Situated Visual Alarm Displays Support Machine Fitness Assessment for Nonexplainable Automation

Michael F. Rayo, Chelsea R. Horwood, Morgan C. Fitzgerald, Marisa R. Grayson, Mahmoud Abdel-Rasoul, and Susan D. Moffatt-Bruce

Abstract*—Determine if situated visual alarm displays can support machine fitness assessment (MFA), facilitating improved hazard recognition and alarm accuracy assessment in the presence of inaccurate alarms. Poor performance of opaque automation is more difficult to detect, which increases the likelihood of cascades resulting in overall system failure. MFA reduces the negative impact of poor automation performance. Integrated alarm visualizations were shown to 32 nurses for 10 cases focused on patient outcome and 17 focused on alarm quality, all using real patient data. Five of the ten outcome cases would ultimately result in an emergency (unbeknownst to the nurse). Alarm cases ended with a true, false, or unnecessary alarm. Responses for nurses' concern, confidence, alarm quality, and intended response were recorded. Qualitative analysis of interviews was performed. Using the situated visual alarm displays, nurses reported less confidence (6.5 vs. 9.1, $p < 0.001$), more concern (5.4 vs. 1.6, $p < 0.001$), and more urgent responses for emergency cases. Their alarm event detection was better than the alarms' detection (0.608 vs. 0.438, $p < 0.001$), as was their interpretation accuracy (0.453 vs. 0.243, $p < 0.001$). Nurses showed differentiated concern for emergency cases, nonemergency cases with alarms, and those without alarms (5.4 vs. 3.8 vs. 1.6, $p < 0.001$). Situated visual alarm displays combining visual trends with alarm signals improves detection of hazardous events and mitigates the negative effects of poor opaque automation performance.

***Index Terms*—Alarm design, computer interface, design display principles, explainable artificial intelligence (AI), human systems integration, trust in automation.**

I. Introduction

THE need for design strategies to integrate opaque artificial intelligence (AI) technologies into high-performing joint human–machine systems will be critically important for the foreseeable future. This may seem to contradict the increasing consensus in the AI community about the inherent limitations of AI and the associated need for explainable, observable, and transparent AI [1]. However, current opaque AI technologies regularly outperform explainable alternatives in most settings in terms of accuracy and efficiency. For this reason, the vast majority of recent AI implementations are opaque, not explainable [1]. Replacing these technologies will not be trivial. Not only will it take time for organizations to replace all of these technologies, but each replacement decision, which is a type of sacrifice judgment [2], will require that the explainable alternative must unambiguously outperform its opaque counterpart in order to justify the expenditure and risk to implement it. This will take some time, and may never be fully completed.

Utilizing automation to sustain system performance in high-stakes, high-uncertainty, high-complexity settings requires that it contributes positively to joint human-automation activities, even when the automation is performing poorly [3], [4]. This is more difficult for opaque automation because it violates the majority of guidelines for joint human-automation work [5], leaving only one strategy left, which is to continuously optimize the automation itself. However, the upper bound of accuracy of current automation is not sufficient to guarantee performance over the full range of possible or even likely scenarios. When opaque automation fails, these failures are more likely to cascade into overall system failures [6]–[9], which is commonly referred to as system brittleness [10], [11]. Some of these brittleness studies clearly show that the presence of poorly performing automation resulted in system performance worse than that of unaided operators [8], [9]. Simply reducing the frequency of machine failure is not sufficient, and has been shown to result in "robust yet fragile" systems in which the increased complexity resulting from machine optimization also increases the likelihood of rapid and catastrophic collapse [12].

A good example of this is in the history and present state of threshold alarms in the healthcare industry. The inability of modern threshold alarms to support the cognitive functions

Manuscript received 19 July 2021; revised 14 October 2021, 9 December 2021, and 24 January 2022; accepted 30 January 2022. Date of publication 24 March 2022; date of current version 15 September 2022. This work was supported in part by the Institute for the Design of Environments Aligned for Patient Safety, which is sponsored by the Agency for Healthcare Research & Quality under Grant P30HS024379. This article was recommended by Associate Editor M. Hou. *(Corresponding author: Michael F. Rayo.)*

Michael F. Rayo and Morgan C. Fitzgerald are with the Department of Integrated Systems Engineering, The Ohio State University, Columbus, OH 43210 USA (e-mail: rayo.3@osu.edu; fitzgerald.205@osu.edu).

Chelsea R. Horwood is with the Department of Surgery, The Ohio State University, Columbus, OH 43210 USA (e-mail: chelsea.horwood@osumc.edu).

Marisa R. Grayson is with the Department of Integrated Systems Engineering, The Ohio State University, Columbus, OH 43210 USA, and also with the Mile Two LLC, Dayton, OH 45402 USA (e-mail: mbigelow@miletwo.us).

Mahmoud Abdel-Rasoul is with the Department of Biomedical Informatics, The Ohio State University, Columbus, OH 43210 USA (e-mail: mahmoud.abdel-rasoul@osumc.edu).

Susan D. Moffatt-Bruce is with the Department of Surgery, The Ohio State University, Columbus, OH 43210 USA, and also with the Royal College and the University of Ottawa, Ottawa, ON K1S 5N8, Canada (e-mail: smoffat-tbruce@royalcollege.ca).

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of The Ohio State University under Application No. 2013H0419.

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/THMS.2022.3155714.

Digital Object Identifier 10.1109/THMS.2022.3155714

required to dynamically direct attention from "something that is less important to something that is more important" [13] is still a pressing problem. The most consequential contributor to these alarms' problems is their chronic lack of informativeness [14], [15]. Alarm informativeness, or the amount of information carried with a given alarm signal, was originally defined and calculated as the proportion of times an alarm system correctly detects a targeted event that has been predetermined to be hazardous. This is traditionally defined as an alarm's number of true positives (TP) divided by the sum of TP and false positives (FP) [16]. This definition was later revised by separating out unnecessary alarms (U) from TP. Unlike FP, unnecessary alarms correctly identify alarm events. However, they refer to alarm events that are not hazardous in the immediate context, but still interpret and convey them as hazardous or urgent. Said another way, these alarms correctly identify events that do not require operator attention. The new informativeness calculation therefore becomes TP/(TP+FP+U) [17], and decreases as the proportion of false and unnecessary alarms increases [14]. Low informativeness reduces trust [18], increases response time and likelihood to disregard alarms in general [14], [16], [19], [20], and increases the likelihood of responding to inaccurate alarms [21]. Together, these result in delayed or insufficient responses to emerging hazardous conditions [22]. Even though there has been sustained investment in opaque alarm technologies that have been shown in simulations to increase informativeness [23]–[26], these innovations have not translated into an appreciable reduction in the alarm problem observed in hospitals. At the heart of this problem are two issues that clinicians face: the inability to determine whether or not an alarm is true, and whether a patient event is occurring that requires an urgent response [27]. Current alarm technologies are still not well-suited to address these issues [28].

This article explores the effectiveness of a new design strategy, situated visual alarm displays, to address these issues. This strategy transforms low informativeness, opaque alarms into high informativeness, interrogatable visual displays by visually situating alarms with other relevant environmental data, and by expanding the definition of informativeness. This new definition includes not only what the automation *detects* or *interprets*, but what it effectively *conveys* to other agents [14]. It sidesteps the distinction between opaque and explainable automation, instead focusing on supporting machine fitness assessment (MFA; i.e., assessing the machine's fitness to perform a particular task for a specific situation). MFA is the goal of explainable, transparent, and observable automation [3]. Situated displays convey more than the embedded automation can detect or interpret, allowing operators to visually detect gaps and other discrepancies between the alarm output and other data streams. They support the interrelated goals of supporting hazard management, providing operational context, and supporting alarm prioritization [29]. This facilitates MFA support even for automation that cannot provide its own explanations or allow other agents to explain it (i.e., projective causal reasoning, described in [30]).

In our article, situated alarm displays comprised of alarm data overlays of physiological data of a series of real patients that were presented to registered nurses. Display designs were influenced by our domain knowledge attained through previous observational research [27] and multiple representation aiding techniques. We evaluated MFA in two ways. The first is by outcome: we measured performance of detecting hazards and calibrating urgency of response as the results of effective MFA. The second is more direct: we measured participants' ability to discern accurate from inaccurate alarms. We sought to answer five research questions. With the assistance of a situated alarm display, can clinicians (RQ1) differentiate between urgent and nonurgent patients in terms of level of concern and urgency of response? Also, can they (RQ2) outperform the alarm system in correctly detecting events of interest (i.e., alarm events) and (RQ3) correctly interpret which of these events are patient hazards? Finally, (RQ4) are they influenced to unduly increase their concern for nonurgent patients if false or unnecessary alarms are present? And (RQ5) will that influence be so strong that they cannot differentiate between nonurgent patients with alarms and urgent patients with alarms?

II. Methods

A. Integrated Alarms Display Design

The design of our situated alarm display was highly influenced by multiple representation aiding techniques to maximize the data available to the participants without having them suffer from data overload, a common complicating factor of data-driven displays. These techniques included Woods's representation design [31], the ecological interface design process [32], and the Gestalt Laws [33]. We first selected a frame of reference, using findings from our previous article [27]. These findings emphasized the importance of understanding trends over time of heart rate (HR), SpO_2 (i.e., the level of oxygenation of the blood), and blood pressure, to detect patient decompensation, which is a rapid deterioration in the patient's ability to maintain physiological function. They also emphasized the need for quick perceptual pickup of these trends, as decompensation can result in morbidity or mortality in minutes. In a separate study, we found that current visual displays and auditory alarms do not display these data at all necessary timescales. There are displays for second-to-second dynamics, such as heart rhythm, and hour-to-hour, such as validated vital signs in electronic health records (EHR), but nothing that captures and conveys the minute-to-minute dynamics [28]. This minute-to-minute perspective of physiological measures, now possible through a newly available data source, became the primary frame of reference for our displays.

We used these and other data to set the alarm data in context [31]. This allowed for direct visual comparison of the alarms to the physiological data and the alarm thresholds (see dashed lines, Fig. 1). These techniques are effective strategies to mitigate the risk of operators getting lost in crowded data displays by increasing informativeness as data density increases [14], [15]. With the new vital sign and alarm datasets displayed simultaneously but independently on the same display, using primarily a point encoding for alarms and a line encoding for vital signs, we expected to realize the benefits of a configural display: reduced attentional costs, increased conveyance of data

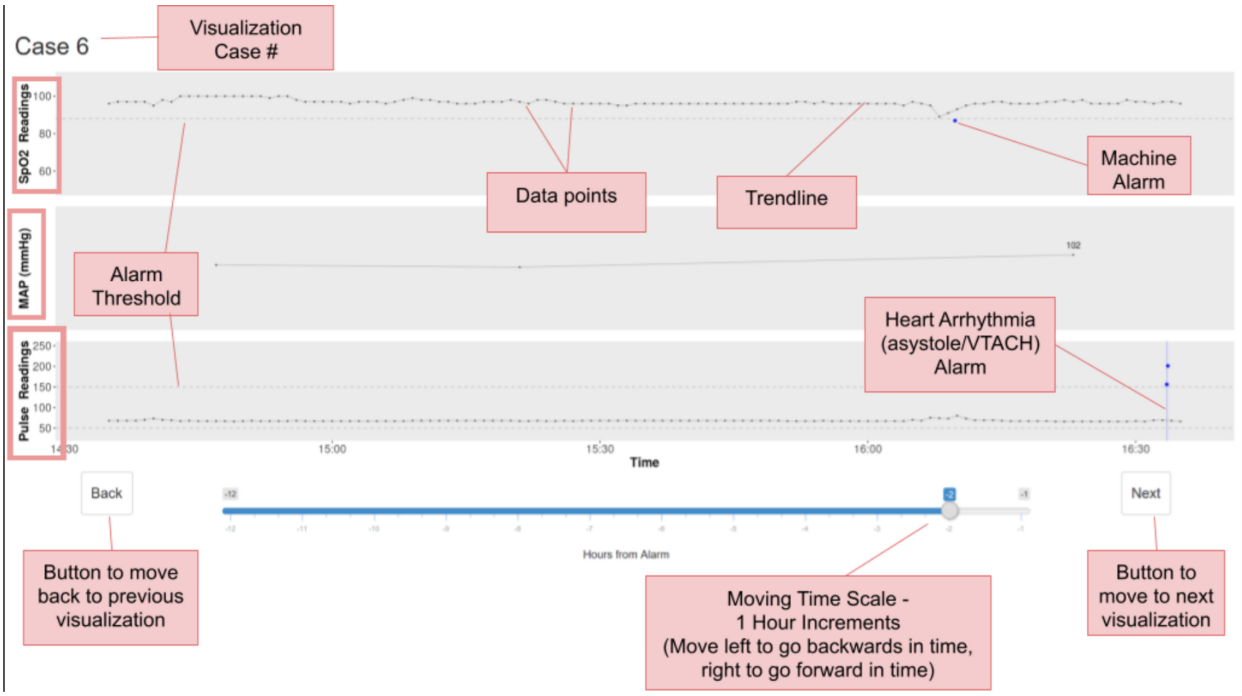


Fig. 1. Example of integrated alarm display, with callouts describing different sections (Note: light red boxes are annotations, not present during experiment).

interactions, and increased detection of emergent phenomena [34]. Finally, our display used the Gestalt Law of Proximity [33] to provide instant visual association to support human–machine common ground [35] on the agreement between alarms and vital sign values relative to alarm thresholds. Although there is a paucity of research on the comparative design benefits of using proximity to encode similarity, it seemed to be an acceptable choice in the context of the overall display design. The display used primarily analogical representations to support quick perceptual pickup of key patterns, as they are effective in redirecting attention in situations of both low and high uncertainty and complexity [36], [37].

The situated alarm display, shown in Fig. 1, shows synchronized timelines for HR, mean arterial pressure (MAP), and SpO_2. HR and SpO_2 values were recorded and displayed in 1-min intervals (where data were available) as small black dots. MAP was displayed more sporadically, wherever available, also as small black dots. Each vital sign mark was connected to the previously recorded mark by a thin black line. Alarm marks that corresponded with vital sign values were drawn as overlays of the vital sign displays related to their function (e.g., heart alarms were overlays for the HR display panel), synchronized in time, and placed where the vital sign reading was recorded at the time of the alarm (e.g., HR = 47). Any alarms that were not associated with a vital sign value (e.g., no heart rhythm detected) were represented as a vertical line at the time that the alarm was recorded. Alarm thresholds for both HR and SpO_2 were represented as horizontal dashed lines. The amount of historical information displayed was an adjustable parameter under the participant's control with the default view showing the last hour of recorded data for the patient. An interactive slider allowed the participant to adjust the timeline and view up to 12 h into the past. The feature was included as a means for the nurses to explore more contextualized information about patients they are concerned about, which may in turn change their interpretation. It is this type of interaction that Thomas and Cook [38] advocate for in supporting rich interactivity between humans and machine-produced data.

The data visualizations were developed in RStudio (version 1.1.447) and displayed in a web-based browser app powered by R Shiny (version 1.0.5).

B. Patient Case Selection

A total of 27 patient cases were selected from a dataset of 270 deidentified patients available to the research team. Cases contained between 2 and 12 h of patient data. They also included all auditory cardiac and respiratory (nonventilator) alarms that occurred during the case. To answer research question 1, urgent cases were defined as those resulting in an emergency response team (ERT) being called. To ensure the best like-to-like comparison, five cases were chosen that resulted in an ERT. These cases were called ERT. Additional five were selected that did not result in ERT, but matched an ERT case based on the care unit, time period, and initial diagnosis (i.e., non-ERT cases). These 10 cases were labeled as outcome cases. This matching was done so that ERT and non-ERT cases captured time windows in which the patients were comparable in acuity and stability. To answer research questions 2 and 3 regarding human–machine accuracy relative to the alarm system, the other 17 cases were drawn from a pool of cases where alarm quality was previously annotated by an experienced physician (i.e., alarm cases). Research questions 4 and 5, regarding the influence of alarms on concern level, were answered by studying the differences between ERT, non-ERT,

TABLE I
CASE ATTRIBUTES AND ORDER

Case #	Type	Subtype	Alarm event	# Alarms	# Alarms visible	Difficulty
1	Alarm	False	Arrhythmia	119	64	Easy
2	Alarm	Unnecessary	HR low	3	3	Easy
3	Alarm	True	Oxygen low	23	15	Easy
4	Outcome	Non-ERT	N/A	1	0	Easy
5	Outcome	ERT	N/A	3	2	Hard
6	Alarm	True	HR low	22	14	Hard
7	Alarm	Unnecessary	Oxygen low	9	2	Easy
8	Alarm	False	Arrhythmia	4	1	Hard
9	Outcome	ERT	N/A	3	3	Easy
10	Outcome	Non-ERT	N/A	4	0	Easy
11	Alarm	False	HR high	6	4	Easy
12	Alarm	Unnecessary	Oxygen low	8	8	Hard
13	Alarm	False	Oxygen low	5	2	Easy
14	Outcome	Non-ERT	N/A	3	0	Easy
15	Outcome	ERT	N/A	32	6	Easy
16	Alarm	False	Oxygen low	18	13	Easy
17	Alarm	True	Oxygen low	50	15	Hard
18	Alarm	False	Oxygen low	49	22	Easy
19	Outcome	ERT	N/A	53	1	Hard
20	Outcome	Non-ERT	N/A	0	0	Hard
21	Alarm	True	Oxygen low	69	59	Hard
22	Alarm	False	Arrhythmia	6	5	Easy
23	Alarm	False	Arrhythmia	19	1	Hard
24	Outcome	ERT	N/A	0	0	Hard
25	Outcome	Non-ERT	N/A	0	0	Hard
26	Alarm	False	Oxygen low	29	19	Hard
27	Alarm	False	Oxygen low	7	6	Hard

and false/unnecessary alarm cases. Details of each case are shown in Table I.

ERT case data started at the beginning of the patient encounter or 12 h before the ERT, whichever was shorter, and ended 5 min before the ERT was called. Non-ERT cases were a randomly selected 12-h period or the length of the encounter, whichever was shorter. ERT cases contained between 0 and 53 total alarms, with 0 to 6 initially visible. Non-ERT cases contained between 0 and 4 total alarms, with 0 alarms initially visible.

Alarm cases were designated as true, false, or unnecessary (i.e., alarm quality) based on the quality of the last alarm on the display. False alarms incorrectly detected an alarm event [17]. For example, a heart arrythmia alarm resulting from a patient moving and jostling the sensors, or inadvertently touching or

brushing the sensors, would be considered false. Unnecessary alarms incorrectly interpreted an alarm event as being hazardous [17]. For example, a low HR alarm triggered when HR drops below 50 beats per minute (bpm) is unnecessary for a patient with a resting HR of 48 bpm. Unnecessary alarms have also been called nonactionable [22] and nuisance [39] alarms. All remaining alarms were true alarms [17]. Determination of alarm quality was made by an experienced physician through real-time, direct observation of the alarms on the patient care unit [40]. The detailed list of true, false, and unnecessary alarm criteria can be found in Appendix 1 in the online addenda. Of the 17 alarm cases, four were true, ten were false, and three were unnecessary. The start and end time of the alarm case matched the start and end of the direct observation period. Alarm cases contained between 3 and 119 alarms, of which a range of 1–64 were initially visible to the clinician.

Cases were selected so that the set contained a sufficient range of intended difficulty and had sufficient power for outcome and alarm analyses. Difficulty was characterized as high or low based on a trained clinical researcher's assessment of the ability to predict future patient status based on current data. These determinations were aligned with classic patterns of barriers and facilitators to event detection and sensemaking, including aspects of data overload [41] and change blindness [42] resulting from imperceptibly small but compounding changes over a prolonged timescale. The number of each case type (i.e., true, false, unnecessary, ERT, and non-ERT) was determined to get as close as possible to the proportions found in the clinical setting, and also provides a sufficient number for analysis. The order in which cases were presented to clinicians was chosen so that no two consecutive cases would be of the same type and to mitigate any learning effects from early cases on decisions for subsequent cases. Alarm cases were interspersed with outcome cases. Details of the cases are shown in Table I.

C. Participants

Over a 1-month period, 32 nurses with bedside patient responsibilities in a midwest multidisciplinary tertiary care medical center participated in this article. Nurses were recruited from multiple work shifts from intensive care, medical/surgical, cardiac medical/surgical, and progressive care units. There were 24 female and 8 male participants. The age of the participants ranged from 25 to 59 years old. The years of experience ranged from 3 weeks to 31 years of direct patient care responsibilities.

D. Experimental Protocol

Each participant session began with a training case presented on an iPad device. The interviewer described the visualization and each of its subcomponents. After the training, the study began by having the participants imagine that they were just beginning a new shift. They were asked to briefly review and assess each patient with the novel display as they would at the beginning of a new shift. This article workflow was meant to mimic how clinicians might use this new technology to familiarize themselves with new patients or periodically monitor them amid other duties, much like they currently do using EHR. It also explored how this technology could benefit decision-making if it were to be dynamically shown to a clinician when one of their patients' auditory alarms was triggered. Each case consisted of the visualization created with the patient's data. No other patient information was shared with the participant. Because it was meant only to simulate the brief snapshot in time directly after viewing the visualization, no auditory alarms were used, even though in the real environment they would be present. Historical clinical alarms were presented in the visual displays. All participants responded to all patient cases, unless their clinical duties interrupted their session and they could not be rescheduled. After each case, the following questions were asked regarding the display:

1) Do you think the alarm is true, false, or unnecessary? Why?
2) What do you think is happening to the patient? Why?
3) On a scale from 1 to 10 (10 being most confident) how confident/sure are you that this is happening to the patient?
4) On a scale from 1 to 10 (10 being most worried) how nervous/worried/concerned are you about this patient?
5) How would you respond to this patient? (A—drop current task and go immediately to the patient, B—finish current task and then go to the patient, C—eventually go see the patient, or D—do not need to see the patient).
6) How did the display help or hinder you in answering the questions above?

One interviewer and one recorder were present during all interviews. All interviews were recorded via Echo Smartpen© and transcribed.

E. Data Analysis

1) Distinguishing ERT From Non-ERT Cases (RQ1): Linear mixed models with random intercepts were fit to estimate differences in the concern and confidence scores between ERT and non-ERT cases (the primary research outcome of interest) and, separately, between true, false, and unnecessary alarm cases (a secondary research outcome of interest). The random intercepts were included to account for repeated measures within participant. All models were visually checked for the assumption of normally distributed residuals using *q-q* plots and no other transformations were necessary as the model assumptions were not determined to be violated. Urgency of response outcomes were compared for patient outcome cases and alarm cases using chi-square tests.

2) Performance Differences in Event Detection and Interpretation (RQ2,3): Alarm accuracy for the alarms in this article was measured in two ways: correct detection of alarm events (i.e., detection accuracy) and correct interpretation that the alarm event is a hazard (i.e., interpretation accuracy). For alarms, detection accuracy was defined as

$$\frac{\mathrm{TP} + \mathrm{U}}{\mathrm{TP} + \mathrm{FP} + \mathrm{U}}$$

where TP is the number of TP, FP the number of FP, and U the number of unnecessary alarms, as defined above. Interpretation

TABLE II
SUMMARY OF RESULTS FOR ALL RESEARCH QUESTIONS

	Research question for nurses with situated alarm (SA) displays	Statistical technique	Results
1.	Differentiation of concern for ERT vs. non-ERT patients	Linear mixed models with random intercepts	5.4 vs. 1.6, p < .001
2.	Nurse/SA vs. alarm in detecting alarm event	One sample binomial tests of proportion	.608 vs. .438, p < .001
3.	Nurse/SA vs. alarm in interpreting hazardous event	One sample binomial tests of proportion	.453 vs. 243, p < .001
4.	Undue alarm prioritization: concern for non-ERT patients with false and unnecessary alarms vs. non-ERT with no visible alarms	Linear mixed models with random intercepts	3.8 vs. 1.6, p<.001
5.	Correct emergency prioritization: concern for ERT patients vs. non-ERT patients with false and unnecessary alarms	Linear mixed models with random intercepts	5.4 vs. 3.8, p < .001

accuracy was defined as

$$\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}+\mathrm{U}}.$$

Human–machine detection accuracy was defined as

$$\frac{\mathrm{CI}_{\mathrm{FP}}+\mathrm{CI}_{\mathrm{U\ or\ TP}}}{\text{total cases}}$$

where $\mathrm{CI}_{\mathrm{FP}}$ is the correct identification of false alarm cases and $\mathrm{CI}_{\mathrm{U\ or\ TP}}$ is the sum of the correct identification of TP cases, correct identification of U cases, and cases where TP or U cases were identified as the other. Human–machine interpretation accuracy was defined as

$$\frac{\mathrm{CI}_{\mathrm{FP}}+\mathrm{CI}_{\mathrm{U}}+\mathrm{CI}_{\mathrm{TP}}}{\text{total cases}}$$

where $\mathrm{CI}_{\mathrm{FP}}+\mathrm{CI}_{\mathrm{U}}+\mathrm{CI}_{\mathrm{TP}}$ is the sum of the correct identification of false, unnecessary, and true alarm cases. Human–machine accuracy was compared to alarm accuracy for both detection and interpretation using one sample binomial tests of proportion.

3) Effect of Presence of Alarms (RQ4,5): To answer these research questions, patient cases were redefined and combined. This was performed because none of the non-ERT cases in the available case dataset contained initially visible alarms, and no participants increased the viewable timescale to see alarms on these cases. Therefore, the nurses' level of concern could have solely been influenced by the presence or absence of alarms. ERT cases were relabeled as ERT/alarm. Non-ERT cases were relabeled as non-ERT/nonalarm. False alarm and unnecessary alarm cases were combined, and relabeled as non-ERT/alarm. True alarm cases were not included in this analysis because even though there was no ERT, it may have been the result of nurse response to the true alarm.

A linear mixed model with random intercepts (to account for repeated measures within participant) was fit to estimate the effect of the presence of alarms in obscuring the difference between hazardous and nonhazardous patient conditions.

Hypothesis testing was conducted at a 5% type one error rate (alpha = 0.05). All statistical analyses were conducted in SAS version 9.4 (SAS Institute, Cary, NC). Results of the post-test interview were used to generate preliminary explanations for the statistical findings. This article complied with the American Psychological Association Code of Ethics and was approved by the Institutional Review Board at The Ohio State University. Informed consent was obtained from each participant.

III. RESULTS

Participants spent, on average, 0.95–1.74 min per patient case, which is commensurate with a brief patient review at the beginning of a new shift. They correctly identified the alarm quality in 56.0% of the true alarm cases, 30.1% of the unnecessary alarm cases, and 43.2% of the false alarm cases. Participant age, participant experience, and patient case duration were not associated with any of the performance measures. A summary of the most relevant research findings is shown in Table II.

A. Distinguishing ERT From Non-ERT Cases (RQ1)

Fig. 2 shows the results of the linear mixed model analysis for nurse confidence and concern scores comparing ERT and non-ERT cases. Concern was less for non-ERT cases (mean score 1.6, 95% CI 1.0 to 2.2) than for ERT cases (mean score 5.4, 95% CI 4.8 to 6.0) (mean difference −3.8 95% CI −4.4 to −3.3, $p < 0.001$). Confidence for non-ERT cases (mean 9.1, 95% CI 8.6 to 9.7) was higher than for ERT cases (mean 6.5, 95% CI 5.9 to 7.0) (mean difference 2.7, 95% CI 2.2 to 3.2, $p < 0.001$).

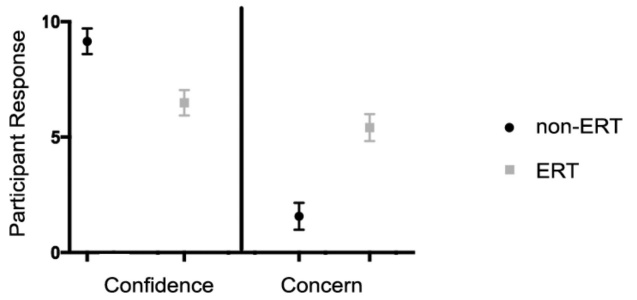


Fig. 2. Confidence and concern scores for ERT and non-ERT cases.

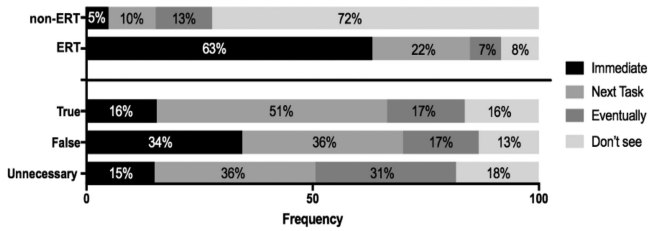


Fig. 3. Urgency responses for all cases.

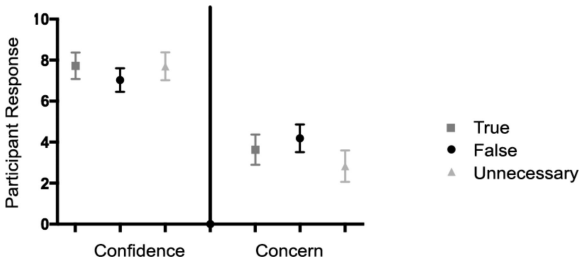


Fig. 4. Confidence and concern scores for alarm cases.

Fig. 3 shows response urgency for all cases. Nurses reported they did not have to see the patient for 72% of the non-ERT cases. This was in contrast to the ERT patients where 63% of nurses reported that they would drop everything and assess the patient immediately ($p < 0.001$). Hard ERT cases were less concerning than easy ERT cases (3.9 ± 0.63 vs. 7.5 ± 0.75). Hard non-ERT cases were no more concerning than easy non-ERT cases (1.5 ± 0.41 vs. 1.6 ± 0.6). Case difficulty was not associated with changes in confidence.

B. Distinguishing True, False, and Unnecessary Cases (RQ1)

Fig. 4 shows the results of the linear mixed model analysis for nurse confidence and concern scores for alarm cases, comparing true, false, and unnecessary alarms. Alarm type showed an overall effect ($p < 0.0001$). In pairwise comparisons, concern was higher for false alarm cases (mean score 4.2 95% CI 3.5 to 4.9) than true alarm cases (mean score 3.6, 95% CI 2.9 to 4.4) (mean difference 0.6, 95% CI 0.1 to 1.1, p-value 0.028) and unnecessary alarm cases (mean score 2.8, 95% CI 2.1 to 3.6) (mean difference 1.4, 95% CI 0.8 to 1.9, p-value < 0.001). Concern was also higher for true alarm cases than unnecessary cases (mean difference 0.8, 95% CI 0.2 to 1.4, p-value 0.011). Details of response urgency are shown in Fig. 3.

TABLE III
NURSES' LEVEL OF CONCERN BASED ON THE PRESENCE/ABSENCE OF ALARMS AND THE PRESENCE/ABSENCE OF A RESULTANT EMERGENCY EVENT (I.E., ERT)

Condition	Avg Concern (95% CI)	DF	p (between all pairs)
Non-ERT / non-alarm	1.6 (1.2-1.93)	64.1	<.0001
Non-ERT / alarm	3.8 (3.51-4.09)	39.2	
ERT / alarm	5.4 (5.07-5.73)	64.1	

Confidence for false alarm cases (estimated mean score 7.0, 95% CI 6.5 to 7.6) was less than that of true alarm cases (estimated mean score 7.7, 95% CI 7.1 to 8.4) (mean difference -0.7, 95% CI -1.2 to -0.2, p-value 0.006) and unnecessary alarm cases (estimated mean score 7.7, 95% CI 7.0 to 8.4) (mean difference -0.7, 95% CI -1.2 to -0.1, p-value 0.015). There was no difference in the reported confidence between unnecessary and true alarms.

C. Performance Differences in Event Detection and Interpretation (RQ2,3)

Human–machine event detection accuracy was greater than alarm event detection (0.608, CI 0.56–0.65 vs. 0.438, $p < 0.001$) for the alarms in the article. Human–machine interpretation accuracy was greater than alarm interpretation accuracy (0.453, CI $0.41-0.50$ vs. 0.243, $p < 0.001$). In total, hard true cases were harder to identify (56 of 106 correctly identified, 42%) than easy true cases (25 of 32, 78%, $p < 0.01$). Hard unnecessary cases were also harder to identify (2 of 32, 6%) than easy unnecessary cases (26 of 64, 41%, $p < 0.001$). Hard false cases were not harder to identify (39 of 118, 30%) than easy false cases (77 of 192, 40%, $p = 0.66$). Case difficulty was not associated with changes in confidence.

D. Effect of Presence of Alarms (RQ4,5)

Nurses were more concerned about ERT/alarm cases than non-ERT/alarm cases (5.4 vs. 3.8, $p < 0.0001$), and more concerned about non-ERT/alarm cases than non-ERT/no-alarm cases (3.8 vs. 1.6, $p < 0.0001$). There were no ERT/no-alarm cases. Details of this analysis are shown in Table III.

IV. DISCUSSION

Our results indicate that situated visual alarm displays mitigated the negative effects of imperfect alarming, allowing nurses to discern urgent from nonurgent events and outperform alarms in detecting and interpreting hazards. These results are promising not only for alarm displays, but represent a valuable strategy for representing the outputs of any opaque automation.

A. Discerning Hazardous From Nonhazardous Events

Nurses were able to discern the difference between ERT and all other cases, reporting a higher level of concern and urgency for ERT cases and a higher level of confidence for

non-ERT cases. They also effectively prioritized their response to these cases, reporting that they would respond urgently (e.g., immediately or as their next task) to the majority (85%) of the ERT cases and would prioritize other work (e.g., not responding or responding eventually) for the majority (85%) of non-ERT cases. Together, these suggest that nurses were able to quickly pick up signals from the dense data display, form and reform their clinical picture of the patient, and effectively assess the patients' stability to determine their clinical response. One notable example that illustrates this is that nurses responded urgently to abrupt changes in one patient's condition, but responded nonurgently to small, self-correcting changes in a second patient, even though both patients' condition had been fluctuating, and their current state (e.g., HR and absence of recent alarms) was similar. These observed benefits are consistent with the general benefits consistently found by the use of analogical displays designed to convey context through the use of empirically determined frames of reference [31], [32].

It was interesting that participant age and years of experience did not influence performance, and case difficulty did not always influence performance. This is promising in that it appears that situated alarm displays do not require extensive clinical or technology training to realize the inherent benefits. Intended case difficulty only affected performance for ERT, true alarm, and unnecessary alarm cases (44% of total cases). This is a warning to study designers that reducing case difficulty to a unidimensional measure likely obscures the true dimensions of complexity and difficulty.

B. Mitigating Negative Effects of Inaccurate Alarms

Although the operators' performance was far from perfect in correctly identifying true, false, and unnecessary alarms, situated alarm displays reduced the negative impact of inaccurate alarming in event detection and event interpretation. They also reduced (but did not remove) the alarms' influence in overestimating hazards. Our findings suggest that false alarms unnecessarily increased nurses' concern and urgency, but did not elevate either to the levels reported for ERT cases. Through the situated alarm displays, nurses were able to better detect and interpret alarmable events than alarms could on their own. This is notable as the literature suggests that nurse performance would have been equal or less than the alarm-only performance due to the high proportion of false and unnecessary alarms both in the experiment and in their operational settings [20]. These displays allowed the nurses to recalibrate quickly, if not immediately, in the midst of inaccurate alarming.

Although it is a laudable goal to improve an individual alarm's performance and strive for 100% reliability, binary alarm systems will never achieve this goal [45]. This makes these findings of improved human–machine performance with imperfect, opaque automation even more timely and important. Even though the nurses repeatedly asked for additional information in order to be more confident in their assessment (which the researchers did not provide), they were able to use these rudimentary situated alarm displays to improve their assessment of the alarms' fitness, leading to better patient assessment.

An initially surprising finding was that nurses were more concerned and responded more urgently to false alarms than to true or unnecessary alarms. However, the nurses' interview responses revealed that they had lower confidence in the false alarm cases, which resulted in high concern and higher urgency responses. False alarm cases, because they were often the result of imperfect physiological sensing because of patient movement or poor patient-sensor connection, often had abrupt changes in trends of the vital sign parameters due to how sensor artifact manifests within the displays. These displays were noted to be "busy" and "cluttered," which also contributed to reduced confidence. By contrast, true alarm cases had less abrupt changes, resulting in higher confidence and lower urgency. This heightened awareness, concern, and urgency toward patients that do not need it is a mark of overall system brittleness that is not fully addressed by this design, and deserves future consideration.

One additional finding from the nurse interviews shows how this joint human–machine system mitigated the negative effects of inaccurate alarms. Nurses reported an ability to "see past" the alarms in some cases to form a correct diagnosis of the patient based on their interpretation of the display, resulting in a well-calibrated response relative to the patient's actual status.

C. Limitations

This article has a number of limitations. The collected patient cases were all from a single medical center. Our results may not be generalizable to other institutions. However, the cases had alarms that were applicable to all nursing specialties tested in the article, arguably making them relevant to all acute care nurses. There is also the possibility of interviewer and response bias given the study design. However, anonymity was maintained to mitigate some concern of response bias. Two meetings were held among researchers to maintain a consistent study process with all participants.

Since participants were real practitioners from multiple units, their prior experiences likely influenced their responses. Prior articles have shown variability in safety culture among different units within the same hospital system that can lead to different ideas and responses to patient changes [46], [47]. There were not enough participants to make comparisons between nursing units. Further analysis will need to be done with a larger cohort of nurses to evaluate differences based on unit type.

Another limitation of this article is the absence of a control condition to compare to the situated alarm display performance. Future articles that directly measure performance with and without situated alarm displays could further validate and strengthen this article's findings, even though, as noted above, the alarm literature suggests that alarm-only performance is a valid if not slightly optimistic predictor of joint alarm-operator performance. Future articles could also explore the separate benefits of visually displaying alarms over time and alarms being situated with environmental data.

Finally, this article was conducted experimentally, not *in situ*. As such, there may be real-world dynamics that would affect its generalizability. We do not believe that the displays designed for this article were optimal, nor do we believe that we observed behaviors and decisions that would be perfectly replicated in

real-world settings. However, we believe that situated visual displays coupled with auditory alarms in real clinical settings may improve performance relative to our findings due to the complementary attentional directing qualities of the two sensory stimuli, the high bandwidth of vision, and the reduced cognitive burden of using multimodal cues in high-workload settings [48]. The general benefits of configuring complementary signals across multiple sensory modalities should be explored further.

V. Conclusion

Situated alarm displays supported operators' abilities to predict decompensation events and discern alarm accuracy. By incorporating visual trends, they were able to evaluate departure from norms and have more confidence in interpreting system status and changes. Our novel displays also succeeded in mitigating the negative effects of false alarms, which continue to plague automation in high-complexity, safety-critical fields.

Acknowledgment

The authors would like to thank the nurses and physicians at The Ohio State University Wexner Medical Center for their participation in this article. The authors would like to thank Todd Yamokoski, Jaclyn Buck, Morgan Reynolds, Vikrin Wang, Ryan Gifford, and Dane Morey for sharing their clinical perspective, providing assistance in recruiting participants, assisting with interviews, and assisting with data analysis. The authors' views do not necessarily represent the views of AHRQ.

References

[1] A. B. Arrieta *et al.*, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inform. Fusion*, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.

[2] D. D. Woods, "Essential characteristics of resilience," Ashgate Publishing Company, 2006, pp. 21–34. [Online]. Available: http://www.worldcat.org/title/resilience-engineering-concepts-and-precepts/oclc/901309999

[3] M. F. Rayo *et al.*, "The need for machine fitness assessment: Enabling joint human-machine performance in consumer health technologies," *Proc. Int. Symp. Hum. Factors Ergonom. Healthcare*, vol. 9, no. 1, pp. 40–42, 2020, doi: 10.1177/2327857920091041.

[4] M. Rayo, P. J. Smith, E. Roth, N. Sarter, K. L. Mosier, and C. A. Miller, "Making brittle technologies useful," *Proc. Hum. Factors Ergonom. Soc. Annu. Meeting*, vol. 61, no. 1, pp. 198–201, Oct. 2017, doi: 10.1177/1541931213601533.

[5] G. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich, "Ten challenges for making automation a team player in joint human-agent activity," *IEEE Intell. Syst.*, vol. 19, no. 6, pp. 91–95, Dec. 2004, doi: 10.1109/MIS.2004.74.

[6] R. D. Sorkin and D. D. Woods, "Systems with human monitors: A signal detection analysis," *Human-Comput. Interact.*, vol. 1, no. 1, pp. 49–75, 1985, doi: 10.1207/s15327051hci0101_2.

[7] C. Wickens and A. Colcombe, "Dual-task performance consequences of imperfect alerting associated with a cockpit display of traffic information," *Hum. Factors: J. Hum. Factors Ergonom. Soc.*, vol. 49, no. 5, pp. 839–850, 2007. [Online]. Available: http://hfs.sagepub.com/content/49/5/839.short

[8] C. D. Wickens and S. R. Dixon, "The benefits of imperfect diagnostic automation: A synthesis of the literature," *Theor. Issues Ergonom. Sci.*, vol. 8, no. 3, 2007, pp. 201–212, doi: 10.1080/14639220500370105.

[9] P. J. Smith, C. E. McCoy, and C. Layton, "Brittleness in the design of cooperative problem-solving systems: The effects on user performance," *IEEE Trans. Syst., Man, Cybern. - Part A: Syst. Humans*, vol. 27, no. 3, pp. 360–371, May 1997, doi: 10.1109/3468.568744.

[10] D. D. Woods, "Essentials of resilience, revisited," in *Handbook on Resilience of Socio-Technical Systems*, M. Ruth and S. Goessling-Reisemann, Eds. Cheltenham, U.K.: Elgar, 2019, pp. 52–65, doi: 10.4337/9781786439376.00009.

[11] D. D. Woods, "The theory of graceful extensibility: Basic rules that govern adaptive systems," *Environ. Syst. Decis.*, vol. 38, no. 4, pp. 433–457, 2018, doi: 10.1007/s10669-018-9708-3.

[12] D. L. Alderson and J. C. Doyle, "Contrasting views of complexity and their implications for network-centric infrastructures," *IEEE Trans. Syst. Man Cybern. Part Syst. Humans*, vol. 40, no. 4, pp. 839–852, Jul. 2010, doi: 10.1109/tsmca.2010.2048027.

[13] M. Logan and S. A. Colburn, *A Siren Call to Action: Priority Issues From the Medical Device Alarms Summit*. Washington, DC, USA: Association for the Advancement of Medical Instrumentation, 2011.

[14] M. F. Rayo and S. D. Moffatt-Bruce, "Alarm system management: Evidence-based guidance encouraging direct measurement of informativeness to improve alarm response," *BMJ Qual. Saf.*, vol. 24, no. 4, pp. 282–286, Apr. 2015, doi: 10.1136/bmjqs-2014-003373.

[15] D. D. Woods, "The alarm problem and directed attention in dynamic fault management," *Ergonomics*, vol. 38, no. 11, pp. 2371–2393, Nov. 1995, doi: 10.1080/00140139508925274.

[16] J. Meyer and Y. Bitan, "Why better operators receive worse warnings," *Hum. Factors*, vol. 44, pp. 343–353, 2002. [Online]. Available: http://hfs.sagepub.com/content/44/3/343.short

[17] M. N. Lees and J. D. Lee, "The influence of distraction and driving context on driver response to imperfect collision warning systems," *Ergonomics*, vol. 50, no. 8, pp. 1264–1286, Aug. 2007, doi: 10.1080/00140130701318749.

[18] J. Lee and K. See, "Trust in automation: Designing for appropriate reliance," *Hum. Factors*, vol. 46, no. 1, pp. 50–80, 2004, doi: 10.1518/hfes.46.1.50_30392.

[19] J. P. Bliss and M. C. Dunn, "Behavioural implications of alarm mistrust as a function of task workload," *Ergonomics*, vol. 43, no. 9, pp. 1283–1300, Sep. 2000, doi: 10.1080/001401300421743.

[20] J. P. Bliss, R. D. Gilson, and J. E. Deaton, "Human probability matching behaviour in response to alarms of varying reliability," *Ergonomics*, vol. 38, no. 11, pp. 2300–2312, 1995, doi: 10.1080/00140139508925269.

[21] M. Maltz and J. Meyer, "Use of warnings in an attentionally demanding detection task," *Hum. Factors: J. Hum. Factors Ergonom. Soc.*, vol. 43, pp. 217–226, 2001. [Online]. Available: http://hfs.sagepub.com/content/43/2/217.short

[22] M. Cvach, "Monitor alarm fatigue: An integrative review," *Biomed. Instrum. Technol.*, vol. 46, no. 4, pp. 268–277, 2012.

[23] S. L. Hyland *et al.*, "Early prediction of circulatory failure in the intensive care unit using machine learning," *Nature Med.*, vol. 26, no. 3, pp. 364–373, 2020, doi: 10.1038/s41591-020-0789-4.

[24] Q. Li and G. D. Clifford, "Signal quality and data fusion for false alarm reduction in the intensive care unit," *J. Electrocardiol.*, vol. 45, no. 6, pp. 596–603, 2012, doi: 10.1016/j.jelectrocard.2012.07.015.

[25] Q. Gui, X. Wang, B. Liu, Z. Jin, and Y. Chen, "Finding needles in a haystack: Reducing false alarm rate using telemedicine mobile cloud," in *Proc. IEEE Int. Conf. Healthcare Inform.*, 2013, pp. 541–544, doi: 10.1109/ichi.2013.84.

[26] W.-T. M. Au-Yeung, A. K. Sahani, E. M. Isselbacher, and A. A. Armoundas, "Reduction of false alarms in the intensive care unit using an optimized machine learning based approach," *NPJ Digit. Med.*, vol. 2, no. 1, 2019, Art. no. 86, doi: 10.1038/s41746-019-0160-7.

[27] C. R. Horwood, S. D. Moffatt-Bruce, M. Fitzgerald, and M. F. Rayo, "A qualitative analysis of clinical decompensation in the surgical patient: Perceptions of nurses and physicians," *Surgery*, vol. 164, no. 6, pp. 1311–1315, 2018, doi: 10.1016/j.surg.2018.06.006.

[28] C. R. Horwood, M. F. Rayo, M. Fitzgerald, E. A. Balkin, and S. D. Moffatt-Bruce, "Gaps between alarm capabilities and decision-making needs: An observational study of detecting patient decompensation," in *Proc. Int. Human Factors Healthcare Symp.*, vol. 7, pp. 112–116, Jun. 2018.

[29] R. J. Mumaw, E. M. Roth, and E. S. Patterson, "Lessons from the glass cockpit: Innovation in alarm systems to support cognitive work," *Biomed. Instrum. Techn.*, vol. 55, no. 1, pp. 29–40, 2021, doi: 10.2345/0899-8205-55.1.29.

[30] R. R. Hoffman, S. T. Mueller, and G. Klein, "Explaining explanation, part 2: Empirical foundations," *IEEE Intell. Syst.*, vol. 32, no. 4, pp. 78–86, Aug. 2017, doi: 10.1109/mis.2017.3121544.

[31] D. D. Woods, *Toward a Theoretical Base for Representation Design in the Computer Medium: Ecological Perception and Aiding Human Cognition*, J. M. Flach, Peter A Hancock, J Caird, and K. J. Vicente, Eds. Hillsdale, NJ, USA: Erlbaum, 1995.

[32] C. M. Burns and J. R. Hajdukiewicz, *Ecological Interface Design*. Boca Raton, FL, USA: CRC Press, 2004.

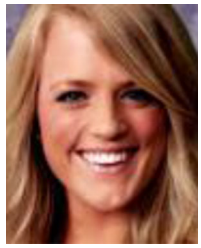
[33] C. Ware, *Information Visualization*. Morgan Kaufmann, 2004, pp. 187–226, doi: 10.1016/b978-155860819-1/50009-1.

[34] K. B. Bennett and J. M. Flach, "Graphical displays: Implications for divided attention, focused attention, and problem solving," *Hum. Factors: J. Hum. Factors Ergonom. Soc.*, vol. 34, no. 5, pp. 513–533, 1992, doi: 10.1177/001872089203400502.
[35] D. D. Woods and E. Hollnagel, *Joint Cognitive Systems*. Boca Raton, FL, USA: CRC Press, 2006. [Online]. Available: http://books.google.com/books?id=CzaG96osYSMC&printsec=frontcover&dq=joint+cognitive+systems&cd=1&source=gbs_api
[36] M. F. Rayo, N. Kowalczyk, B. W. Liston, E. B.-N. Sanders, S. White, and E. S. Patterson, "Comparing the effectiveness of alerts and dynamically annotated visualizations (DAVs) in improving clinical decision making," *Hum. Factors: J. Hum. Factors Ergonom. Soc.*, vol. 57, no. 6, pp. 1002–1014, Sep. 2015, doi: 10.1177/0018720815585666.
[37] X. Wu, M. She, Z. Li, F. Song, and W. Sang, "Effects of integrated designs of alarm and process information on diagnosis performance in digital nuclear power plants," *Ergonomics*, vol. 60, no. 12, pp. 1653–1666, Dec. 2017, doi: 10.1080/00140139.2017.1335884.
[38] J. J. Thomas and K. A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.
[39] K. Graham and M. Cvach, "Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms," *Amer. J. Crit. Care*, vol. 19, no. 1, pp. 28–34, Jan. 2010, doi: 10.4037/ajcc2010651.
[40] M. F. Rayo, J. Mansfield, D. Eiferman, T. Mignery, S. White, and S. D. Moffatt-Bruce, "Implementing an institution-wide quality improvement policy to ensure appropriate use of continuous cardiac monitoring: A mixed-methods retrospective data analysis and direct observation study," *BMJ Qual. Saf.*, vol. 25, pp. 796–802, Nov. 2015, doi: 10.1136/bmjqs-2015-004137.
[41] E. S. Patterson, E. M. Roth, and D. D. Woods, "Chapter 14: Facets of complexity in situated world," in *Macrocognition Metrics and Scenarios*, 1st ed. Farnham, U.K.: Ashgate Publishing Company, 2010.
[42] D. J. Simons and D. T. Levin, "Change blindness," *Trends Cogn. Sci.*, vol. 1, no. 7, pp. 261–267, 1997.
[43] S. H. Koch *et al.*, "Evaluation of the effect of information integration in displays for ICU nurses on situation awareness and task completion time: A prospective randomized controlled study," *Int. J. Med. Inform.*, vol. 82, no. 8, pp. 665–675, Aug. 2013, doi: 10.1016/j.ijmedinf.2012.10.002.
[44] P. M. Sanderson, M. O. Watson, and W. J. Russell, "Advanced patient monitoring displays: Tools for continuous informing," *Anesth. Analg.*, vol. 101, no. 1, pp. 161–168, Jul. 2005, doi: 10.1213/01.ane.0000154080.67496.ae.
[45] D. Manzey, N. Gérard, and R. Wiczorek, "Decision-making and response strategies in interaction with alarms: The impact of alarm reliability, availability of alarm validity information and workload," *Ergonomics*, vol. 57, no. 12, pp. 1833–1855, Sep. 2014, doi: 10.1080/00140139.2014.957732.
[46] D. T. Huang *et al.*, "Perceptions of safety culture vary across the intensive care units of a single institution∗," *Crit. Care Med.*, vol. 35, no. 1, pp. 165–176, 2007, doi: 10.1097/01.ccm.0000251505.76026.cf.
[47] R. F. Moody, D. J. Pesut, and C. F. Harrington, "Creating safety culture on nursing units," *J. Patient Saf.*, vol. 2, no. 4, pp. 198–206, 2006, doi: 10.1097/01.jps.0000242978.40424.24.
[48] C. D. Wickens, "Multiple resources and mental workload," *Hum. Factors: J. Hum. Factors Ergonom. Soc.*, vol. 50, no. 3, pp. 449–455, Jun. 2008, doi: 10.1518/001872008x288394.

Michael F. Rayo was born in Columbus, OH, USA, in 1975. He received both the B.A. degree in music performance and the B.S. degree in chemical engineering from Case Western Reserve University, Cleveland, OH, USA, in 1998, and the M.S. degree in industrial and systems engineering and the Ph.D. degree in health and rehabilitation sciences from The Ohio State University, Columbus, OH, USA, in 2007 and 2013, respectively.

From 2013 to 2016, he was a Research Scientist with the Cognitive Systems Engineering Laboratory. Since 2016, he has been an Assistant Professor with the Integrated Systems Engineering Department, The Ohio State University, Columbus, OH, USA. He has authored 18 articles, 2 book chapters, and 35 proceedings papers. His research interests include visual analytics for sensemaking, human–machine teaming, proactive safety, and societal scale engineering.

Dr. Rayo was the recipient of the IBM Faculty of the Year Award in 2019 and the Human Factors and Ergonomics Society Best Student Proceedings Paper in Healthcare Award in 2007. He is an Associate Editor of the journal *Human Factors in Healthcare*.



Chelsea R. Horwood was born in Cleveland, OH, USA, in 1989. She received the B.A. degree in biology and the Doctor of Medicine degree from Saint Louis University, St. Louis, MO, USA in 2011 and 2015, respectively.

She then completed the General Surgery residency training with The Ohio State Wexner Medical Center. During this time, she also obtained an MPH in clinical and translational science with The Ohio State University. She is currently undergoing her Fellowship training with the University of Colorado in Trauma, Acute Care, and Critical Care. To date, she has authored 14 articles, two proceedings papers, and two book chapters. Her research interests include outcomes-based care in Acute Care surgery patients and patient safety in healthcare.

Morgan C. Fitzgerald, photograph and biography not available at the time of publication.

Marisa R. Grayson was born in Pittsburgh, PA, USA, in 1993. She received the B.S. degree in systems engineering from the University of Virginia, Charlottesville, VA, USA, in 2016, and the M.S. degree in industrial and systems engineering from The Ohio State University, Columbus, OH, USA, in 2018.

From 2016 to 2018, she was a Research and Teaching Assistant with the Cognitive Systems Engineering Laboratory. Since 2018, she has been a Lead Cognitive Systems Engineer with Mile Two LLC, Dayton, OH, USA. She has authored two articles and five proceedings papers. Her research interests include human–machine teaming, resilience engineering, representation aiding, and design process fundamentals.

Grayson was a recipient of the HFES 2018 Healthcare App Design Competition.

Mahmoud Abdel-Rasoul, photograph and biography not available at the time of publication.

Susan D. Moffatt-Bruce was born in London, ON, Canada, in 1967. She received the B.Sc. degree in biochemistry from McGill University, Montreal, QC, Canada, in 1990, the M.D. degree from Dalhousie University, Halifax, NS, Canada, in 1994, the Ph.D. degree in immunology from the University of Cambridge, Cambridge, U.K., in 2001, and the MBA degree from the Ohio State University, Columbus, OH, USA, in 2015.

From 1994 to 2004, he was trained in General Surgery, Cardiothoracic Surgery, and Transplantation Surgery with Dalhousie University and Stanford University, Stanford, CA, USA. From 2006 to 2000, she was a Member of the Department of Surgery Faculty with the Ohio State University, Columbus, OH, USA. Additionally, between 2010 and 2017, she served as the Chief Quality Officer for the Ohio State University Wexner Medical Center and the Executive Director of the University Hospital between 2017 and 2020. In January 2002, he joined the Royal College of Physicians and Surgeons of Canada, Ottawa, ON, Canada and joined the Department of Surgery, University of Ottawa, ON, Canada. She has authored more than 180 articles and her ongoing funded research has been focused on quality and patient safety implementation science within healthcare.