy and response bias to be analyzed independently in the acknowledgement stage. This was advantageous because most alarm interactions are divided into two stages: recognition/acknowledgment and corrective/evasive action. More research into the second stage of corrective or evasive action (post alarm-acknowledgement) is forthcoming.

The results of this experiment, showing that decision making accuracy can be enhanced with LAT, apply to many practical domains. For instance, implementing LAT into the cockpits of commercial and military aircraft may reduce risk because of its positive impact on situational awareness. Tumer & Bajwa (1999) conducted a survey of Engine Health Monitoring (EHM) literature and concluded that automated EHM is hindered primarily by too much uncertain monitoring data and too many false alarms that cause humans' reluctance to rely on the system. This study implies that EHM systems could specifically benefit from likelihood alarm technology because of the higher decision making accuracy, which translates to more hits and less false alarms, displayed by EICAS users in the experiment. Additionally, display monitors interacting with LAT in any similar function would trust alarms more and better resist the cry wolf effect because of the higher PPV. Given high degrees of display integration with current automated systems, LAT's mitigation of out-of-the-loop performance has been demonstrated further with this experiment's first stage analysis of Dr. Bustamante's *a-b* SDT decision making model of accuracy and bias. Decision support tools in all domains requiring similar human monitoring roles of dynamic fault management would better serve if outfitted with LAT. Considering the delicate and crucial delegation of tasks between humans and automation, discussed in the introduction, it is not a far leap to acknowledge the positive role that LAT may play in better coalescing human monitors and alarm systems.

One potential limitation of this study lies in the low-fidelity flight simulation, which had to be so because of participant pool limitations. Training student participants to interact with EICAS in a high fidelity flight simulation is unreasonable because most are unfamiliar with commercial aviation practices. Additionally, this study only examined the first stage of the DM model. Analyzing how humans make decisions under the conditions of this experiment is part of a larger line of research in the University of Idaho's Cognitive Engineering and Decision Making Laboratory, under direction of principle investigator Dr. Ernesto A. Bustamante. One intention along this research line is to examine the DM model's second

stage when gauging performance during an engine malfunctions-correction task. Another suggestion for future research is to directly assess the attentional costs of LAT versus BAT and task-critical information provisions.

References

- Bailey, N.R., and Scerbo, M.W. (2005). Concurrent monitoring for multiple critical signals in a complex display: A vigilance perspective, In *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1518-1522). Santa Monica, CA: Human Factors and Ergonomics Society.
- Banks, W.W., & Boone, M.P. (1981). *Nuclear control room annunciators: Problems and recommendations* (NUREG/CR-2147). Springfield, VA: National Technical Information Service.
- Bliss, J.P., and Dunn, M. (2000). Behavioral implications of alarm mistrust as a function of task workload. *Ergonomics*, 43, 1283-1300.
- Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics*, 38(11), 2300-2312.
- Breznitz, S. (1984). *Cry wolf: The psychology of false alarms*. Hillsdale, NJ: LEA.
- Bustamante, E.A. (2005). A signal detection analysis of the effects of workload, task-critical and likelihood information on human alarm response, In *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 1513-1517). Santa Monica, CA: Human Factors and Ergonomics Society.
- Bustamante, E. A. (under review). Theoretical Implications and Practical Applications of the *a-b* Signal Detection Theory Model. *Ergonomics*.
- Bustamante, E. A., Anderson, B. L., & Bliss, J. P. (2004). Effects of varying the threshold of alarm systems and task complexity on human performance and perceived workload. In *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting* (1948-1952). Santa Monica, CA: Human Factors and Ergonomics Society.
- Bustamante, E.A., and Bliss, J.P. (2004). Effects of varying the threshold of alarm systems on human performance. Manuscript submitted for publication.
- Bustamante, E.A., and Bliss, J.P. (2005). Effects of workload and likelihood information on human response to alarm signals, In *Proceedings of the 13th International Symposium on Aviation Psychology* (pp. 81-85). Oklahoma City, OK: Wright State University.
- Bustamante, E.A., Bliss, J.P., & Anderson, B.L. (2007). Effects of varying the threshold of alarm systems and workload on human performance. *Ergonomics*, 50(7), 1127-1147.
- Bustamante, E. A., Bliss, J. P., & Newlin, E. T. (2008). A two-stage signal detection model of decision making while interacting with integrated alarm systems. In *Proceedings of the 8th Annual Symposium on Human Interaction with Complex Systems*. Norfolk, VA, April 3-4.
- Bustamante, E.A., Fallon, C.K., & Bliss, J.P. (2005b). Improving decision-making with a likelihood alarm display [Abstract]. *American Psychological Association Divisions 19 & 21 Midyear Symposium*. Fairfax, VA, March 3-4.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Comstock, J.R., and Aamegard, R.J. (1992). The multi-attribute task battery for human operator workload and strategic behavior research. (Technical Memorandum No. 104174). Hampton, VA: National Aeronautics and Space Administration.
- Cooper, G.E. (1977). *A survey of the status and philosophies relating to cockpit warning systems*. (NASA-CR-152071). NASA Ames Research Center, Moffet Field, California.
- Dorfman, D. D. (1969). Probability matching in signal detection, *Psychonomic Science*, 17(103).
- Drury, C.G., Ghylin, K.M. and Holness, K. (2006). Error analysis and threat magnitude for carry-on bag inspection, In *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 1189-1193). Santa Monica, CA: Human Factors and Ergonomics Society.
- Edworthy, J., and Adams, A. (1996). *Warning design: A research prospective*. London: Taylor & Francis.
- Fallon, C.K., Bustamante, E.A., Ely, M.R., & Bliss, J.P. (2005). Improving user trust with a likelihood alarm display. Manuscript submitted for publication.
- Getty, D.J., Swets, J.A., Picket, R.M., & Gonthier,

- D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied*, 1, 19-33.
- Gupta, N., Bisantz, A. M., & Singh, T. (2002). The effects of adverse condition warning system characteristics on driver performance: An investigation of alarm signal type and threshold level. *Behaviour and Information Technology*, 21(4), 235-248.
- Kantowitz, B.H. & Casper, P.A. (1988). "Human workload in aviation" in E.L. Wiener & D.C. Nagel (eds.) *Human factors in aviation*, Academic Press. San Diego, CA
- Kragt, J., & Bonten, J. (1983). Evaluation of a conventional process-alarm system in a fertilizer plant. *IEEE Transactions on Systems, Man, and Cybernetics*. SMC-13. 586-600.
- Lawless S. T. (1994). Crying wolf: False alarms in a pediatric intensive care unit. *Critical Care Medicine*. 22(6), 981-985.
- Li, C.R., Lin, W., & Chang, H. (2004). A psychophysical measure of attention deficit in children with attention-deficit/hyperactivity disorder, *Journal of Abnormal Psychology*, 113, 228-236.
- Meyer, J., & Ballas, E. (1997). A two-detector signal detection analysis of learning to use alarms. *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting* (Santa Monica, CA: Human Factors and Ergonomics Society), 186-189.
- Parasuraman, R., Hancock, P.A. & Olofinboba, O. (1997). Alarm effectiveness in driver-centered collision-warning systems, *Ergonomics*, 40, 390-399.
- Parasuraman, R., & Riley V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253.
- Robinson, D.E., & Sorkin, R.D. (1985). A contingent criterion model of computer assisted detection. In R. Eberts & C. G. Eberts (Eds.) *Trends in Ergonomics/Human Factors*, Vol. II. North-Holland: Amsterdam, 75-82.
- Sheridan, T.B. and Hennessy, R.T., Eds (1984). *Research and Modeling of Supervisory Control Behavior: Report of a Workshop*. Washington, D.C.: National Academy Press.
- Sheridan, T. (2002). *Humans and Automation: System Design and Research Issues* (New York: Wiley Interscience).
- Sheridan, T. and Parasurman, R. (2006) Human automation interaction. In *Proceedings of the Human Factors and Ergonomics Society 46th Annual Meeting*, R. Nickerson (Ed) (Santa Monica, CA: Human Factors and Ergonomics Society).
- Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors*, 30(4), 445-459.
- Sorkin, R. D., & Woods, D. D. (1985). Systems with human monitors: A signal detection analysis. *Human Computer Interaction*, 1, 49-75.
- Tumer, I. Y., & Bajwa, A. (1999). A survey of aircraft engine health monitoring systems. In *Proceedings of the 35th Joint Propulsion Conference, Aerospace Systems Condition Monitoring Session*, (pp. 1-7). Los Angeles, CA: American Institute of Aeronautics and Astronautics.
- Wickens, C.D., & Carswell, M.C. (1995). The proximity compatibility principle: its psychological foundation and relevance to display design, *Human Factors*, 37.
- Wickens, C.D., & Dixon, S.R. (2007). The benefit of imperfect diagnostic information: A synthesis of the literature, *Theoretical Issues in Ergonomics Science*, 8, 201-212.
- Woods, D.D. (1982). The cognitive coupling of man and machine. *Second International Workshop on Safety and Reliability of Industrial Computer Systems*, West Lafavette, IN.
- Woods, D.D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, 38(11), 2371-2394.
- Xiao, Y., & Seagull, F.J. (1999). An analysis of problems with auditory alarms: Defining the roles of alarms in process monitoring tasks. *Proceedings of the Human Factors and Ergonomic Society 43rd Annual Meeting*, 256-260.

Table 1

Means and Standard Deviations of Participants' Alarm Response Accuracy

| Probability of Engine Malfunctions | Alarm System | | | |
|------------------------------------|--------------|-----------|----------|-----------|
| | BAS | | LAS | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| .10 | 0.51 | 0.10 | 0.77 | 0.09 |
| .90 | 0.50 | 0.00 | 0.50 | 0.01 |

Table 2

Analysis of Variance Source Table for Participants' Alarm Response Accuracy

| Source | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | Partial η^2 |
|--|-----------|-----------|-----------|----------|------------------|
| Probability of Engine Malfunctions (P) | 1.00 | 0.20 | 0.20 | 39.63* | .52 |
| Alarm System (AS) | 1.00 | 0.16 | 0.16 | 31.90* | .47 |
| P x AS | 1.00 | 0.17 | 0.17 | 34.20* | .49 |
| Error | 36.00 | 0.18 | 0.01 | | |

* $p < .01$

Table 3

Means and Standard Deviations of Participants' Alarm Response Bias

| Probability of Engine Malfunctions | Alarm System | | | |
|------------------------------------|--------------|-----------|----------|-----------|
| | BAS | | LAS | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| .10 | 0.41 | 0.37 | 0.43 | 0.12 |
| .90 | 1.00 | 0.00 | 1.00 | 0.01 |

Table 4

Analysis of Variance Source Table for Participants' Alarm Response Bias

| Source | <i>df</i> | <i>SS</i> | <i>MS</i> | <i>F</i> | Partial η^2 |
|--|-----------|-----------|-----------|----------|------------------|
| Probability of Engine Malfunctions (P) | 1.00 | 3.31 | 3.31 | 89.13* | .71 |
| Alarm System (AS) | 1.00 | 0.00 | 0.00 | 0.03* | .00 |
| P x AS | 1.00 | 0.00 | 0.00 | 0.06* | .00 |
| Error | 36.00 | 1.34 | 0.04 | | |

* $p < .01$

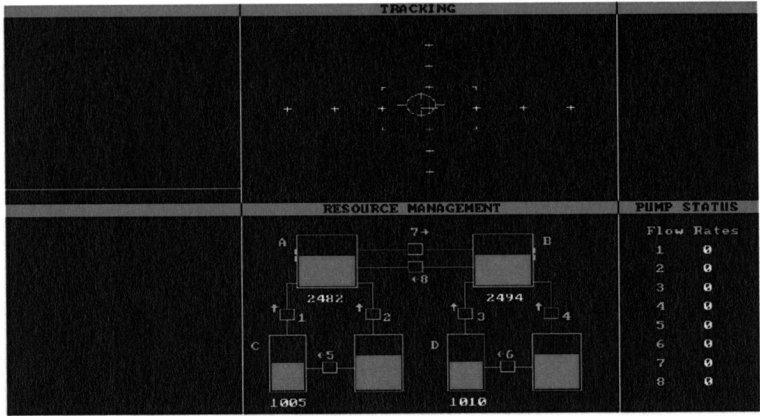


Figure 1. Multi-attribute task battery.

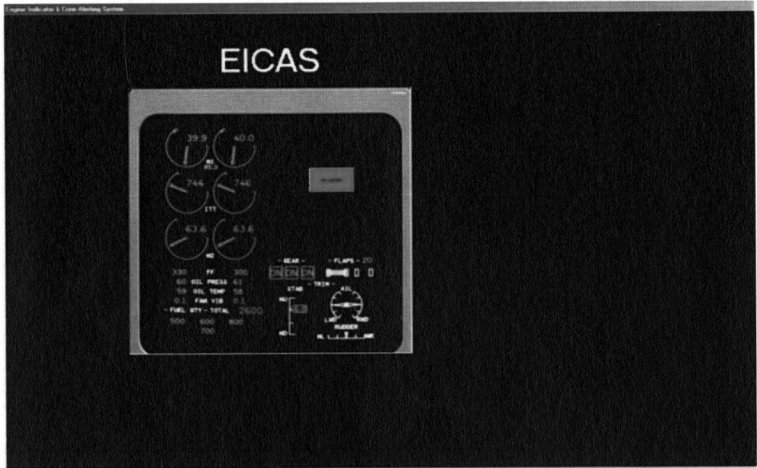


Figure 2. Simulated EICAS display.

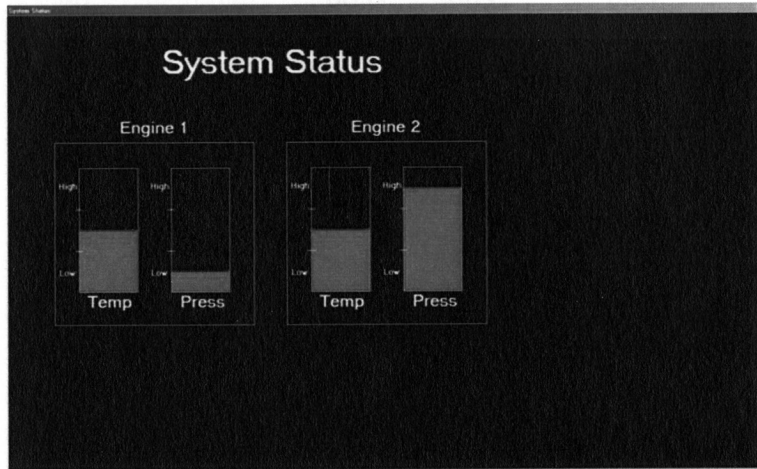


Figure 3. System status information.

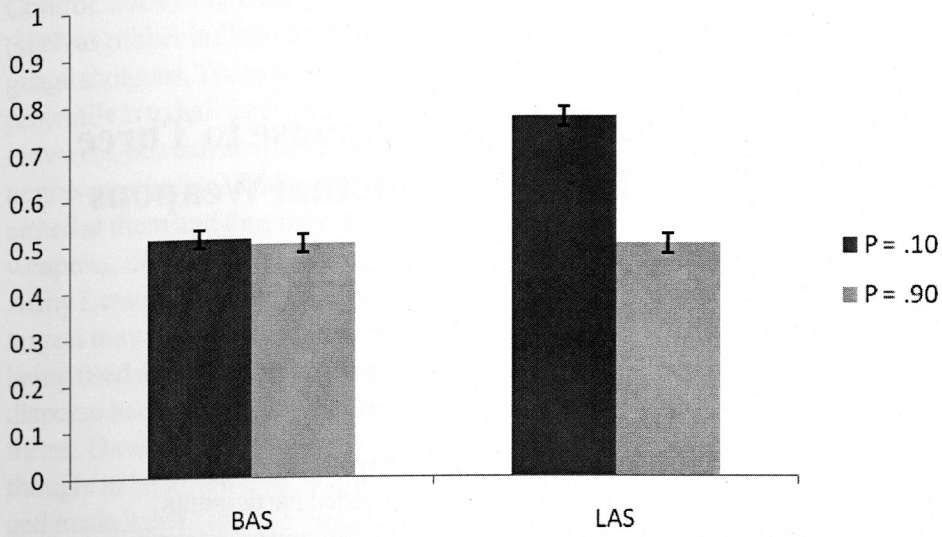


Figure 4. Participants' alarm response accuracy as a function of the probability of engine malfunctions and the type of system.

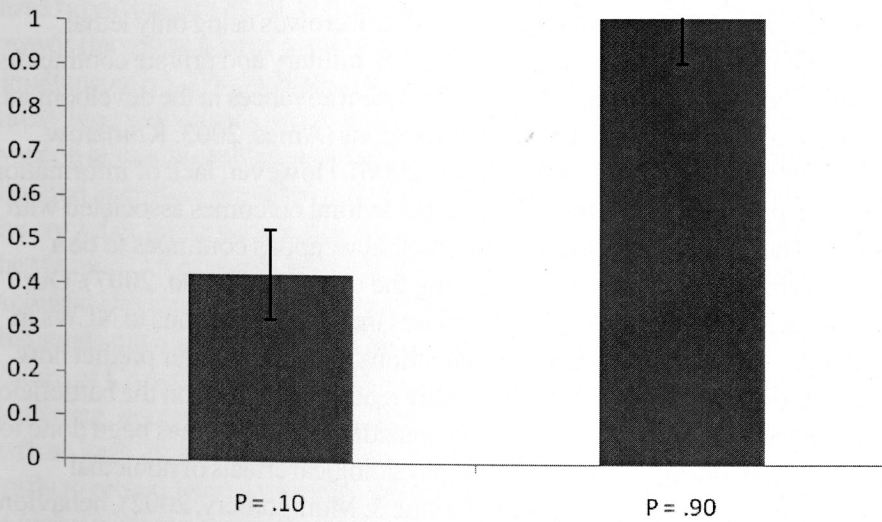


Figure 5. Participants' alarm response bias as a function of the probability of engine malfunctions.