Ocean Engineering 299 (2024) 117257

Contents lists available at ScienceDirect

Ocean Engineering

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

Research paper

Human factor influences on supervisory control of remotely operated and autonomous vessels

Erik Veitch [a,*], Ole Andreas Alsos [a], Tingting Cheng [b,c], Kristin Senderud [d], Ingrid Bouwer Utne [c]

[a] NTNU, Department of Design, Kolbjørn Hejes Vei 2b, 7491, Trondheim, Norway
[b] Wuhan University of Technology, School of Transportation and Logistics Engineering, Heping Ave 1178, 430063, Wuhan, China
[c] NTNU, Department of Marine Technology, Otto Nielsens Veg 10, 7491, Trondheim, Norway
[d] University of Bergen, Department of Psychosocial Science, Christies Gate 12, 5015, Bergen, Norway

ABSTRACT

Autonomous ships require remote supervision from a human operator to ensure safety. However, there are knowledge gaps concerning human factor influences on remote supervisory control. We investigate the influence of five factors on remote supervisory control during simulated intervention scenarios: (i) Skillset, represented by gamers and navigators; (ii) Monitoring Time, represented by either 5 or 30 min of passive monitoring; (iii) Number of Vessels, represented by either one or three vessels; (iv) Available Time, represented by 20- or 60-s critical time windows; (v) Decision Support System (DSS), represented by availability of a DSS. The experiment was a randomized factorial design where participants (n = 32) completed two interventions: first a handover (automation detects a critical event and hands over control) and then a takeover (operator detects a critical event and takes over control). We observed: (i) gamers and navigators both demonstrated transferrable skillsets, but neither group excelled over the other; (ii) monitoring time affected boredom, but this translated to minor performance effects. Moreover, performance was reduced under conditions of (iii) supervising three vessels, (iv) low time availability, and (v) unavailable DSS. These outcomes contribute to the empirical basis for assessing maritime human factors in remotely controlled and autonomous ship design.

1. Introduction

Advances in navigation technology are heralding a new age of remotely controlled and autonomous vessels. Guided by sensors and control algorithms, autonomous ships break with the conventions of crewed navigation and may vastly improve safety, efficiency, and logistics at sea. Autonomous ships can perform path planning and collision avoidance automatically (Öztürk et al., 2022; Vagale et al., 2021), and can even perform delicate docking and undocking sequences on-par with experienced human operators (Martinsen et al., 2020; Suyama et al., 2022). While in operation, a fleet of autonomous ships can be overseen remotely from a single remote control center, where human operators are able to take preventative action, if needed. Several real-life application of autonomous vessels have been demonstrated, including in open water environments (e.g., Kim et al., 2022), inland waterways (e.g., Peeters et al., 2020), and constrained urban waterways (e.g., Brekke et al., 2022). However, despite constituting one of the major trends in ocean engineering research today (Tavakoli et al., 2023), relatively little is known about the role of the human operator of remotely controlled and autonomous vessels (Negenborn et al., 2023). Specifically, knowledge gaps persist concerning maritime human factor influences on human supervisory control.

One such knowledge gap presents itself in the risk sciences, where practitioners model human error probabilities in human-system integration of autonomous ships (Guo and Utne, 2022; J. Liu et al., 2022; Ramos et al., 2020). In this domain, charting the limits of human performance translates to investigation of Risk-Influencing Factors (RIFs). In turn, risk-based design guidelines lack critical details related to human factors, such as the amount of time needed to safely take over control and gain situation awareness under various conditions (Hoem et al., 2021; Rødseth et al., 2022). In the design of Decision Support Systems (DSSs) for navigation, too, a gap has emerged between designers' and operators' expectations of human-computer interaction, which may lead to inappropriate designs if left unaddressed (Aylward et al., 2022; Veitch et al., 2022). Important gaps have also been highlighted concerning recruitment, certification, and training of control room personnel (Emad et al., 2022; Sharma and Kim, 2021). Indeed, human factors have been explicitly highlighted by a considerable number of researchers as a critical knowledge gap in the development of remotely controlled and autonomous vessels (Kari and Steinert, 2021; MacKinnon et al., 2020; Öztürk et al., 2022; Ramos et al., 2018; Wróbel et al., 2021).

* Corresponding author.
E-mail addresses: erik.a.veitch@ntnu.no (E. Veitch), oleanda@ntnu.no (O.A. Alsos), Trista_tt07@whut.edu.cn (T. Cheng), kristin.g.senderud@gmail.com (K. Senderud).

https://doi.org/10.1016/j.oceaneng.2024.117257
Received 17 April 2023; Received in revised form 20 February 2024; Accepted 21 February 2024
Available online 5 March 2024

In the maritime domain, advances in automation have historically tended to underscore human factors, rather than circumvent them (MacKinnon et al., 2015). This is demonstrated in recent applications of machine learning in path planning (e.g., Deraj et al., 2023; Rongcai et al., 2023; Wang et al., 2022), whose results, despite utilizing methods like neural networks that claim to mimic the human brain, have not yet been shown to fully comply with the Collision Regulations (COLREGs) (Öztürk et al., 2022). In this regard, one of the greatest challenges is modeling decidedly human-orientated concepts like "good seamanship," "proper look out," and "making a full appraisal of the situation" (IMO, 1972). Maritime navigation is, after all, a complex social activity, where decisions are based as much upon in-the-moment intuitions as on calculated plans (Hutchins, 1995; Suchman, 2007).

The International Maritime Organization (IMO), for their part, has highlighted regulatory gaps concerning the role of remote operators for autonomous ships. In their "Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships" (IMO, 2021), the role of operators at the Remote Operating Center (ROC) was listed as among the most "high priority issues" to be addressed. A better understanding of these themes may serve to guide a new code for autonomous ships, expected to enter into force in 2028 (IMO, 2022).

1.1. Research question

Motivated by these knowledge gaps, the research question we address in this paper is, "What factors influence human supervisory control of highly automated vessels?" To this aim, our focus is on interventions marking the transfer of control between the autonomy and the operator, and upon experimentally assessing intervention performance under various conditions. The outcomes of this investigation will provide fundamental knowledge that will enable safe and timely monitoring and intervention by human operators, which ultimately is necessary to achieve an acceptable risk level of autonomous vessels.

There are three major contributions: (i) the hypothesized effects of five selected human factor influences on remote supervisory control are tested, (ii) a method is described for rigorous testing of remote supervisory control systems using a simulation experiment, and (iii) results are presented that build a better understanding of the role of human operators in autonomous ship systems.

This article is organized as follows: In Section 1, we introduce five factors that we selected for experimental investigation; in Section 2, we describe the methodology we used to design the experimental campaign and the simulator used to host it; in Section 3, we present the result of the experiments; in Section 4, we discuss the implications of the results for fields of human factors and risk management; and in Section 5, we present our concluding remarks.

1.2. Simulator test platform

We used a simulator for our experiments that re-created a real autonomous research vessel called "milliAmpere2" (Fig. 1), designed and operated by the Norwegian University of Science and Technology (NTNU). Built in the Unity game platform (Version, 2022.2.8; Unity Technologies, 2022), the simulator architecture was based on the open-source *Gemini* platform (https://github.com/Gemini-team/Gemini). Building on the original architecture, we built the simulator to re-create operations that the real "milliAmpere2" underwent during field trials in 2022 at a 100-m canal crossing in Trondheim, Norway. This version, which was used in this study, is available for download with some licensing restrictions (https://github.com/mikael-rh/ScenarioBuilder). The simulator architecture design is documented in Hansen (2022) and its interface design is documented in Ek (2022). The simulator is hosted at the NTNU Shore Control Lab (Alsos et al., 2022).

1.3. Selection of five factors for investigation

Based on a review of the literature, five factors were selected for investigation. These five factors are described below. Also described are the knowledge gaps associated with each of the five factors as well as how we planned to address these gaps experimentally. These five factors do not cover all possible maritime human factor considerations in remote supervisory control; however, they do represent a relevant subset where empirical contributions are currently needed.

Skillset (Factor A). The skillsets possessed by remote operators will likely influence their performance. A significant amount of research has investigated what skillsets – or, more specifically, what accumulated knowledge, understanding, and proficiencies (KUPs) (IMO, 2017, Table A-II/1) – are needed for remote operators (Veitch and Alsos, 2022). Most researchers agree that remote operators should possess the essential skillsets of conventional navigators, including ship-handing and seamanship (Baldauf et al., 2019; Saha, 2021; Yoshida et al., 2020). Indeed, during public trials of the "milliAmpere2," the Norwegian Maritime Authority required a "safety host" onboard with a minimum certificate of competency of "D5" (Norwegian Maritime Authority, 2011). While safety hosts are not the same as remote operators, their roles and responsibilities are both defined by completing manual interventions, which implies that their required skillsets may be similar. However, researchers also point out that remote operators represents a fundamentally new role and as such will require new skillsets (Lutzhoft et al., 2019; Sharma and Kim, 2021). To date, the potentially transferrable skillsets of groups other that navigators have not been investigated. To frame this debate, we introduced a second group; namely, video game players, whom we hereafter refer to as "gamers." The question of whether gamer skillsets cross over to remote supervisory operations has been investigated in drone research (e.g., Lin et al., 2015; McKinley et al., 2011). However, this question has not yet been empirically tested in the context of supervisory control of maritime autonomous vessels. We hypothesized that both navigators and gamers would demonstrate relevant skillsets for the remote operator role, with one groups' skillset complementing the other. To test this, recruited gamers and navigators and assessed how their unique skillsets transferred to the role of remote operator.

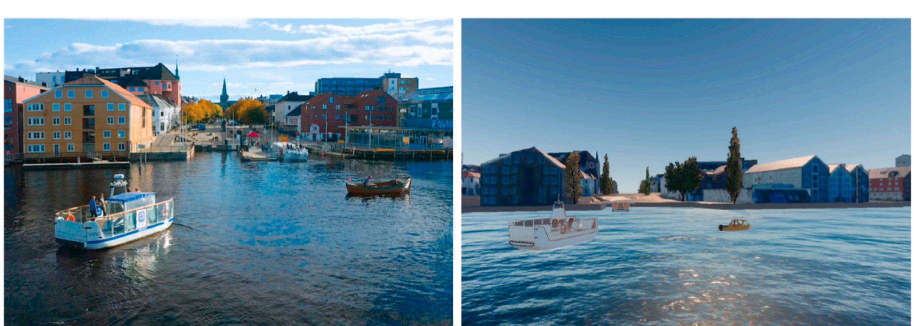

Fig. 1. "MilliAmpere2" autonomous ferry during public trials (left); "milliAmpere2" digital twin in virtual simulator (right).

Monitoring Time (Factor B). The amount of time in a passive monitoring role affects vigilance (Warm et al., 2008), which in turn, is known to affect performance of operators working with automated systems (Parasuraman and Riley, 1997). In psychology experiments, a drop in vigilance has been shown to set in somewhere between 20 and 30 min, traditionally by detecting increases in reaction times or error rates while assigned a tedious monitoring task (Cummings et al., 2016; Grier et al., 2003; Molloy and Parasuraman, 1996). For supervisory operators of highly autonomous ships, vigilance decrement is potentially dangerous if it undermines operators' ability to take timely and decisive preventive actions (Ramos et al., 2018). Since timely and decisive preventive action is precisely what is required of remote operators during intervention, it is important to understand how vigilance decrement might affect its execution. In this study, we leverage the known association between vigilance and boredom to gain insights into vigilance decrement (Pattyn et al., 2008). Boredom can be measured using a questionnaire approach by assessing boredom-related affects (van Tilburg and Igou, 2012) and by direct appraisal (i.e., asking to what extend they feel bored). We hypothesized that a time-induced boredom experience may lead to poorer remote supervisory performance. To test this, we designed intervention scenarios that fell on both sides of the 20- to-30- minute "boundary" separating uncompromised attention from a vigilant decrement, selected 5- and 30-min scenarios as sufficient to instill a clear difference.

Number of Vessels (Factor C). Simultaneous monitoring of multiple assets is known to affect performance in remote supervision tasks, as shown in tests involving drone simulations (e.g., Cummings et al., 2013; Dixon et al., 2005). Since the early days of autonomous ship research, the question of "how many vessels per operator" has been raised (Man et al., 2015), suggesting that multi-asset supervision will, as for drones, affect operators' capacity to supervise them safely. At least one test on the subject has been conducted, but with inconclusive results (Man et al., 2018). The lack of empirical studies about multivessel supervision for maritime remote supervision may lead to speculative claims about how many vessels should be allotted per operator, which, in turn, may lead to inappropriate designs (Dybvik et al., 2020). Some of the uncertainly may stem from the general expectation that maritime vessels, which move relatively slowly, are natural candidates for a format where one remote operator is responsible for many vessels. While this "one to many" format may address a contemporary lack of skilled mariners (Laugaland, 2022), it must be formally tested to shore up our understanding of how multivessel supervision influences maritime supervisory control. We hypothesized that multiple vessels would produce a measurable reduction in performance of remote supervisory duties, given the interface designs that are currently available. To test this, we tested remote operators under conditions of supervising either a single vessel or three vessels.

Available Time (Factor D). The amount of time available to attend to a critical event is, like multivessel supervision, known to affect remote supervision performance in drone applications (Cummings and Mitchell, 2006; Gutzwiller et al., 2016; D. Liu et al., 2016). However, the topic has not been formally investigated in the case of autonomous maritime vessels. Given that all critical events impose some sort of response window, an operator's ability to take timely and decisive preventative actions may be considered their single most important role. Correspondingly, the vessel's "operational envelop," which is defined partly by the amount of time needed for an operator to obtain situation awareness during intervention tasks, plays an important role in the approval of the vessel's overall design (Rødseth et al., 2022). We hypothesized that lower available time would hamper operators' capacity to fulfill interventions, and vice versa. To test this, we varied the amount of available time to resolve an intervention to either "low" or "high" levels, in this case corresponding to 20 and 60 s, respectively.

Decision Support **System (Factor E).** The presence of a Decision Support System (DSS) will likely affect the performance of maritime remote supervision. A DSS distills large amounts of information in a complex system into an integrated interface to help guide experts' decision-making, often with prompts and warnings to guide their attention. DSSs follow the maxim of supporting human decision-making, rather than supplanting it, and as such fit well into the paradigm of human-machine teaming in automation applications (C. Liu et al., 2022). Examples of DSS applications in supervisory control are most common in drone and aviation applications, and attempts have been made to transfer applicable design aspects to the maritime industry (Turan et al., 2016). In the maritime industry, some applications of DSSs already exist; examples include auto-crossing and auto-docking consoles on passenger ferries (e.g., Kongsberg, 2020) and DSSs based on publicly available Automatic Information System (AIS) data (e.g., Wu et al., 2022). Some research has assessed navigators' use of DSSs using both qualitative methods (e.g., Aylward et al., 2022; Costa et al., 2017) and quantitively methods (Man et al., 2015; Soper et al., 2023). However, there is a knowledge gap concerning empirical data that may shed light on the extent to which a DSS can support remote supervisory performance for autonomous ships. In our aim to address this gap, we sought to test whether the availability of a generic DSS yielded a significant influence on performance compared to when the DSS was not available. To accomplish this, we designed a DSS that contained two generic features: (i) object detection, inspired by recent work on automation transparency techniques developed by Helgesen et al. (2022); and (ii) warning prompts with a sound alert that appeared prior to the critical event, inspired by AI concepts developed by Wu et al. (2022). The generic DSS we tested is not based on any existing product. For more details about the DSS design, see Section 2.3.

Taken together, the factors A through E, although not an exhaustive list, represent important considerations in the effort towards understanding human factor influences in remote supervisory control. To illustrate this with a real-world example, consider the roles of Factors A through E in the Helge Ingstad (HI) accident in 2018. The HI was a Norwegian frigate which collided at night with an oil tanker with 137 people onboard (Norwegian Safety Investigation Authority, 2021). This accident could have led to several fatalities and a devastating oil spill potentially damaging large parts of Norway's coastline. Fortunately, only a few injuries occurred, but the whole frigate had to be scrapped afterwards. The main findings of the accident investigation pointed to multiple contributing factors: namely, the bridge crew lacked sufficient skills and training; the Vehicle Traffic Services (VTS) failed to monitor HI on their screens; the navigator in charge (with limited experience themselves) was responsible for training two persons at the same time as they were navigating the ship; HI was sailing with a speed of 16–17 knots in a congested area leaving a short reaction time for everybody involved; and, finally, the DSS on the bridge was inadequately used and had an inefficient design and location. This example serves to underline the importance of maritime human factor influences on ship safety.

1.4. *Two types of interventions: handover and takeovers*

To approach the research question, we first defined two types of interventions in maritime supervisory control: (i) *handovers*, where the automation hands over control to the operator (or vice versa), and (ii)

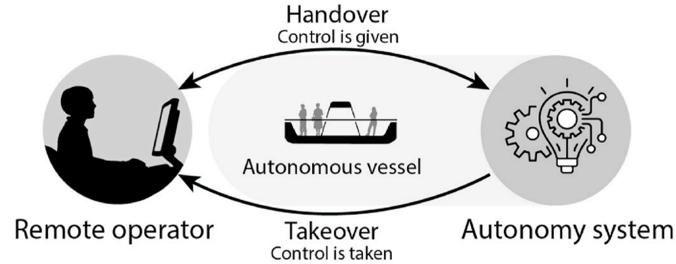

Fig. 2. Illustration of handover and takeover interventions.

takeovers, where the operator takes over control from the automation (Fig. 2). In this study, we frame the investigation of influencing factors on supervisory control around these two types of interventions.

It is important to consider that interventions are not considered failures of the system. Rather, interventions are part of the system design functionality, whose intention is to combine human and autonomous control in way that extends operational capabilities and makes the system more resilient to real-world conditions. A description of the autonomy system and its integration with redundant hardware and the ROC is out of the scope of this work; for more information on this, see Brekke et al. (2022).

2. Method

In the experiment, 32 volunteer participants without any prior knowledge of the experiment or simulator individually completed two virtual scenarios. These were completed in the "Remote Operating Center" (ROC) at the NTNU Shore Control Lab. The two scenarios were initiated with the ferry in autonomous mode, whereupon it crossed back and forth a 100-m-wide urban canal at a top speed of 3 knots, with passengers embarking and disembarking. The choice of the vessel and operational areas was based on the public trials of the "milliAmpere2," which took place from September to October 2022 (Fig. 1). Each participant was assigned the role of a "remote operator" whose role was to "ensure the safety of the ferry and its passengers" and whose primary task was to "monitor the ferry and control it remotely, if necessary."

The key events in the two virtual scenarios were interventions, defined by either a "handover" or a "takeover" (Fig. 2). When under remote control, the operator uses a joystick controller to maneuver the ferry. When appropriate, the remote operator can hand control back to autonomous mode, marking the end of the intervention. Data latency (i.e., the network lag between vessels and the ROC) was not modelled in the simulator and is out of the scope of this experiment.

The NTNU Shore Control Lab test facility allowed the simulator experiment to be undertaken in controlled conditions. The lab is based on a standard simulator set-up: the experimenter orchestrates the simulator from a separate room (Fig. 3a) and the test participant uses the simulator at the ROC (Fig. 3b). Further details about the lab, including its technical specifications and its design process, are available in Veitch et al. (2021).

For gameplay in the simulator, two controllers are available: (i) a customized button panel with basic controls (dock/undock, manual control on/off, switch between vessels, change camera view) and (ii) a joystick controller for maneuvering the vessel and changing camera angle and input source. Details about the simulator and its controllers and configuration are found in Hansen (2022).

2.1. Recruitment of participants and demographics

Sixteen individuals were recruited from each of two groups: (i) gamers and (ii) navigators. Basic selection criteria were for individuals between 18 and 65 years of age who, to permit use of eye-tracking, did not use eyeglasses (corrective contact lenses were permitted). Demographic information is presented in Table 1.

i. **Gamers.** These consisted of individuals who self-identified as gamers, without any strict criteria for how much or how often they played, or what type of games they played. Recruitment channels for gamers included: (i) poster with QR code displayed around the campus of NTNU, (ii) posting on Facebook groups for local interest groups in gaming, (iii) posting in Discord channels for forums about gaming, (iv) snowball sampling through recruits once experiments began. Upon registering, we collected preliminary information about

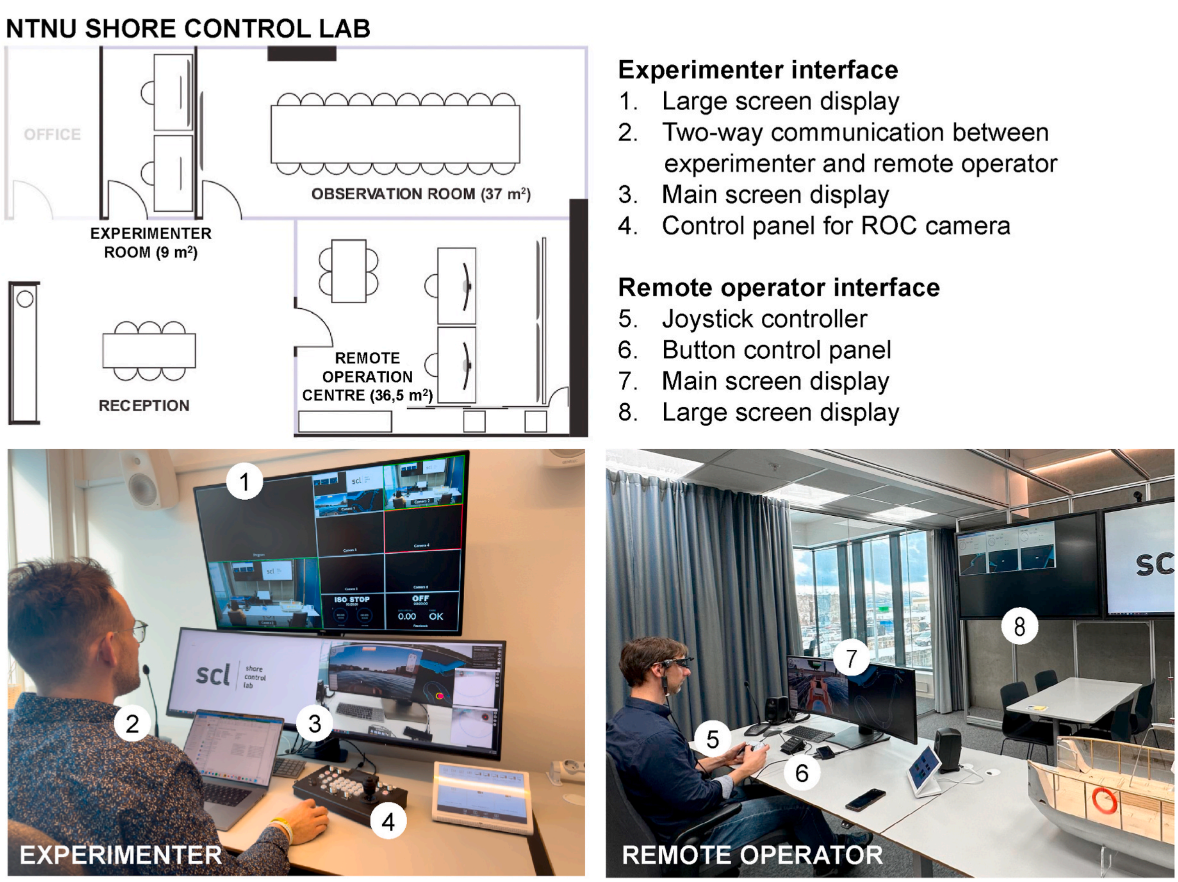

Fig. 3. NTNU Shore Control Lab layout with separate experimenter room (bottom left) and Remote Operating Center (bottom right).

Table 1
Demographic information and skillset characterization for gamer and navigator samples (n = 16 per group).

Gamers				Navigators			
Gender (n/16)				**Gender (n/16)**			
3 females		13 males		2 females		14 males	
Age (years)				**Age (years)**			
Mean	*St. Dev.*	*Min*	*Max*	*Mean*	*St. Dev.*	*Min*	*Max*
25	5.0	20	41	43	9.5	28	62
Total gaming experience (years)				**Total seafaring experience (years)**			
Mean	*St. Dev.*	*Min*	*Max*	*Mean*	*St. Dev.*	*Min*	*Max*
15	5.6	7.0	30	14	8.1	3.0	28
Gaming frequency (hours/week)				**Deck officer certificate[a] (n/16)**			
Mean	*St. Dev.*	*Min*	*Max*	*Class 1:*		11	
16.5	8.9	4.5	35	*Class 2:*		1	
Self-reported skill level (n/16)[b]				*Class 3:*		1	
Expert:		8		*Class 4:*		2	
Advanced:		4		*Class 5:*		1	
Intermediate:		4		*DP certificate:*		10	
Novice:		0		*Auto-crossing:*		2	

[a] Class 1–5 based on Norwegian certificates for seafarers (Norwegian Maritime Authority, 2011).
[b] Based on top three games played in the past three years (highest skill level is reported).

recruits' gaming skills, frequency, and experience, and sampled the most skilled to participate.

ii. **Navigators.** These consisted of individuals with a valid certificate for maritime navigation in Norway, ranging from Class 1 to Class 6 according to the Norwegian Maritime Directorate (Norwegian Maritime Authority, 2011). Recruitment channels for navigators included: (i) email invitations to managers at local companies involved in relevant business sectors (ferry operator, tug services, high-speed ferry operator, harbor authority), (ii) email invitations to managers of local organizations with members potentially consisting of individuals with maritime background (e.g., yacht club, maritime training facility, local traditional boatbuilding club), (iii) snowball sampling through recruits once experiments began.

2.2. Experiment protocol and data collection

A detailed plan for ethical conduct of the experiment and management of personal data was approved by the Norwegian Agency for Shared Services in Education and Research (Sikt) (Project Number 493904). Informed consent was provided by all participants, and all participants were rewarded with a gift card.

Every trial underwent the same procedure according to a prescribed experimental protocol (Fig. 4). The instructor (first author) was present for all trials; the co-instructor varied (either the second or fourth author or a graduate student) but the latter was always briefed on the protocol prior to testing. This helped to control experimental conditions. Additional control variables included the lighting in the room, time of day, participants' skillsets, and participants' baseline familiarization with the simulator. Trials took approximately 1–2 h to complete, depending on the length of the scenarios.

Each trial started with the participant providing informed consent and filling out an "Experience Questionnaire." Following this, participants completed two 10-min training scenarios designed specifically to familiarize participants with the interface, controls, and tasks. Most importantly, this included familiarization with transitions between autonomous mode and manual mode. After completing the training scenarios, the participant completed a "Simulator Sickness Questionnaire" (SSQ; Kennedy et al., 1993) to record their "before test" condition. The SSQ was used to gauge participants' level of comfort before and after data collection, helping to monitor for signs of simulator sickness and to help explain any potential data anomalies.

Next, all data collection devices were calibrated and initiated along with the first of the two test scenarios. After completing Scenario 1, the participant was asked to fill out the "Boredom Questionnaire" (adapted from van Tilburg and Igou, 2012). Then, after a short break, they completed Scenario 2 and filled out a second SSQ to record their "after test" condition.

Table 2 lists the data collected for this study. All data collected

Table 2
Description of collected data.

Data	Description	Source
Demographics	Information about participants, with a focus on relevant skillsets and experience	"Experience Questionnaire" (Veitch, 2023)
Boredom	Participants' subjective feeling of boredom and related affective experiences	"Boredom Questionnaire" (adapted from van Tilburg and Igou, 2012)
Ferry position and heading	Position and heading for all vessels (4 Hz)	Simulator (Hansen, 2022)
Ferry status	Docking status, collision status, boarding status, camera status (4 Hz)	Simulator (Hansen, 2022)
Button presses	All inputs registered on button panel and handheld joystick controller (4 Hz)	Simulator (Hansen, 2022)
Screen recording	Video recordings of all screens used by participants in the simulator	Simulator (Hansen, 2022)

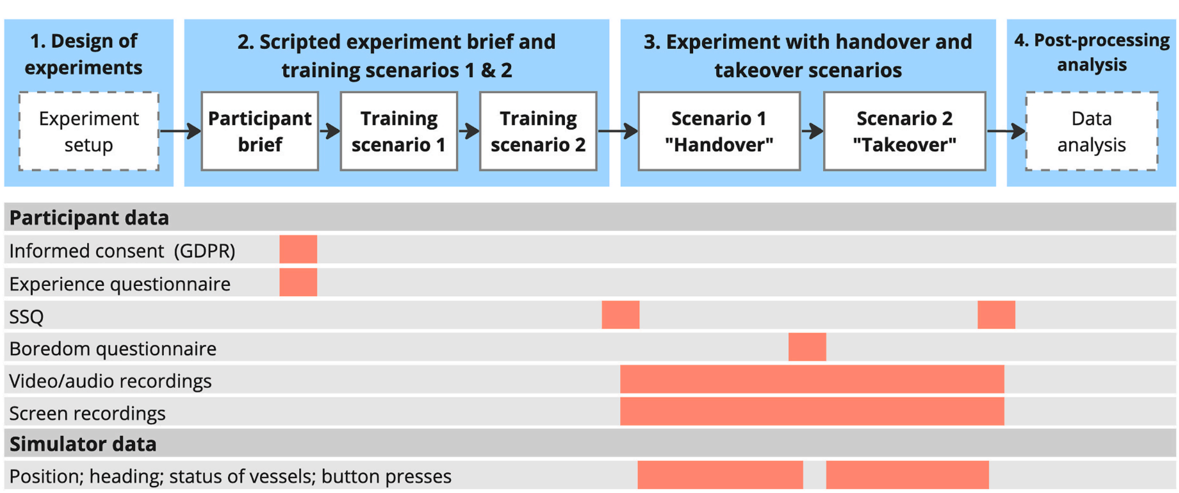

Fig. 4. Experimental protocol and data collection.

during the experiment was archived in the NTNU DataverseNO (available in Veitch, 2023), including copies of the original questionnaires. All numerical recordings were Unix time-stamped to allow for synchronization during post-processing.

2.3. Scenario design & factor settings

In preparation for the experiment, two scenarios were designed in the simulator. Both scenarios were designed to investigate performance for manual intervention in a safety-critical situation.

The design of the scenarios was reached iteratively and was influenced by the Scenario Analysis in the Crisis Intervention and Operability study (CRIOP) framework (Johnsen et al., 2004). Hoem et al. (2021) described this process in detail, using early versions of the simulator scenarios used herein as case studies. The two scenarios were designed to be realistic enough to represent takeover and handover situations in supervisory control, yet straightforward enough to allow drawing "most likely" conclusions from the results. Prior to testing, iterations of scenario designs were tested internally among students, graduate students, and faculty members, as well as by experienced navigators.

From the two simulation scenarios, all test trials were developed. Table 3 lists the five factors and the two levels at which they were tested, which in combination yielded $2^5 = 32$ unique permutations, or "treatments" (for the complete run log, see Veitch, 2023). Each treatment was tested first on Scenario 1, then on Scenario 2—for a total of 64 sets of outcomes.

2.3.1. Intervention scenario 1: "handover"

In this scenario, the autonomous mode hands over control to the remote operator following a failure in the autonomy system. The operator thereafter manually drives the ferry to one of the two ferry terminals using the controller (Fig. 5). The prompt to take over control is an on-screen alert that flashes, accompanied by a beeping sound. Variations of the scenarios according to factor level are listed in Table 3. When the DSS is available, an additional alert message appears in the upper center of the screen 10 s prior to the critical event, reading "Fatal system error! Prepare for POSSIBLE manual control."

Table 3
Factor settings for simulated scenarios.

Factor	Level	Description of Scenario 1: "Handover"	Description of Scenario 2: "Takeover"
Skillset (*A*)	**Gamer**	Participants are individuals with video game-playing skillsets	
	Seafarer	Participants are individuals with maritime navigation skillsets	
Monitoring Time (*B*)	**5 min**	Five minutes passes before handover intervention is initiated	Five minutes passes before takeover intervention is initiated
	30 min	Thirty minutes passes before handover intervention is initiated	Thirty minutes passes before takeover intervention is initiated
Number of Vessels (*C*)	**1 ferry**	Single autonomous vessel in operation for monitoring	
	3 ferries	Three autonomous vessels in operation for simultaneous monitoring	
Available Time (*D*)	**20 s**	20 s available to complete handover before scenario ends	20 s available to take over control and avoid collision
	60 s	60 s available to complete handover before scenario ends	60 s available to take over control and avoid collision
Decision Support System (*E*)	**Available**	DSS interface displayed; warning provided from autonomy system	DSS interface displayed; warning provided about CPA
	Unavailable	Default interface is displayed (e.g., video streams, speed, heading, basic alarms)	

2.3.2. Intervention scenario 2: "takeover"

In this scenario, the autonomous mode cannot resolve a traffic situation and the operator must take over control. The operator needs to identify this situation and take evasive action to avoid the impending collision (Fig. 6). Unbeknownst to the participant, the collision path is programmed so one of the boats will always crash into the ferry unless specific actions are taken to complete the takeover. Variations of the scenarios according to factor level are listed in Table 3. When the DSS is available, an additional orange alert message appears in the upper center of the screen 10 s prior to the collision target leaving its berth that reads "Watch for crossing boat. Target approaching: distance *X* m, speed *Y* knots" (where *X* and *Y* corresponded to the scenario settings).

Note that Scenario 2 (takeover) was always run after Scenario 1 (handover). While learning effects were expected, they were considered negligible due to the novelty of Scenario 2 compared to Scenario 1. For those participants engaged in 30-min scenarios, a short break of about 5–10 min was allotted in between trials.

2.4. Design of experiments

The methodology for structuring the experimental trials and conducting the analysis was based on "Design of Experiments" (DOE; Montgomery, 2017). DOE emphasizes controlled testing and randomized order of test trials and a statistical inference approach to data analysis. The chosen experimental design was a factorial experiment with five factors, each tested at two levels, with a total of $2^5 = 32$ trials (for the complete run log, see Veitch, 2023). No repetition points were conducted, which is acceptable considering the purpose of the experiment is factor screening, not model prediction.

One of the most important features of the experimental design was run order randomization. However, while the factor settings in the simulator were easy to change, the factor setting representing participants' skillsets was relatively hard to change. To handle this "hard-to-change variable," we repeated the experiment in four "blocks:" groups of eight participants in random order. The first two groups were gamers, and the second two groups were navigators. This approach of blocking runs by levels of a hard-to-change variable is called a "split-plot design" (Montgomery, 2017, Chaper 14.5.3). The resulting restricted randomization was accounted for in the data analysis (see Section 2.7).

2.5. Handover scenario metrics: track score and handover time (scenario 1)

The performance metric in Scenario 1 was based on how effectively the operator was able to accept a handover in a timely and controlled way. A perfect score (100%) represents an immediate control transfer with no interval of time passing between automatic control, followed by a chosen track equivalent to the automation track. The track score was computed by comparing the trajectory after the handover to its equivalent in automatic mode (Equation (1)). A score approaching 0% represents a state of remaining stopped (or drifting a small distance towards the dock), having never taken manual control. In the case where a participant reversed direction after handover, their score was computed by comparing to the trajectory in the corresponding reverse direction.

The score is computed at each time step recorded in simulator log data, as follows:

$$\mathrm{Score}_i = 1 - \frac{d_{ai} - d_{mi}}{d_{ai}} \times 100\% = \frac{d_{mi}}{d_{ai}} \tag{1}$$

$$= \sqrt{\frac{(x_{mi} - x_0)^2 + (y_{mi} - y_0)^2}{(x_{ai} - x_0)^2 + (y_{ai} - y_0)^2}} \tag{2}$$

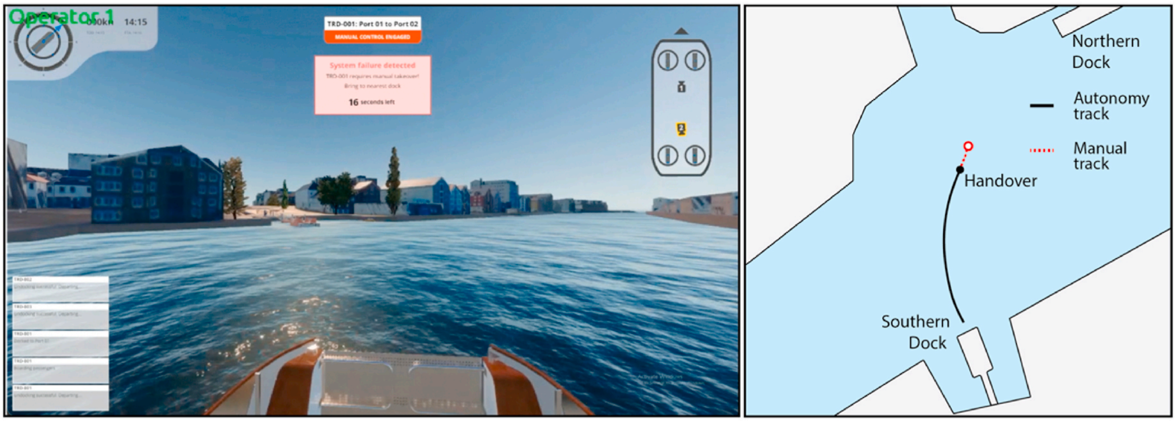

Fig. 5. Handover scenario underway for participant D27 (left) with accompanying scenario illustration (right).

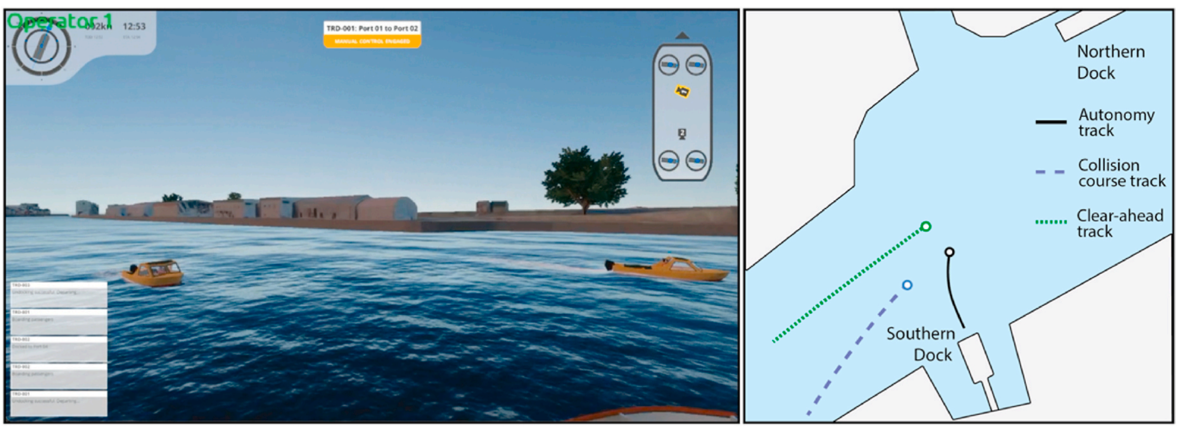

Fig. 6. Takeover scenario underway for participant X77 (left) with accompanying scenario illustration (right).

Where d_{mi} is the absolute distance travelled after handover (manual mode) and d_{ai} is the absolute distance travelled by the baseline (automatic mode). (x_0, y_0) represents the location of the handover event. The final track score was taken as the median of Score_i.

Note that scores of all trials were computed only up to the time series where the baseline trajectory slowed to begin its docking sequence. This was done to fairly compare scores across 20-and 60-s scenarios, the latter of which may obtain higher scores if there is more time to "catch up" to the baseline trajectory at it nears its docking position.

Fig. 7 illustrates the track of participant B03 after handover, where the track score is 63%.

Handover time was computed as the first controller input after the ferry automation failed, measured in seconds. For example, if the first controller input was logged at 1807 s for a scenario where the critical event occurred after 30 min (1800 s), then the reaction time was 7 s.

2.6. Takeover scenario metric: collision outcome (scenario 2)

Takeover performance in Scenario 2 was measured by how effectively the participant was able to avoid a collision. This was done by categorizing each trial as "Collision Avoided," "Near Miss," or "Collision." The categories were chosen by individually analyzing the recorded log files from the simulator in addition to replaying the video from the trial. Fig. 8 shows examples of the three performance categories.

For "collision avoided" and "near miss" categories, the Closest Point of Approach (CPA) was also computed from the log data during the collision event sequence. The CPA accounted for the outermost edge of the two targets using shape approximations of the two hull structures (approximated to the nearest 0.1 m).

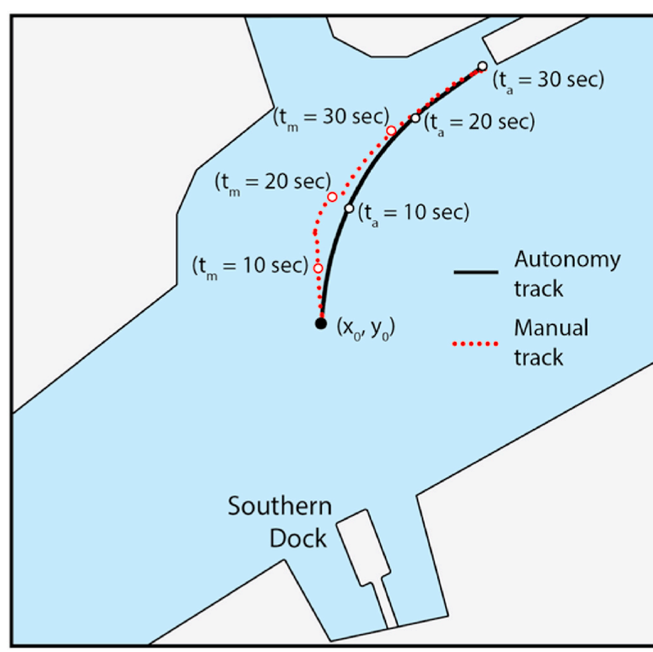

Fig. 7. Track score was computed by comparing manual track to autonomy track; the example presented is participant B03 who achieved a track score of 63%.

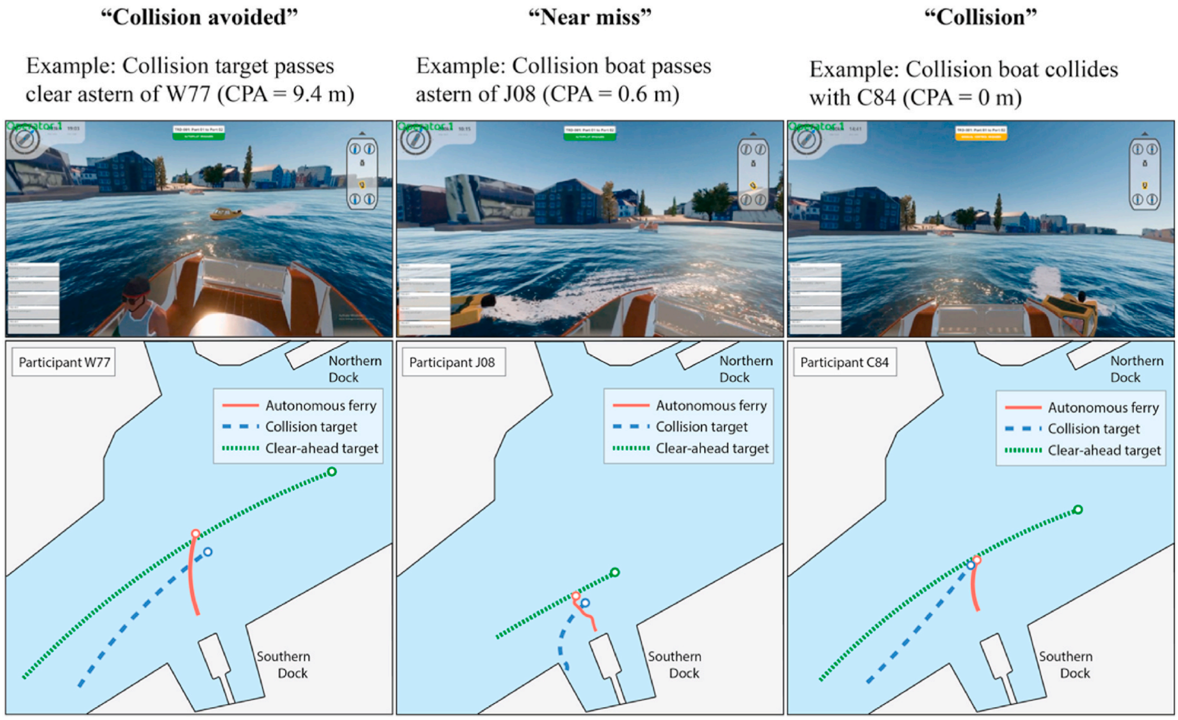

Fig. 8. Collision outcomes for takeover scenario: "collision avoided" (left), "near miss" (center), and "collision" (right).

2.7. Quantitative analysis

The aim of the quantitative analysis was to estimate the extent to which the five factors independently influenced performance metrics. Towards this aim, statistical tests of significance were used to assess the extent to which one could reject the null hypothesis that each factor did not affect the response. In line with the exploratory nature of the study, a significance level of $\alpha = 0.10$ was set as the prescriptive baseline of significance to help minimize risk of Type II errors. Running these statistical tests involved several steps, which were followed strictly to ensure that they were employed appropriately and did not produce misleading results.

i. Compute Analysis of Variance (ANOVA) for the response in question, including all second-order interaction effects. Because we used a split-plot experiment with restricted run-order randomization, a special form of ANOVA was used called Restricted Maximum Likelihood (REML) ANOVA (Corbeil and Searle, 1976).
ii. Remove terms from the model that are not significant, based on a *P*-value greater than 0.10. Hierarchy was maintained during this model selection process, meaning that insignificant first-order effects were included in the case where second-order interaction effects were significant.
iii. The resulting model was checked for appropriateness. First, a plot of model residual errors was visually checked for heteroscasticity (evenness of residual errors versus model predictions). Should the plot not show heteroscasticity, a transformation was applied to the responses and the ANOVA was repeated. Guidance on appropriate transformations was provided by a Box-Cox plot (Box and Cox, 1964).
iv. Once the resulting model showed heteroscasticity (whether a transformation was applied or not), two other diagnostic checks were completed to ensure that all assumptions imposed by ANOVA were appropriately met. First, a normal plot of residuals was visually inspected to ensure that residuals were normally distributed. Second, residuals were plotted against run order to check that the residuals were visually random and independent of run order. Should any of these diagnostics reveal signs that ANOVA assumptions were not met, then the analysis was deemed inappropriate for presentation.
v. The results presented are those that meet the conditions listed above.

3. Results

The results are divided into three sections.

i. The first section presents results of the "Boredom Questionnaires" (Section 3.1).
ii. The second section reports on the handover metrics, assessed in terms of track score and handover time (Section 3.2).
iii. The third section reports on the takeover metrics, assessed in terms of collision avoidance outcome (Section 3.3).

All data reported in this article is available in the DataverseNO repository (Veitch, 2023).

Table 4
ANOVA for boredom responses.

Source of variation	Responses from "Boredom Questionnaire" (Part 1/2)			Responses from "Boredom Questionnaire" (Part 2/2)		
	DF	F-value	*P*-value	DF	F-value	*P*-value
Skillset (*A*)	25.00	1.58	0.2208	26.00	0.0269	0.8710
Monitoring Time (*B*)	25.00	5.68	0.0251*	26.00	3.43	0.0753*
Number of Vessels (*C*)	25.00	0.0002	0.9880	26.00	0.1228	0.7289
Available Time (*D*)	25.00	0.0160	0.9002	26.00	1.39	0.2494
DSS (*E*)	25.00	1.22	0.2600	26.00	0.5375	0.4700

3.1. Boredom results

Table 4 presents ANOVA for the boredom responses collected from the two-part "Boredom Questionnaire." From the results, Monitoring Time (*B*) was the only significant effect at the 90% confidence level ($P <$ 0.10). This result was the same for both parts of the "Boredom Questionnaire."

For the analysis, Likert scale responses from the questionnaires was normalized and an arcsine square root transformation was applied to uphold ANOVA assumptions (see Section 2.7 for details). In the ANOVA table (as in all ANOVA tables presented in this article), we list three test statistics: (i) Degree of freedom (DF; the number of estimated parameters used to compute the effect's sum of squares and variance), (ii) F-value (test for comparing the effect's variance to the residual variance; i.e., signal-to-noise), and (iii) *P*-value (probability of observing the F-value given the null hypothesis is true).

The "Boredom Questionnaire" data is presented in Figs. 9 and 10 in a series of boxplots, arranged by factor and sub-divided by level. Note that one participant did not fill out Questionnaire Part 1 (n = 31). All participants filled out Part 2 (n = 32).

In the boxplots (as with all boxplots presented in this article), thick lines indicate median values and crosses indicate mean values. The upper and lower bounds of the box depict the interquartile range (IQR); the lines extend to the minimum and maximum, unless there are one or more outliers, in which case they extend to 1.5 times the IQR. Outliers are marked as dots.

3.2. Handover scenario results: track score and handover time

Table 5 presents the ANOVA for the track score and handover time metrics, measured in the handover scenario (Scenario 1).

For track score, the ANOVA confirms that the effects of Skillset (*A*), Monitoring Time (*B*), Number of Vessels (*C*), and DSS (*E*) were all significant at the 90% confidence level (*P*-values less than 0.10). The effect of DSS was highly significant (very small *P*-value). Note that a square root transformation was applied to the track score data to adhere to ANOVA assumptions (see Section 2.7 for details).

For handover time, the ANOVA confirms that the effects of Number of Vessels (*C*) and DSS (*E*) were highly significant (very small *P*-value). In addition, three interaction effects were found to be significant (*AC*, *AE*, and *CE*).

The handover metrics are presented in Figs. 11 and 12 as series of boxplots, grouped by factor and sub-divided by level (n = 16 per boxplot).

3.3. Takeover scenario results: collision outcome

Table 6 presents the ANOVA for the collision outcome metric, measured in the takeover scenario (Scenario 2). The ANOVA confirms that the effect of Available Time (*D*) was highly significant (very small *P*-value). Furthermore, the *BD* interaction was highly significant. Note that the data was transformed to a scale between 0 and 1 and an arcsine square root transformation was applied to uphold ANOVA assumptions (see Section 2.7 for details). The transformation mapped "Collision Avoided" to 1, "Near Miss" to 0.5, and "Collision" to 0 (see Section 2.6 for definitions of these three categories).

The takeover scenario performance data is presented in Fig. 13 as a series of stacked bar charts, grouped by factor and sub-divided by level.

4. Discussion

Number of Vessels (*C*) influenced performance during handovers. When the number of autonomous vessels increased from one to three, handover performance diminished. On average, operators with three vessels had a 10% drop in track score (P = 0.06; Fig. 11) and a 3-s increase in handover time (P = 0.01; Fig. 12), compared to those supervising one vessel.

There are several studies in the literature that corroborate the finding that performance of remote operations may be diminished as the number of assets under supervision increases. For example, in one experiment, operators of aerial drones were found to have "significantly degraded" situation awareness when tasked with handling 16 missiles compared to 8 or 12 (Cummings and Guerlain, 2007). In another experiment, air traffic controllers were found to have significantly lower situation awareness when tasked with handling more than 12 aircraft (Endsley and Rodgers, 1996).

That the number of supervised vessels should negatively affect operators' performance runs counter to the promise that autonomous ships will improve overall mission performance, operational efficiency, and safety. This reflects Bainbridge's "ironies of automation," which articulate the paradoxical undermining of human abilities through the use of technologies designed to enhance those very abilities (Bainbridge, 1983). Before remotely operated and autonomous ships can be successfully implemented, the question of how to appropriately support multivessel operations must be addressed.

One way to support multivessel operations is through re-design of the HMI. The interviews with participants after their trials shed light on how this may be done (interview transcripts are available in Veitch, 2023). Specifically, responses to the questions "What can be improved?"

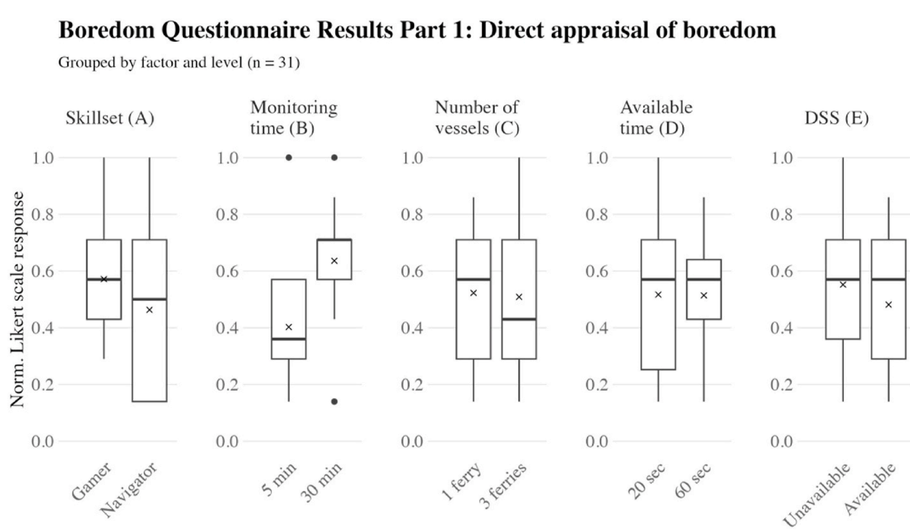

Fig. 9. Boxplots of normalized Likert responses from Boredom Questionnaire Part 1, grouped by factor and sub-divided by level.

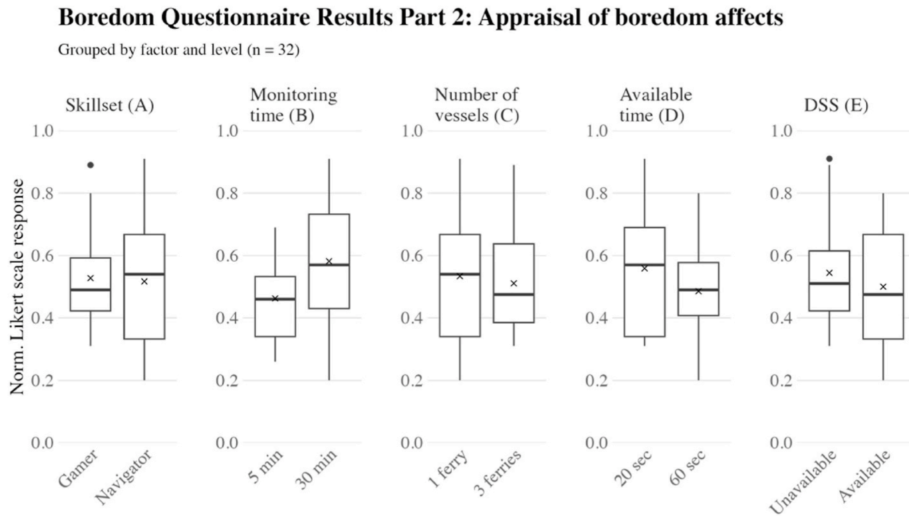

Fig. 10. Boxplots of normalized Likert responses from Boredom Questionnaire Part 2, grouped by factor and sub-divided by level.

Table 5
ANOVA for handover metrics (track score and handover time).

Source of variation	Handover scenario: Track scores			Handover scenario: Handover time		
	DF	F-value	*P*-value	DF	F-value	*P*-value
Skillset (*A*)	26.00	2.93	0.0989*	1.94	4.64	0.1679
Monitoring Time (*B*)	26.00	3.24	0.0836*	17.99	0.3688	0.5513
Number of Vessels (*C*)	26.00	4.01	0.0557*	17.99	8.36	0.0097*
Available Time (*D*)	26.00	2.76	0.1087	17.99	0.3010	0.5900
DSS (*E*)	26.00	21.56	<0.0001*	18.23	21.69	0.0002*
AC	–	–	–	17.99	5.45	0.0314*
AE	–	–	–	18.18	5.51	0.0305*
CE	–	–	–	18.23	8.06	0.0108*

Table 6
ANOVA for takeover metric (collision outcome).

Source of variation	DF	F-value	*P*-value
Skillset (*A*)	2.00	0.1250	0.7575
Monitoring Time (*B*)	23.00	0.2277	0.6377
Number of Vessels (*C*)	23.00	0.9109	0.3498
Available Time (*D*)	23.00	14.57	0.0009*
DSS (*E*)	23.00	<0.0001*	1.0000
BD	23.00	8.20	0.0088*

and "What could have made the scenario easier for you?" uncovered three ideas for improved support of multivessel operations. The most reported suggestion (17 out of 32 participants) was to enhance the tracking functionality of the DSS based on CPA, whereby warnings would we provided for possible collisions. Another common suggestion, (reported by 11 out of 32 participants), was to use 360-degree cameras onboard the vessels. This would remove the need to switch front and back camera angles—a feature of the current interface that was reported to disrupt situational awareness. Another suggestion (reported by 5 out of 32 participants), was to display all vessels simultaneously on a large screen display. Further testing would need to be conducted to verify that these three features, uncovered in exit interviews with participants, would result in measurably improved multivessel operations.

Number of Vessels (*C*) also played an important role in the interaction effects Skillset- Number of Vessels (*AC*; Fig. 14a) and Number of Vessels-DSS (*CE*; Fig. 14c). For example, handover time tripled from 5 to 15 s when the DSS was unavailable for three-ferry operations; by contrast, handover time increased from only 5–8 s when the DSS was unavailable for single-ferry operations (Fig. 14c). This result suggests that the pernicious effects of supervising multiple vessels can largely be circumvented with the availability of a DSS. Considering the importance of time-critical response during handover, the interaction effects *AC* and *CE* underscored the role of decision support in remote supervisory control.

The Decision Support **System (*E*) significantly influenced performance during handovers.** This was reflected in the track scores and handover times for participants in the handover scenario, for which the availability of the DSS had the largest effect on performance of all factors tested (Figs. 11 and 12, respectively).

Interestingly, the DSS did not affect takeover performance. One reason for this may be that takeovers relied more on the operator's own situation awareness-related cognitive processes than on DSS prompts. This is in line Mica Endsley's "Guidelines for the Design of Human-Autonomy Systems," which includes the maxim: "Use automated assistance for carrying out routine tasks rather than higher-level cognitive functions" (Endsley, 2016). From this perspective, the DSS, in attempting to assist the user with high-level cognitive processes involved in collision target diagnosis and avoidance, may have unintentionally been distracting operators more than helping. Compared to takeovers, handovers were a more passive form of intervention that did not summon the full breadth of operators' cognitive powers, which might explain why the DSS helped for handovers, but not for takeovers.

Remote operators experienced boredom after 30 min of passive monitoring, but the effect on intervention performance was minimal. The findings from both questionnaires confirmed that remote operators experienced boredom after 30 min of passive monitoring, but not after 5 min ($P_1 = 0.03$, $P_2 = 0.08$; Figs. 9 and 10). This suggested the presence of a vigilance decrement, in line with the literature (Sections 1.2). The fact that only Monitoring Time (*B*) had a significant effect on boredom affirmed that the vigilance decrement was only a function of time and no other factor tested.

The experiment replicated others' empirical investigations on vigilance decrement, showing that it sets in between 5 and 30 min of passive monitoring (Cummings et al., 2016; Grier et al., 2003; Molloy and Parasuraman, 1996). The main difference between our finding and that of previous research is that previous research has assessed vigilance decrement in terms of secondary task completion while assigned a tedious task. Our approach, by contrast, assessed the subjective experience of boredom in a relatively complex remote operator role with the aid of questionnaires. In addition, our use of questionnaires, which

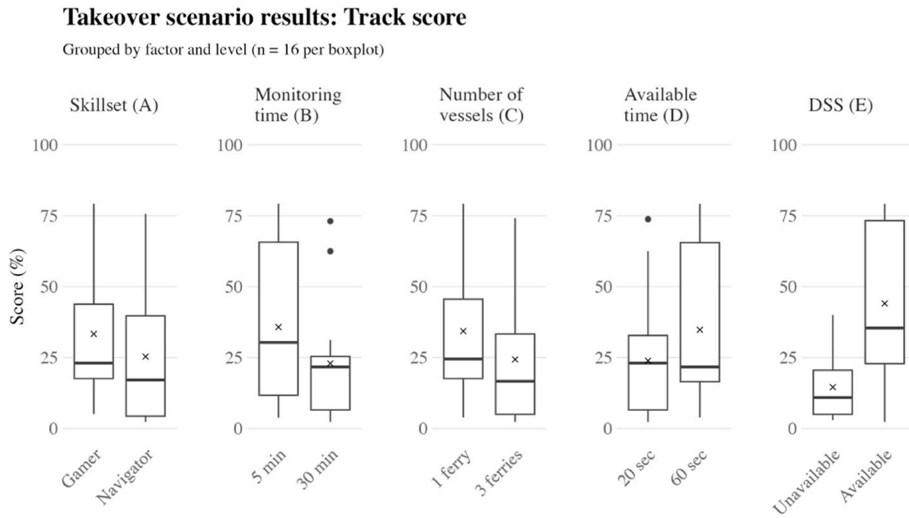

Fig. 11. Boxplot of handover track scores grouped by factor and sub-divided by level.

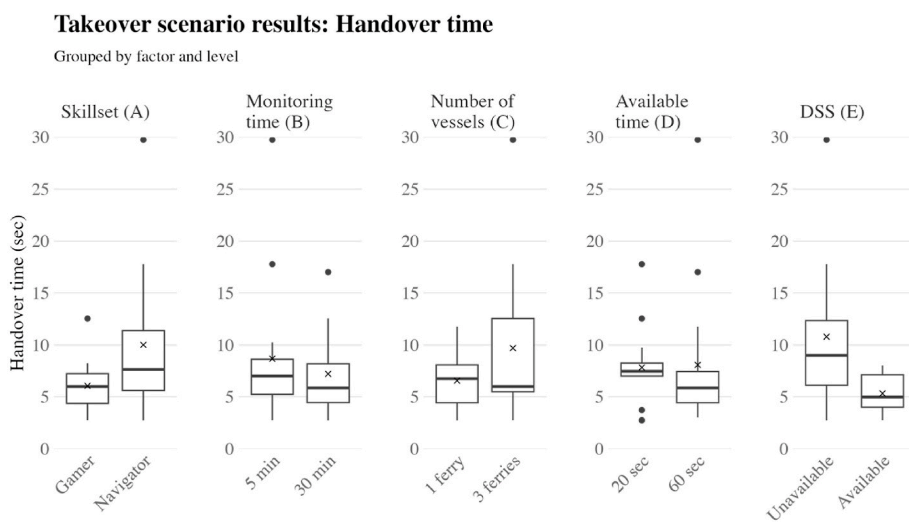

Fig. 12. Boxplot of handover times grouped by factor and sub-divided by level.

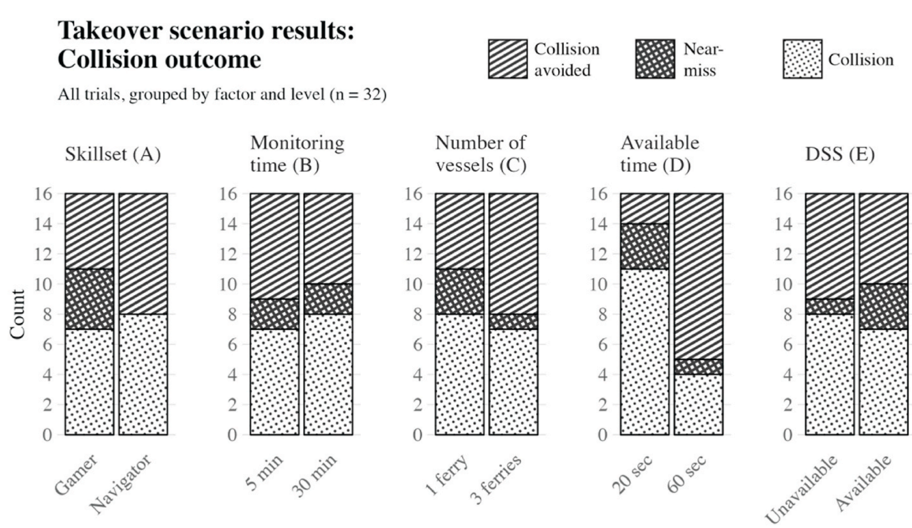

Fig. 13. Stacked bar charts of collision outcomes, grouped by factor and sub-divided by level.

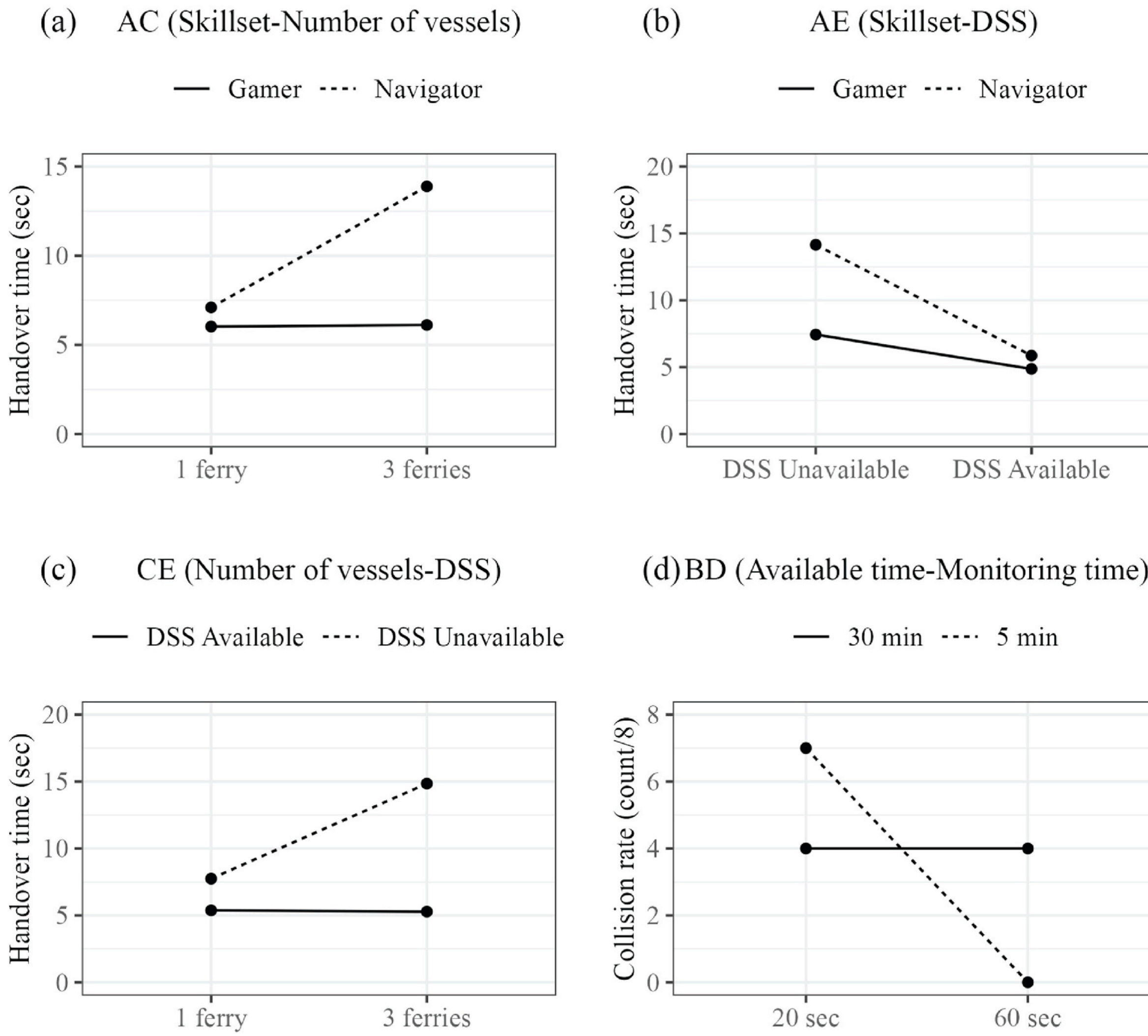

Fig. 14. Interaction plots for handover and takeover results. (a) Skillset-Number of Vessels (*AC*), (b) Skillset-DSS (*AE*), (c) Number of Vessels-DSS (*CE*), (d) Monitoring Time-Available Time (*BD*).

centered the experience of boredom, could be translated to vigilance decrement only to the extent that boredom and vigilance decrement are associated phenomena (Pattyn et al., 2008).

This finding raises the question of whether operator boredom and the associate vigilance decrement led to reduced performance. Indeed, the results showed that operator boredom led to reduced track score in the handover scenario (13% reduction, $P = 0.08$; Fig. 11). Note that while this result was significant ($P \leq 0.10$), it was not highly significant ($P < 0.05$). Moreover, this effect was not detected for other performance metrics (namely, handover time and collision outcome). It follows that we cannot conclude from this experiment that boredom led to reduced performance in general. On the contrary, the findings suggested that remote operators who experienced boredom after extended passive monitoring generally performed similarly to vigilant operators during interventions, despite the clear difference in boredom.

Skillset (*A*) did not influence boredom and the associated vigilance decrement. Prior to this empirical evidence, one might have reasonably argued that gamers, who are accustomed to high-paced video games, might have experienced higher degrees of boredom than navigators during passive monitoring. Navigators, by contrast, are accustomed to sustaining attention during longs periods of relative calm. The evidence counters this prior assumption. Parallels can be drawn from the literature; Warm et al. (2008), for instance, in summarizing research about vigilance, show that vigilance decrement affects experienced and inexperienced operators in equal measure. van Tilburg and Igou (2012), concluded their in-depth studies on boredom by describing boredom as "a chore:" an undesired burden, granted—but one that will be managed irrespective of an individual's skillset or experience.

The interaction of Monitoring Time (*B*) and Available Time (*D*) tipped the balance for collision outcome during takeovers. One instance where the effect of boredom most clearly manifested itself was in the interaction effect Monitoring Time-Available Time (*BD*). This interaction effect was highly significant in terms of collision outcome ($P = 0.009$; Fig. 14d). When interventions occurred after 5 min of monitoring, 20 s was not sufficient time to avoid the collision (100% collision rate) and 60 s was sufficient (100% collision avoidance rate). However, when interventions occurred after 30 min, the amount of available time made no difference to collision outcome (both 20- and 60-s time availability resulted in 50% collision rate) (Fig. 14d). What makes this result striking is that overall, the collision outcomes were impartial to Monitoring Time. So, why were collision outcomes so markedly split in the group with 5 min of monitoring time? And what was it about 5 min of monitoring combined with 20 s available time that led to such dire collision outcomes? One explanation is "cognitive tunneling" (also known as "attentional narrowing"). Cognitive tunneling is an increased selectivity of attention that results from various stressors (Kahneman, 1973). One well known example occurred during the Three Mile Island meltdown, whose investigation revealed that operators had become erroneously fixated on a single indicator showing that water levels were too high in one of the reactors as the explanation for the unfolding event (Rubinstein and Mason, 1979). This fixation effectively "tunneled" their attention and blinded them to information that might have challenged this hypothesis. Similarly, the heightened sense of stress in the 5-min scenarios may have led operators to fixate on features of the interface that effectively blinded them to the impending collision. Indeed, this effect might be amplified by "automation complacency," a state where operators place trust in automatic processes to such an extent that detecting its failures or shortcomings may become challenging (Parasuraman and Riley, 1997; Wickens and Alexander, 2009).

Skillset (*A*) had a minor effect on performance overall, yet its effect shone through during multivessel operations. The difference in performance metrics, when compared across the two Skillset levels (gamer and navigator), was minimal. The only difference was in track score (8% higher for gamers, P = 0.10; Fig. 11). Although, since this difference that was not highly significant ($0.05 < P \leq 0.10$), the evidence is not strong that any significant difference exits between the two groups.

There were two exceptions to this finding, both of which lay in interaction effects on handover time. The first of these was the interaction effect Skillset-Number of Vessels (*AC*) and the second was Skillset-DSS (*AE*). To illustrate the effect of *AC*, consider that when compared to navigators, gamers' handover times in the three-vessel condition were significantly lower than in the single-ferry condition (7-s deviation, P = 0.03; Fig. 14a). As for the effect of *AE*, consider that when compared to navigators, handover times in the DSS-unavailable condition were significantly lower than in the DSS-available condition (6-s deviation, P = 0.01; Fig. 14b). This implies that gamers were quicker to react during multivessel operations and that their swiftness relied less on decision support than navigators.

In the literature, similar observations have been made when comparing gamers and pilots during aerial drone operations. For example, Lin et al. (2015) found that video gaming expertise was associated with better performance during simulated multi-drone operations. Furthermore, McKinley et al. (2011) found that gamers were especially adept at "visually acquiring, identifying, and tracking targets" when compared to conventional pilots. Indeed, cognitive science research has indicated that gamers may have better spatial abilities than non-gamers (Sims and Mayer, 2002), which might explain one mechanism behind their improved performance. Further afield, research on human-machine teaming, which often employs video games as their testing arena, (e.g., Crandall et al., 2018; Shirado and Christakis, 2017) suggests that expert-level gameplay may harbor transferrable skillsets to the unique demands encountered in human-machine interaction applications.

Despite the advantages of lowering handover time under these conditions, though, it is important to consider that neither *AC* nor *AE* interaction effects were detected for the two other performance metrics. This raises the important point that handover time and overall performance were largely independent. Indeed, handover times and track scores were only weakly correlated (Kendall's $\tau = -0.53$, $p < 0.001$, where a strong correlation is represented by $-1.00 \leq \tau < 0.80$). In other words, gamers may have been *faster* in some conditions, but this did not mean they performed *better* than navigators. In fact, the findings may suggest that gamers were characterized as more "trigger happy": quick to react, although not correspondingly more effective in their choice of action. Navigators, by contract, appeared more "level-headed": slower to commit to a course of action, yet more deliberate when they did so.

Taken individually, neither gamers nor navigators portrayed the exclusive skillsets needed for remote operators of autonomous vessels. Instead, both groups possessed unique attributes of relevance to a remote operator. Without a control in our experiment, however, it is impossible to know to what extent gamers and navigators performed better than a group possessing neither skillset. However, we were able to identify that gamers' adeptness at multivessel operations and navigators' deliberateness in intervention execution were both desirable in remote operations. Moreover, it is important to consider that many of the skillsets that the navigators possessed went untested in the experimental setup. For example, the set up did not include standard navigation equipment like radar or chart plotting, nor did the boat traffic in the scenarios adhere to COLREGs—two elements that under most circumstances underpin maritime navigation. These specific navigation skills, which went untested in this experiment, would likely be included in the list of desirable attributes for future remote operators.

Available Time (*D*) influenced takeover performance. The situation awareness demands associated with takeovers appeared to impose a time limit for completion of the intervention. Specifically, it imposed a time limit to complete the three stages commonly associated with situation awareness: namely, information gathering, information processing, and projection (Endsley, 2016). This is analogous to the "maximum response time" defined by Rødseth et al. (2022), expressed as "the maximum time the operator will need to reach the control position, gain situational awareness and be ready to perform actions to maintain safety" (p. 70). In this experimental set up, the "maximum response time" was closer to 60 s (resulting in 4 collisions) than 20 s (resulting in 11 collisions) (Fig. 13). Putting the finding into context, though, one must consider that the experiment took place in a simulated waterway, which featured congested traffic and relatively constrained maximum response time compared to coastal or open water operational envelopes. Still, the finding suggests that there may be ways to effectively increase the maximum response time: for example, by enforcing external constraints like lowering the maximum speed in the waterway.

The results have implications for risk monitoring and management of autonomous vessels. The risks associated with autonomous vessels are of technical, environmental, human, and organizational nature. The outcomes of the study in this paper provide valuable information related to human and organizational risk influencing factors (RIFs). The results constitute a basis for developing safety requirements for systems; for example, related to the time that should be available for the operators for both takeover and handover control. The results are also supported by the findings of Hogenboom et al. (2021), who concluded that not taking the time aspect into consideration for Dynamic Positioning (DP) operations prevents efficient risk mitigation.

In operation, both the autonomous systems and the human operators must be able to control risk (Utne et al., 2017). Situation awareness and decision making is transferred between the human operator and the control system with shifting levels of autonomy. The present study shows, for example, that the availability of a DSS influences the handover performance of the human operators. Furthermore, the results showed that the operators needed sufficient time to react, and that the number of vessels involved influenced the handover time. This underscored the need for having risk monitoring to provide early warnings of potential system deviations outside the operating envelope of the autonomous vessels. This may enhance situation awareness and prepare and support the decision making of operators in critical situations. The RIFs can provide a foundation for the development of risk and safety indicators to be used in risk monitoring, both by the human operators, but also potentially by the control systems of the autonomous systems. Vigilance, for example, is already monitored by some automated systems (e.g., in cars).

The study showed limited positive effect of navigational experience compared to gaming skills in the two scenarios tested. This contradicts findings, for example, in accident investigations, in human reliability studies, and the requirements to certification and training in many high-risk automated industries, such as for DP operations. In light of these contradictory findings, the results may be indicative of the limitations of using a simulator instead of real-life ship operations, especially if the scenarios were not sufficiently complex to heed the full breadth of navigator experience or did not utilize standard navigation tools like radar or ECDIS. Similar tests in real-life ship operations should therefore be investigated in future studies to verify findings in relation to risk monitoring and management.

The results of this study can guide further investigations on maritime human factors. As a starting point, continued research attention should be directed towards those factors producing observable effects on performance (e.g., number of vessels, DSS, available time). However, there are some important limitations that must be considered going forward.

Firstly, due to the design of the experiment, the results are only accurate for determining *whether* the human factor influences investigated existed; any indications of *how much* they influenced performance were limited by the fact that no repetitions of treatments were included in the experimental design. Repetitions are needed for estimates of error and making predictions; our experiment, lacking repetitions, was thus only appropriate for factor screening. Furthermore, the nature of the simulation study means the results represent a *model* of the factor influences. The results must be verified in a real-world setting. At the time of writing, "Degree 3" automation as defined by IMO is still under development; however, as the technology matures, verification tests should eventually be completed to ground our screening study in real-world data. Furthermore, we constrained our model space to five major factors; however, this by no means represents the entire model space of human factor influences on supervisory control. As a final remark on limitations, the applicability of the results must be carefully considered when weighed against other operational envelops and interface designs. Some results may be different in, say, an open-water operational envelop, or if the interface included standard maritime navigation components like radar or ECDIS.

The study also demonstrated some considerable strengths. Firstly, the factorial experimental design approach allowed for a rigorous methodology for investigating factor effects, which can be of value for future experimenters in the field of human factors and HCI. In addition, results are fully transparent given that full dataset is openly available (Veitch, 2023). Moreover, the experiment is repeatable, given that the simulator is available on an open-source license (see Section 1.2). Finally, the two types of interventions adopted in this study (namely, "handovers" and "takeovers") provided a useful typology for studying the distinct mechanisms at play during modal transfers of control. We hope that others may find this typology useful in future studies of human supervisory control.

5. Conclusion

In this study, we screened five influencing factors (labeled *A* through *E*) on human supervisory control performance of highly automated vessels. We defined performance in terms of three metrics assessed during handover and takeover interventions: (i) handover time (sec), (ii) track score (%), and (iii) collision avoidance outcome (collision avoided, near miss, or collision). We found that the effect of Skillset (*A*), defined in this experiment as gamer and navigator skillsets, had only a minor effect on overall performance. However, the interaction effect Skillset-Number of Vessels (*AC*) was significant, with gamer skillsets producing considerably faster handover times under multivessel operations. However, considering that track score and collision avoidance metrics were not affected, this indicated that although gamer skillsets resulted in faster handovers, they did necessarily produce better outcomes. In other words, neither gamer nor navigator skillsets alone produced stand-out performance results in this experiment. This suggests that future remote operators' skillsets should possess a wholly distinct skillset that may lend some attributes from both gamer and navigator skillsets (e.g., multitarget tracking, assessment of traffic regulations).

Monitoring Time (*B*), while significantly affecting the reported levels of boredom experienced during simulated trials, had only minor impact on performance. This showed that although boredom experience was associated with negative affects (e.g., restlessness, the desire to do something more meaningful), the vigilance decrement that was present for operators in 30-min trials yielded comparable intervention performance to those in 5-min scenarios. One interpretation of this result is that boredom and the associated vigilance decrement do not constitute a major hazard in supervisory control. However, this interpretation should be carefully assessed in terms of the potentially long-term effects of boredom, which were not captured in this study.

The effects of Number of Vessels (*C*), Available Time (*D*), and DSS (*E*) were all significant. The effects of *C* and *E* manifested themselves in handover interventions; the effect of *D* was present exclusively in takeover interventions. Overall, these three factors play a significant role in supervisory control of autonomous vessels, which has implication for design and risk management. The interactions of the factors were also important to consider. For example, the interaction effect Number of Vessels-DSS (*AE*) showed that performance-inhibiting effects of supervising three vessels simultaneously was almost entirely circumvented when a DSS was made available.

The factor screening outcomes of this experiment can help guide further research and design activities. For example, aware of the significant effects of a DSS on intervention performance, engineers and designers can be guided towards creating interfaces that leverage these performance enhancements (e.g., during multivessel operations). Towards this aim, the study reifies tenants of human-centered design approaches that encourage designers to test interfaces iteratively and to observe them after implementation to understand how prototypes may produce advantages or trade-offs in performance, efficiency, and safety.

Overall, there are several outcomes of the study. Firstly, it contributes to providing a basis for improved research on risk assessment and modeling of human and organizational factors, as well as to risk mitigation related to human operator performance. Secondly, it contributes to engineering and human-centered design activities aimed at developing remote supervisory infrastructure for autonomous maritime vessels. Finally, it outlines a path forward for continued experimental testing to advance maritime human factors research. Unlocking the potential of reliable, safe, and sustainable automation technology will hinge on the extent to which well-designed interfaces can provide seamless interaction with human supervisors.

CRediT authorship contribution statement

Erik Veitch: Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Conceptualization. **Ole Andreas Alsos:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition. **Tingting Cheng:** Methodology. **Kristin Senderud:** Methodology, Investigation. **Ingrid Bouwer Utne:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset is published on the NTNU DataVerseNO repository (https://doi.org/10.18710/WYFMMP)

Acknowledgments

The authors gratefully acknowledge 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 team behind the Shore Control Lab, especially Felix Petermann for his technical support during experiments, Mikael Hansen for his programming support, and Leander Pantelatos for his assistance with recruitment posters. The authors also acknowledge the support of Dr. Ingrid Bouwer Utne and Dr. Egil Eide for their help in recruiting navigators to the study, as well as that of Dag Rutledal and Kjetil Vasstein for their valuable inputs on the Experience Questionnaires and on recruitment channels and that of Koen van de Merwe for his valuable inputs and assistance with data collection. Finally, we would like to thank the three anonymous reviewers who contributed significantly to revisions of this article.

References

Alsos, O.A., Veitch, E., Pantelatos, L., Vasstein, K., Eide, E., Petermann, F.-M., Breivik, M., 2022. NTNU Shore Control Lab: designing shore control centres in the age of autonomous ships. J. Phys. Conf. 2311 (1), 012030 https://doi.org/10.1088/1742-6596/2311/1/012030.

Aylward, K., Weber, R., Lundh, M., MacKinnon, S.N., Dahlman, J., 2022. Navigators' views of a collision avoidance decision support system for maritime navigation. J. Navig. 75 (5), 1035–1048. https://doi.org/10.1017/S0373463322000510.

Bainbridge, L., 1983. Ironies of automation. Automatica 19 (6), 775–779. https://doi.org/10.1016/0005-1098(83)90046-8.

Baldauf, M., Fischer, S., Kitada, M., Mehdi, R.A., Al-Quhali, M.A., Fiorini, M., 2019. Merging conventionally navigating ships and mass—merging VTS, FOC and SCC? TransNav 13 (3), 495–501. https://doi.org/10.12716/1001.13.03.02. Scopus.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. Roy. Stat. Soc. B 26 (2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Brekke, E.F., Eide, E., Eriksen, B.-O.H., Wilthil, E.F., Breivik, M., Skjellaug, E., Helgesen, Ø.K., Lekkas, A.M., Martinsen, A.B., Thyri, E.H., Torben, T., Veitch, E., Alsos, O.A., Johansen, T.A., 2022. milliAmpere: an autonomous ferry prototype. J. Phys. Conf. 2311 (1), 012029 https://doi.org/10.1088/1742-6596/2311/1/012029.

Corbeil, R.R., Searle, S.R., 1976. Restricted maximum likelihood (REML) estimation of variance components in the mixed model. Technometrics 18 (1), 31–38. https://doi.org/10.1080/00401706.1976.10489397.

Costa, N.A., Holder, E., MacKinnon, S.N., 2017. Implementing human centred design in the context of a graphical user interface redesign for ship manoeuvring. Int. J. Hum. Comput. Stud. 100, 55–65. https://doi.org/10.1016/j.ijhcs.2016.12.006.

Crandall, J.W., Oudah, M., Tennom, Ishowo-Oloko, F., Abdallah, S., Bonnefon, J.-F., Cebrian, M., Shariff, A., Goodrich, M.A., Rahwan, I., 2018. Cooperating with machines. Nat. Commun. 9 (1), 233. https://doi.org/10.1038/s41467-017-02597-8.

Cummings, M.L., Gao, F., Thornburg, K.M., 2016. Boredom in the workplace: a new look at an old problem. Hum. Factors 58 (2), 279–300. https://doi.org/10.1177/0018720815609503.

Cummings, M.L., Guerlain, S., 2007. Developing operator capacity estimates for supervisory control of autonomous vehicles. Hum. Factors 49 (1), 1–15. https://doi.org/10.1518/001872007779598109.

Cummings, M.L., Mastracchio, C., Thornburg, K.M., Mkrtchyan, A., 2013. Boredom and distraction in multiple unmanned vehicle supervisory control. Interact. Comput. 25 (1), 34–47. https://doi.org/10.1093/iwc/iws011.

Cummings, M.L., Mitchell, P.J., 2006. Automated scheduling decision support for supervisory control of multiple UAVs. J. Aero. Comput. Inf. Commun. 3 (6), 294–308. https://doi.org/10.2514/1.19599.

Deraj, R., Kumar, R.S.S., Alam, M.S., Somayajula, A., 2023. Deep reinforcement learning based controller for ship navigation. Ocean Eng. 273, 113937 https://doi.org/10.1016/j.oceaneng.2023.113937.

Dixon, S.R., Wickens, C.D., Chang, D., 2005. Mission control of multiple unmanned aerial vehicles: a workload analysis. Hum. Factors 47 (3), 479–487. https://doi.org/10.1518/001872005774860005.

Dybvik, H., Veitch, E., Steinert, M., 2020. Exploring challenges with designing and developing Shore Control Centers (SCC) for autonomous ships. Proceedings of the Design Society: DESIGN Conference 1, 847–856. https://doi.org/10.1017/dsd.2020.131.

Ek, S., 2022. *Design Of Simulator for Researching Autonomous Marine Vessels* [Master Thesis. NTNU. https://bibsys-almaprimo.hosted.exlibrisgroup.com/permalink/f/13q4kuj/BRAGE11250/2996832.

Emad, G.R., Enshaei, H., Ghosh, S., 2022. Identifying seafarer training needs for operating future autonomous ships: a systematic literature review. Australian Journal of Maritime & Ocean Affairs 14 (2), 114–135. https://doi.org/10.1080/18366503.2021.1941725.

Endsley, M.R., 2016. Designing for Situation Awareness: an Approach to User-Centered Design. CRC Press, Boca Raton, FL, USA.

Endsley, M.R., Rodgers, M.D., 1996. Attention distribution and situation awareness in air traffic control. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 40 (2), 82–85. https://doi.org/10.1177/154193129604000216.

Grier, R.A., Warm, J.S., Dember, W.N., Matthews, G., Galinsky, T.L., Szalma, J.L., Parasuraman, R., 2003. The vigilance decrement reflects limitations in effortful attention, not mindlessness. Hum. Factors 45 (3), 349–359. https://doi.org/10.1518/hfes.45.3.349.27253. Scopus.

Guo, C., Utne, I.B., 2022. Development of risk indicators for losing navigational control of autonomous ships. Ocean Eng. 266, 113204 https://doi.org/10.1016/j.oceaneng.2022.113204.

Gutzwiller, R.S., Wickens, C.D., Clegg, B.A., 2016. The role of time on task in multi-task management. Journal of Applied Research in Memory and Cognition 5 (2), 176–184. https://doi.org/10.1016/j.jarmac.2016.04.003.

Hansen, M., 2022. *Developing a Video Game for Research and Prototyping of Unmanned Maritime Vessels* [Master Thesis, NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3028969.

Helgesen, Ø.K., Vasstein, K., Brekke, E.F., Stahl, A., 2022. Heterogeneous multi-sensor tracking for an autonomous surface vehicle in a littoral environment. Ocean Eng. 252, 111168 https://doi.org/10.1016/j.oceaneng.2022.111168.

Hoem, Å.S., Rødseth, Ø.J., Johnsen, S.O., 2021. Adopting the CRIOP framework as an interdisciplinary risk analysis method in the design of remote control centre for maritime autonomous systems. In: Arezes, P.M., Boring, R.L. (Eds.), Proceedings of the AHFE 2021 Virtual Conferences on Safety Management and Human Factors, and Human Error, Reliability, Resilience, and Performance. Springer Nature Switzerland, pp. 219–227. https://doi.org/10.1007/978-3-030-80288-2_26.

Hogenboom, S., Parhizkar, T., Vinnem, J.E., 2021. Temporal decision-making factors in risk analyses of dynamic positioning operations. Reliab. Eng. Syst. Saf. 207, 107347 https://doi.org/10.1016/j.ress.2020.107347.

Hutchins, E., 1995. Cognition in the Wild. MIT press.

IMO, 1972. International Regulations Preventing Collisions at Sea (COLREGs) (Adopted 20 October 1972, Entered into Force 15 July 1977) 1050 UNTS 16 (COLREGs).

IMO, 2017. International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (STCW) 1978, as Amended in 1995/2010. International Maritime Organisation.

IMO, 2021. Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS). International Maritime Organization. MSC.1/Circ.1638). https://www.imo.org/en/MediaCentre/PressBriefings/pages/MASSRSE2021.aspx.

IMO, 2022. Maritime Safety Committee (MSC 106), pp. 2–11. *November 2022*. https://www.imo.org/en/MediaCentre/MeetingSummaries/Pages/MSC-106.aspx.

Johnsen, S.O., Bjørkli, C., Steiro, T., Fartum, H., Haukenes, H., Ramberg, J., Skriver, J., Asa, N.H., 2004. *CRIOP®: A Scenario Method for Crisis Intervention And Operability Analysis* (SINTEF A4312). SINTEF Digital. https://www.sintef.no/en/publications/publication/1269001/.

Kahneman, D., 1973. Attention and Effort. Prentice-Hall, Inc.

Kari, R., Steinert, M., 2021. Human factor issues in remote ship operations: lesson learned by studying different domains. J. Mar. Sci. Eng. 9 (4), 385. https://doi.org/10.3390/jmse9040385.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G., 1993. Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. Int. J. Aviat. Psychol. 3 (3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3.

Kim, D., Kim, J.-S., Kim, J.-H., Im, N.-K., 2022. Development of ship collision avoidance system and sea trial test for autonomous ship. Ocean Eng. 266, 113120 https://doi.org/10.1016/j.oceaneng.2022.113120.

Kongsberg, 2020. First Adaptive Transit on Bastøfosen VI. February 13. https://www.kongsberg.com/maritime/about-us/news-and-media/news-archive/2020/first-adaptive-transit-on-bastofosen-vi/.

Laugaland, J.M., 2022. Stor Mangel På Sjøfolk – Lovar Bonus På 7000 for Rekruttering. August 24. NRK. https://www.nrk.no/rogaland/stor-mangel-pa-sjofolk-_-lovar-bonus-pa-7000-for-rekruttering-1.16074489.

Lin, J., Wohleber, R., Matthews, G., Chiu, P., Calhoun, G., Ruff, H., Funke, G., 2015. Video game experience and gender as predictors of performance and stress during supervisory control of multiple unmanned aerial vehicles. Proc. Hum. Factors Ergon. Soc. Annu. Meet. 59 (1), 746–750.

Liu, C., Chu, X., Wu, W., Li, S., He, Z., Zheng, M., Zhou, H., Li, Z., 2022. Human–machine cooperation research for navigation of maritime autonomous surface ships: a review and consideration. Ocean Eng. 246, 110555 https://doi.org/10.1016/j.oceaneng.2022.110555.

Liu, D., Peterson, T., Vincenzi, D., Doherty, S., 2016. Effect of time pressure and target uncertainty on human operator performance and workload for autonomous unmanned aerial system. Human Factors in Digital Industrial Systems 51, 52–58. https://doi.org/10.1016/j.ergon.2015.01.010.

Liu, J., Aydin, M., Akyuz, E., Arslan, O., Uflaz, E., Kurt, R.E., Turan, O., 2022. Prediction of human–machine interface (HMI) operational errors for maritime autonomous surface ships (MASS). J. Mar. Sci. Technol. 27 (1), 293–306. https://doi.org/10.1007/s00773-021-00834-w.

Lutzhoft, M., Hynnekleiv, A., Earthy, J.V., Petersen, E.S., 2019. Human-centred maritime autonomy—an ethnography of the future. J. Phys. Conf. 1357, 012032 https://doi.org/10.1088/1742-6596/1357/1/012032.

MacKinnon, S.N., Man, Y., Lundh, M., Porathe, T., 2015. Command and control of unmanned vessels: keeping shore based operators in-the-loop. In: ATENA Conferences System, NAV 2015 18th International Conference on Ships and Shipping Research, pp. 24–25. Milan, Italy.

MacKinnon, S.N., Weber, R., Olindersson, F., Lundh, M., 2020. Artificial Intelligence in maritime navigation: a human factors perspective. In: Stanton, N. (Ed.), Advances in Human Aspects of Transportation. Springer International Publishing, pp. 429–435.

Man, Y.M., Lundh, M., Porathe, T., MacKinnon, S.N., 2015. From desk to field—human factor issues in remote monitoring and controlling of autonomous unmanned vessels. Procedia Manuf. 3, 2674–2681. https://doi.org/10.1016/j.promfg.2015.07.635.

Man, Y.M., Weber, R., Cimbritz, J., Lundh, M., MacKinnon, S.N., 2018. Human factor issues during remote ship monitoring tasks: an ecological lesson for system design in a distributed context. Int. J. Ind. Ergon. 68, 231–244. https://doi.org/10.1016/j.ergon.2018.08.005.

Martinsen, A.B., Bitar, G., Lekkas, A.M., Gros, S., 2020. Optimization-based automatic docking and berthing of ASVs ssing exteroceptive sensors: theory and experiments. IEEE Access 8, 204974–204986. https://doi.org/10.1109/ACCESS.2020.3037171.

McKinley, R.A., McIntire, L.K., Funke, M.A., 2011. Operator selection for unmanned aerial systems: comparing video game players and pilots. Aviat Space Environ. Med. 82 (6), 635–642. https://doi.org/10.3357/ASEM.2958.2011.

Molloy, R., Parasuraman, R., 1996. Monitoring an automated system for a single failure: vigilance and task complexity effects. Hum. Factors 38 (2), 311–322. https://doi.org/10.1177/001872089606380211.
Montgomery, D.C., 2017. Design and Analysis of Experiments, ninth ed. John Wiley & Sons, Inc., Hoboken, NJ, USA.
Negenborn, R.R., Goerlandt, F., Johansen, T.A., Slaets, P., Valdez Banda, O.A., Vanelslander, T., Ventikos, N.P., 2023. Autonomous ships are on the horizon: here's what we need to know. Nature 615 (7950), 30–33. https://doi.org/10.1038/d41586-023-00557-5.
Norwegian Maritime Authority, 2011. Forskrift Om Kvalifikasjoner Og Sertifikater for Sjøfolk. https://www.sdir.no/sjofart/regelverk/rundskriv/kvalifikasjoner-og-sertifikater-for-sjofolk/.
Norwegian Safety Investigation Authority, 2021. *Part Two Report on the Collision between the Frigate HNOMS "Helge Ingstad" and the Oil Tanker Sola TS outside the Sture Terminal in the Hjeltefjord In Hordaland County On 8 November 2018* (Marine 2021/05). Norwegian Safety Investigation Authority. https://www.nsia.no/Marine/Published-reports/2021-05-eng.
Öztürk, Ü., Akdağ, M., Ayabakan, T., 2022. A review of path planning algorithms in maritime autonomous surface ships: navigation safety perspective. Ocean Eng. 251, 111010 https://doi.org/10.1016/j.oceaneng.2022.111010.
Parasuraman, R., Riley, V., 1997. Humans and automation: use, misuse, disuse, abuse. Hum. Factors 39 (2), 230–253. https://doi.org/10.1518/001872097778543886.
Pattyn, N., Neyt, X., Henderickx, D., Soetens, E., 2008. Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? Physiol. Behav. 93 (1), 369–378. https://doi.org/10.1016/j.physbeh.2007.09.016.
Peeters, G., Kotzé, M., Afzal, M.R., Catoor, T., Van Baelen, S., Geenen, P., Vanierschot, M., Boonen, R., Slaets, P., 2020. An unmanned inland cargo vessel: design, build, and experiments. Ocean Eng. 201, 107056 https://doi.org/10.1016/j.oceaneng.2020.107056.
Ramos, M.A., Thieme, C.A., Utne, I.B., Mosleh, A., 2020. Human-system concurrent task analysis for maritime autonomous surface ship operation and safety. Reliab. Eng. Syst. Saf. 195, 106697 https://doi.org/10.1016/j.ress.2019.106697.
Ramos, M.A., Utne, I.B., Mosleh, A., 2018. On Factors Affecting Autonomous Ships Operators Performance in a Shore Control Center. September 16. PSAM 14 - Probabilistic Safety Assessment and Management, Los Angeles, CA, USA.
Rødseth, Ø.J., Wennersberg, L.A.L., Nordahl, H., 2022. Towards approval of autonomous ship systems by their operational envelope. J. Mar. Sci. Technol. 27, 67–76. https://doi.org/10.1007/s00773-021-00815-z.
Rongcai, Z., Hongwei, X., Kexin, Y., 2023. Autonomous collision avoidance system in a multi-ship environment based on proximal policy optimization method. Ocean Eng. 272, 113779 https://doi.org/10.1016/j.oceaneng.2023.113779.
Rubinstein, E., Mason, J.F., 1979. The accident that shouldn't have happened: a narrative account of what is believed to have occurred, based on reports from many experts. IEEE Spectrum 16 (11), 33–57. https://doi.org/10.1109/MSPEC.1979.6368289.
Saha, R., 2021. Mapping Competence Requirements for Future Shore Control Center Operators. Maritime Policy and Management. https://doi.org/10.1080/03088839.2021.1930224.
Sharma, A., Kim, T., 2021. Exploring technical and non-technical competencies of navigators for autonomous shipping. Marit. Pol. Manag. 49 (6), 831–849. https://doi.org/10.1080/03088839.2021.1914874.
Shirado, H., Christakis, N.A., 2017. Locally noisy autonomous agents improve global human coordination in network experiments. Nature 545 (7654), 370–374. https://doi.org/10.1038/nature22332.
Sims, V.K., Mayer, R.E., 2002. Domain specificity of spatial expertise: the case of video game players. Appl. Cognit. Psychol. 16 (1), 97–115. https://doi.org/10.1002/acp.759.
Soper, J., Veitch, E., Thistle, R., Smith, J., Veitch, B., 2023. An analysis of factors influencing ice management performance in an experimental marine simulator and their application to decision support system design. J. Offshore Mech. Arctic Eng. 1–21. https://doi.org/10.1115/1.4063617.
Suchman, L.A., 2007. Human-machine Reconfigurations: Plans and Situated Actions, second ed. Cambridge University Press.
Suyama, R., Miyauchi, Y., Maki, A., 2022. Ship trajectory planning method for reproducing human operation at ports. Ocean Eng. 266, 112763 https://doi.org/10.1016/j.oceaneng.2022.112763.
Tavakoli, S., Khojasteh, D., Haghani, M., Hirdaris, S., 2023. A review on the progress and research directions of ocean engineering. Ocean Eng. 272, 113617 https://doi.org/10.1016/j.oceaneng.2023.113617.
Turan, O., Kurt, R.E., Arslan, V., Silvagni, S., Ducci, M., Liston, P., Schraagen, J.M., Fang, I., Papadakis, G., 2016. Can we learn from aviation: safety enhancements in transport by achieving human orientated resilient shipping environment. Transport Research Arena TRA2016 14, 1669–1678. https://doi.org/10.1016/j.trpro.2016.05.132.
Unity Technologies, 2022. Unity (2022.2.8) [Computer Software].
Utne, I.B., Sørensen, A.J., Schjølberg, I., 2017. *Risk Management of Autonomous Marine Systems and Operations*. ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering, Trondheim, Norway. September 25. https://doi.org/10.1115/OMAE2017-61645.
Vagale, A., Oucheikh, R., Bye, R.T., Osen, O.L., Fossen, T.I., 2021. Path planning and collision avoidance for autonomous surface vehicles I: a review. J. Mar. Sci. Technol. 26 (4), 1292–1306. https://doi.org/10.1007/s00773-020-00787-6.
van Tilburg, W.A.P., Igou, E.R., 2012. On boredom: lack of challenge and meaning as distinct boredom experiences. Motiv. Emot. 36 (2), 181–194. https://doi.org/10.1007/s11031-011-9234-9.
Veitch, E., 2023. Dataset for: Human Factor Influences on Supervisory Control of Remotely Operated and Autonomous Vessels (Version 2). DataverseNO. https://doi.org/10.18710/WYFMMP.
Veitch, E., Alsos, O.A., 2022. A systematic review of human-AI interaction in autonomous ship systems. Saf. Sci. 152, 105778 https://doi.org/10.1016/j.ssci.2022.105778.
Veitch, E., Dybvik, H., Steinert, M., Alsos, O.A., 2022. Collaborative work with highly automated marine navigation systems. Comput. Support. Coop. Work. https://doi.org/10.1007/s10606-022-09450-7.
Veitch, E., Kaland, T., Alsos, O.A., 2021. Design for resilient human-system interaction in autonomy: the case of a shore control centre for unmanned ships. Proceedings of the Design Society 1, 1023–1032. https://doi.org/10.1017/pds.2021.102.
Wang, S., Yan, X., Ma, F., Wu, P., Liu, Y., 2022. A novel path following approach for autonomous ships based on fast marching method and deep reinforcement learning. Ocean Eng. 257, 111495 https://doi.org/10.1016/j.oceaneng.2022.111495.
Warm, J.S., Parasuraman, R., Matthews, G., 2008. Vigilance requires hard mental work and is stressful. Hum. Factors 50 (3), 433–441. https://doi.org/10.1518/001872008X312152.
Wickens, C.D., Alexander, A.L., 2009. Attentional tunneling and task management in synthetic vision displays. Int. J. Aviat. Psychol. 19 (2), 182–199. https://doi.org/10.1080/10508410902766549.
Wróbel, K., Gil, M., Chae, C.-J., 2021. On the influence of human factors on safety of remotely-controlled merchant vessels. Appl. Sci. 11 (3), 1145. https://doi.org/10.3390/app11031145.
Wu, B., Li, G., Zhao, L., Aandahl, H.-I.J., Hildre, H.P., Zhang, H., 2022. Navigating patterns analysis for on-board guidance support in crossing collision avoidance operations. IEEE Intelligent Transportation Systems Magazine 14 (3), 62–77. https://doi.org/10.1109/MITS.2021.3108473.
Yoshida, M., Shimizu, E., Sugomori, M., Umeda, A., 2020. Regulatory requirements on the competence of remote operator in maritime autonomous surface ship: situation awareness, ship sense and goal-based gap analysis. Appl. Sci. 10 (23), 1–27. https://doi.org/10.3390/app10238751.