Title
Analysis of human errors in human-autonomy collaboration in autonomous ships operations through shore control experimental data

Authors
Cheng, Tingting
Veitch, Erik A
Utne, Ingrid Bouwer
et al.

Publication Date
2024-06-01

DOI
10.1016/j.ress.2024.110080

Reliability Engineering and System Safety 246 (2024) 110080

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Analysis of human errors in human-autonomy collaboration in autonomous ships operations through shore control experimental data

Tingting Cheng [a,b,e], Erik A. Veitch [c], Ingrid Bouwer Utne [b,*], Marilia A. Ramos [d], Ali Mosleh [d], Ole Andreas Alsos [c], Bing Wu [a,f]

[a] State Key Laboratory of Maritime Technology and Safety (SKL MTS), Wuhan, China
[b] Department of Marine Technology, Norwegian University of Science and Technology, Norway
[c] Department of Design, Norwegian University of Science and Technology, Norway
[d] The B. John Garrick Institute for the Risk Sciences, University of California Los Angeles, USA
[e] School of Transportation and Logistics Engineering, Wuhan University of Technology, China
[f] Intelligent Transportation System Center, Wuhan University of Technology, Wuhan, China

ARTICLE INFO

Keywords:
Autonomous ships
Maritime safety
Virtual experiment
Human-autonomy collaboration
Human performance

ABSTRACT

Human-autonomy collaboration plays a pivotal role in the development of Maritime autonomous surface ships (MASS), as Shore control center (SCC) operators may engage in the control loop by directly operating the MASS, or, in the supervisory loop, monitoring the MASS and taking over control when needed. Thus, effective human performance during takeover control and operation is crucial for the safety of MASS operations. However, since the MASS is still in the early phase of development, the mechanism of human errors is unknown, and the data on human-autonomy collaborative operation is scarce. Human reliability analysis (HRA) aims to assess human errors qualitatively and quantitatively, and is widely used in various complex systems to help safety analysis. This study is dedicated to incorporating advanced HRA methods elements to identify and quantify human errors during taking over control and operation of a MASS in collision avoidance scenarios. It presents virtual experimental results, combined with theoretical human error identification and assessment methods. At first, we apply the Human-System Interaction in Autonomy (H-SIA) method to identify potential human errors; secondly, we identify relevant Performance Shaping Factors (PSFs) including Experience, Boredom, Task complexity, Available time and Pre-warning, and performance measures of the human errors, and implement them in the virtual experiment based on a full-scale autonomous ferry research vessel called milliAmpere2. Finally, we build a Bayesian Network (BN) to present causal and probabilistic relationships between PSFs and human errors through experimental data. The results show that available time has the highest impact on takeover performance of operators, followed by task complexity and pre-warning. Boredom does not present a significant sole impact unless combined with available time. Experience does not show a significant impact on human performance. In addition to the relevance of the human errors analysis to the safe development and operational design of MASS, the developed method benefits other human-autonomy collaborative systems. The developed BN model shows adaptability to assess human error probabilities, and the practical significance of integrating experimental data into the existing HRA methodologies for complex systems.

Abbreviations: ANOVA, Analysis of variance; AS, Autonomous system; BN, Bayesian network; CoTA, Concurrent task analysis; CPT, Conditional probability table; DAG, Directed acyclic graphs; DoA, Degree of autonomy; ESD, Event sequence diagram; HCI, Human-computer interface; HAC, Human-autonomy collaboration; HEP, Human error probability; H-SIA, Human-system interaction of autonomy; HRA, Human reliability analysis; mA2, milliAmpere2; MASS, Maritime autonomous surface ship; IDA, Information-decision-action; NDRT, Non-driving-related-task; NPP, Nuclear power plant; PSF, Performance shaping factor; PM, Performance measure.

* Corresponding author.
E-mail address: Ingrid.b.utne@ntnu.no (I.B. Utne).

https://doi.org/10.1016/j.ress.2024.110080
Received 1 September 2023; Received in revised form 27 February 2024; Accepted 13 March 2024
Available online 14 March 2024

1. Introduction

1.1. Development of MASS and HAC

With the development of increasing automation, information technology and artificial intelligence, the development of maritime autonomous surface ships (MASS) is emerging rapidly due to their potential to improve safety and efficiency. MASS is defined as several degrees of autonomy (DoA), considering its functional levels, decision location etc. [1]. Compared to the highest DoA - "Highly autonomous", in which MASS can navigate without human intervention, the DoAs involve "remote operation" are believed to be more feasible and realistic in current studies [2,3]. Hence, humans will still participate in MASS operation, and human errors may remain, even be in new forms [1,4]. To cope with this challenge, in the current early phase of MASS development, researchers have been dedicated to developing the assistance systems for human operators, as well as to carry out safety analysis serving to the human-oriented and safety-critical strategies for MASS design [5,6].

Through the experience on conventional ships, some common human errors and innate limitations were concerned when developing intelligent navigation assistance systems. For example, humans are usually hard to percept the targets with a long distance or poor visibility [7], and understand external operation conditions [8], ship motions under the impact of hydrometeorological conditions [9]; they are also hard to always choose the optimal timing and path to take evasive actions [10,11], etc. Accordingly, researchers have proposed many methods for intelligent assistance systems to complement the limitations of human. Xu et al. (2023) developed an image detection algorithm to augment humans' perception [12]. Zhang et al. proposed novel big data-driven methods to identify optimal timing of accident avoidance for humans taking evasive actions, and developed methods to quantify the impact of hydrometeorological conditions on the timing [9,13]. In their further studies, advanced machine learning methods were developed to predict ship motion trajectories considering ship dynamics effected by waves, wind, and currents [10,11]. These models are capable of capturing features of ship maneuvering reflecting hydrometeorological conditions. Moreover, a new assistance system for emergency operation based on accident consequences were proposed by Zhang et al (2022, 2023) [11,14], and an online risk-based decision assistance system was proposed by Johansen et al. (2023) [15]. These emerging assistance systems, as the advanced automations, help different cognitive activities of humans during ship operations [7–15]. The integration of these automated functionalities would further contribute to develop comprehensive autonomous systems (AS). AS is further defined as a system's ability of integrated sensing, perceiving, analyzing, communicating, planning, decision-making, and acting, to achieve its goals as assigned by its human operator(s) through designed human-computer interface (HCI). Hence, in MASS development introducing AS, new interaction modes which is between humans and AS occur [16], and are called Human-autonomy collaboration (HAC) in this paper.

In the conventional ships equipped by automated functionalities, humans are always in the control loop and take control authority, and the automation execute the commands from the humans. It can be deemed as a collaboration between humans and automation [3,17]. In the MASS, however, the AS is capable of finishing complete tasks without human intervention, and humans would be in the supervisory loop. That means the AS takes control authority, and the humans are in the Shore control centers (SCC) and only monitor the AS unless they were requested to takeover control. Once humans take over control, the AS "degrades" to an automation, which means the humans go back to the control loop [18,19]. Hence, during a MASS operation, humans may switch between these two loops by interacting with the AS, due to the rapidly changing environments or complex tasks. This switch brings additional complexity to the interaction between the humans and AS [17], and it is imperative to explore the new human errors and the failure mechanism.

This study aims to identify and quantify human errors of taking over control and operation of MASS in collision avoidance scenarios by incorporating advanced Human reliability analysis (HRA) methodology and the experimental data.

1.2. Human safety issues in MASS: human performance, influencing factors

Human errors in remote operation of MASS were investigated. Man et al., (2018) discovered the remote supervisory tasks may restrict human performance on situation awareness [19]. Ramos et al. (2019) explored human tasks when a MASS should avoid collision [20], and further proposed a Human-System Interaction in Autonomy (H-SIA) framework to model the tasks of human, autonomy, and their interaction [21]. These works contribute to identifying human errors of MASS operation in a collision avoidance scenario. Zhang et al. (2020) defined and assessed human errors in a MASS emeregency operation based on expert judgments [22]. Liu et al., (2021) analyzed human tasks of a remotely operated MASS and assessed human failure probabilities based on expert judgments [23].Our previous work modeled the HAC process of remotely operated MASS, and identified human errors in the cognitive level during the interaction between humans and AS [24].

Table 1 presents the human errors identified in current studies for MASS operation, along a comparison with the conventional vessels operation. Overall, in conventional vessels' operation, human errors were identified by searching from the historical data of accidents and incidents. The procedure that human operators should follow is given, and the automations that have been being used are relatively mature. Task allocation between humans and automations is clear, so the collaborative way of humans and automations can be deemed as "static". By contrast, in MASS operation, the functionalities of AS are varying from different DoAs, so, and the scenarios that AS can handle are different. Once the AS encounters a scenario it cannot handle, even the AS does not know that, humans have to take over control and perform recovery actions to prevent accidents. Hence, within a voyage period, in different scenarios that the AS can or cannot handle, the tasks of humans are unclear, and the collaborative ways of humans and AS are considered "dynamic".

In case of the AS knows itself cannot handle the scenario, it would alert and request humans for takeover control; however, in case of the AS does not know itself failed, humans have to perform additional tasks to observe the real time situations, in order to take over control timely. To this point, the human error in both case is failure of takeover control, but the mechanism behind them is different, and the relevant influencing factors need to be further explored. Table 1 presents the influencing factors effect on human errors in MASS operation in current studies. Among them, adverse HCI design, insufficient vigilance/boredom, tasks assigned for operators, experience, situation awareness, information overload, and trustness on automation were highlighted in current studies.

1.3. Human reliability analysis (HRA) methodologies and relevant experiment

To investigate what human errors may occur, and how the influencing factors may effect on human performance, the Human reliability analysis (HRA) descipline have been widely used in aid of the development of various control rooms in process industries, such as nuclear power plants (NPPs) [32], chemical plants [33], manufacturing plants [34], etc. The use of HRA encompasses modeling the human system interaction, identifying the possible human errors and its causes, and quantifying the human error probabilities. Applying HRA during systems design and operation phases allows for developing risk management strategies for avoiding human errors, leading to an overall safer operation [35].

Table 1
Human errors and influencing factors in conventional vessels and MASS operations

Human errors in Conventional vessels operation	Human errors in MASS operation	Influencing factors effect on human errors in MASS operation
Navigation errors; Supervision errors, Traffic monitoring errors; Voyage planning errors; Communication errors; Takeover errors [25]	Situation awareness[18]	Adverse HCI design [26]
Pilot monitoring failure; Pilot failure to remember adequate parameters for maneuver; Pilot failure to plan the control actions; Captain situation assessment failure; Helmsman control action execution failure[27]	Failure to check information, recognize objects, identify alert source, visualize information; Misreading, misundersdanding, not or incorrectly demanding information; Not/ untimely give commands; Give wrong commands; Commands given to wrong ship. The total is 22 human errors [20]	Negligence, insufficient vigilance, uncoordinated HCI, fatigue, information overload, unsufficient sense of responsibility, poor physical and mental conditions, automation-induced complacency, lack of experience, understanding, situation awareness, insufficient training [22]
Master failure to use engine control panel effectively; Master failure to detect the ship's heading by radar; Interpretation error of bridge team; Improper lookout; Inability to use echo sounder in restricted waters; Deviation from the safe route for demonstration purposes. The total is 32 types of human errors [28]	Inadequate supervision; Planned inappropriate operation; Failure to correct known problem; Supervisory violations; Skill-based errors; Rule-based mistakes; Knowledge-based mistakes; Routine violations; Exceptional violations. [17]	Information overload, situation awareness, skill degradation, boredom, fatigue[20]
Failure to push the button to spot the target vessel; Failure to turned the helm on the bridge to change course and make an evasive maneuver [29]	Untimely perception; Incorrect decision; Operation failure. [22]	Poor HCI design, Poor procedures and task allocation[30]
Not maintaining Continuous visual lookout; Inadequate use of aids Communication interrupt; Comparison error; Wrong judgment; Select wrong items; Wrong action; Wrong time, Improper implementation [31]	Information-processing phase: Do not percept the information of other ships correctly or timely; Do not percept the information of technical conditions correctly or timely; ... Decision-making phase: Do not estimate the risk of navigational situation correctly or timely; Do not evaluate effects of environmental conditions on autonomy feasibility correctly or timely; ... Action-taking phase: Do not send the command for changing operational mode correctly or timely; Do not send the maneuvering command correctly or timely; ... The total is 190 human errors[24]	Experience of conventional ship operation, and novel HCI [7] The needed time for situation awareness for taking over control [18]

HRA procedure

Current HRA methods prescribe the following main steps [36]: (i) problem definition, (ii) task analysis, (iii) human error identification, (iv) representation, (v) human error quantification, (vi) impact assessment, (vii) error reduction analysis, and (viii) documentation and quality assurance. Herein, the process of human-system interaction is illustrated through the first step, in aid of analyzing tasks and identifying human errors in the second and third steps; the identified human errors are integrated into the overall level of failure events of the system in the forth step; the factors that may influence human performance are defined as Performance Shaping Factors (PSFs) in the fifth step, and are used for assessing why certain failures could occur, and calculating the Human Error Probability (HEP); the impact of human errors on the overall risk and reduction measures are analyzed and documented in the final three steps. Many HRA methods were developed after serious accidents in NPPs that involved human errors. Some of them contain the above all steps, and some focus the qualitative or quantitative steps[37]. The core of HRA process is to understand how the operators interact with the system, so that the human errors can be identified, and the PSFs can be determined and used to assess the human errors.

However, as discussed in Section 1.1 and 1.2, the AS brings significant changes of the interaction mode between human and the system, and a direct control role of operators is replaced by a dynamic switching between the control role and the supervising role. In this process of dynamic switching, new human errors and PSFs may arise. E.g., humans monitor the AS operation as the supervising role, however, when the AS failed, humans did not recognize the AS failure and take over control timely, maybe due to their loss of vigilance, occupation by too much secondary tasks, overconfidence to the AS, etc. The current HRA methods have not considered the new characteristics of the HAC systems. The limitation is summarized as following: 1) most of current HRA methods have been developed in the context of NPPs operation and other process industries. However, considering the new interaction modes of HAC systems, such as MASS operation, there is very limited exploration in current HRA community. 2) some HRA methods were used in maritime field to identify and quantify human errors by leveraging historical accident data. Yet, the data in MASS operation is extremely scarce, and expert judgement is regarded as the main source [38]. Considering the characteristics of HAC systems and the data scarcity, new HRA methods need to be explored which would be adaptable to HAC systems and compatible to the data other than accidents, e.g., experimental data.

H-SIA and BN

For human error identification, the H-SIA was developed explicitly for human-autonomy collaborative MASS for identifying errors of human, AS, and interaction between them [21]. It presents the potential to combine quantitative analysis tools, such as BN, to contribute to qualitive and quantitative HRA. The H-SIA framework comprises two elements: (1) an Event Sequence Diagram (ESD), and (2) a Concurrent Task Analysis (CoTA). The interaction process between humans and AS along the time can be modeled by using ESD, in which a flowchart was provided with a series of designed questions. It allows the analyst to model different interaction events based on the AS functionalities in different DoA. Thereafter, the CoTA is developed from the ESD, and models the interactions between tasks performed by the operators and AS. The Information, Decision and Action (IDA) cognitive model is applied to be specific stop-rules for tasks' redescription, and finally derives basic tasks [39]. Since the H-SIA method is capable of modeling the tasks of humans and AS considering different DoA's functionalities, and also flexible to couple with other quantitative tools, e.g., fault trees, Bayesian networks (BN), it shows adaptability to develop HRA in HAC systems, especially MASS.

BN is a tool to qualitatively and quantitatively model causal and probabilistic relationships between variables, and provide a framework for reasoning about events, based on available information [40]. BN is widely used in risk and safety analysis, and shows its potential in HRA for various industries. BNs are annotated directed acyclic graphs (DAGs), comprising nodes, arcs, and conditional probability distributions. They are used to represent causal and probabilistic relationships among a set of random variables. In the general steps in HRA, BNs have been used to qualitatively analyze relationships among PSFs and human errors, and used to quantify human errors in NPPs field [41], maritime transportation [42], offshore emergency [43], and others. As for the challenge of data scarcity the HRAs face, especially in HAC systems, BNs show good adaptivity in handling this challenge by leveraging new information from various sources and updating its models. This paper proposes a new scheme combining H-SIA and BN for assessing human performance in HAC systems.

Experiment

The common and critical issue of data scarcity challenges HRAs in various industries, since the real experiment is high-cost and unrealistic [44]. Virtual environment is a computer-aided simulation environment where participants can gain artificial experience, including performing in hazardous scenarios, and empirical data can be collected [43]. Using simulator and experimental research is a good alternative to compensate for the weaknesses of sparse historical measurements and possibly biased expert judgment [32].

In NPP fields, [32] conducted experiment using full-scope simulator to observe the relationship between PSFs and human performance. In their further studies, quantitative analysis between PSFs and HEPs was investigated [45]. Musharraf et al. (2016, 2018) presented several studies for HRA by conducting virtual experiments in offshore emergency operation [43,44]. In chemical fields, virtual experiments were used for data collection to investigate human performance effected by PSFs [33], and demonstrated as a promising way to collect data for HRA in process industries.

With the rapid emergence and the unique characteristics of HAC systems, especially automated vehicles [46], unmanned aerial veihlcles [47], and MASS [24], using simulator and experimental research shows considerable potential to handle the issue of data scarcity.

1.4. New contributions

This study incorporates H-SIA, BN and the experimental data to identify and quantify the human errors in takeover control in MASS operation and investigate the relationship between human errors and PSFs. Specifically, its novelties and contributions include:

- Combination of H-SIA and BN with experimental data to propose a new HRA method for human error identification and quantification, which is adaptable to MASS operation.
- Utilization of experimental data to quantify the human errors by conducting experiment with human operators using the virtual version of milliAmpere2 (mA2), an autonomous ferry developed by Norwegian University of Science and Technology (NTNU) [48].
- Providing insights to carry out HRA research for MASS operation, and can be extended to other HAC systems' operation.

This study reveals the causal relationships between human errors in MASS operation and their PSFs – human experience, boredom, task complexity, available time and pre-warning, and quantifies the human errors using limited experimental data. The proposed approach contributes to reducing the challenge of scarce empirical data available in MASS safety research.

1.5. Outline

The paper is structured as follows: Section 2 elaborates the characteristics of the HAC and MASS operation. Section 3 presents the methodological approach. Section 4 presents the case study and results analysis. Section 5 discusses the method application, followed by the conclusions in Section 6.

2. Characteristics of human-autonomy collaboration (HAC) in MASS

2.1. Human roles in HAC systems comparing with in conventional systems

Conventional systems and HAC systems in this paper are differentiated by considering whether they involve autonomy or automation from the perspective of the capability of adaptation. While automation is a physical technology viable for application in a defined environment [49], autonomy can be defined as the ability of a system to integrate sensing, perceiving, analyzing, communicating, planning, decision-making, and acting to achieve its goals [50].

Fig. 1a shows a general form of conventional systems that involves automation but not autonomy. "Computer" represents the automation in the system, and is regarded as the "tool" of human. Humans send the command through the "controller" to the computer, which is transmitted from the computer to the controlled process through the "actuator". Hereto, this task has been finished. At this moment, the status of the controlled process changes, and the feedback is transmitted to the "computer" through the "sensor", and then is presented to humans through the "display". The computer, as an automation, always executes the commands of humans. In an "active control" role, humans are always aware of operational environments and system conditions and regularly provide control commands.

Fig. 1b shows the HAC systems described in this study. With the increase of DoAs, the autonomy is able to handle ongoing situations without humans' intervention, and is regarded as the "teammate" of human, which is represented by the "computer" in this figure. Humans are always in the supervisory loop rather than in the control loop, and monitor these situations. Therefore, the lines between "human operator" and "computer" are dotted instead of the solid in Fig. 1a. The computer and the task at the bottom constitute a closed control loop. Still, the human operator is in the supervisory loop. Humans can be considered a "backup" role. When the autonomy cannot handle hazardous situations, the human has to switch from the supervisory loop to the control loop, and continue operating until the hazards are eliminated, or the risk is mitigated to an acceptable level. The human is considered the ultimate safety barrier if autonomy fails. Whether the human can avoid accidents depends on whether they can timely take over control and correctly operate after takeover.

In Fig. 1c, the "computer" represents an autonomy which is capable of coping with all situations without human intervention, so, the situation is only displayed to the humans but does not need their intervention. This DoA, which is the highest, is outside of the scope of this study.

2.2. Interaction process of HAC in MASS

To clarify the HAC process, we consider the time period that are before and after humans take over control, as shown in Fig. 2. When the human is in the supervisory loop, the system is in the autonomous mode, and when the human takes control, the system switches to the manual mode. After takeover control, the human operator switches to the control loop from the supervisory loop.

Considering the definition of the high DoA [51], it is assumed that in the autonomous mode, operational tasks can be accomplished autonomously by the AS without intervention of the human operator. Human operators' tasks are limited to regularly monitoring the system's operational status and external environment within a certain time interval, as indicated in Fig. 2. Once a situation that the AS cannot handle occurs, the human operator has to take over control. After taking over control, the system enters into the manual mode, the AS execute commands of the human operator, and the human operator has to appropriately operate for avoiding hazards. Humans' operation performance, as well as takeover performance, may be affected by many issues.

In cases of potential hazards that the AS cannot handle, the AS may

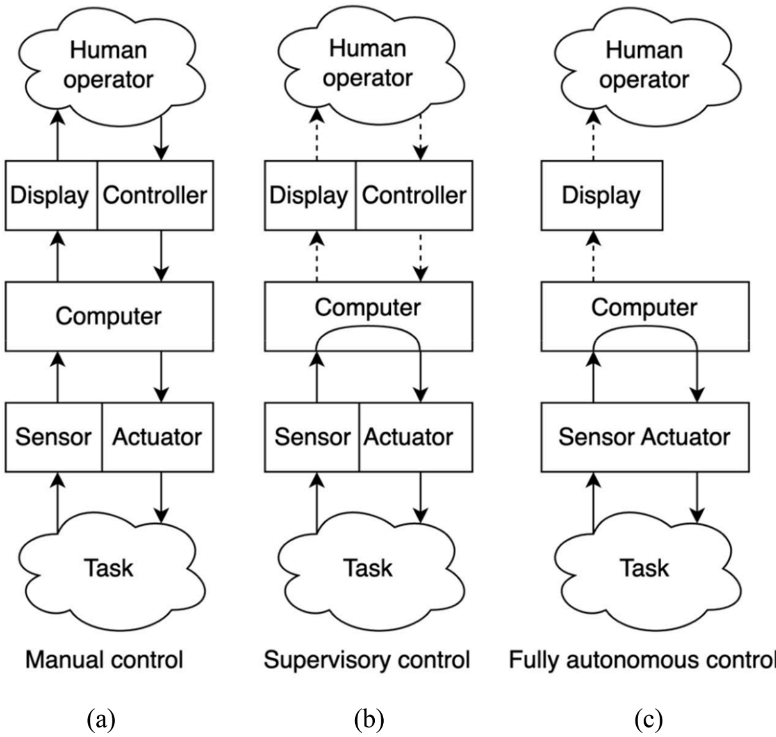


Fig.1. Comparison between conventional systems and human-autonomy collaborative system.

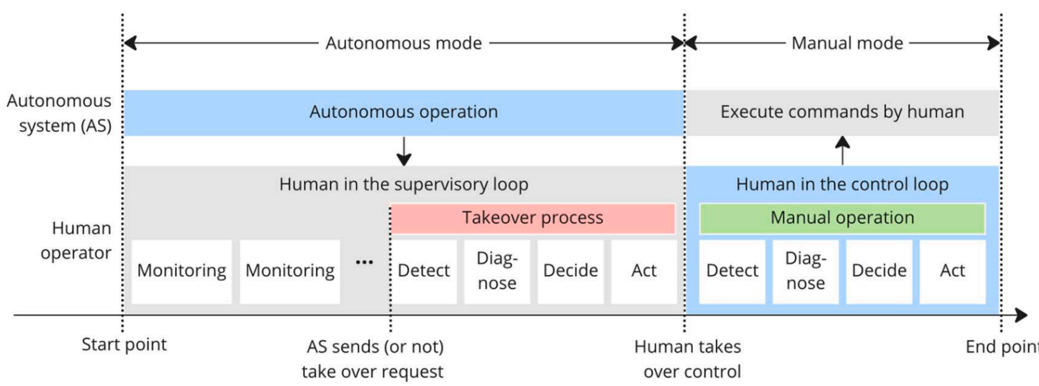


Fig. 2. Illustration of the human-autonomy collaboration before, during and after takeover control.

or may not send a takeover request to the human when the hazard occurs. Hence, when the AS cannot send a takeover request to the human, s/he have to proactively detect and diagnose this hazard, and take over control; when the AS is able to detect the hazard and send the request to th human, s/he may passively take over control. Whether and how the existence of a takeover request can affect human performance is significant for HAC safety [24].

Secondly, compared to conventional systems, the human in HAC systems who is in the supervisory loop for a long period may lose their situation awareness [52], and thus fail to detect hazards timely and take over control. Hence, whether and how the time that the human remains in the supervisory loop will affect their performance needs to be investigated.

Thirdly, in HAC systems, the human may receive more digitalized and visualized information [32], even may take responsibility of monitoring autonomy, compared with in conventional systems. During the takeover process as indicated in Fig. 2, the human must detect the incoming information, diagnose the ongoing situations, and decide whether to take control [46]. More complex tasks may have a higher chance of human errors, since when facing with a high-complexity task, humans employ a non-compensatory decision-making process, meaning they may miss some necessary information [53]. Hence, how the task complexity may effect human performance needs to be considered.

Forthly, since humans may not have to constantly monitor the HAC systems, instead, they may do unrelated work in their workstations [24], until they are requested for takeover control. If the time used to respond to hazardous situations is not long enough, the humans may not be able to finish the cognitive process and take over control timely. In this case, a time sufficiently long for humans' response could mitigate their stress, help diagnose the situation, and provide a solution [54]. This issue is also critical in conventional complex systems, e.g., process industries.

Fifthly, in HAC systems, especially MASS, some studies agree that humans should have essential maneuvering skills and seamanship; however, others believe that more new knowledges and experience, such as in HCI field etc., should be required for the humans in such systems [55]. Whether and how humans' experience may effect human performance is also a new issue in HAC for MASS.

In a sum, considering the forementioned characteristics of HAC in

MASS, it is critical to explore how these issues affect human performance and cause human errors in takeover control and manual operation. This work is beneficial to systematic safety analysis and management for MASS operation and helps the human-oriented design of MASS and its SCC.

3. Methodology

Fig. 3 shows the framework of the proposed approach. Phase 1 involves identifying human errors through the H-SIA method. This paper focuses on Phase 2 and Phase 3, which are detailed below.

3.1. Phase 1: Identify human errors by H-SIA

The H-SIA framework comprises two elements: (1) an ESD, and (2) a CoTA. The ESD presents what can happen in an HAC process, and the CoTA further details how these events occur. The first step is familiarization with the operational scenario (e.g., collision avoidance), the system, and its DoA; the second step is developing the ESD. H-SIA provides a flowchart for ESD development, where the questions are related to the DoA and the system design. Depending on the answers to the questions, specific events are added to the ESD as pivotal events. The possible outcomes of events may be binary, such as success and failure, but also may be several states (adequate, inadequate, failure).

The pivotal events can be transcribed into tasks to be performed by the human and the AS through the CoTA, which builds on Task Analysis theory to model the tasks different agents must perform and their interaction. The process of CoTA includes: (i) define agents to be analyzed (human or the AS). Each of them will have a hierarchy of task analysis; (ii) define top task the agents must perform, e.g., avoid collision when the ship is in the collision course; (iii) re-describe the top task, defining high-level tasks by transcribing the events of the ESD; (iv) identify parallel tasks that need to be executed all the time, e.g., monitoring, performing data collection from surroundings; (v) decompose tasks into basic tasks until they are associated with only one of the IDA phases and until the dependency with another agent's tasks is clearly modeled (specifics please see [21]). The basic tasks of the operator can eventually be identified and converted into human errors.

3.2. Phase 2: Collect data from experiments

In Phase 2, the human errors identified in Phase 1 are made measurable, and possible PSFs are determined and implemented by conducting experiments. It serves to estimate the causal and probabilistic relationships between human errors and PSFs.

Step 2.1: Define performance measures (PMs) – response variables

Performance measures (PMs) conception originates from psychology and is used to quantify psychological behavior in experiments [56]. This step aims to determine which PMs will be reasonable and feasible to measure the identified human errors in an experiment. Three aspects of PM are usually considered in experiment design [29,57]: (1) behavioral PMs. These involve observable behavioral indicators from the simulation participants undergo, such as the objective values in the simulator (e.g., distances, speeds, headings, accelerations, etcetera); (2) self-reported PMs. These involve questionnaires and interviews to learn participants' background information and some unobservable information, such as situation awareness and vigilance levels; (3) physiological PMs. These involve observable physiological indicators from human bodies (e.g., heart rate, pupil diameter etc.) to measure elements such as humans cognition and emotion,.

The PMs in this paper are defined based on the identified human errors from Phase 1, and refer to the literature related to human operational experiments involving conventional systems or other HAC systems, such as automated vehicles. The determined PMs are treated as response variables and made observable in the experiment design.

Step 2.2: Identify Performance Shaping Factors (PSFs) – independent variables

The relevant PSFs for a task depend on the conditions or circumstances in which an event occurs [44]. PSFs affecting human performance in a complex system are usually broadly categorized into external and internal groups [58]. The external group involves tasks and equipment characteristics in the system (e.g., equipment design, procedures, etc.); while the internal involves humans' characteristics resulting from internal influences (e.g., motivations, emotions, etc.), or from external influences (e.g., experience, knowledge, etc.) [59].

In efforts to develop and design an SCC of MASS, the most concerned PSFs that have been being discussed in existing studies should be considered and included in the experimental investigation. For example, external factors, such as whether task complexity or loads would influence operator performance [22], whether or how much in advance should the pre-warning be sent [60], etcetera; and internal factors, such as whether experience on maneuvering conventional ships [61] and vigilance of the operator would influence her/his performance[22], etcetera. The PSFs and their states in this paper are defined based on the characteristics of human limitation and the needs of MASS development. They are able to be manifested in the experiment to distinguish the contexts of human operating a MASS.

Step 2.3: Design experiment

In this step, at first, the operational scenario determined in Phase 1 is designed in the simulator, and the PSFs and their states determined in Step 2.2 are manifested in the operational scenario in different experimental sessions. Secondly, the PMs defined in Step 2.1 are considered the basis for developing the ports of the simulator to collect data in the experiment. The tasks of the human operator identified in Phase 1 are assigned to the participants in the experiment. The experimental procedure refers to [62], consisting of four stages for each participant: (1) Recruitment stage, in which participants are invited based on the experimental needs; (2) Preparation stage, in which researchers introduce to participants with the purpose, procedure of the experiment, and the data anonymity and confidentiality; (3) Training stage, in which researchers have the participants get familiarized with the simulator and operation; (4) Test stage, in which the participants finish the testing for the designed accident scenarios, and the data will be collected; (5) Exit stage, in which the participants finished the experiment, and will finish the interviews. Finally, the collected data are processed and used to next phase.

3.3. Phase 3: Quantify human errors by bayesian network (BN) method

BNs are annotated directed acyclic graphs, comprising nodes, arcs, and conditional probability distributions. They represent causal and probabilistic relationships among a set of random variables. The Bayesian formula given in Formula (1) serves as the theoretical basis of the BN and is used to describe the conditional probability inference between two variables.

$$P(A|B) = P(B|A)p(A)/P(B) \tag{1}$$

The BN is mostly dedicated to modeling the uncertainty in miscellaneous systems, which primarily encompass Bayesian inference problems. Bayesian inference problems involve conditional probability reasoning and can be divided into two distinct ways: forward reasoning and backward reasoning. Forward reasoning, a kind of assessing reasoning, entails updating the probability of response variables by transmitting new information about explanatory variables along the arc direction of the BN. Backward reasoning, also known as diagnostic reasoning, begins by determining the expected value (evidence) of the response variables. Subsequently, this value is placed in the BN, and information is transmitted in reverse to derive the value of the explanatory variables.

When a BN consists of n nodes, it is denoted as $\Delta = \{G(V, D), P\}$, where $G(V, D)$ represents an acyclic directed graph G containing n

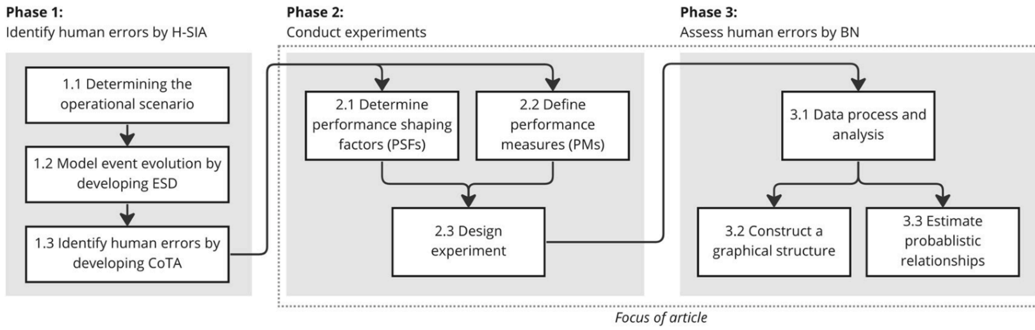


Fig. 3. Research framework.

nodes. The node variables are represented by the elements in the set $V = \{V1, ..., Vn\}$, and D indicates a set of directed links between pairs of the nodes. P shows the joint probability distribution over the set of node variables X_V, and $P(X_V)$ can be factorized as

$$P(X_V) = \prod_{v \in V} P\left(X_V | X_{pa(v)}\right) \tag{2}$$

where $X_{pa.v}$ denotes the set of parent variables of variable X_v for each node $v \in V$. In the BN inference process, suppose there is an event represented by $\beta = \{\beta_1, ..., \beta_n\}$ with n reference values, and we have the observed values $X = \{X_1, ..., X_n\}$, then based on the BN, we can determine the posterior probability distribution table of β using Eq. (3):

$$P(\beta | x_1, ..., x_n) = P(x_1, ..., x_n | \beta) P(\beta) / P(x_1, ..., x_n) \tag{3}$$

Accordingly, Fig. 4 presents a basic example of events A and B represented by nodes in BNs. Node A is deemed as the parent node of the node B - the child node, and affects the occurrence probability of event B. The arrow indicates the causal relationship between the events. Meanwhile, $P(B|A)$ implies the probabilistic dependency between these two events. When constructing a BN, each node may have child nodes and connect to them, but please note that the graph should be acyclic so that there is no closed-loop in a BN.

In a sum, two steps are required to develop a BN: (1) the graphical structure, and (2) the probabilistic relationships. BNs may be constructed by relying on expert knowledge, or be data-driven. Many studies use accident data to generate BNs [42]. However, as data from MASS operations are scarce, experimental data provide opportunities to model BNs and, hence, contribute to investigating PSFs impacts on human performance [41,43,44]. This paper develops the BN structure by leveraging experimental data. The developed BN is used to make forward and backward inferences in this study, to discover the human error probabilities in a given situation, and to investigate the PSFs' influence on human errors.

Step 3.1: Construct the graphical structure based on statistical analysis

To investigate the causal and probabilistic relationships between PSFs and human errors, the PSFs that are identified in Section 3.2 as response variables are considered as the root nodes in the BN, as shown in Fig. 5; the PMs that are defined as independent variables in Section 3.2 serve as the measurement of human errors, and are used to generate intermediate nodes in the BN. The arcs between PSFs and PMs are derived by using Analysis of variance (ANOVA), since in the case of limited experience and knowledge, applying the statistical methods to small dataset shows their potential to build BN graphical structures [41,45]. The kernel of ANOVA applications is if statistically significant relationships exist between two variables, the arcs are directed between these two nodes. It helps to model causal relationships in BNs [63,64]. In our study, ANOVA is used to model the directions from the cause - PSFs to the consequence - PMs.

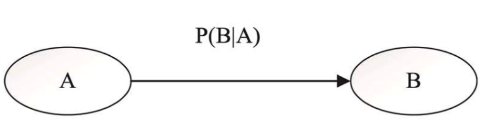


Fig. 4. Basic graphical representation of BN.

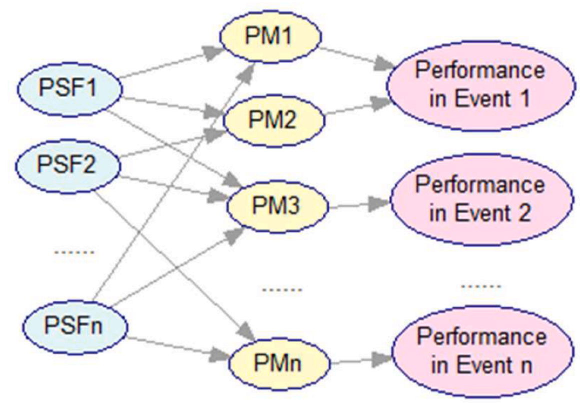


Fig.5. The connections between PSFs, PMs and top nodes.

Finally, the human errors that are identified from the events in Section 3.1 are considered as the top nodes in the BN structure. The arcs between PMs and human error nodes are based on which event in the ESD the PMs are involved, i.e., the event of takeover control and operation after takeover in this study.

Step 3.2: Estimate probabilistic relationships based on experiment data

Once the graphical structure is developed, the next step is to assign each node with a conditional or marginal probability table. These tables are developed based on the corresponding parent nodes' states and experimental data.

The number of discrete states for each node is finite in the BN. Therefore, the sum of the marginal probabilities on all states of the same node should equal 1. For a root node, each possible state is quantified with the marginal probabilities of its states. Nodes with one or more parents are quantified with conditional probability tables (CPTs). The CPTs contain values for every possible combination of states of the node and its parent nodes. For a binary node with n parents, the CPT will contain $2(n+1)/2$ columns. Each column in the CPT should sum to 1. Once the raw data is collcted, the probability required for CPTs are calculated. For instance, the probability of "$PM1 = state1$" in the situation $(PSF1 = state1, PSF2 = state2)$ can be calculated by using $P(PM1 = state1) = n/N$, in which n is the number of experimental trials where the $PM = state1$ with the situation $(PSF1 = state1, PSF2 = state2)$, N is the total number of trials in this situation. In this way, all the CPTs can be obtained.

4. Case study and results

The proposed methodology is applied with data collected from a virtual experiment with a simulator of the autonomous passenger ferry called milliAmpere2 - a real autonomous urban passenger ferry that was designed and developed by the Norwegian University of Science and Technology (NTNU) in 2019–2021. The ferry was commissioned and tested during a three-week operation period in 2022. The mA2 ferry is designed to cross a 100 m canal in Trondheim in Norway, and it is used for research and development.[1] The virtual version of the mA2 in the Gemini platform can be considered its "digital twin", as it shares its precise geometrical features and inertial characteristics [65], as shown in Fig. 6.

The case study in this paper focuses on a collision scenario, which involves a series of events modeled by ESD. The 32 participants in the experiment, of which 16 were experienced navigators, and 16 were gamers, were required to execute corresponding tasks identified by CoTA (cf phase 1), The order of the experimental sessions was randomized.

The data used in this paper, focusing on the risk of collision and HRA, was collected as part of a much larger experiment in the NTNU Shore Control Lab (see[55]).

4.1. Phase 1: Human errors identification by H-SIA

Step 1.1: Description of the operational scenario

The area of operation in the case study is an urban canal in Trondheim in Norway. The traffic mainly involves small boats with lengths not over 10 m. The responsibility of mA2 is crossing with passengers between the two shores. The moment of the hazard occurring is defined as when the target boat occurs on the collision course of mA2. The number of target boats in this study is determined as either one or two.

These characteristics of the operational scenario are used to model the ESD and were implemented in the virtual experiment.

Step 1.2: Model the ESD

According to the current development of mA2, the following assumptions were made to build the ESD:

(1) If there is one Target boat (TB) encountering the mA2 and on a collision course, mA2 is responsible for avoiding collision. If there are two or more TBs encountering mA2 and on collision course, the operator is responsible for taking over control and avoiding collision. Hence, the initial event (IE) is defined as "there is a TB (s) on the collision course of mA2". Then, two paths arise from the states of IE are one TB or multiple TBs, as shown in Fig. 7.

(2) The operator may or may not receive an alarm, depending on the design of the AS. Hence, a question is defined as "whether there is an alarm or not", and two paths arise from the states of this question.

(3) The possible end states involve a *collision, near miss*, and *safety*. The "*safety state*" represents the case where the collision has been avoided successfully, and the system functions normally. *Near miss* in this paper refers to situations when a collision has been avoided, but the mA2 has a problem with its maneuverability or control performance. In this case, even though mA2 avoided the collision, the technical systems will not operate safely. The likelihood of an accident occurring in the next journey increases.

The pivotal events of the ESD are identified as shown in Table 2.

Step 1.3: Model the CoTA

The CoTA is applied to human operator who remotely operates mA2. The top task is to avoid collision when two non-autonomous boats are on collision courses. Table 1 presents the agents (the operator or the AS) acting in each event. The high-level tasks of the operator are identified based on each event. The task analysis performed in this study leverages previous related works on autonomous ships[21], and the authors' extensive knowledge on ships operation, and control room design and operation in different industries. Fig. 8 presents a simplified CoTA for the operator, and Table 3 further describes these tasks. The parallel task of the operator is to keep monitoring screens. Most tasks are re-described by decomposing the high-level tasks until reaching the stop-rule described in Section 3.1. For simplicity, Task 2.1 is not decomposed into understanding, assessing the alarm, etc. The interface tasks are identified as shown in Fig. 8 by using circles. Possible human errors are obtained by converting the basic tasks as shown in Table 3.

4.2. Phase 2: Experimental data

Step 2.1: Define PMs related to human errors

The identified human errors in step 1.3 need to be measured in the experiment, as described in Table 3. Please note that tasks T1.1 and T2.1 are assumed to be successfully accomplished, because in the experiment, we made sure participants can visualize and listen to the takeover request physiologically, and they were required to keep monitoring the screens. Hence, the corresponding E1.1 and E2.1 are excluded from the experiment.

Step 2.2: Determine the PSFs and implementation in the experiment

Considering the characteristics of HAC in MASS, five PSFs were tested during the experiment: experience of operators, boredom, task complexity, available time, and pre-warning. some of these PSFs are included in existing HRA methodologies, such as THERP [36], Spar-H [66], and Petro-HRA [67]. In the following analysis, the description of the PSFs is based on the existing issues illustrated in Section 2, and influenced by the Petro-HRA guideline [67], and related to [55].

Operational views were developed in the experiment as Fig. 9. Fig. 9(a) includes three operational views of Figs. 9(b), 9(c) and 9(d). Fig. 9(b) presents the whole traffic situation in the Nidelva canal. The pink marks represent conventional boats, and the green marks represent the autonomous ferries that are under operator supervision. Fig. 9(c) presents the zoomed-in situations around each ferry at the same time. It also shows the heading of each ferry, and the ports of departure and destination. Fig. 9(d) presents the main view for operator maneuvering. The top left presents the speed of the ferry in real time, the left bottom presents incoming notifications from the AS about the ferry status, such as it is (un)docking, whether there is a boat entering into its detection range, etc.

PSF1: Experience

Experience is defined as how often the operator has experienced the tasks or scenario [67]. The outcome of experience is knowledge and skills. As illustrated in Section 2, since industries have not widely adopted MASS, seafarers do not have extensive experience in maneuvering a MASS. However, whether gamers, who have experience playing computer games and thus with simulated environment, can perform well when maneuvering a MASS remotely would be an interesting issue. Hence, we define two levels for experience as *Seafarer* and *Gamer*. Seafarers are individuals with a navigation license and who are employed as operators aboard ferries operating in coastal Norway. Gamers are individuals who regularly play desktop computer games or console video games.

PSF2: Boredom

Boredom is defined as "an unpleasant, temporary affective state resulting in a human's lack of interest for a specific current activity" by [68]. Boredom can negatively impact human performance. For example, too long waiting time may cause the operator lose their situation awareness and vigilance [69]. Based on studies in psychological area [70], two levels; *5 min* and *30 min*, are selected to represent two levels of boredom [71]. It states that humans would keep vigilant within 5 min,

[1] For more information about the milliAmpere2 urban autonomous ferry, see https://www.ntnu.edu/autoferry

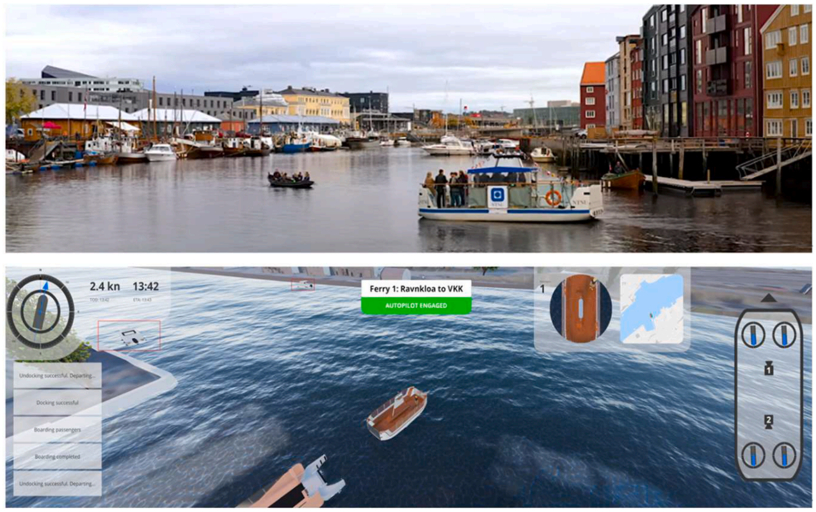


Fig. 6. MilliAmpere2 and its virtual version in Gemini platform.

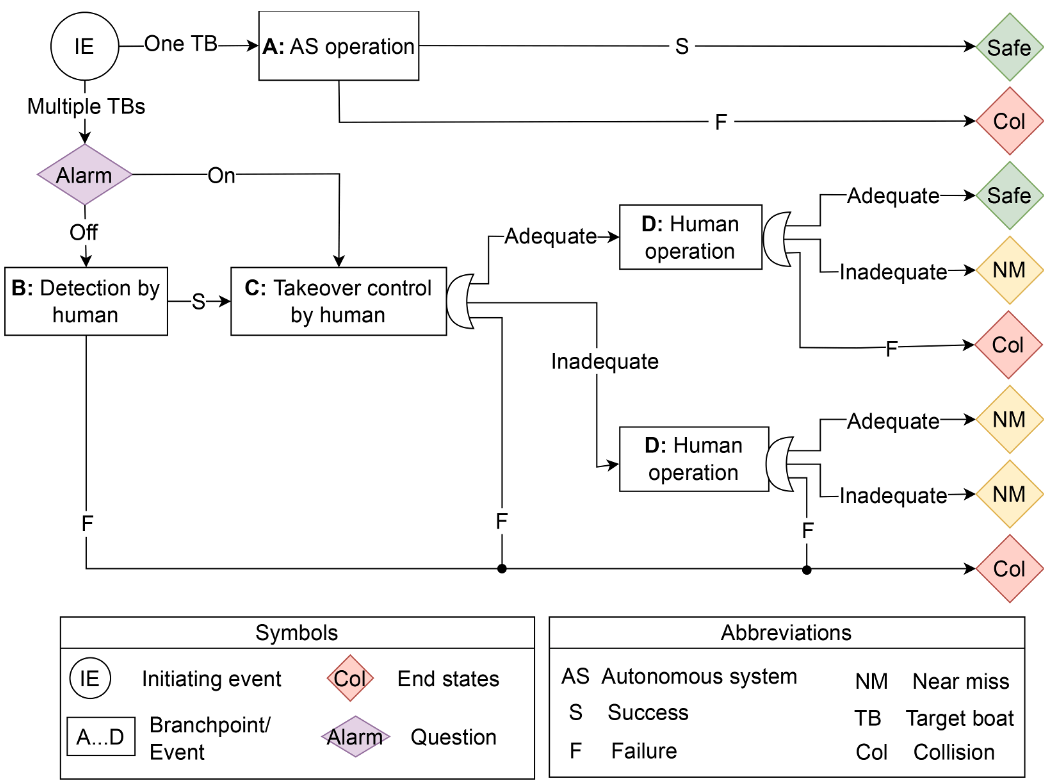


Fig. 7. Event sequence diagram of avoiding collision by human and AS collaboration.

Table 2
Pivotal events in the ESD

Pivotal events	Explanation
Initiating event (IE)	There is one TB or two on collision course.
A	AS detects and operates mA2 for avoiding collision with the TB.
B	Operator detects two TBs.
C. Outcomes: adequate, inadequate, failure	Operator takes over control mA2.
D. Outcomes: adequate, inadequate, failure	Operator maneuvers mA2 to avoid collision with the two TBs.

but they probably lose vigilance after 20–30 min. In the experiment, we implement these two levels by using the amount of time from the participant starts to monitor the screens in an autonomous mode to the two target boats lead to multi-boats collision avoidance scenario.

PSF3: Task complexity

Task complexity refers to how difficult the task is to perform in the given context, involving the complexity of goal, size, step, and others [67]. With the concerns mentioned in Section 2.2, we design two levels of Task complexity: *nominal* and *complex*. *Nominal* involves the operator being responsible for supervising and steering one autonomous ferry and monitoring light traffic with less than ten boats. *Complex* involves supervising three autonomous ferries, while monitoring heavy traffic with fifteen boats. When the operator is responsible for three ferries, s/he has

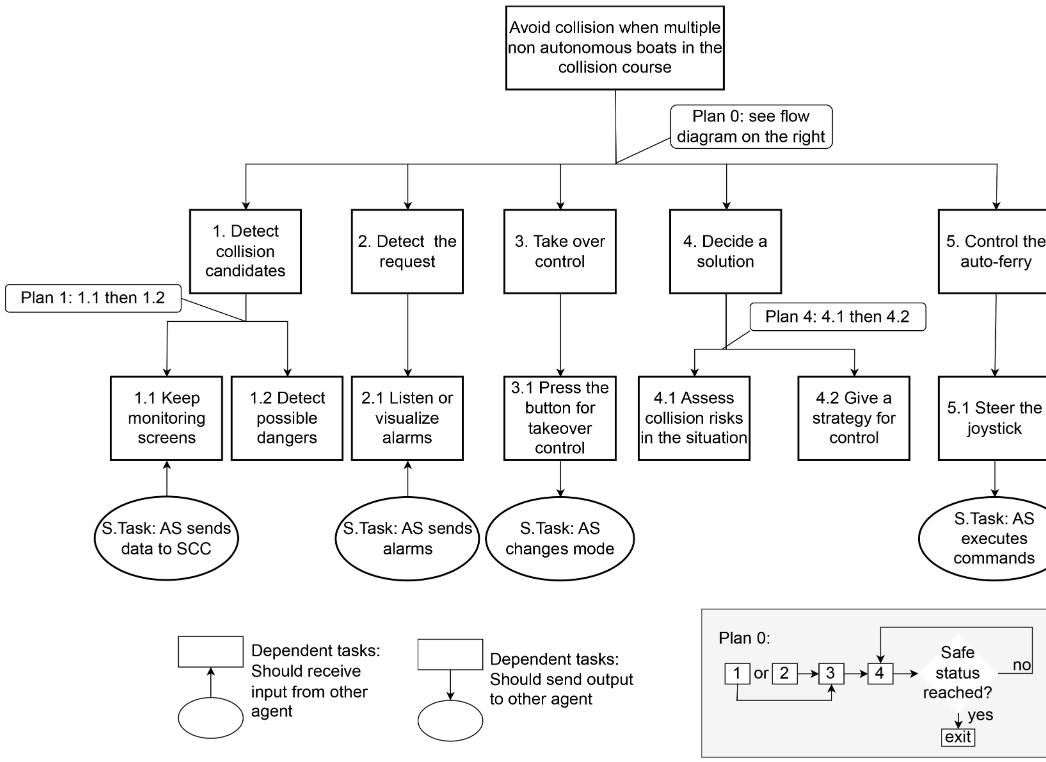


Fig. 8. Concurrent task analysis for the operator to avoid collision.

to switch the main view for monitoring among three ferries (Fig. 9(d)) and diagnose for each ferry whether there is a hazardous situation (Fig. 9 (c)) and which one should be taken over control at first, and maneuver it if needed. Meanwhile, the operator has to monitor the congested traffic in the canal (Fig. 9(b)). However, in the situation of one ferry, the operator can keep staying in the main view and focus on one ferry.

PSF4: Available time

Available time represents the amount of time an operator has for detection, diagnosis, and action upon an abnormal event [67]. Since available time is critical for human operation based on the characteristics of HAC systems, in this paper, it is defined as for takeover behavior, which means the amount of time available from when the hazard occurs (i.e., two boats occur in the collision course of mA2) to when the collision is about to happen. The collision will happen if the operator does not take over control within the available time. Based on pre-testing on volunteers, two levels of available time are assumed, i.e., 20 s and 60 s in the experiment.

PSF5: Pre-warning

The pre-warning request is based on the system's assessment of whether the situation is over the operational limits the system can handle [52,72]. Considering the characteristics of HAC systems mentioned in Section 2.2 that the AS may or may not send a takeover request to the operator when it cannot handle the hazardous situation, these two states are implemented in the experiment with two states: *On* and *Off*. It means the pre-warning will or will not be sent when a multiple-encountering scenario occurs. The second yellow box from the top down shown in Fig. 9(d) represents the takeover request.

Step 2.3: Experimental protocol

The details of the experiment include the following:

- *Subjects*

Thirty-two participants including five females and twenty-seven males were recruited between 18 and 65 years of age. Herein, The experienced navigators are individuals with a valid navigation license in Norway, as defined by the Norwegian Maritime Authority. The number of navigators holding the certificates is shown as following: Class 1: 11; Class 2: 1; Class 3: 1; Class 4: 2; Class 5: 1. The certificates define their skills in terms of allowed vessel tonnage and what title they can have. For example, for Class 1 the individual has no tonnage limit and they can have the title of office of the watch, chief mate, or master [73]. Sixteen gamers were individuals who regularly play desktop computer games or console video games.

- *Devices*

The virtual version of the mA2 in the Gemini platform was used to design the collision scenario between autonomous ferry(s) and target boats. A joystick was provided to participants to control the course and speed of the ferry, as shown in Fig. 10(a). The camera buttons were adjustable for observing passengers on the ferry and are not further discussed in this study. A control panel was provided for changing the operational modes (manual or autonomous control), as shown in Fig. 10 (b). The buttons on the first and second lines were used to adjust the status of batteries and cameras. These are not discussed in this study.

- *Collision Scenario*

The collision avoidance between the autonomous ferry and two target boats is designed as the testing scenario. Each test starts with the ferry(s) crossing the canal back and forth in autonomous mode. The virtual milliAmpere2 ferry is 8.4 m long, and operates at a speed of 3 knots. Other boats in the scenario are 7 m long, and sail both directions along the canal at around 5 knots.

According to the functions of the real mA2, the collision scenario was designed as: when TB1 is 40 m away from the ferry, which is about three times the sum of lengths of mA2 and TB1, mA2 can autonomously detect TB1 and gradually slow to a full stop to let it pass. However, in this case, mA2 enters a collision course with TB2, which is approaching. If the operator does not detect the hazard, intervene and take preventative action, TB2 will collide with the mA2 at speed.

The position that TB2 was leaving depends on the level of PSF4:

Table 3
Basic tasks, human errors, performance measures and data sources of the operator

Basic tasks and descriptions	Human errors	Performance measures	Data sources
T1.1: Keep monitoring screens. *Checking information on the screens. Depends on AS sending data to SCC.*	**E1.1:** Operator do not monitor screens regularly.	-	-
T1.2: Detect possible dangers. *Detect possible TBs are on collision course.*	**E1.2:** Operator do not detect possible dangers timely and correctly.	**PM1.2:** Question on "Whether they noticed the possible danger, and please describe it"	Interview data
T2.1: Listen or visualize alarms. *Depends on AS sending alarms to SCC.*	**E2.1:** Operator does not listen or visualize alarms timely.	-	-
T3.1: Press the button for takeover control. *Leads to AS changes the operational mode.*	**E3.1:** Operator does not press the button for takeover control timely.	**PM3.1:** Status and time of pressing the button for takeover control	Simulator data
T4.1: Assess collision risks in the situation	**E4.1:** Operator does not assess the collision risks in the situation timely and correctly.	**PM4.1.1, PM4.1.2:** Distances between mA2 and TB1 (TB2) when takeover control"	Simulator data
T4.2: Give a strategy for control.	**E4.2:** Operator does not give a proper strategy for avoiding collision.	**PM4.2:** Question on "whether s/he have a strategy for avoiding collision and execute it"	Interview data
T5.1: Steer the joystick. *Leads to AS executing the commands of the operator.*	**E5.1:** Operator does not steer the joystick properly.	**PM5.1.1, PM5.1.2:** The shortest distances between mA2 and TB1 (TB2) after takeover control	Simulator data

Available time. The distance from this position TB2 was leaving to the position of potential collision between TB2 and mA2 equals the available time multiples the speed of TB2. Thus, two positions that TB2 was leaving corresponding to two levels of available time were designed in the scenario. The levels of PSF3: Task complexity and PSF5: Pre-warning were designed in the scenario and shown as Fig. 9(b), 9(c) and 9(d). The levels of PSF1: experience and PSF2: boredom were arranged to different testing trials.

- Experiment protocol

A split-plot experimental design was used to structure the factor treatments and randomize test trial order. Compared to the completely randomized experimental designs, it can reduce costs and improve precision and efficiency in the estimates of effects of factors [62].

- Data collection and transformation

The collected data include simulator log files with sampling frequency 5 Hz, background questionnaire, interviews, and camera recordings. The *pandas* and *matplotlib.pyplot* in the Python packages were used to calculate, classify, and visualize the simulator data. Table 4 presents the categories of the collected data.

4.3. Phase 3: Eliciting BN model

4.3.1. Generate the graphical structure

The BN structure comprises five PSFs as root nodes, and the human operators' performances of different events in the ESD as top nodes. PMs are integrated as intermediate nodes as below.

PM 1.2 is defined as the node "Detect dangers in advance" with two states: *yes* and *no*. It is fed by using participants' self-reported data from the interview.

PM3.1 is expressed as the node "Takeover time" to describe the duration from when the hazard occurs to when the operator takes over control. It was defined as two states: *timely* and *untimely*. Along the elapsed time, by checking the positions, speeds and headings of each boat, at first, we recognized "the time when the collision candidate was unberthing" by using the speeds and heading data; secondly, we subtracted it from "the time of switching auto to be manual"; finally, we can obtain the Takeover time by calculating the duration from the collision candidate unberthing to the participants take over control. If the Takeover time is more than available time, the collision occurred and the state of Takeover time is "untimely"; if the takeover time was less than available time, the collision did not occur, and the state of Takeover time is "timely".

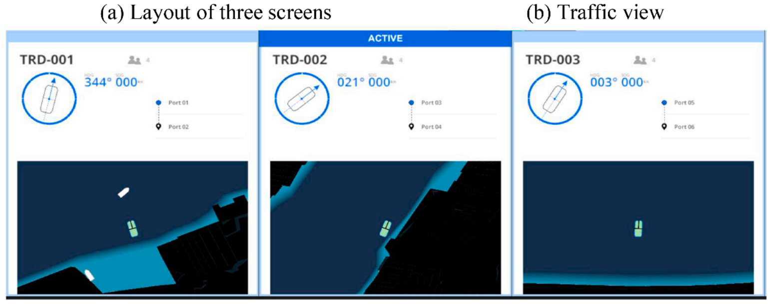
PM 4.1.1 and PM 4.1.2 are integrated into the node "Takeover quality". It is a term commonly used in the field of automated vehicles [74]. The distances between the automated vehicle and the encountering objects can be used to describe this quality. In this study, Takeover quality is defined as two states: *adequate* and *inadequate*. Based on the navigational experience and previous studies [75,76], the sum of lengths of two encountering boats can be considered as a minimum safe distance for collision avoidance. So, we define that when the operator takes over control, if the distances between mA2 and two TBs are both over the minimum safe distance, the takeover quality is deemed as adequate. Otherwise, it is inadequate. The distance when takeover control was calculated by using the positions data at the moment of takeover control, and this moment of takeover control is the timepoint of "switching auto to be manual", as shown in Table 3.

PM 4.2 is represented by the node "Navigational strategy" with two states: *yes* and *no*. This refers to whether the operator had a strategy for avoiding the collision and followed it to execute the avoidance maneuver. It is assumed that once the operator had a strategy for avoiding collision, their would execute it. This variable is fed by using participants' self-reported data from the interview.

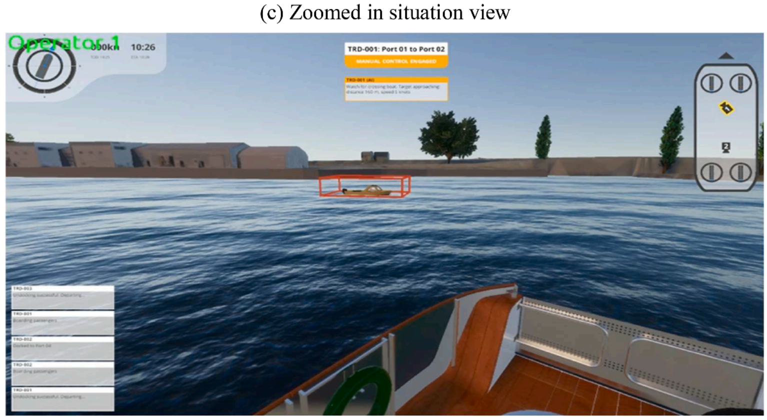
PM 5.1.1 and PM 5.1.2 are integrated into the node "Navigational situation during manual operation" with three states: *good, moderate*, and *bad*. The distances between encountered vessels can be used to determine the benchmark for navigational situations [76,77]. During manual operation by the operator, if the shortest distances between the ferry and two TBs are both more than the sum of lengths of the ferry and TBs, this situation is considered as a good operation. If one of the shortest distances is less than the benchmark, it is considered as a moderate situation. If both the shortest distances are less than the benchmark, it is a bad situation. In the similar way with calculating PM4.1.1 and 4.1.2, We obtained the distances by using position data. Futhermore, in the duration after takeover control, we screened out the shortest distance between the ferry and TBs from the distances at over several thousands timepoints by programming a minimum function.

The ANOVA was used to investigate the relationship between each intermediate node with the five PSFs, and conducted by using the Design-expert software[78]. When the ANOVA results indicated a significant relationship between the PM and the PSF, an arrow was directed from the PSF to the relevant intermediate node. The *F*-test was used to calculate the *p*-value between each PSF and PM. A *p*-value less than 0.05 indicates there is a significant relation between the PSF and the PM. For example, the results of ANOVA between Takeover time and five PSFs are shown in Table 5. It presents available time, task complexity and pre-warning have the significant effect on Takeover time. In this way,

(a) Layout of three screens (b) Traffic view



(c) Zoomed in situation view



(d) Main view for operator maneuvering

Fig. 9. Layout and three operational views in the shore control room.

(a) Joystick (b) Control panel

Fig. 10. Joystick and the control panel for steering autonomous ferry(s) in the experiment (c.f. Hanssen (2022)).

the results of each PM were obtained and are summarized in Table 6. The BN graphical structure which then is derived is shown in Fig. 11.

4.3.2. Estimate CPTs

Based on the collected data, the probabilities required for the CPTs are calculated using probability theory. For instance, the probability of Takeover time being less than the benchmark time in the situation (task complexity = nominal, pre-warning = on, available time = 20 s) can be calculated using Eq. (4).

$$\mathrm{P}\ (\text{takeover time} < \text{benchmark time}) = \frac{\mathrm{n}}{\mathrm{N}} \tag{4}$$

Herein, n is the number of trials where the takeover time is less than benchmark time with the situation (task complexity = nominal, pre-warning = on, available time = 20 s). *N* is the total number of trials in this situation.

Table 4
Information of the data collected from experiment

Data types	Data	Description	Source
Behavioral data	Output signal from simulator	Elapsed time: /s	Simulator
		Vessels' coordinates relative to the initial position (x,y,z)	
		Vessels' speed m/s	
		Vessels' heading °	
		Whether is (un) berthing. (y/n)	
		Whether the passengers are (dis) embarking. (y/n)	
		Whether the collision (allision) happened. (y/n)	
	Input signal from participants	Switch between auto and manual	Control panel
		Switch between main monitoring views of operator	
		Switch between views of the cameras onboard	
		Drive-reverse oepration	Joystick
		Steering operation	
		(Un)berthing operation	
	Logfile of collision accident	Collision object: ID of target boats or ports	Simulator
		Collision time (s)	
		Speed of vessels' when collision (m/s)	
Information of participants	Interview	The designed questions were asked by researchers.	Recorded by wearable micphone
	Demographic info	Years of experience, gender, age, etc.	Questionnaire
	Video	Video of the process of participants operating (front and back views)	CCTV

Table 5
ANOVA results of Takeover time

PSFs	P-value	Significance
Experience	0.9937	-
Boredom	0.1561	-
Task complexity	0.0353	+
Available time	<0.0001	+
Pre-warning	0.0072	+

Table 6
ANOVA results summarization

PSFs Effects on:	Experience	Boredom	Task complexity	Available time	Pre-warning
Detect dangers in advance	-	-	-	-	-
Takeover time	-	-	+	+	+
Takeover quality	-	+	-	+	-
Navigational strategy	-	+	-	+	+
Navigational situation during manual operation	-	+	-	+	-

CPTs for all situations are figured out in the same way shown in Eq. (4). Two CPT examples, Takeover time and Navigational situation during manual operation, are shown in Table 7 and 8.

The CPT of Takeover performance can be derived directly. When Takeover time is "untimely", Takeover performance must fail; When Takeover time is "timely", the state of Takeover performance is the same as that of Takeover quality.

The CPT of Operation performance is estimated depending on the collected data, as shown in Table 9. Only when the situation was good and the operator had and executed the navigational strategy, the operation performance was considered as adequate. Once the collisions happened after takeover control, it was considered operation performance is failure. It should be mentioned that this operational phase excludes ten sessions where the collision happened before takeover. In addition, there are three sessions where the collision happened after takeover. All CPTs are integrated into the BN graphical structure, and Fig. 12 shows the final BN model. Other CPTs please see Appendix.

4.4. *Results and sensitivity analysis*

4.4.1. *Examples of BN results*

Based on Eq. (3) and Fig. 12, the probability of human performance and PSFs can be inferred. Take the Fig. 13 as an example of forward inference, when setting evidence *Complex* to the node *Task complexity*, the probability of *takeover failure* becomes 0.31, and *operation failure* becomes 0.15. The full occurrence probability of performance nodes related to each PSF can be calculated in this way.

Moreover, take the Fig. 14 as an example of backward inference, when setting evidence *Failure* to the nodes *Takeover performance* and *Operation performance,* it is observed that the probability of Available time = 20 s becomes 1; the probability of Task complexity = Complex and Pre-warning = Off increase as 0.85 and 0.75, respectively; the probability of Boredom = 5 min becomes 0.55.

Due to the limited space, we do not present all the results through forward and backward inference in this study, but the human error probabilities and impact of PSFs in human performance are discussed in the following subsections. The limitation of calculation on human error probabilities due to the small dataset in this study will be discussed in Section 5.

4.4.2. *Sensitivity analysis*

Sensitivity analysis is often used to check whether the BN model is robust, and the results are reasonable, and also to determine the degree of effects of the input parent node on the output child node [79].

To investigate the effects of different PSFs on the performance nodes, we treat performance nodes as the target nodes in a sensitivity analysis by changing the occurrence probability of each PSF, and observe the response of the target nodes. This can be done by using GeNIe to generate Tornado plots. Tornado plots visualize the range of outputs expected from a variety of inputs, or alternatively, the sensitivity of the output to the range of inputs. Due to the limited space, three examples of the results are presented in Figs. 15, 16, and 17. The green bar means positive effects, and the red bar means negative effects.

Fig. 15 illustrates the results of "Takeover performance is adequate". We select the range of parameter spread from 0 to 1, which means the probability of relevant PSFs can be adjusted to be 0 and 100 %. It is observed that when setting a probability of "Available time = 20 s" to be 0 % (keeping the other PSFs constant), and the probability of "Takeover performance = adequate" is 0.6875; When setting 100 % probability to "Available time = 20 s", the probability of "Takeover performance = adequate" is 0.195313. The influence ranking of the PSFs on Adequate takeover performance is: Available time > Boredom > Task complexity > Pre-warning. Moreover, we observed when the available time is certain, takeover quality has more significant effects on takeover performance when the boredom level is 5 min than 30 min.

Fig. 16 presents another example of the results of "Takeover

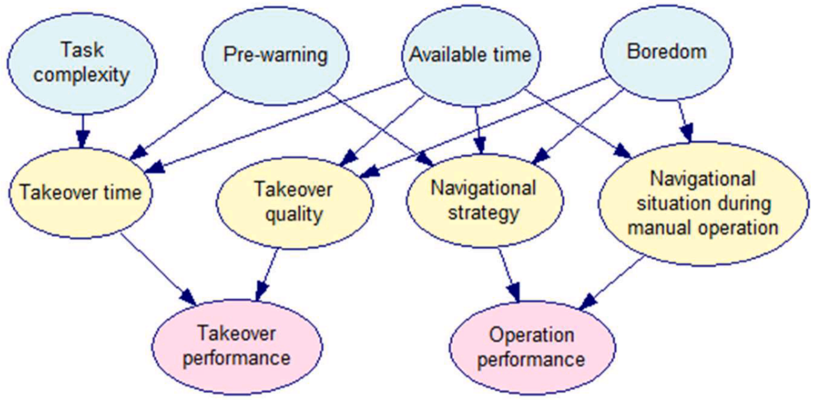


Fig.11. BN structure of PSFs, PMs and operator performances.

Table 7
CPT data collected for Takeover time

Available time	20 sec				60 sec			
Task complexity	Nominal		Complex		Nominal		Complex	
Pre-warning	On	Off	On	Off	On	Off	On	Off
Takeover Timely	1	0.75	0.75	0	1	1	1	1
Takeover Untimely	0	0.25	0.25	1	0	0	0	0

Table 8
CPT data collected for Navigational situation during manual operation

Available time	20 sec		60 sec	
Boredom	5 min	30 min	5 min	30 min
Situation: Good	0	0.5	0.25	0.333
Situation: Moderate	1	0.5	0.625	0.5
Situation: Bad	0	0	0.125	0.167

Table 9
CPT data collected for Operation performance

Navigational strategy	Yes			No		
Navigational situation during manual operation	Good	Moderate	Bad	Good	Moderate	Bad
Adequate	1	0	0	0	0	0
Inadequate	0	0.889	1	1	0.75	0
Failure	0	0.111	0	0	0.25	1

performance is failure". The top five of impact magnitudes of PSFs and intermediate nodes are: Available time; Task complexity; Pre-warning; Takeover time conditional upon available time is 20 s, task complexity is nominal, pre-warning is off; Takeover time conditional upon available time is 20 s, task complexity is complex, and pre-warning is on.

Fig. 17 presents an example of operation performance. The first two bars show when the available time is 60 s, and the boredom level is 5 min, the situation during manual operation significantly affects operation performance. Looking at the first and the third bars, they have the same condition of available time (i.e., 60 s). When the boredom level is 5 min, to adjust the probability of "situation during manual operation is good" has more significant effects on the probability of "the operation performance is adequate", comparing to when the boredom level is 30 min. Their effects are both positive, which means the higher probability of "situation during manual operation is good" leads to the higher probability of "the operation performance is adequate". The fourth bar shows the available time – 20 s has negative effects on the adequate operation performance. The whole plot shows the available time can significantly influence operation performance alone, but the boredom and pre-warning do not show their sole significant impact unless combing them together with available time.

5. Discussion

5.1. Findings of the PSFs' impact

The results of the BN model show that 'available time' ranks as the top one PSF in whether adequate or failed takeover performance of

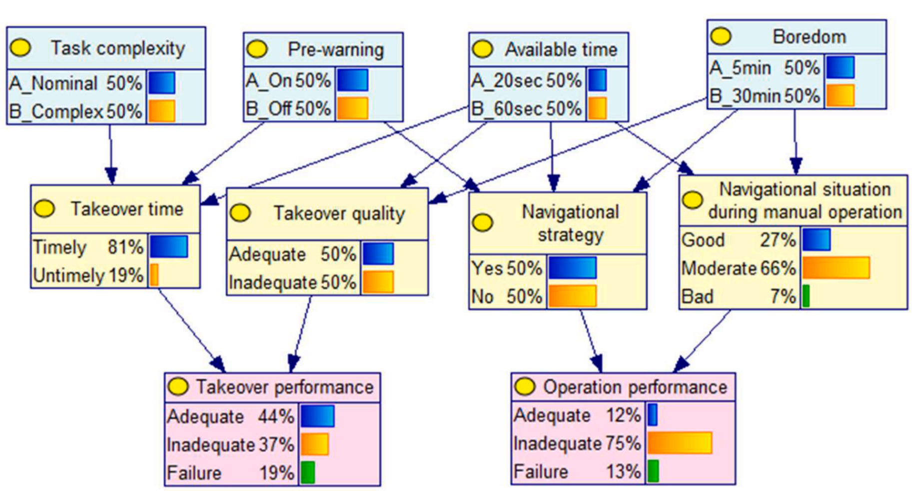


Fig. 12. BN model digram.

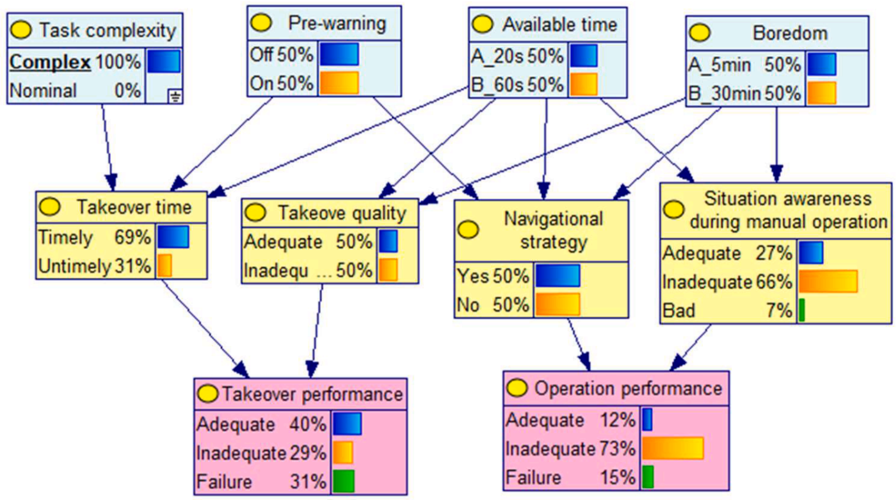


Fig. 13. An example of BN calculation results of setting evidence to the PSF Task complexity.

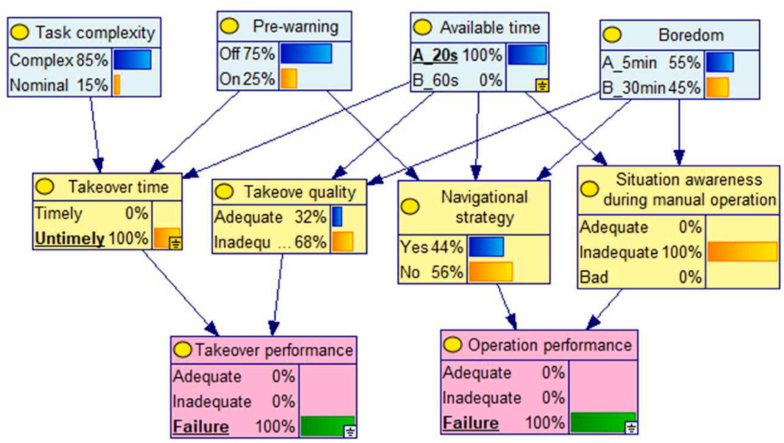


Fig. 14. An example of BN calculation results of setting evidence to Takeover failure and Operation failure.

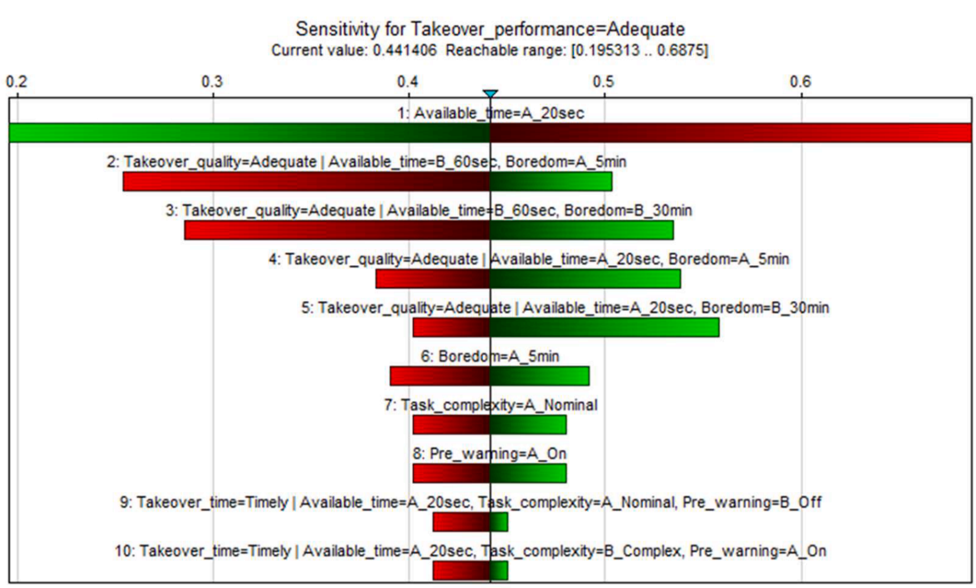


Fig. 15. Tornado plot of sensitivity analysis for "Takeover performance is adequate".

Fig. 16. Tornado plot of sensitivity analysis for "Takeover performance is failure".

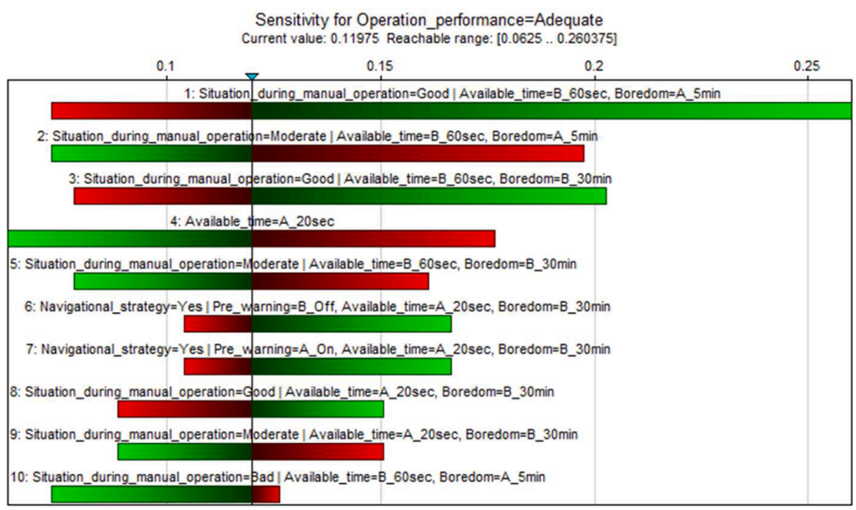


Fig. 17. Tornado plot of sensitivity analysis for "Operation performance is adequate".

operators. This implies that the operator needs long enough time to "recover" situation awareness to accomplish the takeover cognitive process (as in Fig. 2). By contrast, too short available time may cause the operator to not accomplish the takeover cognitive process, leading to takeover failure. This study selected two levels of 'available time' aiming to boats in a canal which are 20 s and 60 s. Considering different types of vessels and waterway conditions, levels of 'available time' can be redefined.

It is worthy mentioning that boredom is the sixth significant factor for 'takeover performance is adequate' but not such significant for 'takeover failure'. We infer that during a long time in autonomous mode, the high level of boredom would decrease operators' perception and situation awareness and further affect their cognition. In the field of automated vehicles, appropriate non-driving-related-tasks (NDRT) are potentially beneficial to activate the attentional resources of the operator and further improve their takeover performance [80]. Therefore, we suggest designers to consider introducing appropriate activities for activating the attentional resources of operators to prevent the loss of vigilance. Moreover, task complexity and pre-warning are significant for "takeover failure". This finding is in line with the studies in automated vehicles [80]. Moreover, no (or improper) pre-warning was also demonstrated to be able to result in significantly worse takeover performance [81].

It should be mentioned that takeover quality was evaluated by using the distances between mA2 and the encountering boats. This paper is the starting point to emphasize the significance of takeover quality for human performance and how this may influence human errors and their probability. Timely takeover and adequate takeover quality can be both regarded as essential for adequate takeover performance. Hence, this paper suggests it is necessary to distinguish takeover time and takeover quality when assessing takeover performance, and to develop corresponding tools for facilitating timely takeover and adequate takeover quality, such as alarm systems, and vision-augment systems.

Interestingly, the PSF *Experience* was not involved in the BN model, because it does not significantly affect any PM based on ANOVA results. It means seafarers and gamers did not present significant differences in their performance on takeover and operation. Many studies discussed the importance of the experience of operators to handle situations [7,19,22,61]. The experience investigated in this paper includes not only conventional maneuvering, but also new knowledge and skills on the new technology, especially HCI. Hence, we suggest investigating the effects of specific experience on human performance by decomposing experience into more subsets, such as knowledge of navigational situations, proficiency in operation and interaction with interfaces, and automated equipment. Moreover, the PM 2.1 *Detection* did not show significant relationships with each PSF. The limitation may be because the adopted data for this PM was all from interviews, and it is necessary to introduce more objective data to improve the accuracy in future works.

5.2. H-SIA, the development of the BN and human performance quantification

The use of H-SIA, with a structured development of the scenarios through the ESDs and task decomposition through the CoTA, helped identify the performance measures that should be assessed in the experiments.

The BN in this paper is developed by leveraging two sources of data, i.e., simulator data and interview data, and using the ANOVA method. The BN shows insight of how PSFs affect human errors and provides a way to quantify human errors in MASS operation.

Intermediate nodes, described by the PMs measured in the experiment, are defined as several states in the BN. This work presents a starting effort to make measuring human performance possible in MASS operation. However, in future works, to improve accuracy of the BN, more information observed in experiment could be fused to consider intermediate nodes, such as the real-time speeds, heading of the MASS, and the physiological index of operators. The utilization of ANOVA method presents the availability of statistic techniques for modeling BN structures. The CPTs of all intermediate and top nodes are derived from the experimental data, and would be updated by introducing additional data in the future.

We estimated the probabilities of takeover and operation performance when the probabilities of the PSFs are 0.5 and 0.5 based on the experimental design, as shown in Fig. 12. It is worthy noting that *bad* "navigational situation during manual operation (situation)" does not directly lead to operation failure, but can increase the probability of operation failure. For example, in the case of Fig. 12, the probability of "operation performance" is *failure, inadequate* and *adequate operation* is 0.13, 0.75 and 0.12, respectively. When given evidence to *bad* situation, the probability of "operation performance" is *failure, inadequate* and *adequate operation* changes to 0.39, 0.61, and 0, respectively. In addition, we do not quantify prior probabilities of PSFs, and the obtained CPTs contain uncertainty, since the probabilities are derived from simulator data and interview data that involve limited data quantity. These issues would be focused on in future works by introducing a larger quantity and more various forms of data.

5.3. Significance of the results for HRA

This paper demonstrates the value and compatibility of H-SIA, BN, and experiment data to identify and quantify human errors in HAC systems, and enables limited amount of experimental data to enhance the technical basis of future HRA method.

The paper identifies and measures the human errors in HAC systems. Moreover, the constructed causal structure enables HRA analysts to explain why human errors happen and what can be done to prevent them in different contexts in MASS operation. In addition, this paper estimates CPTs, which contribute to quantifying human performance probabilities, and help present what contexts may more likely cause takeover failure and operation failure. This quantitative model provides an empirical basis for the HRA in MASS operation.

There is still space to improve the results' accuracy. Some PMs are quantified based on interview data, such as navigational strategy which is used to represent whether the operator had and executed a navigational strategy in manual operation. However, to be more accurate, more objective data should be included to generate PMs to measure the human errors. This could reduce bias from subjective data and improve the model performance, and facilitate the HRA development in HAC systems.

5.4. Simplification in the experimental simulator and procedure

The virtual experiment shows its adaptivity to collect empirical data and helps quantify human performance in MASS application. However, the case study has a few limitations due to the simplification of the experimental simulator and procedure. At first, the case study considers a simplified environment and not all conditions the mA2 would experience, such as various boat types and speeds in the canal. To validate the achieved outcomes in the real mA2 operation is out of the scope of this paper and will be included in future works. Second, each PSF was distinguished by two states and implemented in the experiment, since we expect it is feasible to measure them through the experiment. However, in real situations, PSFs may have various states, such as task complexity. Therefore, defining PSFs' states more realistically should be explored in the future. Thirdly, human performance may differ due to the difference between virtual and real environments. How 'real' the virtual environment appears and feels is called the fidelity of the virtual environment [82]. How the fidelity issue may cause difference of human performance in virtual and real environment needs to be further explored.

This paper, as a starting point, aims to analyze human errors and explore PSFs' effect on human performance in MASS operation, so we used the virtual version of mA2. In future works, the results can be compared with a similar analysis in real MASS operation. This observation would be a breakthrough and potentially make the results of HRA more accurate and realistic.

5.5. Significance of this study for development of intelligent assistance systems

Intelligent assistance systems are capable of helping humans by augmenting perception, understanding, decision-making of humans, and gruadully replace humans to finish some tasks [9,10,12,15,18]. Nonetheless, what and how likely human errors may occur, what scenarios humans may make errors more likely, also effect on the design and development of intelligent assistance systems. As our results showed, e.g., the human error probability is the largest when humans do not have enough available time to take over control. Hence, researchers and developers can explore intelligent assistance systems from two aspects.

The first aspect is to investigate how long should be enough for humans to recover situation awareness for a certain situation. It most likely is more than 20 s since the results show a positive effects of 20 s on human errors; moreover, it will be significant to investigate how often should humans regain situation awareness, considering the available time to takeover control. The other aspect is to develop intelligent assistance systems which would help humans regain situation awareness more quickly. E.g., an effective alarm that may provide accurate and intuitive information to humans; and some intelligent assistance systems that are capable of helping humans predict ship trajectories, assess the impact of hydrometeorological conditions on ship dynamics, identify the optimal timing to avoid collision, etc. They are beneficial to compensate for humans' innate limitation. Another result in our study is that improper task complexity would increase human error probability. Hence, decision-making assistance systems that can handle complex scenarios should be explored in the development of MASS, such as emergency operational decision assistance, etc. [13,21]. This paper provides potential directions to prevent human errors for developing intelligent assistance systems in MASS.

6. Conclusion

To identify human errors and quantify the human performance probabilities during takeover control and manual operation of SCC operators in MASS operation, this paper proposes an approach by combining H-SIA, virtual experiment, and BN. The results show a theoretical significance of making it feasible to calculate the probability of human errors. These probabilities provide an empirical basis for applying HRA to MASS operation. This work is helpful to systematic safety analysis and management for autonomous ships and benefits other HAC systems.

The results of analysis shows the practical significance of this approach: available time has the most significant effects on takeover and operational performance; task complexity and pre-warning are significant to takeover failure. Boredom and pre-warning show their effects on operational performance. The participants' experience did not significantly affect human performance. These findings can be utilized as input to safety analysis and management, and further facilitate human-oriented design and development of SCC.

Future works can be expected to apply the proposed approach to other operational scenarios, such as different navigational scenarios and task allocations in HAC. In addition, more PSFs and their levels, such as operators stress should be considered. More performance measurements can be constructed by using physiological data. It can help more accurately illustrate the effects of PSFs on human performance, and further improve model accuracy for human error assessment.

CRediT authorship contribution statement

Tingting Cheng: Conceptualization, Methodology, Validation, Formal analysis, Data curation, Visualization, Writing – original draft. **Erik A. Veitch:** Software, Formal analysis, Investigation, Data curation, Writing – review & editing. **Ingrid Bouwer Utne:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing – review & editing. **Marilia A. Ramos:** Conceptualization, Methodology, Validation, Writing – review & editing. **Ali Mosleh:** Conceptualization, Writing – review & editing. **Ole Andreas Alsos:** Resources, Visualization, Funding acquisition, Writing – review & editing. **Bing Wu:** Funding acquisition, Writing – review & editing.

Declaration of competing interest

We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Data availability

The authors do not have permission to share data.

Acknowledgements

The authors acknowledge the funding from the International Cooperation and Exchange of the National Natural Science Foundation of China (Grant No. 51920105014, 52071248, 52272422); the financial support provided by the Norwegian Research Council under the MAROFF-2 project "Land-based Operation of Autonomous Ships" (LOAS, Project Number 296527). The authors also acknowledge the support of the Department of Design at NTNU and the entire team behind the Shore Control Lab.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ress.2024.110080.

References

[1] Negenborn RR, Goerlandt F, Johansen TA, Slaets P, Valdez Banda OA, Vanelslander T, et al. Autonomous ships are on the horizon: here's what we need to know. Nature 2023;615:30–3.
[2] Veitch E, Alsos OA. A systematic review of human-AI interaction in autonomous ship systems. Saf Sci 2022;152:105778.
[3] Huang Y, Chen L, Negenborn RR, Van Gelder P. A ship collision avoidance system for human-machine cooperation during collision avoidance. Ocean Eng 2020;217:107913.
[4] Liu C, Chu X, Wu W, Li S, He Z, Zheng M, et al. Human–machine cooperation research for navigation of maritime autonomous surface ships: a review and consideration. Ocean Eng 2022;246:110555.
[5] Wang Y, Chen P, Wu B, Wan C, Yang Z. A trustable architecture over blockchain to facilitate maritime administration for MASS systems. Reliab Eng Syst Saf 2022;219:108246.
[6] Fan C, Montewka J, Zhang D. A risk comparison framework for autonomous ships navigation. Reliab Eng Syst Saf 2022;226:108709.
[7] Baldauf M, Fischer S, Kitada M, Mehdi R, Al-Quhali M, Fiorini M. Merging conventionally navigating ships and MASS-merging VTS, FOC and SCC? TransNav: Int J Marine Navigat Saf Sea Transp 2019;13:495–501.
[8] Liu C, Kulkarni K, Suominen M, Kujala P, Musharraf M. On the data-driven investigation of factors affecting the need for icebreaker assistance in the Baltic Sea. Cold Regions Sci Technol 2024:104173.
[9] Zhang M, Montewka J, Manderbacka T, Kujala P, Hirdaris S. A big data analytics method for the evaluation of ship-ship collision risk reflecting hydrometeorological conditions. Reliab Eng Syst Saf 2021;213:107674.
[10] Zhang M, Kujala P, Musharraf M, Zhang J, Hirdaris S. A machine learning method for the prediction of ship motion trajectories in real operational conditions. Ocean Eng 2023;283:114905.
[11] Zhang M, Taimuri G, Zhang J, Hirdaris S. A deep learning method for the prediction of 6-DoF ship motions in real conditions. Proc Inst Mech Engr, Part M: J Engr Maritime Envr 2023;237(4):887–905.
[12] Xu X, Wu B, Xie L, Teixeira ÂP, Yan X. A novel ship speed and heading estimation approach using radar sequential images. IEEE Trans Intell Transp Syst 2023;24:11107–20.
[13] Zhang M, Kujala P, Hirdaris S. A machine learning method for the evaluation of ship grounding risk in real operational conditions. Reliab Eng Syst Saf 2022;226:108697.
[14] Zhang M, Conti F, Le Sourne H, Vassalos D, Kujala P, Lindroth D, et al. A method for the direct assessment of ship collision damage and flooding risk in real conditions. Ocean Eng 2021;237:109605.
[15] Johansen T, Blindheim S, Torben TR, Utne IB, Johansen TA, Sørensen AJ. Development and testing of a risk-based control system for autonomous ships. Reliab Eng Syst Saf 2023;234:109195.
[16] Sheridan TB. Human supervisory control of automation. Handb. Human Fact. Ergon. 2021:736–60.
[17] Wróbel K, Gil M, Chae C-J. On the influence of human factors on safety of remotely-controlled merchant vessels. Appl. Sci. 2021;11:1145.
[18] Rødseth ØJ, Lien Wennersberg LA, Nordahl H. Towards approval of autonomous ship systems by their operational envelope. J. Marine Sci. Technol. 2022;27:67–76.
[19] Man Y, Weber R, Cimbritz J, Lundh M, MacKinnon SN. Human factor issues during remote ship monitoring tasks: an ecological lesson for system design in a distributed context. Int. J. Ind. Ergon. 2018;68:231–44.
[20] Ramos M, Utne IB, Mosleh A. Collision avoidance on maritime autonomous surface ships: operators' tasks and human failure events. Saf Sci 2019;116:33–44.
[21] Ramos MA, Thieme CA, Utne IB, Mosleh A. Human-system concurrent task analysis for maritime autonomous surface ship operation and safety. Reliab Eng Syst Saf 2020;195:106697.
[22] Zhang M, Zhang D, Yao H, Zhang K. A probabilistic model of human error assessment for autonomous cargo ships focusing on human–autonomy collaboration. Saf Sci 2020;130:104838.
[23] Liu J, Aydin M, Akyuz E, Arslan O, Uflaz E, Kurt RE, et al. Prediction of human–machine interface (HMI) operational errors for maritime autonomous surface ships (MASS). J Marine Sci Technol 2021:1–14.
[24] Cheng T, Utne IB, Wu B, Wu Q. A novel system-theoretic approach for human-system collaboration safety: case studies on two degrees of autonomy for autonomous ships. Reliab Eng Syst Saf 2023;237:109388.
[25] Graziano A, Teixeira A, Soares CG. Classification of human errors in grounding and collision accidents using the TRACEr taxonomy. Saf Sci 2016;86:245–57.
[26] Chang C-H, Kontovas C, Yu Q, Yang Z. Risk assessment of the operations of maritime autonomous surface ships. Reliab Eng Syst Saf 2021;207:107324.
[27] Abreu DT, Maturana MC, Droguett EL, Martins MR. Human reliability analysis of conventional maritime pilotage operations supported by a prospective model. Reliab Eng Syst Saf 2022;228:108763.
[28] Yildiz S, Uğurlu Ö, Wang J, Loughney S. Application of the HFACS-PV approach for identification of human and organizational factors (HOFs) influencing marine accidents. Reliab Eng Syst Saf 2021;208:107395.
[29] Fan S, Blanco-Davis E, Zhang J, Bury A, Warren J, Yang Z, et al. The role of the prefrontal cortex and functional connectivity during maritime operations: an fNIRS study. Brain Behav 2021;11:e01910.
[30] Porathe T, Hoem Å, Rødseth Ø, Fjørtoft K, Johnsen SO. At least as safe as manned shipping? autonomous shipping, safety and "human error." safety and reliability – safe societies in a changing world. 1st ed. London: CRC Press; 2018. p. 417–25.
[31] Xi Y, Yang Z, Fang Q, Chen W, Wang J. A new hybrid approach to human error probability quantification–applications in maritime operations. Ocean Eng. 2017;138:45–54.
[32] Park J, Lee D, Jung W, Kim J. An experimental investigation on relationship between PSFs and operator performances in the digital main control room. Ann Nucl Energy 2017;101:58–68.
[33] Iqbal MU, Srinivasan R. Simulator based performance metrics to estimate reliability of control room operators. J Loss Prev Process Ind 2018;56:524–30.
[34] Velmurugan K, Saravanasankar S, Venkumar P, Sudhakarapandian R, Di Bona G. Hybrid fuzzy AHP-TOPSIS framework on human error factor analysis: implications to developing optimal maintenance management system in the SMEs. Sustain Futures 2022;4:100087.

[35] Patriarca R, Ramos M, Paltrinieri N, Massaiu S, Costantino F, Di Gravio G, et al. Human reliability analysis: exploring the intellectual structure of a research field. Reliab Eng Syst Saf 2020;203:107102.
[36] Kirwan B. A guide to practical human reliability assessment. CRC press; 1994.
[37] Di Bona G, Falcone D, Forcina A, De Carlo F, Silvestri L. Quality checks logit human reliability (LHR): a new model to evaluate human error Probability (HEP). Math Prob Eng 2021;2021:1–12.
[38] Ramos MA, Mosleh A. In: Human Role in Failure of Autonomous Systems: A Human Reliability Perspective. 2021 Annual Reliability and Maintainability Symposium (RAMS). Orlando, FL, USA: IEEE; 2021. p. 1–6.
[39] Chang Y.H.J., Mosleh A. Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents Part 1: overview of the IDAC Model 2007.
[40] Xu S, Kim E, Haugen S, Zhang M. A Bayesian network risk model for predicting ship besetting in ice during convoy operations along the Northern Sea Route. Reliab Eng Syst Saf 2022;223:108475.
[41] Groth KM, Mosleh A. Deriving causal Bayesian networks from human reliability analysis data: a methodology and example model. Proc Inst Mech Eng, Part O: J Risk Reliab 2012;226:361–79.
[42] Fan S, Blanco-Davis E, Yang Z, Zhang J, Yan X. Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network. Reliab Eng Syst Saf 2020;203:107070.
[43] Musharraf M, Smith J, Khan F, Veitch B, MacKinnon S. Incorporating individual differences in human reliability analysis: an extension to the virtual experimental technique. Saf Sci 2018;107:216–23.
[44] Musharraf M, Smith J, Khan F, Veitch B, MacKinnon S. Assessing offshore emergency evacuation behavior in a virtual environment using a Bayesian Network approach. Reliab Eng Syst Saf 2016;152:28–37.
[45] Kim Y, Park J, Jung W, Choi SY, Kim S. Estimating the quantitative relation between PSFs and HEPs from full-scope simulator data. Reliab Eng Syst Saf 2018; 173:12–22.
[46] Wu H, Wu C, Lyu N, Li J. Does a faster takeover necessarily mean it is better? A study on the influence of urgency and takeover-request lead time on takeover performance and safety. Acc Anal Prev 2022;171:106647.
[47] Lin J, Wohleber R, Matthews G, Chiu P, Calhoun G, Ruff H, et al. Video game experience and gender as predictors of performance and stress during supervisory control of multiple unmanned aerial vehicles, 59. Los Angeles, CA: SAGE Publications Sage CA; 2015. p. 746–50.
[48] Vasstein K, Brekke E, Mester R, Eide E. Autoferry gemini: a real-time simulation platform for electromagnetic radiation sensors on autonomous ships, 929. IOP Publishing; 2020, 012032.
[49] Kaber DB. Issues in human–automation interaction modeling: presumptive aspects of frameworks of types and levels of automation. J Cogn Eng Decis Mak 2018;12: 7–24.
[50] Utne IB, Sørensen AJ, Schjølberg I. Risk management of autonomous marine systems and operations, 57663. American Society of Mechanical Engineers; 2017. V03BT02A020.
[51] IMO. IMO takes first steps to address autonomous ships 2018;http://www.imo.org/en/MediaCentre/PressBriefings/Pages/08-MSC-99-MASS-scoping.aspx.
[52] Endsley MR. Designing for situation awareness: an approach to user-centered design. CRC press; 2016.
[53] Payne JW. Task complexity and contingent processing in decision making: an information search and protocol analysis. Organ Behav Hum Perform 1976;16: 366–87.
[54] Hogenboom S, Parhizkar T, Vinnem JE. Temporal decision-making factors in risk analyses of dynamic positioning operations. Reliab Eng Syst Saf 2021;207:107347.
[55] Veitch E, Alsos OA, Cheng T, Senderud K, Utne IB. Human Factor Influences on Supervisory Control of Remotely Operated and Autonomous Vessels. Ocean Engr 2024;299:117257.
[56] Muckler FA, Seven SA. Selecting performance measures:" Objective" versus" subjective" measurement. Hum Factors 1992;34:441–55.
[57] Tomaka J, Blascovich J, Kelsey RM, Leitten CL. Subjective, physiological, and behavioral effects of threat and challenge appraisal. J Pers Soc Psychol 1993;65: 248.
[58] Bona G, Falcone D, Forcina A, Silvestri L. Systematic human reliability analysis (SHRA): a new approach to evaluate human error probability (HEP) in a nuclear plant. Int J Math, Eng Manag Sci 2021;6:345–62.
[59] Lin Y-L, Hwang S-L. The application of the loglinear model to quantify human errors. Reliab Eng Syst Saf 1992;37:157–65.
[60] Hoem ÅS, Veitch E, Vasstein K. Human-centred risk assessment for a land-based control interface for an autonomous vessel. WMU J Marit Aff 2022;21:179–211.
[61] Wróbel K. System-theoretic approach to safety of remotely-controlled merchant vessel. Ocean Eng. 2018.
[62] Montgomery DC. Design and analysis of experiments. John wiley & sons; 2017.
[63] Uğurlu Ö, Kartal ŞE, Gündoğan O, Aydin M, Wang J. A statistical analysis-based Bayesian Network model for assessment of mobbing acts on ships. Marit Policy Manag 2023;50:750–75.
[64] Dey S, Stori J. A Bayesian network approach to root cause diagnosis of process variations. Int J Mach Tools Manuf 2005;45:75–91.
[65] Alsos OA, Veitch E, Pantelatos L, Vasstein K, Eide E, Petermann F-M, et al. NTNU Shore Control Lab: designing shore control centres in the age of autonomous ships. J Phys: Conf Ser 2022;2311:012030.
[66] Gertman D, Blackman H, Marble J, Byers J, Smith C. The SPAR-H human reliability analysis method. US Nucl Regul Commiss 2005;230:35.
[67] Bye A., Laumann K., Taylor C., Rasmussen M., Øie S., Van de Merwe K., et al. The petro-HRA guideline 2017.
[68] Fisherl CD. Boredom at work: a neglected concept. Human Relat 1993;46:395–417.
[69] Ramos M, Utne IB, Vinnem JE, Mosleh A. Accounting for human failure in autonomous ships operations. Saf Reliab-Safe Soc Chang World ESREL 2018;2018: 355–63.
[70] Molloy R, Parasuraman R. Monitoring an automated system for a single failure: vigilance and task complexity effects. Hum Factors 1996;38:311–22.
[71] Senderud K. Kjedsomhet og vigilans hos operatører i kontrollrom for autonome ferger. 2022.
[72] Porathe T. Autonomous ships: a research strategy for human factors research in autonomous shipping. Springer; 2021. p. 479–86.
[73] Norwegian Maritime Authority. (2011). Forskrift om kvalifikasjoner og sertifikater for sjøfolk. https://www.sdir.no/sjofart/regelverk/rundskriv/kvalifikasjoner-og-sertifikater-for-sjofolk/n.d.
[74] Du N, Zhou F, Pulver EM, Tilbury DM, Robert LP, Pradhan AK, et al. Examining the effects of emotional valence and arousal on takeover performance in conditionally automated driving. Transp Res Part C: Emerging Technol 2020;112:78–87.
[75] Cheng T, Wu Q, Wu B, Yan X. A Probabilistic Decision-Making System for Joining Traffic Lanes Within an Inland Traffic Separation Scheme. Mar Technol Soc J 2021; 55:44–63.
[76] Cheng T, Wu Q, Wu B, Zhang M. Prediction of rationality for ferry crossing behavior based on multi hidden markov model. IEEE; 2019. p. 1382–8.
[77] Wu B, Cheng T, Yip TL, Wang Y. Fuzzy logic based dynamic decision-making system for intelligent navigation strategy within inland traffic separation schemes. Ocean Eng. 2020;197:106909.
[78] Stat-Ease. Handbook for experimenters. 2019.
[79] Shabarchin O, Tesfamariam S. Internal corrosion hazard assessment of oil & gas pipelines using Bayesian belief network model. J Loss Prev Process Ind 2016;40: 479–95.
[80] Gold C, Happee R, Bengler K. Modeling take-over performance in level 3 conditionally automated vehicles. Acc Anal Prev 2018;116:3–13.
[81] Agrawal S, Peeta S. Evaluating the impacts of driver's pre-warning cognitive state on takeover performance under conditional automation. Transp Res Part F: Traff Psychol Behav 2021;83:80–98.
[82] Liu D, Yu J, Macchiarella ND, Vincenzi DA. Simulation fidelity. human factors in simulation and training. CRC Press; 2008. p. 91–108.