Doctoral theses at NTNU, 2023:281

Erik Veitch

Designing for Land-based Control of Autonomous Vessels

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Architecture and Design
Department of Design

NTNU
Norwegian University of
Science and Technology

Erik Veitch

Designing for Land-based Control of Autonomous Vessels

Thesis for the Degree of Philosophiae Doctor

Trondheim, September 2023

Norwegian University of Science and Technology
Faculty of Architecture and Design
Department of Design

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Architecture and Design
Department of Design

ISBN 978-82-326-7262-2 (printed ver.)
ISBN 978-82-326-7261-5 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:281

Printed by NTNU Grafisk senter

To Brian Veitch

"There are some enterprises in which a careful disorderliness is the true method."

—Ishmael in *Moby Dick*, describing a whaling voyage (Herman Melville)

Abstract

Background: Autonomous ships can navigate without crew but require human supervisory control from land. This hybrid form of maritime navigation allows remote supervisors to control many ships at once. However, the success of this new form of navigation depends on the extent to which humans can monitor vessels and intervene when needed. Today, relatively little is known about how best to design for land-based control of autonomous ships.

Aim: The research aim is to explore relevant aspects of human-machine collaboration in land-based control of autonomous ships. Four research questions are investigated:

(1) What is the research state-of-the-art?
(2) What is the design state-of-the-art?
(3) How can design principles be applied to remote control center design?
(4) What factors influence remote supervisory control of autonomous ships?

Method: The study adopted four methods corresponding to the four research questions:

(1) Systematic literature review (review of 42 peer-reviewed articles)
(2) Qualitative analysis (interviews with 14 experts)
(3) Interaction design techniques (field study, workshops, prototyping)
(4) Design of experiments (test of five factor effects in a simulator experiment)

Contributions: The thesis outlines four contributions:

(1) Synthesis of the research state-of-the-art
(2) Description of design challenges facing remote control center design
(3) Design of a full-scale prototype remote control center
(4) Empirical results on important factors for human supervisory performance

Conclusions: Design for human supervisory control is a growing research topic and will be central to assuring the safety of autonomous ship operations. The most important design challenge constitutes aligning the technology with in-situ operational demands. Interaction

design can contribute to this aim and help define the role of control center operators. Testing using a simulator experiment has proven an effective empirical tool, highlighting the importance of, among other things, supporting multitasking, time pressure, and decision support in the human-machine interface.

Preface

I discovered as a master's student that a person drives a ship. This fact might strike some readers as obvious—even banal. But the truth is that after five years of studying Ocean and Naval Architectural Engineering, I had never once learned this. Nor did I learn this during my two-year stint as an ocean engineering consultant upon graduating. No, it took a simple experiment conducted in a simulator during the first year of my master's program to discover the obvious: you can have one vessel, but you'll have exactly as many ways to drive it as you have people driving it.

That fascinated me. All that time studying hydrodynamic boundary layers and bending moments of structural steel had obscured a simple fact: the ship is a human system. It was Shackleton who explored Antarctica, not the *Endurance*; it was Nansen who explored the North Pole, not the *Fram*. My master's thesis attempted to "reverse engineer" the captain's role aboard a ship, using a simulator to find out how skilled navigators maneuvered their ships during challenging operations and subsequently trying to reduce their panache to a science. If was backpedaling as a naval architect, I was at least starting my research career.

When I heard the news about the *MV Yara Birkeland* in the summer of 2018, just after having submitted by master's thesis, I was in intrigued. Were the Norwegians really trying to build an autonomous ship? A ship with no captain, no crew? For someone just having discovered that a person drives a ship, something seemed awry. I applied for a PhD at NTNU.

In a sense, this PhD thesis is a re-discovery of the obvious: a person drives a ship. (Yes, even an autonomous one.) Looking back, the PhD has been fun and immensely challenging. Inevitably, though, looking back on challenges is like looking the wrong way through binoculars: everything seems innocuously small. Edmund Hillary said he climbed Everest "because it was there," but that was in hindsight. A better answer would've been, "because it was huge." This thesis is my Everest, looked at through the wrong end of binoculars.

Acknowledgements

It may take a village to raise a child, but it takes a whole town to raise a PhD student. First, I would like to thank my supervisors Ole Andreas Alsos and Martin Steinert. Their contributions stand like two icebergs, ninety percent under the surface. Ole taught me to enjoy the PhD experience and to get to know one's own "superpowers." Martin taught me that a good researcher is not measured by their successes, but in how well they handle setbacks. I am also grateful to my original supervisor, Thomas Porathe, whose preparations for retirement do not conceal the fact that he laid the groundwork for this PhD. I am also indebted to the guidance of Morten Breivik and Egil Eide whose visions for autonomous ship research at NTNU have made room for a multidisciplinary team that included Design.

I would also like to acknowledge all my co-authors and to those who provided feedback on manuscripts along the way, the latter including Jon Bernhard Høstmark, Eric Monteiro, Elena Parmiggiani, and Mariann Villa. I am also grateful to my colleagues at the Department of Design, including Åsa Snilstveit Hoem, Brit-Eli Danielsen, Dag Rutledal, Felix-Marcel Petermann, Mina Saghafian, Taufik Akbar Sitompul, Leander Pantelatos, Jooyoung Park, and Thomas Kaland. I also want to acknowledge the support of my colleagues at the Trondheim Maritime Center ("Nyhavna") and the team behind the *milliAmpere* ferries and NTNU Shore Control Lab. I am also grateful to the support of the LOAS[1] project group led by Gudbrand Strømmen at Kongsberg Maritime.

Finally, thank you to my family and my friends for their support. Thank you, Iphigenia, for your proud gaze when I replied to each of the one thousand people who've asked what my PhD was about. (If you only knew what that meant.) Finally, thank you to my mother Leena for teaching me to be optimistic, and to my father Brian for setting footprints to follow.

[1] This project was funded by the Norwegian Research Council project "Land-based Operation of Autonomous Ships" (project number 296527).

Table of Contents

List of Figures

List of Tables

Part I

Introduction and Synthesis

1. Introduction

Sometimes it seems as though each new step towards AI, rather than producing something that everyone agrees is real intelligence, merely reveals what intelligence is not.

—Douglas Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*

Trondheim: A cold October morning in 2022. I was waiting by a canal in the city's colourful downtown, where old wooden buildings line the waterfront and dark, cold water flows swiftly out of the city and into the vast fjord beyond. I was waiting to cross the canal on the autonomous ferry *milliAmpere2*, developed and built at NTNU. I had spent a good part of the past four years assisting in the project, and now its results were real, tangible. The crossing was just one hundred meters, but it may as well have been an ocean. Boarding the ferry, I was greeted by a man with officer's stripes on his navy-blue sweater. "I'm your safety host," he said, welcoming me aboard. As we set off, gliding across with the faint whirring of electric motors and precise, calculated motions, the safety host retired to a makeshift tent on the ferry. As the initial excitement wore off, I took a second look at this tent. It had never been in the drawings for the *milliAmpere2*, nor had its safety host occupant been part of the operational planning. We had accomplished building the world's first autonomous ferry, but something was missing. Or, more precisely, something was present: a human operator. Eventually, as more autonomous vessels entered the water, it was clear that safety hosts would need to be displaced to a remote control center, where they could keep a watchful eye on many vessels at once. As we arrived, the safety host opened the gangway, and I exited the ferry. The crossing was a major accomplishment, but it also eloquently captured a new problem: one that suggested autonomous vessels may be more complex than first meets the eye…

This thesis consists of four research articles that chronicle my four years of PhD research. You, the reader, could skip this text and dive straight into the articles. However, by doing this you'd miss out on some vital part of the PhD journey—a kind of secret ingredient or glue that holds everything together. What I want to get across in this thesis is that vital part, the part that isn't expressed in the lines of the peer-reviewed journal articles that are appended. A PhD is, after all, a lot more than that. It's first and foremost an *experience*, a story of confronting a challenge and rising to the occasion, of developing from a fledgling researcher into a humble expert in the field. It's also carried forward as much by one's own force of will as by the whims of fate and circumstance, producing, seemingly against all odds, a constellation of ideas, a gestalt from the endless tapestry of possible research directions. In this thesis, I wish to set the stage for my article collection, describing the context of the research effort as a whole, rather than through rote recall of its isolated parts.

The chapters of this thesis follow the typical storyline of a research report, chronicling the introduction, background, method, results, discussion, and conclusion. In doing so, the chapters synthesize my PhD as a single research project, rather than through punctuated publishing efforts.

A thesis is very much a public document, and it's the public audience I had in mind when I wrote it. One does not need specialized knowledge to comprehend my PhD thesis nor to understand its main contributions. With that in mind, though, it is worthwhile taking you on a brief tour of some of its key terms. ("Define your terms!" said 17th century philosopher Voltaire, "Or else we will never understand one another.") The first of these terms is *Artificial Intelligence*, or AI. There is no universal definition of AI, and many practitioners in the field of autonomous maritime transportation (including some of my own colleagues) vehemently refuse to subscribe to the term. Despite this, I use the term AI and all it stands for throughout this thesis. John McCarthy, one of the pioneers in the field, defined AI as "the science and engineering of making intelligent machines," and intelligence as "the computational part of the ability to achieve goals in the world."[2] *Navigation*—the process of moving a vehicle from one place to another—exemplifies the primary goal of computational intelligence: to execute planned action, as if by its own agency. Thus, when I discuss maritime navigation executed by machines, it is the term AI I use. When I discuss the human's role in autonomous maritime navigation, it is correspondingly under the umbrellas of *human-AI interaction*, referred to variously as *collaborative control* or *supervisory control*. Okay, let's dive in.

[2] John McCarthy, "What Is Artificial Intelligence?" (Stanford, CA. USA: Stanford University, November 12, 2017), http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.

1.1. The problem

The overarching research question I explore in this work is: **how do we design for land-based control of autonomous vessels?** To make this question more approachable, I break it down into four concomitant parts:

(1) *What is the research state-of-the-art?* What do we know about the topic? What methods do we have at our disposal? What obstacles remain in our quest to integrate human control into maritime autonomy?
(2) *What is the design state-of-the-art?* What are designers doing today? What methods address design of collaborative control systems? What design work is needed?
(3) *How can design principles be applied to remote control center design?* What might a real remote control center look like if we went ahead and built one? What tools and approaches might help us in this process?
(4) *What factors influence remote supervisory control of autonomous ships?* To what extent can we determine important performance-influencing factors in a controlled experiment? Can we measure reaction times and cognitive workload associated with takeovers?

1.2. Why design?

The reader might rightfully wonder: what makes this a PhD about design? Why not about engineering or one its myriad offshoots or hybridizations? This question will inevitably come up, so I will attempt to answer it here before jumping into the core material.

In trying to explain what design is all about, I tend to adopt Herbert Simon's definition: "Everyone designs who devises courses of action aimed at changing situations into preferred ones."[3] By this logic, engineers are designers, and so are architects and even business professionals, lawyers, and doctors. This generalization raises another question: if anyone who earns a living molding preferred situations can be called a designer, then what distinguishes design alone as a scientific discipline?

There is no straightforward answer to this. But considering you're about to read my thesis on the topic, I will articulate my own answer, the one that has guided by work. My take on design as a science is that its preoccupation lies more with the change process towards preferred situations than with the preferred situations themselves. If engineers are taught "applied sciences," then designers are taught the "science of applying." At the university level, the discipline of design attempts to distill this science of applying into overarching

[3] Herbert A. Simon, *The Sciences of the Artificial*, Third edition (Cambridge, MA, USA: MIT Press, 1996). Chapter 5, page 11.

theories, part conceptual and part empirical, presenting its students with intellectual doctrines worthy of academic rigor and debate.

There is no one curriculum for design, just as there are no overarching design theories or methodologies in any science. I have come to consider design more like a complement to the many fields of engineering whose practitioners, so concerned with matters of science and mathematics, have partly forgotten how to apply their trade to real problems. In this way, design is a part of engineering as much as engineering is part of design.

I have also come to consider the role of complexity in design. Designers, as a rule, are drawn to complex problems that do not have an obvious solution (or at times even an obvious problem!). Dealing with complexity also demands an antireductionist attitude that, instead of trying to reduce explanations to axioms and laws, accepts that a system made up of component parts may, in fact, end up producing *more* than the sum of its parts. Dealing with complexity requires a multi-disciplinary approach that moves beyond the compartmentalizing of traditional sciences that is so visible on the university campus. In this light, I like to think of the designer as a de-compartmentalizer, as a shepherd guiding a flock of disparate disciplines through landscape of complexity to a common destination.

Take this thesis as an example. The question of how to control autonomous ships has no obvious answer—the world of maritime navigation is complex. In attempting to guide us towards a preferred situation, you will encounter in this thesis a multitude of disciplines. This includes cognitive science to understand human behaviour, risk science to understand uncertainty, social theory to understand human interactions, statistics to formalize empirical inferences, computer science to understand the human-computer interface, and many more.

In our preoccupation with how things *ought* to be, though, designers may never be contented with how things *are*. But as we will see, how things are today are, for the designer, just how things ought to have been yesterday. And so, the cycle continues: designers are forever shaping preferred situations.

1.3. Contributions

Below is a short list of the main contributions of this thesis, organized roughly to correspond to the four articles I append:

(1) A systematic review of all peer-reviewed scientific literature on the topic of human-AI interaction in the maritime domain covering roughly the decade 2011-2021 and synthesizing the state-of-the-art knowledge.
(2) A qualitative study taking aim at how designers are accounting for human cooperation when integrating AI systems in today's transitioning navigation work.
(3) A report on the process of designing and constructing a remote control center using human-centered design principles. This remote control center became the

NTNU Shore Control Lab, a flexible research infrastructure for exploring supervisory control.

(4) Empirical results from a major experimental campaign targeting five factors and their respective influences on remote supervisory performance. These five factors included skill, vigilance, multitasking, time pressure, and decision support. The experiment was conducted at the NTNU Shore Control Lab. The data from this experiment are openly available.[4]

These contributions have implications for how we think about, design, and evaluate supervisory control of remotely operated and autonomous vessels. This means that the contributions are relevant for (1) *researchers* in marine engineering, design, and Human-Computer Interaction (HCI), (2) *designers* of remote control centers and autonomous ship system infrastructure, and (3) *engineers* working on solutions for autonomous navigation and system integration.

1.4. Thesis overview

This thesis consists of two independent parts:

- **Part I** presents the introduction to this work and provides an overview of important concepts, methods used, results collected, and a discussion about the implications of the results.
- **Part II** includes four research articles that represent the dissemination of this PhD research.

The rest of **Part I** is organized as follows:

- **Chapters 2, 3, 4, and 5** provide an overview of relevant literature. Chapter 2 introduces the concept of machine autonomy in maritime navigation; Chapter 3 provides background on human-AI collaborative work; Chapter 4 examines methods for designing remote control centers, and Chapter 5 introduces the five factors that are selected for experimental study.
- **Chapter 6** describes the methods used to approach the research questions.
- **Chapter 7** presents the results of the thesis.
- **Chapter 8** presents a discussion about the implications of the results.
- **Chapter 9** concludes this thesis.

[4] Erik Veitch, "Dataset for: Human Factor Influences on Supervisory Control of Remotely Operated and Autonomous Vessels" (DataverseNO, March 14, 2022), https://doi.org/10.18710/WYFMMP.

Part II contains the four research articles. Throughout the thesis, I refer to these four articles by article number, instead of by its full title (e.g., *Article 1*).

Article 1. Veitch, Erik, and Ole Andreas Alsos (2022). A systematic review of human-AI interaction in autonomous ship systems. *Safety Science*, vol. 152, 105778.

Article 2. Veitch, Erik, Henrikke Dybvik, Martin Steinert, and Ole Andreas Alsos (2022). Collaborative work with highly automated marine navigation systems. *Computer Supported Cooperative Work* (CSCW).

Article 3. Veitch, Erik, Thomas Kaland, and Ole Andreas Alsos (2021). Design for resilient human-system interaction in autonomy: The case of a shore control centre for unmanned ships. *Proceedings of the Design Society*, vol. 1, 1023–1032.

Article 4. Veitch, Erik, Ole Andreas Alsos, Tingting Cheng, Kristin Senderud, Ingrid Bouwer Utne. Human factor influences on supervisory control of remotely operated and autonomous vessels (In Review).

2. Machine autonomy in navigation

He who cannot draw on three thousand years is living hand to mouth.

—Johann Wolfgang von Goethe

In this chapter, I introduce some concepts relevant to the first stage of my PhD journey; namely, exploring and documenting the research state-of-the-art. In this regard, I reflect on three themes: (1) The role of humans in autonomous systems, (2) The multi-disciplinary approaches needed to design for human-AI interaction, and (3) The knowledge gaps remaining in the field today. The contents of this chapter complement *Article 1* in the collection.[5]

2.1. The role of humans in autonomous system

At first glance, it seems paradoxical to open a chapter about machine autonomy by examining the role of humans. If humans are needed, then the system can hardly be called autonomous, right? The history of automation is replete with such paradoxes. Lucille Bainbridge was among the first to bring this to light in her landmark 1982 paper, "Ironies of Automation."[6] In it, she points out the seemingly self-defeating purpose of automating tasks when those automated tasks end up requiring human supervisors. To make matters worse, the responsibilities and demands of these operators are increased at the same time as their skills decrease with disuse. These ironies are not only academically bemusing, but also potentially dangerous considering their real-world consequences.

---

[5] Erik Veitch and Ole Andreas Alsos, "A Systematic Review of Human-AI Interaction in Autonomous Ship Systems," *Safety Science* 152 (August 1, 2022): 105778, https://doi.org/10.1016/j.ssci.2022.105778.
[6] Lisanne Bainbridge, "Ironies of Automation," *Automatica* 19, no. 6 (1983): 775–79, https://doi.org/10.1016/0005-1098(83)90046-8.

Another paradox presents itself with what is considered AI. As lamented by leading AI researchers in the widely cited "AI 100" report published by Stanford University,[7] "AI brings a new technology into the common fold, people become accustomed to this technology, it stops being considered AI, and a newer technology emerges." Shortly after its acceptance, it seems, AI is no longer considered AI. At one point in time, a pocket calculator must have amazed us for its computational intelligence the same way that ChatGPT did in early 2023. Today, though, even the magic of ChatGPT is wearing off as people begin to accept it for everyday use. Sometimes it seems that the whole field of AI is chasing its own shadow.

But as we will see in this chapter, the apparent paradox that human operators are needed for AI systems like those used for maritime navigation is not really a paradox at all. Specifically, it is not a paradox if we frame autonomous ships as being designed specifically to have human supervisors. That is, if we accept autonomous ships as *collaborative* systems, designed to seamlessly merge human and machine control, rather than as purely computational systems, destined to chase their own shadow.

Navigation by people and machines

The history of AI in navigation can be said to have started in 1898, in a pond in New York City's Central Park. It was here that Nicola Tesla demonstrated his newly patented radio-wave-controlled-actuation using a handheld controller and a model boat. Onlookers stared in astonishment at the boat, which could navigate the pond's contours as if by its own accord. The machine was intelligent! Little did they know that Tesla was controlling the boat's actuators remotely.

The topic of navigation has long attracted researchers interested in both human-human and human-machine collaboration. In his seminal work "Cognition in the Wild," Edwin Hutchins, an anthropologist, defines navigation as "the process of directing the movements of a craft from one point to another."[8] As he illustrates with in-depth case studies of Navy frigates and Micronesian sailors, navigation transcends its computational basis, weaving elements of history, culture, and social organization in a fundamentally social activity. Even if it's locating small atolls in the middle of the South Pacific with only the "mind's eye" and star bearings, navigation is portrayed as a uniquely human activity, a crowning achievement of intelligence and of our ability to work together. It's no wonder, then, that accomplishing navigation by purely computational means has attracted the attention of AI developers.

---

[7] Michael L. Littman et al., "Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report" (Stanford, CA., USA: Stanford University, September 16, 2021), http://ai100.stanford.edu/2021-report.
[8] Edwin Hutchins, *Cognition in the Wild* (MIT press, 1995). Chapter 2, page 49.

Competitions and AI technology have a closely tied history. Indeed, some of the biggest breakthroughs in modern AI techniques have emerged from competitive spirit (take, for instance, AlexNet, the 2012 winner of the ImageNet Challenge, whose technique of using Graphical Processing Units for neural network computation has since revolutionized the field). Perhaps the most well-known AI navigation competition is arranged by the Defense Advanced Research Projects Agency (DARPA) in what has become known as the DARPA Grand Challenge. In early versions of the competition, competitors vied to create a vehicle that could successfully navigate a desert obstacle course entirely on its own. Since its inception in 2004 the competition has generated attention, not only for its spectacular failures and impressive finishes with large cash prizes, but also for a more surprising contribution to the history of AI navigation. This is because the same organizers met a new problem when examining techniques used by the contestants: they consisted of inexplicable programming code—so-called "black boxes." In a competition meant to bridge the gap between research and application, shouldn't the engineers at least be able to explain what is under the hood? To meet this problem, DARPA launched a new program called "Explainable AI" or XAI for short, coining a term that was soon widely adopted in computer science. The technical challenges represented in AI navigation were thereafter linked to the broader design challenges associated with making technologies explainable, transparent, and safe. XAI also marked a broader trend to incorporate human values into AI systems. As Brian Christian writes in his 2022 book "The Alignment Problem," "Research on bias, fairness, transparency, and the myriad dimensions of safety now forms a substantial portion of all the work presented at major AI and machine learning conferences."[9] He continues, writing that these topics "are the most dynamic and fastest growing areas arguably not just in computing, but in all of science." Recent advances in computational techniques used in autonomous navigation appear only to have underscored the importance of human oversight and control.

Remote control centers

The remote control center is well positioned for studying human-AI interaction in maritime navigation. At first, the autonomous ship took the limelight, and the remote control center played second fiddle. But more and more, remote control centers are emerging as a critical part of the autonomous ship infrastructure. This is reflected in impressive commercial initiatives like Massterly, a spin-off from technology giant Kongsberg Maritime and the ship management company Wilhelmsen, who, following in the wake of *MV Yara Birkeland's* unveiling, announced plans to position themselves as the world's first company offering remote control services to autonomous ships. According to Massterly, autonomous vessel

[9] Brian Christian, *The Alignment Problem: Machine Learning and Human Values*, 1st edition (New York, NY, USA: W. W. Norton & Company, 2020). Conclusion, Page 313.

operations, which includes monitoring and control intervention, are among their core services.[10] Almost overnight, new questions materialized. Who are these remote controllers? What types of skills do they have? What types of interfaces do they need at their workstations? Massterly, settling in its role as an industry trailblazer, was leading the conversation about remote control centers for autonomous ships.

Remote control centers, of course, have a history that predates autonomous ships. Vessel Traffic Services (VTS) serve an important role in ports and shipping lanes around the world, managing and overseeing maritime traffic from centralized control centers often equipped with powerful radar and advanced surveillance technology. In light rail applications, controllers oversee operations from a centralized control room like that investigated in seminal work by Christian Heath and Paul Luff.[11] Perhaps most famous of all control centers comes from aerospace in the form of NASA's Mission Control in Houston, Texas. In commercial aviation, too, control rooms have a pedigree in the form of Air Traffic Control (ATC). These buildings are instantly recognizable at airports around the world and serve as perhaps the most ubiquitous example of control rooms in transportation.

With this last point in mind, I travelled to Bodø International Airport in February 2020 to witness the construction of the Remote Control Towers, Norway's first remote control center for ATC (**Figure 1**). As anyone who's spent time in Norway knows, Norwegians tend to seek out the remotest areas on the map to call home. The result is that there are small airports dotted across the land, some of which might land a single airplane over the course of several days. The Remote Control Towers collects isolated air traffic controllers under one roof, where they can remotely direct air traffic at airports across the country. During my visit, I observed pilot testing of the technology. The operator whom I shadowed sat at a workstation surrounded by large screens that stitched together a scene reminiscent of the panoramic view from an ATC tower. She had on a headset and talked to flight crew. Before coming to work at the Remote Control Towers, she worked at the airport on Røst, a small group of islands at the extreme weather end of the Lofoten archipelago. "It's great because my husband is from Bodø, and I can be here with the family," she told me. Before taking a break for lunch, she landed a plane at Røst. Shortly upon returning, she landed a plane at Vardø, a small town at the very Northern tip of the country where thick snowflakes were falling, 850 km away from Røst. Today, the Remote Control Towers operations center is fully operational, running fifteen airports with plans to add more.

---

[10] Massterly, "What We Do," Making autonomy a reality, 2023, https://www.massterly.com/what-we-do.
[11] Christian Heath and Paul Luff, "Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms," *Computer Supported Cooperative Work (CSCW)* 1, no. 1–2 (1992): 69–94, https://doi.org/10.1007/BF00752451.

Figure 1. Avinor Remote Towers (photo by Siri Margrethe Gulliksrud, 25 Feb 2020).

The success of the Remote Control Towers buoyed my own enthusiasm for developing remote control centers for autonomous ships. Autonomous ships are not the same as airplanes, but they share some things in common. Like airplanes, autonomous ships will likely have crew onboard and carry passengers, despite utilizing highly automated navigation aids. Like the Remote Control Towers, the ships' remote control center is the guiding beacon, a bulwark of control and oversight tethered to dry land. What was especially encouraging was that the operators clearly knew what their roles were and appeared at least superficially to enjoy the benefits of remote work—the equivalent of a lighthouse keeper, I imagined, moving to town. I returned from Bodø with sense of optimism and with a clearer vision of what a remote control center might one day look like for autonomous ships.

Levels of automation

In this PhD, I use the word *autonomous* to describe ships that use sensor technology and advanced computational processing to make decisions by their own accord. The word autonomy comes from the Greek "autos," meaning "self," and "nomos" meaning "law" or "rule." Originally used to describes citizens in a democracy or nation-state, the word autonomy reflects the ability to self-govern. Using this definition, an entity is autonomous if it has a direct causal role in the actions taken to fulfill its goals.

When discussing machine computation, the question often arises of "how autonomous" a technology is. One tool that aids this discussion is "Levels of Autonomy," first described in an academic context by Thomas Sheridan in his 1992 book "Telerobotics, Automation, and Human Supervisory Control."[12] As an early pioneer in the study of human-machine systems, Sheridan was convinced that automation essentially meant that humans were doing more supervisory control work. Levels of Autonomy (LOAs) were meant to formally define this. He suggested placing automation on a 10-point scale where 0 meant fully human-controlled and 10 meant fully automatic. A useful analogy may be Alan Turing's 1948 *Turochamp* algorithm for automated chess, which Turing devised before computers existed. Lacking a computer, he plugged in the inputs to his algorithm using a pencil and paper and moved the chess pieces by hand when he arrived at an answer. This is Level 0 autonomy. Today, fully automated chess algorithms can be downloaded for free (and are regularly made to play one another to improve their own algorithms). This is Level 10 autonomy. In the world of maritime navigation, we are at least a few steps away from full autonomy—and with good reason. Navigation is much more complex than chess and, most importantly, is not a game. An autonomous ship's voyage is not constrained by clearly defined rules and by a winner and loser. Rather, an autonomous ship must comply to the nuanced and at times ambiguous stipulations of the "Collision Regulations" (known as COLREGs) and parry the infinite array of obstacles met in the messy real world. Moreover, the marine environment, unlike a chessboard, is *reactive*: the behaviour of other ships will change as soon as they realize they are sharing the water with an autonomous ship. What's more, one could imagine that a fully autonomous ship, like the ill-fated spaceship in "2001: A Space Odyssey," might blithely ignore any human commands in an override attempt. Should a fully autonomous ship eventually emerge, I doubt anyone would want to use it. There are many different taxonomies for LOAs available, distinguishable by how many intervals are defined and by what degree the roles of human supervisors are defined. The International Maritime Organization (IMO), for their part, proposed a four-level LOA taxonomy in 2018, which is depicted in **Figure 2**.

There are several issues with the LOA approach, though. Firstly, who says that a system must adhere to just one LOA? What LOA does an airplane using "fly-by-wire" technology have when it routinely transfers control to its human pilot and co-pilot, shifting from high to low levels of automation on a single flight? Three decades after defining LOAs, Sheridan addressed this problem by distinguishing LOAs from "adaptive automation." Adaptive automation essentially means that the allocations of tasks may shift between human and machine, depending on conditions of the environment, human, or machine. However, adaptive automation also presents a semantic stumbling block: to what extent can a system

[12] Thomas B Sheridan, *Telerobotics, Automation, and Human Supervisory Control* (MIT press, 1992).

be both highly automated and highly human-controlled? Why frame human and machine control as a sliding scale where one predicates a trade-off for the other?

This trade-off conundrum represents the second main issue with LOAs. One solution, as explained by computer scientist Ben Shneiderman in his article "Human-Centered Artificial Intelligence," is to consider instead "stages of automation."[13] The key insight here is to consider human control and machine control *in tandem*, rather than in series. Shneiderman goes further, arguing that the only way to have truly reliable, trustworthy, and safe automation is if they have high human control paired with high automatic control. I will return to stages of automation later, showing how this two-dimensional approach to automation taxonomy is applicable to design of remote control centers.

At this stage, though, it is still reasonable to ask, why call them "autonomous" ships at all? This is a good question, and one that I will return to in a later chapter when I introduce the term "supervised autonomy" as a more apt term.

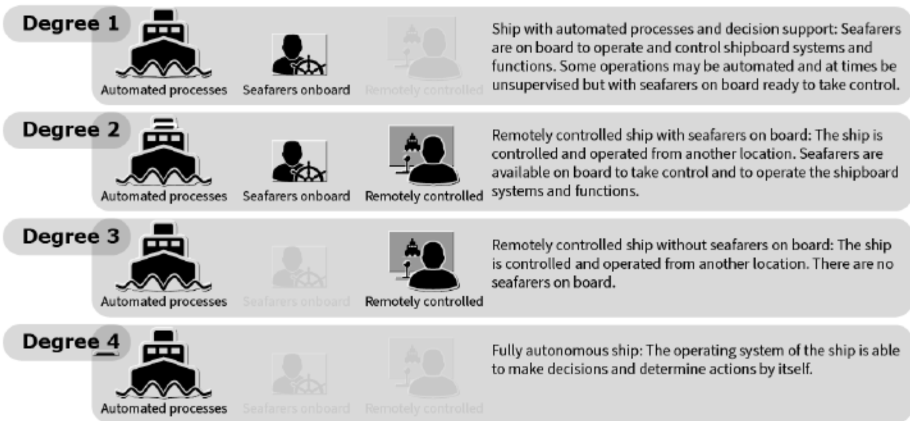


Figure 2. IMO Degrees of Automation.

Situation awareness

Eventually, in any discussion about remote control centers, you will meet the term "situation awareness." As a scientific concept, the term has been debated, but there is no doubt that it raises a fundamentally important concept: that humans, unlike their silicon-and-steel counterparts, are aware of what is going on around them. Early in my PhD, the concept of situation awareness fascinated me. What is it? Where is it? Before I knew it, I was swept up

[13] Ben Shneiderman, *Human-Centered AI* (Oxford University Press: Oxford, UK, 2022).

in works of charismatic non-fiction by computational neuroscientists like Christof Koch[14] and Anil Seth[15]. What this genre offered was plain-English scientific accounts on the origins of consciousness and lived experience—in other words, situation awareness. Unfortunately, I emerged from my foray into consciousness science with more questions than when I had entered. (I admit that I read Anil Seth's book *two* times and still don't fully grasp it.) The take-home message of these books, however, stayed with me: humans have awareness and machines do not, because awareness has more to do with *being alive* than with being intelligent.

In this thesis, situation awareness skirts the cliff edge of explanation and instead captures a compelling problem in the design of control rooms: how do you provide enough information to operators to support situation awareness? Having accepted that situation awareness is at least a useful concept for guiding remote control room design, we may turn to assigning it a practical definition. The most well-known definition of situation awareness was proposed by Mica Endsley in a 1995 paper.[16] In it, she proposed a three-level model encompassing perception, understanding, and projection that together make up situation awareness. According to Endsley, these levels occur in order, and all must be present for situation awareness to be present. There are many similar models that describe situation awareness. For example, a different model called "Information-Decision-Action" (IDA) is commonly cited by risk scientists interested in the probabilities of human error propagation in human-machine systems. This model was described in 1997 by the three researchers in a series of papers in the journal "Reliability Engineering & System Safety."[17] The authors demonstrated the IDA model for analyzing behaviour of nuclear power plant operators during accident response. Its three components (information gathering, decision-making, and action) are recognizable in the three levels of situation awareness described by Endsley (perception, understanding, and projection). Among the various models generally used to describe operators' time- and safety-critical semantic apprehension of their surroundings, Endsley's situation awareness concept has stood the test of time. It has received more citations that any other similar model in fields as diverse as "air traffic control, military operations, transportation, power systems, law enforcement, emergency management, health care, space,

14 Christof Koch, *Consciousness: Confessions of a Romantic Reductionist* (MIT press, 2012); Christof Koch, *The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed* (Mit Press, 2019).
15 Anil Seth, *Being You: A New Science of Consciousness* (Penguin, 2021).
16 Mica R Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors* 37, no. 1 (1995): 32–64, https://doi.org/10.1518/001872095779049543.
17 C. Smidts, S.H. Shen, and A. Mosleh, "The IDA Cognitive Model for the Analysis of Nuclear Power Plant Operator Response under Accident Conditions. Part I: Problem Solving and Decision Making Model," *Reliability Engineering & System Safety* 55, no. 1 (1997): 51–71, https://doi.org/10.1016/S0951-8320(96)00104-4.

transportation, education, mining, and oil and gas operations," as Endsley herself writes in a poignant defense of the concept.[18]

Adopting Endsley's model of situation awareness for remote control centers, we could begin by framing the design of a digital interface in terms of situation awareness Levels 1 through 3. For example, we may select salient colours to draw attention (Level 1); we may leverage schema for rapid processing (e.g., red means warning) (Level 2), and we may use vectors to indicate direction and therefore future states (Level 3). At least at first glance, the model appears to serves us well, providing us with a framework for designing an appropriate interface. However, as we continue to explore how people make decisions under conditions of uncertainty, we quickly arrive at the limits of the situation awareness approach to support human cognitive processes.

Decision making under uncertainty

The concept and theory surrounding Endsley's situation awareness have not been without its detractors. You do not have to dig very deeply into the cognitive psychology literature to discover, for instance, that our decision-making follows a process far removed from the neat sequence of boxes and arrows proposed by Endsley. As explained by Nobel-prize winning behavioural economist and psychologist Daniel Kahneman, our decision making is heavily skewed by biases and heuristics.[19] Kahneman explains that cognitive biases arise from low availability of information, forcing us to use mental shortcuts to arrive at a conclusion—a phenomenon he labels "What You See Is All There Is." One result is the so-called *confirmation bias*, which is of particular interest to the field of control center design. The confirmation bias leads people to seek out information that confirms a hypothesis, rather than seeking information that may support an alternative hypothesis. One might imagine a control room operator faced with a crisis who, after diagnosing the problem, becomes effectively blind to information that may contradict their diagnosis. (In fact, exactly this happened in the high-profile Three Mile Island nuclear power plant meltdown, whose investigators revealed that operators had become erroneously fixated on the notion that the crisis was the result of high water levels in one of the reactors.[20]) In the context of human-machine teaming, a similar effect is known as *automation bias*. This bias describes the tendency to favour information coming from an automation decision-making system over potentially contradictory

---

18 Mica R Endsley, "Situation Awareness Misconceptions and Misunderstandings," *Journal of Cognitive Engineering and Decision Making* 9, no. 1 (2015): 4–32. Page 4.
19 Daniel Kahneman, *Thinking, Fast and Slow* (Farrar, 2000).
20 Ellis Rubinstein and John F. Mason, "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, no. 11 (November 1979): 33–57, https://doi.org/10.1109/MSPEC.1979.6368289.

information arriving via our own senses.[21] To date, hundreds of cognitive biases and heuristics have been catalogued and described in the scientific literature. Interestingly, far from hamstringing our decision-making abilities, our imperfect ways of reaching decisions have likely evolved naturally alongside human cognition. A certain amount of irrationality, at least in an evolutionary sense, may have proven an advantageous trait.

Another issue with situation awareness concerns how we perceive the world around us. This time, the concern lies not with our cognition, but with our very windows into the world: our senses. The archetypal story linking perception to our senses is Plato's Cave Allegory. In Plato's story, prisoners tied to a chain in a cave since birth see only shadows of things cast onto a wall by a fire behind them, believing these shadows to be reality. Today, we know that human vision responds only to a narrow band of the electromagnetic spectrum between infra-red and ultra-violet, yet every colour we experience in this thin slice of reality makes up the totality of our experience. To make matters worse, we may not even perceive what we sense. In a phenomenon known as *inattentional blindness*, humans are prone to "filtering out" objects in our visual field if our attention is focused on something else. The most famous demonstration of this was made by Daniel Simons and Christopher Chabris.[22] In it, they recount an experiment where naïve participants were asked to count basketball passes in a video clip. In the middle of the clip, a person in a gorilla costume walks through the frame, stops in the middle, thumps their chest, and walks off the frame. Astoundingly, when asked if they saw anything strange, only half recounted seeing a gorilla. Studies have since repeated Simon's finding. (My favourite of these is a study of expert lung cancer diagnosticians who did not see large gorilla-shaped tumors added to lung scans, showing that even experts are susceptible.[23]) Later in this thesis, I will report on my own experiment on inattentional blindness in a remote control center setting. Using eye-tracking goggles, I was able to show that supervisory controllers may even look directly at the incongruous object without being able to recall it.

Inattentional blindness clearly does not bode well for models of situation awareness that treat information gathering as a straightforward apprehension of one's surroundings using one's senses. Sure, we may be looking, but who's to say we are really *seeing*? As anyone who has ever experienced talking with a garrulous uncle at a family reunion, the same may be said for of the medium of sound: sure, we may be hearing, but are we really *listening*?

---

21 Christopher D Wickens et al., *Engineering Psychology and Human Performance* (Psychology Press, 2015). See Chapter 12: Automation and Human Performance, page 392-393 for a good summary of automation bias.
22 Daniel J Simons and Christopher F Chabris, "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events," *Perception* 28, no. 9 (1999): 1059–74.
23 Trafton Drew, Melissa L.-H. Võ, and Jeremy M. Wolfe, "The Invisible Gorilla Strikes Again: Sustained Inattentional Blindness in Expert Observers," *Psychological Science* 24, no. 9 (September 1, 2013): 1848–53, https://doi.org/10.1177/0956797613479386.

Aside from cognitive biases and inattentional blindness, intuition also plays a role in decision making. Research psychologist Gary Klein argues for a *naturalistic* approach to studying decision making.[24] Klein makes the case that to truly understand decision making, one must head out into the real world, instead of confining oneself to controlled laboratory studies. His work introduces us to expert firefighters who make split-second decisions in burning buildings and to chess grandmasters who commit to the first move they think of under extraordinary pressure. Francisco Varela, Evan Thompson, and Eleanor Rosch make a similar case in their unique treatment of body and mind as it relates to cognitive science, phenomenology, and Buddhist psychology boundaries.[25] In it, they argue that decision making is as much the product of the body as the brain, and that cognition is "enacted" insofar that decisions emerge from having a body with sensorimotor capabilities in a psychological and cultural context. Anthropologist Lucy Suchman argues along similar lines in her seminal studies of technology use, using the compelling example of canoeing down a river to illustrate her point.[26] We might have a well-thought-out plan for running a rapid, but the true test of judgment will occur when that plan inevitably goes awry—when a log is jamming the river, or when a boulder must be negotiated—calling for *situated action*, the very antitheses of a conscientious decision. Whether one calls it naturalistic decision making, embodied cognition, or situated action, the phenomenon is recognizable by most, yet seemingly at odds with the sequential steps and mechanistic processing put forward by the cognitivist paradigm. As we will see, both cognitivist and naturalistic approaches have their merits when applied to designing remote control rooms, precluding an obstinate subscription to just one or the other.

Computation and the human touch

I recall being at a conference in 2021 and my former supervisor Thomas Porathe raising his hand after a speaker's presentation. "The issue with this," he pointed out, "is that you're saying the human brain is like a computer—but the human brain is not a computer. It's a brain!" I had to admit, his argument had a ring of logic. Like Kahneman, Suchman, and Varela and his colleagues, Thomas was pointing out that models of human cognition tend to lack that little bit extra—the human touch. What Thomas was trying to convey was that treating human operators as computation machines can lead to pitfalls when designing systems where humans and real computation machines are meant to cooperate.

---

24 Gary A Klein, *Sources of Power: How People Make Decisions* (MIT press, 2017).
25 Francisco J Varela, Evan Thompson, and Eleanor Rosch, *The Embodied Mind: Cognitive Science and Human Experience* (Cambridge, MA: MIT press, 2016).
26 Lucy A. Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions*, Second Edition (New York, N.Y., USA: Cambridge University Press, 2007).

It came as surprise, then, when I encountered this cognition-as-computation time and time again. In scientific and popular literature alike, the human brain is often cast as a computer, complete with inputs, language processing, logic, memory, and outputs. Even colloquially, people talk about "rebooting" or "glitching" or "crashing" as if their daily functions worked on some sort of operating code. Having grown up with films like "The Matrix" and "Terminator," I sometimes catch myself using such language. The human-as-computer metaphor is compelling because it does just what a good metaphor should do: it renders a complex concept understandable by comparing it to something tangible. As George Lakoff and Mark Johnson point out,[27] metaphors are not just the domain of highbrow fiction. Metaphors are woven into the very fabric of everyday speech. (Just consider the metaphors "woven" and "fabric" in the proceeding sentence.) Given the universality and explanatory power of metaphors, then, we may be remiss to cast aside the human-as-computer metaphor as inappropriate for designers just yet.

Interestingly, the word "computer" originally referred to humans. During the early 1940s, computers were regularly employed in various roles, including ballistics calculations, cryptography, artillery trajectory calculations, and code-breaking. These human computers worked with pen, paper, slide rules, and mechanical calculators to perform complex mathematical operations. With the advent of electronic programmable computers, however, the term began to change meaning. Alan Turing, who is credited with the advent of the first electronic programmable computer, introduced the idea of the "Turing Machine" as a kind of "thinking machine." Its power lay in its universality. It was not just capable of solving one type of computation, or *definite procedure*, but could be designed to emulate a machine capable of solving *any* definite procedure.[28] The design involved encoding the machine with a type of logical language: programmable machine code. Today, computers are so much a part of our everyday lives that it is hard to imagine that Turing's inspiration came from how human brains and bodies "compute."

Uncertainty and information

When engineers working at AT&T Bell Labs in the United States heard of Turing Machines, it was not long before a similar approach was applied to communication technology. At the time, Claude Shannon was tackling the problem of how people communicate over telegraph and telephone channels. More specifically, he was working on a way to compress this communication to allow for increasing tele-communication demands. Shannon's breakthrough was in demonstrating that the amount of information in a message was

[27] George Lakoff and Mark Johnson, *Metaphors We Live By* (London: University of Chicago Press, 2003).
[28] See Melanie Mitchell, *Complexity: A Guided Tour* (Oxford university press, 2009). for an excellent summary of Turing Machines in the historical context of entropy and Information Theory.

inversely proportional to its entropy: its degree of uncertainty, or randomness. Before Shannon adopted the concept, entropy was firmly in the domain of physics, enshrined in the second law of thermodynamics. This law stated that a system's entropy tends to increase unless work is applied to it. Shannon's innovation was to apply the entropy concept to human communication, which he outlined in his now-famous 1948 paper.[29] Take the example message of "Can you hear me?" The answer, which could be either "yes" or "no," could be said to contain just one "bit" of information as it may be reduced to a simple "yes" for positive, with silence meaning "no." If the answer came back gibberish, on the other hand, the answer contained no information at all, because the uncertainty in the message's response remained the same. This is the reason so many have thrown aside James Joyce's "Finnegan's Wake" in frustration (me included). Despite being written over 17 years by one of the most celebrated authors of all time, the book, which starts with the line "riverrun, past Eve and Adam's, from swerve of shore to bend of bay, brings us by a commodius vicus of recirculation back to Howth Castle and Environs," contains almost no information. Whatever message we seek to find in novels, we find only entropy in "Finnegan's Wake." (The critic Clifton Fadiman in a 1939 review in "The New Yorker," complained that the novel "communicates nothing.") Shannon's greatest insight was that communication, something that seems so ineffably human, could be disarmed by the same explanatory theories that reduced heat transfer and particle motion from alchemy to physics. Throughout this thesis, the concepts of information, communication, uncertainty, and computation feature heavily in discussions about remote control centers design. As we move on in our exploration of this topic, it is worthwhile keeping in mind how these concepts are related to one other, and how in our treatment of complex systems we are compelled to find order.

Bayesian inference

In the optimistic spirit of post-war science and mathematics, scientists began to tackle difficult, open-ended problems where clear answers were not immediately available. This included, for example, constructing actuarial models for insurance, estimating the probability that nuclear war would break out between two nations, and computing the likelihood that a medical diagnosis was correct. The trick was to use a long-forgotten technique called Bayes Theorem, developed by the 18th century Reverend Thomas Bayes. The innovation with the Bayes's approach is that it turned observation—the most fundamental part of science—on its head. Observation, said Bayes, is partly subjective and can therefore be treated as a belief. Thus, instead of depending on lots of data as is normal in inferential statistics, we can instead use opinion, intuitive judgment, or belief to make a best guess, and then adjust that best-

[29] Claude E Shannon, "A Mathematical Theory of Communication," *The Bell System Technical Journal* 27, no. 3 (1948): 379–423.

guess model as more data becomes available. When Bayesian statistics began catching on in the 1950s, critics in the frequentist statistics establishment were horrified.[30] How can one reach a scientific conclusion, they asked, based on no more than a hunch? But the Bayesians prevailed, and rapid advances in computational power of computers helped their case. Today, Bayesian statistics is used in applications ranging from medical diagnosis and fraud detection to weather forecasting and machine learning. Even the algorithms used to produce recommendations in Netflix is based on Bayesian techniques. What started as "best guess" statistics is now a part of our daily lives.

The key insight to the Bayesian approach is that it mimics the way humans make decisions. Confronted with a problem, we do not consider a hypothesis and then proceed to rigorously calculate the probability of observing it in a sample given the hypothesis is true, as is modus operandi in frequentist statistics. This approach takes effort and is rather unintuitive. Instead, we tend make decisions (and sometime very big ones) based on very little data. This initial knowledge is what Bayesians call an *a-priori* belief. Essentially, an a-priori belief is a gamble; then, having committed to it, we adjust our decision based on new experience—what Bayesians call updating a prior belief. As this cycle repeats, each new prediction gives way to a *posterior* belief. The success of the Bayesian approach is not that we start with a gamble; rather, it's that we adjust that gamble over time when new information becomes available. This probabilistic approach to making predictions was inspired by how humans make decisions. It is perhaps not surprising, then, that the field of AI, which is based on mimicking the human brain, has embraced Bayesian techniques in machine learning. Neuroscientists, too, use Bayes Theorem to model how humans perceive the world with our senses to construct best guesses of the reality around us and how, in neurological conditions, these processes may break down.

At this point, let's return to the concept of situation awareness. To what extent can we reduce situation awareness to a practical model in control room design? Bayes, for one, tells us that understanding one's surroundings depends as much on *interaction* with one's environment—a constant cycle between incoming perception and past beliefs. We may therefore be remiss to model situation awareness as following a linear arch from perception through to action since action itself will presuppose the subsequent perception. And what about the notion that situation awareness is something only humans can have? We have seen that computation can be done by computers and humans alike, and that data transfer and communication are one and the same. As designers, then, we may also be remiss to frame situation awareness as

[30] Sharon Bertsch McGrayne, *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy* (Yale University Press, 2011). This book tells an entertaining history of Bayesian statistics in applied mathematics and computer science, including the (now defused) conflicts between so-called "Bayesians" and "frequentists."

something exclusively human. To the same extent that the human brain can inspire computational techniques for perception and prediction of one's surroundings, designers should be open to assigning computational techniques a degree of situation awareness. As designers, we may do well to update our prior beliefs.

Human error and safety

Conferences can be a bit like going to gym class at grade school: one minute you're quietly learning, and the next you're dodging balls whizzing past your head. In 2019, at the very beginning of my PhD, I attended my colleague Åsa Hoem's talk at the Human Factor in Control forum in Trondheim. As soon as she was done, a hand shot up in the room. "You're saying most maritime accidents are caused by human error? But human error doesn't really exist. It's just the result of bad design." Before Åsa could respond, other hands shot up. Eventually, audience members started challenging one another, schisms forming where just moments before there had been a uniform audience. Poor Åsa, I thought, as I made a mental note: never use the term "human error" in a conference hall full of human factors specialists, lest you're prepared to dodge well-aimed verbal projectiles.

The statistic that is thrown around is 80%. That is, 80% of accidents are due to human error. What's confounding about this number is that its origin is hard to track down. In 2021, the Polish risk scientist Krzysztof Wróbel tried to track down its origin and came up short.[31] We can accept that humans make mistakes, but how often should the blame really be laid upon the shoulders of operators? At what point is it the employer's fault for not providing better training, or the safety regulator's fault for not providing better procedures, or the designer's fault for not designing a more user-friendly interface, or some combination of all these shortcomings? Arguing about what percentage of accidents are due to human error is a red herring. If we want to get serious about human error, it serves designers well to drop the "80% rule" and move on.

In the study of human error, there is a historical precedent of examining formal accident investigations. With this in mind, let's consider the Helge Ingstad accident.[32] In the early hours of November 8, 2018, the Norwegian Naval frigate HMoMS Helge Ingstad collided with the oil tanker Sola TS in Hjeltefjorden outside the coast of Bergen. That morning, sipping my coffee in my Trondheim apartment, I saw reporters' images of a badly damaged frigate listing hopelessly on its starboard side, grounded in shallow water. It was my second

[31] Krzysztof Wróbel, "Searching for the Origins of the Myth: 80% Human Error Impact on Maritime Safety," *Reliability Engineering & System Safety*, 2021, 107942.
[32] Norwegian Safety Investigation Authority, "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" (Lillestrøm, Norway: Norwegian Safety Investigation Authority, April 2021), https://www.nsia.no/Marine/Published-reports/2021-05-eng.

week as a research assistant at the Department of Design, and I had the feeling of what my Grade 12 history teacher called "velvet history": one could feel the historical significance of this unfolding event. Almost immediately, there were rumours that human error was the cause of the accident: an Officer-Of-The-Watch was brand new to the job and didn't know what they were doing. I even heard whispers that the accused was a woman![33]

Blame is a knee-jerk reaction. The accident investigation report, in contrast, which was issued in two parts (one in November 2019 and the other in April 2021), was conscientious and measured. Neither part blamed any individual, nor was there even a mention of "human error." According to the two-part report, the accident stemmed from many interrelated factors. Foremost among them was lack of situational awareness among bridge officers. It was shown that the team failed to use available human and technical resources to keep adequate lookout. But this wasn't all. Linked to lack of situation awareness were communication failures among Helge Ingstad's bridge officers and between them and the tanker's, leading to misinterpretations of signals and intentions. The Helge Ingstad bridge crew also failed to verify their own assumptions with how the Sola TS was, counter to their expectations, behaving. Furthermore, inadequate training in the use of navigational systems contributed to the accident, as did organizational pressure to grant officers' clearance quickly due to shortage of qualified navigators. Finally, because the Helge Ingstad was sailing with Automatic Information System (AIS) in passive mode, it could not be identified on the screens at the Fedje VTS nor onboard Sola TS. The Helge Ingstad did not compensate for this added risk despite operating in a maritime traffic system known to depend on AIS as its primary (and often only) source of information. The Helge Ingstad accident investigation reminds us that accidents in complex systems are not caused by human error, as we are inclined to believe. Rather, they are the result of a propagation of events set in motion when risks associated with systematic vulnerabilities are not managed appropriately. This way of thinking has influenced the way I treat control room design in this thesis. Throughout my thesis, I am wary of using human error as scapegoat for system failures and wary of using it as a justification to automate human tasks.

The Helge Ingstad suggests three guiding principles for assuring safety in complex systems involving humans, organizations, and advanced technologies. Firstly, all system uncertainties must be identified, assessed, and managed over time. Second, the influence of factors related to human performance must be understood, quantified, and related to risk. Finally, the extent to which these influencing factors *interact* must be understood, quantified, and related to risk. Understanding risk in a complex system is a tall order, requiring the convergence of many

[33] In March 2023 came the news that the Officer-of-the-Watch was charged with criminal negligence in a civilian court. The 33-year-old man who had had eight months of training before taking command of the frigate on the fateful night of the accident, received 60 days probation for his alleged crime.

disciplines. Next, I will explore the multidisciplinary aspects in design in general and in risk-based design specifically in the context of autonomous maritime systems.

2.2. Multi-disciplinary approaches

Control rooms represent a locus of control, a coming together of people and technology under the unifying structure of an organization. Naturally, the question of how to design a control room meets at the crossroads of many different disciplines. Here, I discuss some of the key contributing fields to control room design relevant for the maritime context. With an eye towards practical applications, I sample three prominent approaches: human-centered design, human-computer interaction, and risk-based design.

Human-centered design

Human-centered design (HCD) bills itself as human-oriented and in tune with the needs and capabilities of a technology's users. This contrasts with technology-oriented design, which ostensibly permits engineers' abstractions of design needs and various constraints to guide the process. HCD is perhaps most strongly associated with Apple's immensely popular line of products launched in the decade roughly spanning 2005 to 2015. The iPod, iPhone, MacBook, and iPad were sleek, intuitive, and fun to use. HCD helped revolutionize consumer tech products by paying attention to the end users' needs and experiences when interacting with a product.

The principles of HCD are perhaps most vividly laid out in Don Norman's book "The Design of Everything Things,"[34] which, since publication of its revised edition in 2013, has become canonical in the field of design. Interestingly, Don Norman is a cognitive psychologist as well as an engineer—a relatively rare combination. Norman's background in psychology shines through in his insights about how people interact with and experience technology. His introduction of "affordances" is a good example of how he walks the line between psychologist and engineer. Affordances, writes Norman, are the perceived and actual properties of an object that determine how it is used—its "what-is-it-for-ness."[35] At the same time, Norman's experience as an engineer shines through in his focus on careful observation, prototyping, and feedback, as well as on the role of constraints in technical design. What's interesting about the book is that its original title was "The Psychology of Everyday Things," and that under this title, the book did not sell as well. With the subtle

[34] Don Norman, *The Design of Everyday Things: Revised and Expanded Edition* (Basic books, 2013).
[35] Although Norman is often credited for the term "affordances" in design, its use in psychology can be traced back to the 1970s, when psychology James Gibson introduced it to explain how perception and action are linked.

change of name, its contributions shifted from introspections in cognitive psychology to outward aspects of design.

Human-computer interaction

As a sub-discipline of computer science, the field of HCI has made a science out of the study of human interaction with computer interfaces. More than any other field, HCI has shown that getting a handle on human interaction can yield better design and has formalized this claim in theoretical frameworks and methodologies. This makes HCI relevant for the design of remote control centers, which host the interface to the autonomous system as well as its locus of oversight and control.

Human-centered design (HCD) and human-computer interaction have an overlapping history. Recently, the original ideas behind HCD have been reified in design approaches tailored to AI systems. Today, human-centeredness is becoming widely acknowledged as a cornerstone of AI-based service design. Google has an excellent guide available on human-AI interaction on their website[36], as do Microsoft[37] (whose guidelines have been enshrined in a well-cited conference paper[38]), and IBM.[39]

In a previous chapter, I briefly introduced *stages of automation* as an alternative to the one-dimensional sliding scales offered by level of automation taxonomies (Chapter 2.1). Emerging from the "Human-Centered Artificial Intelligence" (HCAI) framework proposed by computer scientist Ben Shneiderman, the stages of automation approach considers human control as an independent element in human-computer interaction, rather than as the inverse of machine control. Specifically, says Shneiderman, for high-level machine control to be reliable, trustworthy, and safe, one must combine it with a high level of human control. Shneiderman developed his thoughts about HCAI, which started as a series of lectures and a research article,[40] into a book of the same name.[41] In the book, Shneiderman is among the first I've seen to place remote control rooms in the spotlight for designing AI systems. The reason for this is not to cover weaknesses in the AI system, explains Shneiderman, but rather to create stronger human-machine teams. The key insight is that machines do not mimic

---

36 "People + AI Research," accessed June 9, 2023, https://pair.withgoogle.com.

37 "Guidelines for Human-AI Interaction," *Microsoft Research* (blog), accessed June 9, 2023, https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/.

38 Saleema Amershi et al., "Guidelines for Human-AI Interaction," in *Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems (Glasgow)*, CHI '19 (New York, NY, USA: Association for Computing Machinery, 2019), 1–13, https://doi.org/10.1145/3290605.3300233.

39 "AI Ethics," May 22, 2023, https://www.ibm.com/artificial-intelligence/ethics.

40 Ben Shneiderman, "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy," *International Journal of Human–Computer Interaction* 36, no. 6 (2020): 495–504, https://doi.org/10.1080/10447318.2020.1741118.

41 Shneiderman, *Human-Centered AI*.

human abilities. If they did, we could simply look to successful human teams to understanding the mechanism of what makes a good human-machine team. In other words, the mechanisms underpinning human-computer interactions are distinct and worthy of their own special treatment. Shneiderman presents the concept of *parallel autonomy* that underscores that machine autonomy supports, not supplants, human autonomy. This shift in thinking is perhaps best captured in aviation with the recent change in the name "Autonomous Unmanned Vehicles" (AUVs, commonly known as drones) to "Remotely Piloted Aircraft System" (RPAS).[42] Today, the US Air Force, among others, have adopted the new term, as if to clear up any confusion about who is really in charge. Remote control centers, explains Shneiderman, are the organizational tools that enable parallel autonomy—a way of implementing supervisory control for oversight and intervention of autonomous technologies.

What all practical design approaches seem to have in common is that they are *goal-based*. That is, they are dedicated to the process of mapping functional requirements to system goals. For all our focus on human-centeredness, however, we have not yet mentioned how human error fits into the picture. As we will see next, that is the special domain of a subset of goal-based design called risk-based design.

Risk-based design and risk assessment techniques

Many risk assessment studies begin by stating that autonomous ships have the potential to reduce human error. By now, I hope the reader appreciates the issues with this line of reasoning. However, these studies approach human error in a way that is potentially valuable in that they attempt to uncover exactly what human errors consist of, where they emerge, and how we might get a handle on them.

Risk-based design, broadly speaking, is a goal-based design approach oriented towards functional requirements of safety controls. All complex systems are overshadowed by the possibility that things can go wrong. Safety controls, while unable to stop all things from going wrong, can at the very least prevent harm to people and to the environment when things do go wrong. Consider the example of seatbelts in a car. They may not prevent car accidents, but they do save lives when accidents happen. To design safety controls, one must identify and enumerate all the ways things can go wrong in a system. This rather morbid detective work is the domain of many different risk assessment approaches, some of which I will discuss below.

---

[42] Federal Aviation Administration, "FAA Safety Briefing - May/June 2021" (U.S. Department of Transportation: Washington, D.C., USA, May 1, 2021), https://www.faa.gov/newsroom/safety-briefing/archived-faa-safety-briefing-magazine-issues.

Risk-based design has emerged as one the most common approaches to design of autonomous ship systems in the scientific literature. Risk scientists make it their mission to orient designers towards the most relevant hazards and towards building safety controls to manage the level of uncertainty (i.e., risk) associated with them. In their quiver, risk scientists have many different risk assessment methods. While these methods differ, what they have in common is that they generally express risk using categories (i.e., failure mode, consequence) and a likelihood (i.e., expected rate of occurrence). Hazards are often identified exhaustively, and their interconnectedness is formally mapped in a network where human errors, much like a toppling domino, can propagate through the network. The goal of risk assessment methods can be said to identify where safety controls, like domino-stoppers, can be most effectively placed. One commonly used tool in maritime analyses is System-Theoretic Process Analysis (STPA), which has been praised for its ability to capture organizational and human interaction processes in addition to the standard technological components of a system.

Once hazards have been identified, it is important to consider the extent to which they are associated with human error. Risk-based design is therefore also concerned with assessing Performance Shaping Factors (PSFs; also known as Risk Influencing Factors). Typical PSFs might include stress or skill—what engineers tend to call "human factors" (more on this later). Remote control centers, for their part, seem to be ideally positioned to be both the source of human error propagation and the source of managing it. Another risk assessment method commonly applied to autonomous ship systems is the Bayesian Network (BN). BNs, as I have introduced them before (Chapter 2.1), are ideal when limited data are available and when scientists wish to formalize initial best guesses of how likely different failure modes may be and how they might interact with each other. They are also useful in that they can be updated as more information becomes available. However, BNs also come with drawbacks. Firstly, the nature of basing quantitative models on subjective prior knowledge can introduce significant biases. Second, BNs can be difficult to interpret for non-experts, especially as networks grow to resemble spaghetti and conditional likelihood tables become convoluted. Research papers that apply BNs to autonomous ship systems often conclude with the caveat that more data are needed to validate their results. Overall, though, they provide a flexible way to formally assess risk in the face of complexity and limited data, effectively side-stepping the red herring of "80% human error." Towards the end of this thesis, I show how results of an original experiment I led was used to link PSFs to BNs in risk modelling for remote control centers operations.

Resilience Engineering

While digging into the literature of risk-based design, it wasn't long before I encountered the term *Resilience Engineering*. As far as I can tell, this isn't really a field of engineering (I have yet to meet someone with the title "Resilience Engineer"), but more of a paradigm in the broader field of complex systems design. Complex systems generally involve both self-organizing social factors (e.g., management, employees) and rigid technical factors (e.g., functional performance of an autonomous ship) and often exist under the pretext of low margin for error. Resilience Engineering (RE for short) is concerned with making such complex systems—whether airplanes or space shuttles or hospitals—work. RE is loosely made up of three tenants: (i) productivity must be balanced with safety by the appropriate use of safety controls (ii) risks must be monitored and managed over time, and (iii) humans are the source of resilience in a complex system. The latter is worth highlighting, because it in effect turns human error on its head, implying that humans are a system's most important safety control, as opposed to their greatest liability. Disciples of RE like Erik Hollnagel have long backed up this claim, pointing to a simple observation as evidence: for the most part, things go *right*.[43] Instead of getting upset when things go wrong and going through painstaking investigations to understand the root causes of an accident (only to produce more red-tape and ignored procedures), why not study when things go *right* and foster *those* conditions? A suitable analogy to RE may be positive reinforcement in the process of learning behaviour, where its counterpart, negative reinforcement, represents the traditionalist approach of analyzing system safety breakdowns.

The idea of RE is intuitive enough, but its practical application has led to some surprising results. Foremost among these is the technique known as Functional Resonance Analytic Modeling (FRAM), which can be used to describe the ways that work can be carried out in a complex system. The FRAM model is a collection of nodes representing the functions or activities that make up a system, each connected according to different relationships (e.g., inputs, outputs, time, control, preconditions, and resources). Seeing FRAM in action is something to behold. The first time I encountered it was at a student conference at Memorial University when I was a master's student. My colleague Doug Smith was presenting his FRAM model of an offshore operation[44] developed from a computer program that he helped develop.[45] I looked like something from microbiology class, with hexagons representing system functions and lines connecting them to other functions via nodes. The whole thing had a time dimension and could pinpoint in real time where the greatest risks of an operation lay and when they occurred. Since then, I've heard many people mention FRAM at conferences or in stray sentences of research papers. But Doug is one of the rare researchers who has actually built one. I get the impression FRAM more often discussed than used, for

43 Erik Hollnagel, David D Woods, and Nancy Leveson, *Resilience Engineering: Concepts and Precepts* (Ashgate Publishing, Ltd., 2006).

44 This was later published in another conference proceedings: Faisal Khan et al., "Visualizing and Understanding the Operational Dynamics of a Shipping Operation" (SNAME Maritime Convention, Providence, RI, USA: Society of Naval Architects and Marine Engineers, 2018).

45 Doug's software, "DynaFRAM," is available for download at https://www.engr.mun.ca/~d.smith/

the simple reason that they tend to be a bit unwieldy. At the very least, RE provides us with a perspective for system design that considers human error only in tandem with its counterpart: human ingenuity. RE accepts that humans allow a complex system to bend, not break, under pressure.

Work on risk-based design for autonomous ships is ongoing and will continue to inform much more than just the physical infrastructure of the ship and control room. Risk-based design will also inform assurance, training, certification, safety management, safety procedures, standardization, and regulation. Promising risk-based design approaches developed specifically for autonomous ships include the Operational Design Domain,[46] which borrows from autonomous car design guidelines, and Human-System Integration in Autonomy (H-SIA),[47] which borrows its core principles from the tenants of RE. Risk-based design for autonomous ships is also the topic of a recent PhD thesis,[48] whose main contribution is a "human centered" risk assessment for remote control centers.

2.3. Knowledge gaps

There is still a lot we don't know about how to design land-based control for autonomous ships. One way to proceed is to trace the leading edge of our understanding and use this to guide the way forward. This is what I intend to do here. Later in this thesis, it will become clear how this exercise guided my research towards the most impactful contributions for control room design and especially in my design of experimental testing.

The role of operators

In 2020, the IMO released its much-anticipated report addressing regulatory gaps for operation of autonomous ships. Scanning through the document, my eyes were drawn to a list entitled "high-priority issues." These were aspects that, according to their gap analysis of current regulations, represented the most glaring shortcomings. I had expected some technical showstoppers concerning, for example, perennial complaints of "black-box" methods used in AI. Perhaps for this reason it caught me off guard when I read that the first

46 Ørnulf Jan Rødseth, Lars Andreas Lien Wennersberg, and Håvard Nordahl, "Towards Approval of Autonomous Ship Systems by Their Operational Envelope," *Journal of Marine Science and Technology* 27 (2022): 67–76, https://doi.org/10.1007/s00773-021-00815-z.

47 Marilia A. Ramos et al., "Human-System Concurrent Task Analysis for Maritime Autonomous Surface Ship Operation and Safety," *Reliability Engineering & System Safety* 195 (2020): 106697, https://doi.org/10.1016/j.ress.2019.106697.

48 Åsa Hoem, "Risk Assessment in the Design Phase of Maritime Autonomous Ships – A Human-Centered Approach" (Doctoral Thesis, Trondheim, Norway, NTNU, 2023), https://hdl.handle.net/11250/3063036.

item on the list was "Meaning of the terms master, crew or responsible person."[49] The report went on to explain that above all else, the operator's role in autonomous ship operations had to be clarified.

> "The role, responsibility, and definition of master, especially for degrees of autonomy Three and Four where personnel on the shore side might control the ship, were considered to be a common theme identified in several instruments as a potential gap."[50]

The second item on the list was "Remote control station/centre." Acknowledging that autonomous ships may be operated remotely, the report explained that functional and operational requirements at the control center needed to be clarified. Finally, the third item addressed the "Remote operator as seafarer," raising the question of qualifications, responsibility, and role of the remote operator given their possible designation as a seafarer.

Highlighting these three high-priority issues raised the concept of the remote control center to the fore in autonomous ship operations. It linked concerns like qualification, responsibility, and functional and operational requirements as concomitant to the remote control center operator's role. Lacking a clear picture of the operator's role, we are left only to speculate about their assigned tasks and how this may relate to their performance.

Fortunately, a real-life case study may serve to shed light on the role of the operator. Consider the urban autonomous ferry *milliAmpere2*, a project with which I was closely involved during my PhD and which ranks among the first examples of maritime autonomy tested in the field. The test trials of the *milliAmpere2* in the fall of 2022 marked the first public trial of autonomous maritime technology for a passenger ferry (**Figure 3**). The ferry is small—just 8.5 m long and crossing a canal barely 100 m wide—yet during trials the operator's role was clearly manifested. Specifically, onboard all crossings during the three-week trial, a "safety host" was present onboard. The safety host's main role was to maintain passenger safety and take over control from the autonomous system, if necessary. This was made possible by a joystick controller for the Dynamic Positioning (DP) system, which could be overridden from the computer control at the press of a button. From this basic role stemmed several other important responsibilities: the safety host started up the ferry in the morning, welcomed passengers onboard, chatted amiably with passengers, opened and closed gates, initiated charging, marked down notable events in logbook, coordinated with engineers if there was need for maintenance, and turned off the ferry at the end of the day. The role of

[49] IMO, "Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS)" (London, UK: International Maritime Organization, June 3, 2021), https://www.imo.org/en/MediaCentre/PressBriefings/pages/MASSRSE2021.aspx. Page 7.
[50] IMO. Page 8.

the safety host materialized with the need for assuring safety of public passengers and for delivering an enjoyable, reliable, and consistent transportation service. The safety host role will eventually be displaced to the remote control center, and as such it sheds light onto what the remote operator's role will one day entail. At the end of the 2022 *milliAmpere2* trials, I interviewed three of the safety hosts to get a sense of who they were and what they were expected to do. I will return to these interviews in the Results section.

Figure 3. *milliAmpere2* during trial operations (photo by Mikael Sætereid, 2 Oct 2022).

Regulatory gaps

During my PhD, I often encountered the argument that regulators were presenting obstacles in the path of autonomous technology development. Complaints of this variety characterized the maritime industry as especially heavily regulated and slow to adapt to change. "How can we expect to develop autonomous ships," went the common refrain, "when the regulations don't even allow it?" It's a good question, and here I'll outline some ways regulatory agencies appear to present obstacles to autonomous ship operations. (I will also show how this might not be the most accurate way to depict the role of regulators.)

For one thing, COLREGS, Safety of Life at Sea (SOLAS), and the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (STCW) are all predicated upon personnel being physically onboard the vessel. For example, COLREGs states in Rule 5 that "Every vessel shall at all times maintain a proper look-out by sight and hearing as well as by all available means appropriate in the prevailing circumstances so as to make a full appraisal of the situation and of the risk of collision."[51] SOLAS, for its part, states that all vessels must do everything in their power to help other vessels in distress, establishing a kind of code of conduct for the sea—essentially the "I-got-your-back-if-you-got-mine" rule of the seas. STCW is also written with the preconception that seafarers and navigation crew are located on the ship, with training requirements, qualifications, and competency requirements defined as such.

However, this line of reasoning mistakenly treats regulatory safety frameworks as static. The truth is regulatory frameworks are always changing. Part of regulatory agencies' mandate is to align safety frameworks with technology improvements. Seen in this light, they are more concerned with enabling technologies than obstinately blocking their way. By extension, designers, through their influence on technology development, can have an important impact on regulatory revisions. A better way to phrase the question of regulatory gaps is therefore as follows: "How can current regulatory shortcomings guide us towards more impactful technology development?"

Current regulations indeed do not make room for autonomous ships.[52] Consider the examples of COLREGS, SOLAS, and STCW as presenting obstacles to autonomous ship operation. If we consider regulatory frameworks as dynamic instead of static, then behind each of these apparent obstacles lies a potential design impact. Take the COLREGs case: rather than considering remote operations as an obstacle to proper look-out, we may instead ask how audiovisual streaming may *augment* proper look-out by sight and by hearing. Similarly for SOLAS, rather than considering remote operators as a hindrance to rescue operations, orchestrating rescue remotely may well be an important asset, keeping rescuers out of harm's way. For STCW, basic requirements may well be comparable to those of contemporary seafarers, with added specializations in control room equipment, intervention procedures, and coordinated action.

Rather than existing on two different spheres, I have come to think of regulatory development and design research more like two sides of the same coin. Currently, the IMO is working on a new "Maritime Autonomous Surface Ship (MASS) Code" to address some of the concerns highlighted in their 2021 scoping exercise. Research like that presented in this thesis can do a small part in influencing such regulatory development. Identifying

51 IMO, "International Regulations Preventing Collisions at Sea (COLREGs) (Adopted 20 October 1972, Entered into Force 15 July 1977) 1050 UNTS 16 (COLREGs)," 1972.
52 One very helpful analysis of regulatory issues governing autonomous ships is presented by Marel Katsivela, "Unmanned Vessels and Regulatory Concerns," *Journal of International Maritime Law* 26 (2020): 239–52.

regulatory gaps is the first step in such a dialogue, but it is certainly not the last. In this light, it is the designer's role to address how we may begin to re-cast these gaps as potential design improvements, and not just as obstacles.

Lack of empirical data

With autonomous ships, a new risk picture in maritime operations is emerging. In the center of this new risk picture lies uncertainties about human-AI interaction; namely, about how autonomous systems and human operators will team up to meet real-world challenges. What is needed is more observation-based knowledge allowing us to get a handle on how, exactly, this interaction will unfold.

Researchers across the board have complained about the lack of empirical data concerning autonomous ship systems. This complaint is especially palpable in the risk sciences, whose practitioners build models for predicting how human-AI interactions will unfold in safety-critical contexts (in an emergency, say, or in an equipment failure). For designers, these types of models provide valuable input for design of safety controls. However, risk scientists have thus far been obliged to present their models with a caveat: until more data about autonomous ships are available, their model predictions cannot be verified. It can seem like a catch-22: no model until we have data, no data until we have a model.

Part of the issue stems from risk analysis method like Bayesian Networks and STPA. These are favorites among risk scientists studying autonomous ship systems, yet their greatest strength is also their fundamental weakness: they work when limited data are available. Despite being able to pinpoint exactly what types of interactions are most likely, have the severest consequences, and are most sensitive to changes in other risk-influencing factors, all risk models developed to date lack the rock-solid foundation of historical data. They are more like best guesses awaiting subsequent fine-tuning as soon as new data are available.

The obvious follow-up question is: where do we get data? One way to circumvent the catch-22 presented by risk-based design approaches is to use simulation. Simulation has a long historical precedent as a tool for collecting data when data are sparse or difficult to obtain. This is perhaps most visible in aerospace and commercial aviation with flight simulators. In maritime research, ship bridge simulators also have a long history, with applications ranging from ship design, training, and emergency response analysis. These days, simulator research is getting a second wind with the rise of "digital twins." What distinguishes digital twins is that they recreate the automated system and its interaction with the world in addition to the analog system and its interactions. Consider the conventional ship bridge: all that is needed is a simulation of the controls aboard the ship's bridge and the physics of those controls' actuation in an external world. In a remote control center simulation, by contrast, one also needs a simulation of the machine autonomy and its way of sensing the world. A major advantage of the digital twin approach is that it allows us to design for machine autonomy interaction aspects in addition to human interaction aspects.

I will return to the concept of using simulation in testing later in this thesis, describing a remote control center laboratory that I helped design and build expressly for this purpose. Later, I will also explain how an experiment I conducted that leverages simulation sheds light on aspects such as reaction time and multitasking ability for supervisory controllers and provides data for Bayesian Networks used for risk modelling purposes.

Of course, simulation alone is not enough. A useful analogy is the role of thought experiments in the development of scientific concepts.[53] Like simulation, a thought experiment is not bounded by the constraints of real life, be they related to budget, safety, or even physics. Thought experiments leverage the power of the imagination to communicate and develop ideas. However, like thought experiments, simulation is not sufficient to rigorously test how concepts hold up in the real world. I will return to this idea, too, with a discussion about how simulation tests conducted in a laboratory can be verified using field trials of real autonomous vessels like the *milliAmpere2*.

2.4. Summary

The scientific method is compelling because it allows us to build upon one another's research. As I'm reminded every time I visit the Google Scholar homepage, it is our privilege as well as our duty as scientists to "stand on the shoulder of giants."

In this chapter, I have introduced concepts, theories, and methods that are relevant in the rest of the thesis. This includes concepts like levels of automation, situation awareness, human error, as well as methodological frameworks like human-centered design, human computer interaction, risk-based design, and resilience engineering. I have also traced the contours of where important knowledge gaps currently lie, including as they relate to the role of human operators in autonomous ship operations, regulatory shortcomings, and lack of empirical data.

At this stage, the research state-of-the-art has been our guide both in terms of what has been done and what needs be done. The knowledge we've gained so far will serve as guideposts in the rest of this thesis, like exit signs on a highway leading to a destination. As we will see, certain elements, like human-centered design and simulation techniques, will feature prominently on this journey.

---

[53] Nancy J. Nersessian, *Creating Scientific Concepts* (The MIT Press: Cambridge, MA, USA, 2010). This book examines the fascinating process of developing concepts in science, a process that includes the role of thought experiments. In it, Nersessian makes the case that computer simulation is a type of thought experiment—an idea that I carried forward in my own work.

3. Teaming with machines

> **Moss, Norway: A cold March morning in 2020.** I was waiting for the ferry across the Oslo Fjord to Horten and for the arrival of my project team at the ferry quay. The Land-based Operation of Autonomous Ships project (LOAS) had formally kicked off in January and the team was eager to get started. After boarding the ferry, we met in the cafeteria, which must be the world's only one serving hot dogs in a waffle—apparently a Moss specialty. The team consisted of designers and engineers both from IFE, a leading research institute in human-machine interaction, and from Kongsberg Maritime. With notepads in hand, we were escorted by our guide to the ferry's navigation deck. It was not a coincidence that we had boarded this ferry. The ferry had special technology installed, and the LOAS team was eager to see it in action. The technology was somewhat blandly called "auto-crossing" and "auto-docking," and it did just what you might expect: at the press of a button from the officer in charge, the boat docked and crossed, as if by its own accord. Developed by Kongsberg Maritime, the auto-systems represented some of the most advanced automated navigation technology in operation to date. So far, it didn't make any course adjustments based on targets in view; however, it did come with a detailed display that indicated whether the vessel was in "auto" mode, and, if it was, visualized the complex system at work with small graphics of thrusters and motion vectors. Interestingly, though, as we prepared to the dock in Horten, the captain steered in manually, choosing to ignore the auto-docking feature altogether. Were they trying to make a point?

The date of this ferry crossing happened to be March 10, 2020. Shortly after we arrived in Horten, one of our team members got a call informing her that someone in her daughter's preschool had tested positive for Covid. Panic-stricken, she dismissed herself. Shortly thereafter, we were all kindly asked to leave the premises. Maybe we were infected. Soon we received word that Horten VTS, our next stop, was also cancelling our visit. Just two days later the whole country would be in lockdown, marking what would be a new milestone in my PhD journey, and one that was not planned.

When I eventually got settled in my home office back in Trondheim, the short crossing on the automated ferry left a lasting impression. Looking at my detailed notebook, I promised to make the best of a bad situation. I would call the bridge officers on Zoom and interview them about their experiences using automated navigation systems. Combined with my field notes and interviews with technology designers I had done the previous fall with my colleague Henrikke Dybvik (about which we had reported in an article[54]), I had the groundwork in place for a qualitative study. This qualitative study could compare designers' approaches to human-AI interaction to navigators' own experiences, thereby indicating whether adjustments needed to be made. There was only one problem: I didn't know how to *do* qualitative studies of this kind. But it was lockdown, and I figured I had time to learn.

This study eventually became *Article 2* in this collection.[55] In this chapter, I wish to complement *Article 2* with relevant background and conceptual detail—information that will help set the context of this work relative to my overall PhD journey.

3.1. Centers of coordination

The concept of centers of coordination was defined by Lucy Suchman in 1997 in response to a growing body of research examining supervisory control and time-critical, coordinated action across distributed locations. She introduced the concept to provide a framework for professionals and researchers to better orient themselves to this new type of work. Now, in the wake of technologies enabling higher levels of machine autonomy, there is a renewed interest in conceptual frameworks for supervisory control. As Suchman writes in a book chapter called "Centers of Coordination: A Case and Some Themes:"

> "Centers of coordination are characterized in terms of participants' ongoing orientation to problems of space and time, involving the deployment of people and equipment across distances, according to a canonical timetable or the emergent requirements of rapid response to a time-critical situation."[56]

---

54 Henrikke Dybvik, Erik Veitch, and Martin Steinert, "Exploring Challenges with Designing and Developing Shore Control Centers (SCC) for Autonomous Ships," *Proceedings of the Design Society: DESIGN Conference* 1 (2020): 847–56, https://doi.org/10.1017/dsd.2020.131.
55 Erik Veitch et al., "Collaborative Work with Highly Automated Marine Navigation Systems," *Computer Supported Cooperative Work (CSCW)*, October 8, 2022, https://doi.org/10.1007/s10606-022-09450-7.
56 Lucy Suchman, "Centers of Coordination: A Case and Some Themes," in *Discourse, Tools and Reasoning: Essays on Situated Cognition*, ed. Lauren B. Resnick et al. (Berlin, Heidelberg: Springer Berlin Heidelberg, 1997), 41–62, https://doi.org/10.1007/978-3-662-03362-3_3. Page 42.

In the decades leading up to Suchman's proposed definition, control rooms were becoming increasingly complex. With the rise of nuclear power plants in the 1970s and 80s, questions about human factors and cognitive engineering in control room operations were becoming more urgent. Control rooms were the focal point for orchestrating complex, sociotechnical systems that were materializing in a society embracing technological advancement and economic growth. In the 1990s, control rooms also featured in the field of Computer Support Collaborative Work (CSCW), whose practitioners examined the social mechanisms around which computerization hinged. Using techniques originating in ethnography, researchers studied workers in their natural settings like a previous generation of ethnographers might have studied an isolated tribe. They entered hectic line control rooms of the London Underground,[57] tense air traffic control towers,[58] and austere emergency services dispatch centers,[59] immersing themselves in their unique cultures and describing their social and technological interactions. This was the direction of the field when Suchman identified centers of coordination as a new research arena.

Suchman's centers of coordination are as important today as ever before. They remind us that work is primarily a social action. Any designer neglecting to account for social mechanisms of work and its emergent, cultural elements risks designing a work environmental misaligned to social cues, gestures, and myriad forms of communication—a center of incoordination. They also remind us that work in complex systems does not unfold a neatly ordered way, as one might be inclined to believe from behind the ordered desktop of a designer. Rather, control room work invokes heuristic decision making, commonsense judgment calls, and tacit communication that transcend even well-meaning procedures and protocols.

As we will see, the canonical concept of centers of coordination also provides a useful lens for examining human-AI interactions. Interacting with intelligent machines sets the tone for new social mechanisms that are we are just beginning to understand. Consider, for example, the role that physical pointing and verbal exclamations played in the classic studies of London Underground control room work.[60] A parallel may be drawn to current efforts in Explainable AI (XAI) to develop techniques that highlight what pixels an image recognition algorithm is using to reach its classification decisions—in effect, pointing and exclaiming

57 Heath and Luff, "Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms."
58 Richard Bentley et al., "Ethnographically-Informed Systems Design for Air Traffic Control," 1992, 123–29.
59 Jack Whalen, "A Technology of Order Production: Computer-Aided Dispatch in Public Safety Communications," *Situated Order: Studies in the Social Organization of Talk and Embodied Activities*, 1995, 187–230.
60 Heath and Luff, "Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms."

where it is looking.[61] Control room work is a fundamentally social activity, and this is no less so with the introduction of machine autonomy imbued with its own agency.

3.2. The alignment problem

Whether it be a ferry or a fighter jet, the individuals responsible for the remote control of autonomous vehicles are finding themselves in a supervisory role—a role Brian Christian has provocatively termed "sorcerer's apprentice" in his 2021 book "The Alignment Problem."[62] "We conjure a force, autonomous but totally compliant, give it a set of instructions, and scramble like mad when we realize our instructions are imprecise or incomplete."[63] When I was writing *Article 2*, I was inspired by Christian's book. A philosopher and computer scientist, Christian argues that the greatest challenge faced by designers and engineers wielding the extraordinary powers of AI is aligning them to the values inherent in a human world. Reading his book, I immediately recognized tell-tale signs of the alignment problem in my own PhD research. How do we align machine autonomy with the need for oversight, management, coordination, and responsibility in maritime work? How do we align the end of an autonomous vessel's capabilities with the beginning of a human operator's? In the bigger picture, how do we align machine autonomy with operator autonomy? The alignment problem asks engineers, designers, and policy makers to tune in to human needs when integrating AI technologies and nudge design in a correspondingly appropriate direction.

A similar call has gone up from some of the most influential researchers in the field of AI. Every few years, Stanford University assembles an expert panel to discuss issues in their rapidly developing field, publishing their results in a series known as the "AI100 Report." In 2021, the report expressed that "Perhaps the most inspiring challenge is to build machines that cooperate and collaborate seamlessly with humans."[64] Indeed, this is the mission of a new and active multidisciplinary *alignment community* that involves not just computer scientists and designers, but also anthropologists, sociologists, safety specialists, organizational scientists, and many more.

A good place to start addressing the alignment problem in the context of designing remote control rooms for autonomous ships is to assess the extent a discrepancy exists between

---

61 This is exemplified by "saliency maps" produced by Esteva and colleagues' deep neural network (DNN) for diagnosing skin cancer. The saliency maps highlight which pixels are used to reach its diagnosis, thus helping the doctor understand how the algorithm reaches its decision. Andre Esteva et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," *Nature* 542, no. 7639 (February 1, 2017): 115–18, https://doi.org/10.1038/nature21056.

62 Christian, *The Alignment Problem: Machine Learning and Human Values*.

63 Christian. Page 30-31.

64 Littman et al., "Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report." Page 7.

designers' construals of control room work and operators' own accounts of that work. Later, I will describe my observations aboard two highly automated ferries as well as my interviews with navigators and professional technology designers. From these observations, I describe the discrepancy between how designers construed human-AI collaboration compared to navigators' own accounts in the field. This discrepancy sheds light on how integration of AI systems may be better aligned to human cooperation in navigation. The result can be interpreted as a nudge towards rendering computational activities more visible and towards incorporating social cues in the articulation of work in its natural setting.

3.3. Plans and situated actions

When the manuscript for what eventually became *Article 2* was returned to me from the editor, I was chagrined to read one of the reviewer's baffled comments. Hadn't I read Suchman's classic book "Plans and Situated Actions?"[65] I had to admit I hadn't. I went to the library to check it out, wondering what could be so special about this book.

As it turned out, the ideas presented in the book were immensely helpful in grappling with two fundamental problems in human-AI collaboration. The first of these involved transitions of control, both from machine to operator and vice versa. The second involved the difference between situation awareness and "situatedness." By the time *Article 2* was finally accepted, I was well-versed in Suchman's book.

To understand transitions of control, Suchman framed the human-machine interface as two sides of the same coin: on the one side are plans, enshrined as control algorithms; on the other side are situated actions, stemming from in-situ skills. Previously, I mentioned Suchman's metaphor of negotiating a river rapid in a canoe as an illustrative example of combining well-intended plans and well-executed situated actions. The metaphor, however dramatic, extends to maritime navigation. Navigating a ship is, after all, a complex task, where plans are punctuated by off-the-cuff actions, and where decisions are based as much upon in-the-moment intuitions as on premeditated intentions. This way of thinking about navigation revealed an important insight. When plans are automated, the navigators enact two roles: the first as passive operators when plans proceed as expected, and the second as active operators when these plans are inevitably jettisoned to deal with some situation at hand. The operator's in-situ skills should therefore not be downplayed just because they are in a control room removed from the ship's location; on the contrary, the importance of their in-situ skills is *increased* as their responsibilities distill into critical intervention actions.

I also want to briefly mention the second takeaway from Suchman's book; namely, the difference between situation awareness and *situatedness*. As I have mentioned before,

---

[65] Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions.*

designers routinely invoke the concept of situation awareness when dealing with how operators make decisions. But when examining how decisions are really made in the field, this is inevitably at odds with an ineffable situatedness of decision making, elements of which we have already seen in naturalistic decision making and embodied cognition (Chapter 2.1). When considering control room work, we are forced to consider both strains of decision making at once. The interface represents both the designer's construal of decision making and embodies the operators situatedness in the instant they orient themselves to some emergent situation. The successful design of a control room hangs in the balance of enabling both plans and situated actions.

3.4. Paying attention

At face value, the operator's role in the control center appears straightforward: pay attention to the autonomous ship (or ships) and take over control, if necessary. It follows, then, that paying attention is an important task for the operator—perhaps the most important task of all. By this logic, then, it may behoove us to design the control center in such a way that maximally supports operators' capacity to pay attention. This begs the question of *how* to design for supporting operators' attention.

A reasonable approach is to break down the causal mechanisms behind attention. If attention consists of A, B, and C, then we should make sure to support A, B, and C. Unfortunately, this is where things get a little bit complicated. We may define attention as an orientation of the mind that, at least when oriented towards the outside world and not to our own inner world, is linked to sensory perception. Herein lies the problem: attention is derivative of conscious experience, and any study of consciousness will inevitably meet unanswered questions about the nature of our attentive minds. If it has alluded philosophers from Plato to Descartes to modern neuroscientists wielding brain scanners, then it would be hubris to say we can design control rooms around the mechanisms of attention. Instead of throwing up our hands in despair, though, it may be worthwhile to investigate what scientists do know about what it means to "pay attention."

A good place to start is a paper published in 1998 by Giulio Tonini and Gerald Adleman.[66] In it, they make the simple observation that all conscious experience has two properties: it is *informative* and it is *integrated.* By informative, they meant in the same sense that Claude Shannon meant about communication: it reduces uncertainty. Specifically, any conscious experience reduces an enormous amount of uncertainty in the world around us, producing a particular experience of the *way it is*, unique from all other possible experiences. In other

[66] Giulio Tononi and Gerald M. Edelman, "Consciousness and Complexity," *Science* 282, no. 5395 (December 4, 1998): 1846–51, https://doi.org/10.1126/science.282.5395.1846.

words, paying attention means *not* paying attention to everything else. By integrated, Tonini and Adleman meant that what we experience is a single, unified scene. To illustrate what is meant by this, consider your experience of this sentence: when you read the words, you do not experience the shapes of letters separately from their meaning; rather, they are experience as whole. Paying attention is both singular and unified; it is simultaneously informative and integrated.

These observations may seem trite, but they lead to some useful insights for the designer tasked with supporting operators' attention in a control room. Any aids in reducing uncertainty correspond to supporting an operator's attention, since it reduces the amount of work they must do in the true information sense. Alerts and decision support systems may provide such aids as long as they succeed in reducing uncertainty and not just adding more entropy[67].

In the words of neurologist and author Anil Seth, the concept of integrated information can be interpreted as the "what-is-it-like-ness" of any specific conscious experience.[68] Let's consider the idea that the phenomenon of experience has a counterpart in design. A good place to start is Don Norman's affordances, which, as I described in Chapter 2.2, he defined as the "what-is-it-for-ness" inherent in material objects.[69] Affordances in this light serve an important role in shaping our attentive experience by quickly disposing all things a particular artifact is not—creating information in the true sense.

As for integrated experiences, a useful insight is found in what Aude Oliva and Antonio Torralba call the "gist" of a visual scene.[70] Testing their idea, they found that humans could recognize scene gist in a mere glance (200 milliseconds of exposure)—faster than many current image recognition algorithms. We are, in other words, excellent integrators of information. Research like that presented by Oliva and Torralba shows that such integration depends on sensory perception of the scene, and specifically of a spatial layout of objects that associate them with their semantic category. Consider the scene of a beach: we initially recognize the scene as a single entity, and *then* process information about objects and parts—the circle is a beach ball, the shadows on the beach come from tall palm trees. This suggests that remote control rooms should recreate the local scenes completely, leaving it to the

---

[67] An over-abundance of alarms on ship's bridges is a common complaint among seafarers that can lead to complacency, annoyance, and, in the worst case, so-called "cry-wolf" syndrome. My colleague Brit-Eli Danielsen examined this issue (among others) in a recent PhD thesis. Brit-Eli Danielsen, "Usability in Ship Bridge Design - A Mission Impossible? A Qualitative Study of Maritime Stakeholders' Perspectives on Usability in Ship Bridge Design" (NTNU, 2023), https://hdl.handle.net/11250/3069488.
[68] Seth, *Being You: A New Science of Consciousness.* Chapter 2, page 53.
[69] Norman, *The Design of Everyday Things: Revised and Expanded Edition.* Page 11.
[70] Aude Oliva and Antonio Torralba, "Building the Gist of a Scene: The Role of Global Image Features in Recognition," in *Progress in Brain Research*, ed. S. Martinez-Conde et al., vol. 155 (Elsevier, 2006), 23–36, https://doi.org/10.1016/S0079-6123(06)55002-2.

operator to integrate them into a unified scene laden with meaning. As we will see later, operators in a simulator test I conducted complained that their job would have been easier if they had been presented with a 360-degree field of view—an arrangement that would have allowed for panoramic scene gist. A control room should allow both for reducing uncertainty and for experiencing the scene as a single entity, without sacrificing one for the other.

Later in this thesis, we will explore phenomena related to attention, including inattentional blindness, multitasking, and vigilance decrement. But first, we will turn to pragmatic approaches in designing a control room to support operators' work: the topic of the next chapter.

3.5. Summary

At this point in the thesis, I have hopefully given you an appreciation for how machine autonomy can improve the effectiveness and safety of maritime navigation. What remains unclear, though, is how to realize this well-meaning vision. One of the most pressing questions remains: how do we account for human cooperation when integrating AI systems into navigation work? In this chapter we have explored concepts and generated insights about how human-AI interaction ought to unfold in the largely social activity of navigating a ship, a topic explored in detail in *Article 2*.

Having wrestled with some important concepts, we now move on to more practical matters of design. In the next chapter, I will outline the plans for the design, testing, and building of a real remote control center which eventually became the "NTNU Shore Control Lab." I will also outline my strategy for linking the control room to the *milliAmpere2* test ferry, starting by building a virtual simulator.

4. Remote control center design

Far better an approximate answer to the right question, than an exact answer to the wrong question.

–John Tukey

In April 2020 came the announcement that the Department of Design had been awarded funding to build a laboratory space for investigating remote supervisory control of autonomous ships. "As research and development gathers pace," I had written in the funding proposal, "one outstanding question remains: how are human operators expected to safely and efficiently supervise, monitor, and manually intervene in the control of autonomous vessels?" The proposal was underwritten by four departments at NTNU, making the case that control room design shouldn't be pigeonholed into a single discipline but rather should be expanded into a multidisciplinary arena. Shortly after the news broke that we'd been successful in our bid, plans started in earnest to commission what would be eventually be the NTNU Shore Control Lab.

In *Article 3* in this collection,[71] I present the process of designing and building the NTNU Shore Control Lab. In this chapter, I put the contents of this article into context, explaining some of the background behind important design decisions, including why I chose to collaborate with the *milliAmpere2* project and with the open-source virtual simulation project *Gemini*. The Shore Control Lab marked an important milestone in my PhD journey: a middle ground between the theoretical foundations of designing a control room and the practical

[71] Erik Veitch, Thomas Kaland, and Ole Andreas Alsos, "Design for Resilient Human-System Interaction in Autonomy: The Case of a Shore Control Centre for Unmanned Ships," *Proceedings of the Design Society* 1 (2021): 1023–32, https://doi.org/10.1017/pds.2021.102.

aspects of building one. As we will see later in this thesis, it also opened the doors to empirical testing.

4.1. NTNU Shore Control Lab

In June 2020, my supervisor Ole Andreas Alsos and I stood on a bare concrete floor in the newly built Trondheim Maritime Center at the Nyhavna neighbourhood in Trondheim. This was to be a new NTNU office for students and faculty working on maritime autonomy, bringing together several disciplines under one roof. The bare floor symbolized something momentous: a chance to shape the future of autonomous ship research at NTNU—a blank slate.

Ole and I quickly got to work with initial sketches. Our vision conveyed an orientation towards user needs, featuring generous space, abundant natural light, and quality air conditioning, as well as an orientation to the professionalism of maritime navigation work. To ensure that we met the needs of our research collaborators, we made flexibility the core of the lab's design. Practically, this meant that features in the control room could be easily taken down and reconfigured, allowing for researchers to test different hypotheses as easily as testing furniture arrangements in a room. To enable customizability, an aluminum frame was installed around the room with special slots allowing for the rapid installment and dismantling of connectors, joints, and customized mounts. Floor-to-ceiling glass walls were arranged with the building's architect to maximize natural light, and floor-to-ceiling curtains were installed to allow the room's inhabitants to adjust this natural light source as needed. Two large screen displays were installed with sightlines extending to most of the room. Since these displays were installed on sliding mounts and doubled as touch screens, they could also be configured in working mode at a height appropriate to the user. Two workstations were installed near the center of the room on adjustable standing desks, each with their own 60-inch widescreen capable of inputting two separate high-definition video sources. The computational power for the room was sourced from a single computer capable of handling high-fidelity graphical simulation. To maximize flexibility of audiovisual content over all displays in the control room, an Audio-Visual over Internet Protocol (AV-over-IP) network was installed allowing audio and video signals to be encoded and decoded over an Ethernet network and distributed to any device on the network. This included the Instructor Station located in a nearby room, where a researcher could monitor the control room over Closed-Circuit Television (CCTV) and orchestrate user testing at a separate workstation. This enabled flexible deployment of multimedia systems over a network infrastructure that could be easily reconfigured or expanded. The control room was like a Lego set: it looked one way but invited disassembly and rebuilding according to a researcher's own desires.

One thing was clear during the design process: it was not good enough to establish design specifications based purely on our own visions. The design process needed to ground design specifications in researchers' needs while also reflecting the needs of operators that would be working there during technology trials and user testing. Later in this thesis, I will describe the human-centered design methodology I adopted to guide the design process for the NTNU Shore Control Lab, which lent the design process rigour, transparency, and accountability.

4.2. milliAmpere and milliAmpere2 test ferries

At this point in my thesis, I would like to turn back the clock a few years. In fact, I want to return to my very first day at NTNU. On that day, I entered a room at the Department of Electronic Systems one brisk October morning in 2018 as a research assistant. The "Autoferry" project[72] was kicking off, and my first assignment was to assist in this project. After presentations from the PhD students, sveler were served from a little blue model boat. Sveler are Norwegian pancakes traditionally served on ferries. The little blue boat was a model of the *milliAmpere*, the prototype urban autonomous passenger ferry built at NTNU.[73] Towards the end of the meeting, there was an announcement from project coordinator Morten Breivik. Plans were underway to build a second ferry in the wake of *milliAmpere*. My ears perked up. A second ferry?

During the break, svele in hand, Egil asked me about my experience as a naval architect. "Sure," I said, "I know how to design boats." My fate was sealed: my first task as a research assistant would be to support the design process of the second autonomous ferry. This was to be a more polished, advanced prototype based on the first *milliAmpere*. It didn't have a name yet, but it would eventually become the *milliAmpere2*.

When we reassembled in January 2019, I had been given the title "design lead" for the *milliAmpere2* project. I had also been introduced to a master's student in the Industrial Design program named Petter Mustvedt and his supervisor, Einar Hareide, who, in their search for a project topic, had been introduced to the Autoferry group. At first, Petter was skeptical. An autonomous ferry? But his fate was sealed, too. Petter, Einer, and I got to work organizing the first design meetings for the *milliAmpere2*[74] (**Figure 4**).

---

72 This project was funded by NTNU as part of a series called "Digital Transformation," which ran from 2018 to 2023. NTNU, "Autoferry," 2020, https://www.ntnu.edu/autoferry.

73 Edmund F Brekke et al., "MilliAmpere: An Autonomous Ferry Prototype," *Journal of Physics: Conference Series* 2311, no. 1 (July 1, 2022): 012029, https://doi.org/10.1088/1742-6596/2311/1/012029.

74 Petter reported his work in a master's thesis, which turned out to be an excellent account of the design process behind the *milliAmpere2* ferry. Petter Mustvedt, "Autonom ferge designet for å frakte 12 passasjerer trygt over Nidelven" (Master thesis, Trondheim, Norway, NTNU, 2019). (Only available in Norwegian.)

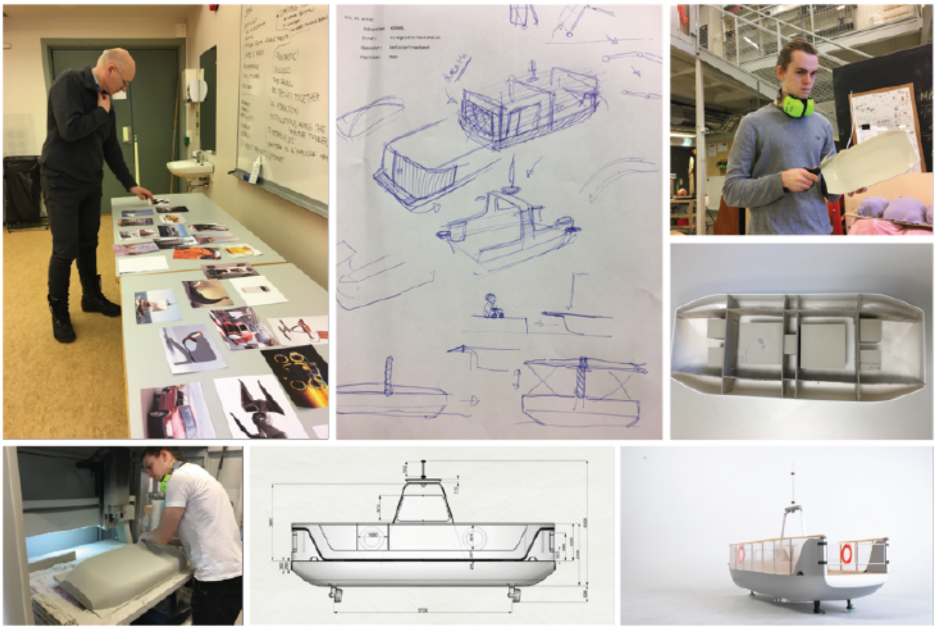


Figure 4. Design process for *milliAmpere2* (clockwise from upper left: Egil Eide in design workshop; early sketches; Petter Mustvedt in workshop; early scale prototype; detailed 1:10 scale prototype; detailed general arrangement; CNC milling for scale prototype).

During the early design phase, little thought was afforded to the remote control center that would inevitably be needed to supervise the *milliAmpere2*. All the attention went to the ferry and especially its complex array of sensors and computer systems. Of course, this was a logical arrangement: build the autonomous ferry first, work out the formalities later.

The design was ready in June 2019, and in the fall of the same year the contract for constructing the aluminium hull was awarded to a small shipyard located a few hours outside of Trondheim. Over the years 2020-2022, the *milliAmpere2* was outfitted with equipment, all the while adhering to a remarkable degree to the design laid out by Petter, Einar, and me (**Figure 5**). Parallel to the physical construction, I also worked on a digital twin of the *milliAmpere2* to be used in simulation testing—a topic I will explore in more detail next.

Figure 5. *milliAmpere2* digital render (left; image by Petter Mustvedt, 2 May 2019) and physical construction (right; photo by Egil Eide, 1 Nov 2021).

4.3. Developing a simulator for the NTNU Shore Control Lab

Standing at the "Ocean Week" stand in May 2019, I asked Kjetil Vasstein if he was nervous. King Harald V would be arriving in a few hours and was planning to stop at a stand featuring our autonomous ferry research. Kjetil's virtual simulator featured among the stand's provisions. It had proven a hit, with people stopping by the stand to play it using a game controller, mostly making the ferry crash into things. "No," replied Kjetil in his characteristic flippant style. "Should I be?"

The *Gemini* simulator was the result of a student project;[75] the brainchild of Kjetil and a few other master's students in 2019. With the release of the first version, the leader of the *milliAmpere2* project, Egil Eide, soon emerged its most enthusiastic player. I recall one afternoon in May 2019, sitting in a room with thirty invited guests where I was leading a critical design review of the *milliAmpere2* ferry as part of my research assistant role. Egil had connected the video game to the main screen and was laughing as he crashed the digital *milliAmpere2* into a sailboat, which promptly sank in the canal. One of the invited guests, Jan Boye Andersen, an expert in aluminum hull design, looked on with consternation. I didn't mention it during that meeting, but what I saw in the simulator was more than just a way to wreak virtual havoc. It was powerful tool for interacting with people and for communicating concepts related to control of autonomous vessels: a thought experiment in action.

[75] What started as a student project was eventually published in a conference paper by Kjetil Vasstein and colleagues in a more in-depth treatment of the potential for digital twins in the maritime domain. Kjetil Vasstein et al., "Autoferry Gemini: A Real-Time Simulation Platform for Electromagnetic Radiation Sensors on Autonomous Ships," *IOP Conference Series: Materials Science and Engineering* 929 (November 27, 2020): 012032, https://doi.org/10.1088/1757-899x/929/1/012032.

With the Shore Control Lab underway, I wanted the *Gemini* simulator to feature in the lab. In the spring of 2020, shortly after we had received news of the lab funding (and newly locked down during the first wave of Covid), I downloaded the repository containing the source code for the *Gemini* platform[76] and starting a few introductory lessons in Unity programming. In a matter of a few short days, I conceded that I needed a bona fide programmer. Before long, I had a job ad distributed to students at NTNU. After a round of interviews, I found my guy: a fourth-year student at the Department of Computer Science named Mikael Hansen.

Mikael immediately demonstrated his skills in Unity programming and soon after took on the simulator challenge as his master's project by adding a "Scenario Builder" to the *Gemini* platform.[77] Working together with another master's project from the design department led by Sondre Ek,[78] the two turned the simulator into a multi-purpose research platform.[79] With a detailed model of Trondheim as its canvas, a user could design their own ferry routes in addition to setting in other vessel traffic, manipulating everything from the location of the ferry's docks to the speed of other boats. The *Gemini* simulator began to resemble Sim City. Emergency situations were added, such as battery fires and autonomy failures, which could be made to occur after a defined amount of elapsed time or after event dependencies triggered them. The key concept was that an operator could immerse themselves in the custom-built scenarios, whether for purposes of research about interface design, behaviour, emergency response, training, or for other aspects of supervisory control research. Whatever could be concocted by the "Scenario Designer" could now be experienced by an operator in our updated version of the *Gemini* simulator.

What this provided for the NTNU Shore Control Lab was a research platform that could be used to collect data about operators' performance and experiences during tailor-made virtual scenarios of supervisory control. What's more, I realized that these virtual scenarios could reproduce the *milliAmpere2* operational domain and thus could be validated at full scale. I set to work designing an experiment. The premise was simple: design scenarios and invite participants to complete them at the Shore Control Lab. Then, adjust independent variables in the simulator settings and measure the corresponding variations in performance. This way,

76 As of June 2023, the *Gemini* platform is available as a Git repository: https://github.com/Gemini-team/Gemini

77 Mikael Hansen, "Developing a Video Game for Research and Prototyping of Unmanned Maritime Vessels" (Master thesis, Trondheim, Norway, NTNU, 2022), https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3028969.

78 Sondre Ek, "Design of Simulator for Researching Autonomous Marine Vessels" (Master thesis, Trondheim, Norway, NTNU, 2022), https://bibsys-almaprimo.hosted.exlibrisgroup.com/permalink/f/13q4kuj/BRAGE11250/2996832.

79 As of June 2023, the Scenario Builder upgrade of the *Gemini* platform is available as a Git reposity: https://github.com/mikael-rh/ScenarioBuilder/

I could begin to assess causal relationships between design elements and supervisory control performance. Later in this thesis, I present the details of just such an experiment and its results.

4.4. Summary

This chapter has outlined three components that came together to allow for the design of a flexible, research-oriented functional prototype of a remote control center for autonomous ships. These three components were (i) the NTNU Shore Control Lab, (ii) the *milliAmpere2*, and (iii) the *Gemini* virtual simulation platform. The combination of these three components would eventually serve a central role in my PhD research by allowing me to cross the divide from theory to testing.

This chapter has provided background to accompany *Article 3* in this thesis, which presents the design process for the NTNU Shore Control Lab. We will revisit the Shore Control Lab again several times throughout this thesis as my main stage for ideation, prototyping, and testing, as well as a meeting ground for the multiple disciplines converging on remote control center design.

5. Factors influencing remote supervisory control of autonomous ships

It's hard to say we're lucky when we face a crisis, but we at least have the luxury of knowing that action is called for, of being forced to move. The true tests of skill and intuition come when everything looks quiet, and we aren't sure what to do—or if we should do anything at all.

—Gary Kasparov, discussing chess

With the NTNU Shore Control Lab built and the virtual simulator set up for research and featuring *milliAmpere2* ferry, it was time launch the next stage of my PhD research: testing.

There were any number of options available for testing. One option was to focus on the user interface, for example by developing a series of prototypes and conducting user tests to identify the best candidates. This approach could certainly have provided valuable contributions; however, I was guided by a suspicion that there was more to supervisory control that just the user interface.

Eventually, I framed my testing campaign around human factor influences on remote supervisory control. Personally, I am a little ambivalent about the term "human factors." What makes something a human factor, anyways? Considering what we know now about how humans perceive the world around them and make decisions, surely a more apt term would be interaction factors? At the very least, I wanted to eschew any parochial blame on "human error," which by now symbolized a surefire pitfall in the design process. What's more, coming from an engineering background, human factors were often presented to me as failures: a hatch placed in the wrong place, a lifting hook welded to the floor that people tripped on. But the more I became involved in design, the more I came to appreciate human factors as a multidisciplinary field contributing to improved safety and better design of complex systems—efforts that fit squarely in my own journey to design a control center for autonomous vessels.

The goal of the testing would be straightforward: select a series of independent factors to study, measure the variation in outputs, and uncover causal relationships between independent factors and outputs. The independent factors were framed as human factors; the outputs were performance metrics collected in a simulator, including response time, performance score, and outcome of a collision avoidance maneuver.

Running an experiment is a lot of work. It's also a major investment of time and resources. That's why I decided, in fall 2021, to enlist some collaborators in my experimental campaign. I ended up inviting two researchers to join: one was a master's student in cognitive psychology from University in Bergen named Kristin Senderud, and the other was a visiting PhD student from Wuhan University of Technology named Tingting Cheng, who was hosted by NTNU's Centre for Autonomous Marine Operations and Systems (AMOS) at the Department of Marine Technology.

Meeting regularly at the Shore Control Lab, the three of us iterated a list of human factors to study. At the time, we didn't call them human factors—they were just independent factors. In fact, each of us bought into the experiment with their own paradigmatic approach to interpreting the relevance of these independent factors. I called them human factors, Kristin called them cognitive factors, and Tingting called them performance-shaping factors. Yet, we were all talking about the same thing: how humans interacted with a technology in the context of a safety-critical system.

The experimental campaign and its results are presented in *Article 4* in this collection.[80] In this chapter, I provide some additional context that will hopefully improve readers' understanding of the contributions of the experimental efforts.

5.1. Which factors to study?

In a complex system, there are seemingly an infinite number of possible factors affecting a supervisor's performance. Perhaps the interface is designed in such a way to make it difficult to locate some crucial piece of information. Maybe the operator's cat ran away recently. It is plausible that both factors may have a strong negative influence on this operator's performance. The multiplicity of influencing factors begs the questions: which ones are important, and how can we isolate them for study in a scientifically rigourous way?

---

[80] Erik Veitch et al., "Human Factor Influences on Supervisory Control of Remotely Operated and Autonomous Vessels (Preprint)" (SSRN, April 17, 2023), https://dx.doi.org/10.2139/ssrn.4437731.

Out of the infinite possible factors to study, my approach was relatively humble: select five relevant and mutually independent factors for examination. Any less than five seemed to generate to narrow a scope for a major experimental effort; any more seemed to stretch our capacity to study them in a single experiment. The question of *which* factors were relevant enough was trickier. The selection process stemmed from a mix of lively discussions with my colleagues, conscientious deduction from scientific literature, and intuitive hunches (later in this thesis, I will formally explain this approach under the label of *abduction*). I introduce all five selected factors formally in *Article 4*, but I wanted to provide readers with some additional context here. In the rest of this chapter, I will outline the selected human factors and draw parallels from the game of chess to help explain what I mean by studying human factor influences on performance.

Remote supervisory skills

Skill is probably the first thing that comes to mind when considering human performance. It is also perhaps the most important. In chess, skill is measured with a numerical score known as the "Elo rating system." The premise is simple: you start with 800 points, indicating a novice level. Two players with equal ratings are expected to score an equal number of wins: a win rate of 50 percent. A difference in ratings is calibrated such that a player with 200 points more is expected to win 75 percent of the time. Of course, the ratings are approximate. People can have good days and bad days (good news for the bookkeepers of chess's popular betting websites). Even Arpad Elo, the rating system's inventor, was skeptical of its accuracy. He once wrote, "The process may be compared to using a meter stick waving in the wind to measure the position of a cork bobbing on the surface of waving water."[81] Still, the rating systems are widely used to classify players from novice (less than 1000 points) and "Class D" (1400 and below) to Experts (2000-2200) and International Grand Masters (2500 and above).

Classifying skill applies to a lot more than just games like chess. In maritime navigation, skill is measured with a license certificate. In Norway, this license varies from Deck Officer Class 5 to Class 1 and is governed by the Norwegian Maritime Authority.[82] In recruiting participants for experiments, I used this as my yardstick for skill in navigating and maneuvering a vessel. But the question remained: did licensed mariners really possess the skills required of a supervisory controller? The conventional thinking was yes, but this claim raised more than a few skeptics' eyebrows.

81 Arpad E Elo, *The Rating of Chessplayers, Past and Present*, Second edition (New York, N.Y., USA: Arco Publishing, Inc., 1978). Page 28.

82 Norwegian Maritime Authority, "Forskrift om kvalifikasjoner og sertifikater for sjøfolk," 2011, https://www.sdir.no/sjofart/regelverk/rundskriv/kvalifikasjoner-og-sertifikater-for-sjofolk/. Pages 9-14 (only available in Norwegian).

One of those skeptics was Jason McFarlane, who when I phoned in January 2020 was working at Kongsberg Maritime and leading the LOAS project (he has since left and founded the start-up Hyke, which is developing urban autonomous ferries in Oslo). I wanted to chat with him about the skills needed to be a remote operator. Jason suggested that video game players likely had the skills needed to be remote operators even more so than navigators. Why wouldn't this be the case? he asked. Gamers were experts at interacting with autonomous agents on a screen and performing the mental gymnastics of handling several visual entities at once, often under time pressure and with teammates in distributed locations—all relevant skills in control room operations. Over the years, I encountered this idea again and again. Checking the research literature, I discovered several studies that formally investigated differences in skilled gamers and skilled pilots in the context of fighter jets and miliary drone pilotage.[83] I found no such studies for maritime pilotage. I figured it was time to test the "gamer hypothesis" formally and contribute empirical findings to the largely speculative discussions about skills in supervisory control for autonomous ships. Who possessed the more transferrable skills to supervisory control? Gamers or navigators?

Vigilance decrement

"Wow! This seems like the most boring job ever!" I've led many tours of the NTNU Shore Control Lab since its opening in 2021, and I've encountered this exclamation more than once. People intuit that remote supervisory control is an inherently boring job. In a way, they are right. There is a significant body of research showing that after about 20 minutes monitoring automated processes, we experience what cognitive psychologists call *vigilance decrement*.[84] Crucially, this is the case even when we are explicitly told to pay close attention. Basically, our mind wanders. Our capacity to pay attention diminishes. We feel bored. We have all seen the videos of meercats trying to stay awake while guarding their nest. They droop and topple over, awakening with a jolt as their recall their mortal assignment.

I knew I wanted to formally address the boredom question in my experiment. Immediately, though, I was faced with a problem: boredom has no universal definition. This risked inviting

---

83 R. Andy McKinley, Lindsey K. McIntire, and Margaret A. Funke, "Operator Selection for Unmanned Aerial Systems: Comparing Video Game Players and Pilots," *Aviation, Space, and Environmental Medicine* 82, no. 6 (June 1, 2011): 635–42, https://doi.org/10.3357/ASEM.2958.2011; Jinchao Lin et al., "Video Game Experience and Gender as Predictors of Performance and Stress during Supervisory Control of Multiple Unmanned Aerial Vehicles," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 59 (SAGE Publications Sage CA: Los Angeles, CA, 2015), 746–50.

84 Robert Molloy and Raja Parasuraman, "Monitoring an Automated System for a Single Failure: Vigilance and Task Complexity Effects," *Human Factors* 38, no. 2 (1996): 311–22, https://doi.org/10.1177/001872089606380211; R.A. Grier et al., "The Vigilance Decrement Reflects Limitations in Effortful Attention, Not Mindlessness," *Human Factors* 45, no. 3 (2003): 349–59, https://doi.org/10.1518/hfes.45.3.349.27253.

weaknesses in any scientific attempt at measuring it. Consulting with Kristin, the cognitive psychology master's student, we decided to use a two-part questionnaire introduced by Wijnand van Tilburg and Eric Igou.[85] The two-part questionnaire asked participants to score their overall boredom on a scale of 1 to 5, as well as scoring related experiences on score of 1 to 7. Related experience included restlessness and the desire to be challenged and to do something more meaningful, among other things. Together, the questionnaires gauged boredom experience and by extension offered a window into vigilance decrement. Problems with semantics aside, I at least had a benchmark with a methodological pedigree by which to assess boredom and its counterpart, vigilance, across the 20-minute threshold.

In my search for quantitative insights into vigilance decrement, I learned that eye-tracking goggles could measure pupil dilation over time and provide a reasonable proxy for cognitive load—not vigilance per se, but potentially related. As we make a mental effort to solve a problem (e.g., working a math problem like 38 x 17), our pupils dilate as if searching for more information in our visual field; as we relax our mental efforts, our pupils contract accordingly. In my experiment, I collected pupil diameters during participants' trials. Later, I will present some of these findings. For the initial analysis, however, I decided to keep things simple: I would approach vigilance decrement as a subjective experience associated with boredom, gauged qualitatively using a questionnaire.

We settled on comparing two groups: one with a 5-minute scenario and the other with a 30-minute scenario. In the experiment, this would serve two purposes: first, it would allow me to confirm that a vigilance decrement really was present after about 20 minutes. Second, it would provide me with the information I needed to uncover whether a causal relationship between vigilance decrement and performance existed. Soon, I would be able to answer the question "Does boredom undermine remote supervisory control?"

Multitasking

Multitasking lies at the very heart of remote supervisory control of autonomous ships. After all, what's the point of having a one-to-one ratio between a remote operator and their autonomous vessel? In this case, you could've saved a lot of trouble by just being onboard.

The one-to-many model is the elevator sales pitch of automation. Consider, for example, the chatbot Eliza, written in the 1960s by M.I.T. computer scientist Joseph Weizenbaum and designed to simulate a form of conversation stemming from Rogerian therapy.[86]

---

[85] Wijnand A. P. van Tilburg and Eric R. Igou, "On Boredom: Lack of Challenge and Meaning as Distinct Boredom Experiences," *Motivation and Emotion* 36, no. 2 (June 1, 2012): 181–94, https://doi.org/10.1007/s11031-011-9234-9.

[86] Joseph Weizenbaum, "ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine," *Communications of the ACM* 9, no. 1 (1966): 36–45.

Weizenbaum had originally doubted that his program could simulate meaningful human interaction, so he was surprised when doctors and psychologists expounded on its potential value. "The human therapist," reported one group of psychologists, "involved in the design and operation of this system, would not be replaced, but would become a much more efficient man since his efforts would no longer be limited to the one-to-one patient-therapist ratio as now exists."[87] Even today, the one-to-many-model for automated therapy persists. New companies like Woebot, founded in 2017 and based on Cognitive Behavioral Therapy, are modelled on the same vision of scalability.[88] Skeptics decry the same criticisms that emerged a generation ago: how can computers, set loose to navigate a human world, provide meaningful interactions with its interlocuters?

Autonomous ships face a similar conundrum to the Eliza and Woebot chatbots. In a world where transportation demand is increasing and interest in maritime careers is diminishing, autonomous ships are quickly billed as our saviors. The human captain, we say, will not be replaced, but will become a much more efficient captain since their efforts would no longer be limited to the one-to-one captain-ship ratio as now exists. What distinguishes autonomous ferries from chatbots, though, is that the former take on a real physical form in the world—not just a digital avatar or persona. Vessels move around and can crash into things. It doesn't take much convincing to concede that we should keep a close watch on them and program the ability to intervene. Instead of chatbots, a more suitable analogy for multitasking in control centers may lie in daycare workers. Specifically, the question of how many operators one needs per vessel may be compared to the question of how many daycare attendants one needs per child. Children are, after all, autonomous agents too, and are prone to various form of intervention, be it diaper changes or tantrum management. In Norway, regulations set the minimum requirement at one daycare employee for every three children under three years old and one daycare employee for every six over three years old. These numbers, however, appear to stem more from conventional wisdom and union parlaying than research. Still, it is an illustrative analogy. Can one-to-*x* guidelines be set, say, for autonomous cargo shuttle vessels? Coastal ferries? Urban passenger ferries in constrained waterways? The experimental campaign would help find out.

---

[87] Kenneth Mark Colby, James B. Watt, and John P. Gilbert, "A Computer Method of Psychotherapy: Preliminary Communication," *The Journal of Nervous and Mental Disease* 142, no. 2 (1966), https://journals.lww.com/jonmd/Fulltext/1966/02000/A_COMPUTER_METHOD_OF_PSYCHOTHERAPY__PRELIMINARY.5.aspx.

[88] As of June 2023, the website for Woebot claims that its chatbot offers solutions to a "severe lack of therapists" and long wait times. "Woebot Health - About Us," *Woebot Health* (blog), accessed June 9, 2023, https://woebothealth.com/about-us/.

Time pressure

If you've ever played chess, then you know that time pressure can significantly affect performance. In fact, world champion titles in chess have separate categories for rapid and even blitz categories, producing their own niche champions in a time-pressure variant of the game.

In real critical situations, people are rarely given the opportunity to sit back, brew a cup of tea, and mull over a problem. On the contrary, centers of coordination, as Lucy Suchman noted, are distinguished by time-criticality (Chapter 3.1). In remote control rooms, problems come yoked to a ticking clock. The question is, how does this time element affect performance of supervisory control?

Ships have an apparent advantage in that they move slowly, allowing the operator much more time to intervene in a course change than, say, an autonomous car speeding along a winding mountain road. However, ships also tend to be underactuated, meaning their control is closer to riding an elephant than to a car. Ship navigators have the unfortunate distinction of being able to watch an accident unfold minutes before impact. An experienced captain of an offshore supply vessel operating on the Grand Banks of Newfoundland once told me "Speed is like salt: easy to put on, hard to take off!" Accident investigators after the Helge Ingstad collision with Sola TS, for example, uncovered that the collision took place on the scale of minutes—not seconds, as we might have been inclined to imagine. The Sola TS first attempted to contact the Helge Ingstad over marine radio shortly after the frigate appeared on their radar 13 minutes before collision.[89] On the Helge Ingstad, meanwhile, bridge officers noticed Sola TS 10 minutes before collision. They also tried radioing Sola TS, also receiving no response. Finally, 2 minutes before collision, radio contact was established, but it was too late to discuss intentions—the ships were doomed for a collision course. This time scale seems generous in retrospect: a disaster unfolding in slow motion. Yet, the report concluded that both bridge crews had very little time to react. What the report did not comment on was how time pressure potentially hindered their ability to obtain situation awareness, establish communication, and adjust the course direction appropriately.

For supervisory control of autonomous ships, time pressure will likely be a defining aspect of performance. While not exactly a "human factor," per se, it translates to something akin to decision-making aptitude under stress. In chess, the very best players, noted Gary Klein, tended also to be the best in rapid chess.[90] Experts distinguish themselves on making

[89] Norwegian Safety Investigation Authority, "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."
[90] Klein, *Sources of Power: How People Make Decisions.* Chapter 10, page 171.

decisions quickly, rather than deliberating on options. This begged the question: does the same hold true in remote supervisory control?

Decision support

I had already seen a Decision Support System (DSS) in action on the bridge of the Horten-Moss ferry on my truncated journey two days before the first Covid lock-down. It looked like an iPad mounted to the officer's control console. On it was displayed a graphical representation of the docking sequence, including the zones in which the navigator should commence various maneuvers and the status of various system components.

There exists no universal definition for DSSs. They simply refer any type of interfacing tool that helps us reach decisions, whether that tool involves visual, audio, or some other sensory output. One might say the canary in the coal mine is a DSS. The same can be said of the stinky chemical compound ethyl mercaptan, which is added to the naturally odourless and highly explosive gas propone. In my chess mobile app, I can turn on decision support and get computer-assisted moves, which appear as little green arrows on the chessboard (this is only permitted when playing other bots). Usually, though, the term DSS is used in an organizational context, where decision making is made under conditions of uncertainty and time pressure, and where display of information can ostensibly support this process. DSSs therefore usually involve displaying large amounts of information on digital displays, with graphical representation of complex systems to help guide the operator's mental model of how various elements relate to one another.

DSSs will be important tools for supervisory controllers at remote control centers. DSSs empower the human operator as the responsible agent in a human-AI collaboration, legally and morally, while also elevating their decision-making, teamwork, and creative problem-solving abilities. One challenge in studying them, though, is that DSSs are never in statis; they are the product of endless design iterations, improvements, and upgrades. It my experiment, I sought to design a DSS system that incorporated the latest thinking in interface design for autonomous ships, while conceding that I couldn't accurately represent all interface design features over all time. I took the challenge anyways, content with comparing two groups: one with a DSS and the other without (details about the DSS, including its appearance and functionality, are in *Article 4*). Later, I will present the results of which group demonstrated better performance, shedding light on just how important the DSS is in remote supervisory control.

5.2. Summary

In this section, I have introduced five factors and their respective hypothesized influences on remote supervisory control of autonomous ferries. These five factors are skill, vigilance, multitasking, time pressure, and decision support. Later in this thesis, I will describe the methodology I adopted to systematically test these hypothesized factor effects as well as the results of the experimental campaign.

So far, we have taken a birds-eye view of important concepts and theoretical frameworks relevant to my PhD research. We have also seen early field observations and the beginnings of a research platform emerging from combining the NTNU Shore Control Lab, *milliAmpere2*, and *Gemini*. Next, I will begin to stitch all these elements together, starting with a description of the methodologies I used to collect data along the way.

6. Methods

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

—Sherlock Holmes, *A Scandal in Bohemia* (Sir Arthur Conan Doyle)

It's time to outline the methods I used in my PhD research. If I want to convince you of anything in this thesis, it's that between the concepts I've presented and the ultimate findings of my work, there lies a deliberate, considered method. It's these methods that distinguish my findings from mere conjecture sound scientific inferences.

You can read detailed descriptions of the methods I used in the individual articles appended in this thesis. The articles alone, however, shed no light on how these individual methods related to and complement one another. In this chapter, I wish to convey the methods I adopted in the context of a single PhD research arch, starting with literature review and ending with experimental testing.

6.1. Systematic literature review

To any new PhD candidate who might be reading this thesis, I cannot stress enough the importance of literature review. Sure, it seems tedious and boring—the antithesis of scientific discovery. I am convinced, however, that literature review pays dividends later when it's time to commit your time and energy to a particular research question and subscribe to particular theoretical frameworks. Literature reviews reduce the randomness of an open-ended research topic by identifying key patterns, synthesizing key findings, and uncovering gaps and inconsistencies. They turn the process of picking a research direction from one based purely on a hunch to one based on an informed decision.

I set out early in my PhD journey to do a literature review. After updating it several times along the way, I eventually published it as *Article 1* in this collection. The methodology I used is called a *systematic literature review*. This may be distinguished from other types of reviews like narrative, meta-analysis, and focused literature reviews. There are no rules for how to complete a literature review of any kind, just as there are no rules for how to produce any particular type of research. However, there are guidelines one can follow. For a systematic review, a popular guideline is provided by the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses," known as "PRISMA" for short.

PRISMA statement

The PRISMA statement[91] was introduced in the field of medicine and psychology to help guide prospective authors achieve the rigour, transparency, and consistency they hoped to convey in their literature reviews. Its value, however, is by no means limited to these fields. Scientists in all sorts of fields who wish to produce high-quality literature reviews refer to PRISMA as a marker of quality. The PRISMA statement takes the form of a checklist. While not all items were applicable for my case, I used the checklist to structure my review in *Article 1*. This included clearly specifying inclusion and exclusion criteria for the review, listing all databases searched, and drawing up a flowchart to outline the study selection. The process was documented carefully and stayed true to the review criteria, without exceptions.

Although it wasn't expressly recommended, I enlisted the help of my supervisor and co-author Ole Alsos to triangulate the abstract screening. We screened studies independently based on the criteria, then reconvened and compared notes. Wherever we disagreed on whether a study should be included, we made our case based on the proposed criteria. This helped to capture aspects any one of us screening alone might have overlooked.

It took more than six months before I heard from the editor of Safety Science, the journal to which I submitted the review article. The recommendation was for major revisions. While this was great news, it also meant my review was quickly becoming outdated. I had selected 2020 as the date cut-off; at this point, 2021 was just around the corner. Fortunately, the PRISMA method allowed me to easily update my search. I simply used my flowchart exactly as I had done before, this time with 2021 set as the date of publishing. As it turned out, almost half (19 of the 42 total papers) I reviewed were published in 2021, a strong indicator that the topic I was investigating was growing rapidly. That the research I was surveying was more frequently appearing in peer-reviewed journals also indicated that the research topic was maturing.

[91] David Moher et al., "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement," *Annals of Internal Medicine* 151, no. 4 (August 18, 2009): 264–69, https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

6.2. Qualitative research methods

My investigations into remote control center design eventually led me to interview professional designers and navigators working aboard highly automated vessels. In fall 2019, I interviewed nine technology designers; later, in the spring of 2020 I interviewed five navigators who worked aboard ferries with automated crossing and docking technologies installed. Note that I used the word "designer" here in the same broad sense Simon Herbert uses it to describe any profession characterized by "transforming a situation into a preferred one."[92] These nine technology designers included researchers and business professionals, as well as professionals with the title "designer."

After conducting the interviews, I intuitively felt they contained a rich source of information. The question remained, though: how to translate this intuition into scientifically rigorous results? Perhaps from my background in engineering, I was engrained with a mistrust of drawing results from anything that wasn't composed of numbers. I had self-reflection enough to realize it was time to challenge this presupposition and I registered for a course called "Qualitative Research Methods." The course was held in the fall of 2021, which was good, because it gave me enough time to learn enough Norwegian to follow along in the lectures. (Graduate level courses at NTNU are meant to be in English, but this course, for reasons that remain unclear, was an exception.) On the first day of class, I got acquainted with my fellow students. Except for one other student, all were from the sociology department. The lecturer, too, was a sociologist, who wrote all her lecture notes in New Norwegian, the second official language of Norway that exists only in the written form. As if the course hadn't already gotten off to a bad enough start, when asked to offer criticism about a study that adopted qualitative methods, I tore it apart in front of the class. I said the method based its findings on hearsay, gossip, and worst of all, trite stereotypes. "Interesting," she replied. "This is one my most esteemed papers using qualitative methods. Does anyone else have any comments?"[93] Mortified, I looked down and figured that that was it, I would fail the course and stick to numbers. However, despite my misgivings (and occasional language troubles), what I eventually learned from the course was immensely helpful in my PhD research. Specifically, it gave me the confidence to analyze the interview data I had collected in the years previous and turn my intuitions and hypotheses into scientifically tested results grounded transparently in observation.

After the course, I was convinced I could trust in my own explanatory inferences based on qualitative data. In *Article 2* I chronicle in detail the process of systematically addressing my

[92] Simon, *The Sciences of the Artificial*. Chapter 5, Page 111.

[93] For the interested reader, I have included the reference to this paper. Olve Krange and Ketil Skogen, "When the Lads Go Hunting: The 'Hammertown Mechanism' and the Conflict over Wolves in Norway," *Ethnography* 12, no. 4 (December 1, 2011): 466–89, https://doi.org/10.1177/1466138110397227.

research questions using field observations and interview data as my sources. For me, adoption of the methodology also represented a shift in how I thought about all observation-based scientific methods. The realization was that making inferences from data was more than an exercise in pure deductive logic. To this aim, one of the great takeaways from the course was learning about *abduction*.

Abduction

One of the heroes of the course on qualitative methods was Charles Peirce. The reference book for the course, "The Art of Social Theory" by William Swedberg,[94] opens with a story about Peirce investigating the problem of who stole his jacket and a gold watch. In a series of insights, he eventually found all his belongings and the thief who took them. What surprised Peirce himself, a lecturer in logic at Johns Hopkins University, was that what lead him to his stolen belongings was essentially of a series of hypotheses, or good guesses. His insight was that a hypothesis emerged as a guess—a process he termed *abduction*. The role of abduction in science, believed Peirce, should not be overlooked. As Sweberg writes, "It is correct that without facts to test the hypothesis or the idea, the guess is of little value. But without the hypothesis or idea, there will be nothing to test and no science at all."[95] Peirce's great insight was that people do not generate hypotheses at random. On the contrary, scientific progress rides on the back of many accurate and insightful guesses. What if we could hone this faculty of guessing? This was Peirce's legacy and the thesis of Swedberg's book: to train abduction-oriented theorizing.

One useful approach in this training process is to be attuned to "surprising" results. As Swedberg explains, "At some point in your research you will find something surprising, something that doesn't fit the current state of knowledge. It is *this* that should be studied."[96] Reflecting on my own work, I recalled it was tempting during data analysis to reinforce ideas I had at the beginning of the research process. This called to mind Kahneman's "What You See Is All There Is" rule of thumb for cognitive biases. I learned to be aware of my own research biases that could end up manifesting as methodological heuristics—shortcuts to results that may come at the cost of more thorough and interesting analysis. Letting oneself be drawn to surprises, by contrary, provides a useful countermeasure. It helps to steer clear of biases stemming from pre-held beliefs and theorize in a way that might challenge the taken-for-granted ideas we hold dear.

Another insight about the abductive reasoning process may be borrowed from an analogy: the measurement of temperature. By using a thermometer in a room, you in fact change the

---

94 Richard Swedberg, *The Art of Social Theory* (Princeton, NJ, USA: Princeton University Press, 2014).
95 Swedberg. Introduction, page 7.
96 Swedberg. Chapter 2, page 39.

room's temperature slightly because some of the molecular energy in the room will be transferred to the thermometer. This same "observation effect" holds for any observation: the act of observing alone will affect the results and therefore any inferences drawn from them.[97]

In his 1962 classic "The Structure of Scientific Revolutions,"[98] Thomas Kuhn reflected on the way in which one views the world—what he termed a *paradigm*. Kuhn argued compellingly that scientific inference was paradigmatic, and that anyone claiming to draw uninterrupted lines from empirical observation to theory, as argued by his philosophical rival Karl Popper, was likely kidding themselves. Like the thermometer measuring temperature, observation itself is a product of the paradigmatic structure of science. Despite Kuhn's compelling arguments, though, most modern scientists still adopt the Popperian way of thinking, claiming strict adherence to objective methods eschewing and even vehemently denying subjectivity.

What's more, scientific judgment is also at the mercy of external pressures like political agendas, funding agencies, or even personal ambitions that may sway scientific judgment. Michael Strevens, in his book "The Knowledge Machine: How Irrationality Created Modern Science,"[99] Strevens recounts the story of Arthur Eddington, an English astronomer who, in 1919, sailed to Africa to photograph the position of stars during a total eclipse of the sun. Eddington's mission was to confirm or falsify Einstein's theory of general relativity, but the images he collected were blurry and almost useless. Regardless, he used them in detailed calculations to confirm Einstein's theory—one that was considered avant-garde at the time. Strevens explained that Eddington "wanted very much for Einstein's theory to be true, both because of its profound mathematical beauty" and because of his "ardent internationalist

97 The "observation effect" has caused quite a row in the field of physics. It has most famously been demonstrated in repeated experiments of Thomas Young's 1801 "double-slit experiment." Young showed that when light passes through two slits, it creates an inference pattern exactly as ripples on a pond would create distinct patterns from two disturbances. Therefore, light consists of waves. However, when repeating this experiment by sending individual particles through the slits, the interference pattern still emerges, confounding the light-as-a-wave theory. Yet *more* confounding is that when observing which slit individual particles pass through, the interference pattern collapses altogether into two slits, as we would expect from particle motion. In other words, just by observing the particles, their behavior is fundamentally changed from wave dynamics to particle motion. I recall learning this in first year physics and feeling that I would no longer understand physics. The observation effect gave rise to the field of quantum mechanics, where the act of observation itself plays a fundamental role in measurement (enshrined in the example of Schrödinger's hapless cat). No universal explanation of the observation effect exists in the quantum era. In this thesis, suffice it to say that observation and measurement are not independent.

98 Thomas S. Kuhn, *The Structure of Scientific Revolutions*, Fourth edition (Chicago, IL, USA: University of Chicago Press, 2012).

99 Michael Strevens, *The Knowledge Machine: How Irrationality Created Modern Science*, Audiobook (Prince Frederick, MD, USA: HighBridge, a division of Recorded Books, 2020).

desire to dissolve the rancor that had some Britons calling for a postwar boycott of German science."[100] Although scientists were skeptical of the blurry photos, they responded not by tearing down Eddington's work, as expected, but by launching data collection efforts of their own.

What I find appealing about abduction is that it accepts subjectivity without casting the whole scientific institution aside as hopelessly irrational. (In our age of misinformation and mistrust, the last thing we need is more undermining of science.) Abductive reasoning, by positioning the researcher *into* the methodology instead of outside it as an objective observer, even serves to empower scientists with more impactful communication. To illustrate this, consider the use of the first-person pronoun "I" and first-person plural "we" in scientific writing. It is more common in scientific writing to circumvent the first-person voice, adopting instead phrases like "the findings show" or "the results suggest." Such phrases imbue an air of objectivity to the data, as if the author were merely a messenger. However, consider what happens when we replace these phrases with "I show" or "we suggest." The resulting text no longer obfuscates the link between the data and the observers' inference. In *Article 2*, I used the first-person plural "we" to present the results. (Indeed, I adopted the first-person voice in *all* articles appended to my thesis, and even in the thesis itself.) I believe this strengthened the validity of my claims by staying true to the abductive-oriented theorizing at the heart of my PhD research and by avoiding passive language that assigns findings a standalone objectivity.

Methods inspired by Grounded Theory

Grounded Theory is based on a laudable premise: that theory can be "grounded" in data even when this data is qualitative. This is possible, maintain Grounded Theorists, as long as the adherent commits to transparently and rigorously tracing these lines from data all the way to theory.

During the peer-review process of *Article 2*, one of my reviewers in an early draft chastised me for my "uncritical" use of Grounded Theory (GT). They complained that the types of questions in the interview guide precluding "pure" GT methods. This was true. With the exception of the navigator interviews, I had learned about GT only *after* heading out into the world with my tape recorder. As a way around this, suggested the reviewer, why not just say you were "inspired" by GT techniques?

---

100 Strevens. Chapter 2, 06:05-06:18.

I did exactly this. Guided by Corbin and Strauss's book "Introduction to Qualitative Methods,"[101] which sets out the tenants of GT in its strictest sense, I pulled out elements that inspired me. If I had reservations about this, they were now tamed after the course I had completed on qualitative research methods and my discovery of abduction-oriented theorizing. The extensive field notes I made during and immediately after the interview analysis process were, under the framework of GT, *memos*—kernels of ideas that represented starting points for subsequent structured analysis. In total, I generated over one hundred memos, which often captured "surprising" elements in the spirit of Swedberg. Under the guise of GT, memos were synthesized into overarching narrative themes through *axial coding*. This process involved conserving the original language of informants when highlighting relevant parts of the data (so-called *codes*), thus helping to nip any pesky biases in the bud. It was not GT in the pure sense, because I didn't have it in mind when defining my research questions and setting out on my data collection efforts back in 2019. However, finding GT was like washing up on an island after being cast adrift in a horizonless sea of qualitative data. It allowed for a strong framework for using myself as a research tool, for trusting my own abductive inferences, and for mitigating the effects of my own biases. As I will discuss later, this process is something that also provided me with more confidence as a designer.

6.3. Human-centered design methods

"How in the world should I go about designing a remote control center?" This was the question I asked myself shortly after receiving the news that our proposal to build what would eventually become the NTNU Shore Control Lab had been approved. The central question of my thesis, once comfortably theoretical and abstract, had now been distilled into a practical one—and a rather pressing one at that. Having been closely involved in the proposal writing process, I was now expected to make true on the promise of ordering the necessary equipment and assembling a lab worthy my collaborators' and my collective vision. If this predicament had a silver lining, though, it was that I could integrate this practical problem into my PhD research. *Article 3* chronicles the details of building the lab using human-centered design (HCD) from start to finish. In it, the question, "How to design a remote control center?" is the research question, HCD is the method, and the completed lab is the result. Here, I want to provide some context to the HCD method I used by presenting some background that justifies its use.

HCD offers a practical approach to complex, open-ended problems. Adding to its appeal is its promotion by household names in the design canon like Don Norman and brothers Tom and David Kelly of the company IDEO. HCD also communicates a clear message: focus on

---

[101] Juliet Corbin and Anselm Strauss, *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Fourth Edition (Thousand Oaks, CA, USA: SAGE Publications, Inc., 2015).

understanding the problem and on paying attention to the needs of users. Compared to other approaches in the design of complex, sociotechnical systems, HCD was approachable and had a low barrier of entry. Alternative camps like systems engineering and resilience engineering had their acolytes and convincing theories and even offered what seemed like a more refined, upper-tier approach. But when I combed books like Nancy Levesen's "Engineering a Safer World: Systems Thinking Applied to Safety"[102] and Erik Hollnagel and colleagues' "Resilience Engineering: Concepts and Precepts"[103] looking for pragmatic guidelines to put these ideas into action, I came up short. While these books described useful ways to handle the conceptual problems involved in designing complex systems, they offered little to the designer eager to roll up their sleeves and *build* something.

In addition to being easy to grasp and practically oriented, HCD also has deep roots in scientific research. This might make it the only method discussed with equal vigour at design firms and editorial boards of scientific journals alike. Perhaps this is the legacy of pioneers like Don Norman, whose background in cognitive psychology lent traces of academic authenticity to HCD. Another appealing aspect is its standardization in the International Standards Organization (ISO). In fact, in the scientific literature, researchers often refer to the standard ISO 9241-201, shortform for the design guideline "9241: Ergonomics of human-system interaction, Part 210: Human-centered design for interactive systems standard."[104] Early in my PhD, I decided to get my hands on this standard. This turned out to be more difficult than I anticipated. Unlike open science articles, ISO standards are protected under strict copyright and have to be ordered directly from ISO, and at no small cost considering they arrive unbound and printed on inexpensive paper. (Compared to my copy of IDEO's free "Field Guide to Human-Centered Design,"[105] which arrived bound and with striking colour imagery, my new ISO standard seemed a dud.) Moreover, the ISO standard was short and written in dull language. At first glance, the only page that seemed interesting was the page depicting HCD as a cycle of four steps, consisting of boxes connected by arrows (**Figure 6**). This cycle started with "Understanding and specifying the context of use" and ended with "Evaluating the design" and then went right back to the start. Despite its lackluster appearance, I had to admit that its concise logic was appealing. As I understood it, the key takeaway was twofold: first, a design is never *done*; rather, it undergoes constant evaluation and improvement based on feedback on the design solutions produced; second, a designer cannot be content with being told the problem; rather, they

102 Nancy G. Leveson, *Engineering a Safer World : Systems Thinking Applied to Safety* (Cambridge, MA, USA: The MIT Press, 2016).
103 Hollnagel, Woods, and Leveson, *Resilience Engineering: Concepts and Precepts*.
104 ISO, "NS-EN ISO 9241-210:2019, Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems" (Standard Norge, September 1, 2019).
105 IDEO, *The Field Guide to Human-Centered Design: Design Kit* (San Francisco, CA, USA: IDEO.org, 2015).

must discover it and *understand* it for themselves. Next, I will outline three practical techniques I used to help me understand the problem, generate solutions, and evaluate the design. These three techniques are design ethnography, prototyping, and workshops.

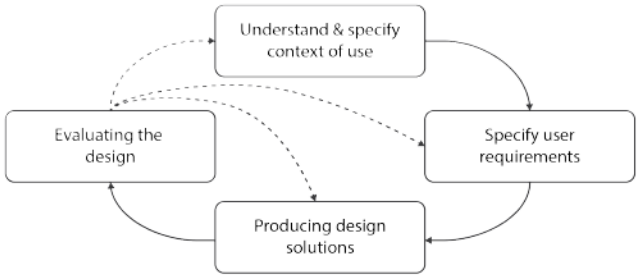


Figure 6. ISO standard for human-centered design (after ISO 9241-201)

Design ethnography

The term *design ethnography* is a mouthful, but its premise is straightforward. You go out into the world with a notepad and paper, you open your eyes, and you take notes.[106] At first, this seems like the antithesis of design. Your colleagues might even wonder why you're leaving the office in the middle of the day to wander around with a notepad. When done right, though, design ethnography can yield dividends by allowing for the discovery of the problem firsthand and by making the abstract problems one is dealing with more tangible. This intention was especially important in the first step of the HCD process for the NTNU Shore Control Lab, which involved "Understanding and specifying the context of use" (**Figure 6**).

With this in mind, I left the university grounds with research assistant Thomas Kaland and headed towards Øya, a neighborhood rumoured to have a new autonomous bus in service. Bright yellow signs had gone up in Øya reading, "Mind the autonomous bus." This drew our attention. Today, we were there to mind the autonomous bus (**Figure 7**).

After downloading an application and ordering the bus, we waited by one of the designated bus stops along its route. After ten minutes, a little green bus approached us. It was moving slowly. So slowly, in fact, that cars behind it drove around. Now, it had arrived, and welcomed us inside by opening its doors. Inside sat a middle-aged man in uniform. We

[106] Here I took inspiration from Martin and Hanington's book "Universal Methods for Design," which defines "design ethnography" as a method that "approximates the immersion methods of traditional ethnography, to… understand the user's world for design empathy and insight." Bella Martin and Bruce Hanington, *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions* (Beverly, MA, USA: Rockport Publishers, 2012).

nodded at him, and he motioned for us to take a seat. Then, he turned a key sticking out of a control box by the window, at the bus lurched into motion. We were off, moving at a speed of about 10 kmph—a slow jogging pace—to the other end of Øya. Suddenly, the bus came to a sudden halt. An alarm went off. The operator, looking nonplussed, picked up a portable control panel with a joystick and put the shoulder strap around his neck. "What stopped the bus?" I asked. "I dunno," shrugged the operator. Thomas suggested that it might have been a seagull that had been standing in the road. "Did the bus stop for a seagull?" Now the operator was maneuvering the bus with a joystick. "This is a tricky spot anyways," he explained, "the bus doesn't like when people park in the street like they do here." After having cleared the mysterious and altogether benign obstacles, the operator turned the key again, transferring control to the computer. Our slow journey continued. Arriving at Øya, approximately 15 minutes later and having traversed a total of 1 kilometer, Thomas and I disembarked and began the short walk back. Debriefing, we agreed that the technology was immature, but that it was probably just a matter of time before the bus was moving smoothly in traffic as the technology improved. Most interesting was the onboard operator. Even as the technology improved, it was hard to imagine the bus without the supervisory operator. Granted, they might eventually be displaced to a centralized control center where they might oversee several buses at once, but the fact remained that somewhere an operator must remain responsible for the buses. The operator represented a broker between two parties: idealized technological efficiency and real-world complexities. As of today, the gap between the two parties was large, but it would eventually shrink. But would it ever disappear entirely? Probably not. The context of the problem, at least, was taking shape.

Figure 7. Safety host aboard autonomous bus (photos by Erik Veitch, August 2020).

Prototyping

Another useful technique was scale model prototyping. Personally, I am not very good in the workshop. (After snapping a drill bit on the boring machine while trying to make a piece of equipment for rock-climbing, I began to make myself scarce at the Department of Design workshop.) Luckily, our research assistant Thomas Kaland was a natural talent. In fact, Thomas seemed to eat wood dust and metal shavings for breakfast. In retrospect, I suspect the only reason he accepted the research assistant position was to get access to the well-outfitted workshop where he could tinker away in his free time. When I asked him if he thought he could create a prototype of the control center, he replied without hesitation. "Sure!"

The prototype turned out to be a brilliant tool for both communicating our ideas and for collecting feedback (**Figure 8**). Instead of relying on words to describe the room ("The room is 36.5 square meters, with two 75-inch displays in the front of the room"), it was a lot more effective to just *show* them using a scale model ("Here's the room and the screens!"). The prototype was always under construction as we incorporated the most recent input or advice and then headed out to collect more input and advice. The prototype also served as a type of scapegoat, allowing someone to direct their misgivings at a physical object instead of a person who might take it as a personal affront. This was good, because we welcomed criticism more than we feared discouraging feedback.

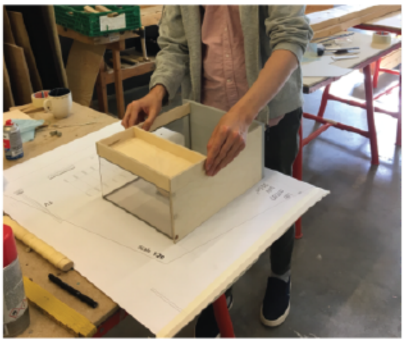


Figure 8. Early prototype (left; photo by Erik Veitch, Aug 2020) and detailed prototype of the control room (right; photo by Thomas Kaland, Sept 2020).

Wherever we went, the prototype was the centerpiece of attention. This was also a good thing, because often the most outspoken person in the room is the center of attention, which can lead to a common form of design myopia known as "groupthink." The prototype served us well at two design workshops: the first involving the team behind the Trondheim Maritime Center and experienced faculty members, and the second involving invited experts on topics of machine autonomy and remote control centers. Today, the prototype sits atop my bookshelf: where old prototypes go to die. But, taking it down, I see the current NTNU Shore Control Lab in miniature, even down to small details like where network and power points are located. I recalled all the iterations and adjustments we had made to the plastic and cardboard components in the effort to get the design just right. Luckily, we had made those iterations and adjustments in miniature.

Workshops

Workshops are the last of the HCD techniques I want to highlight. The premise of the workshops was simple: invite guests with valuable knowledge and good ideas, then try to extract that valuable knowledge and those good ideas. This sounds easy enough, but uncovering, collecting, and documenting participants' knowledge and ideas turned out to be more art that science.

One of our workshops was arranged by me and Thomas at "Skiboli," a small building that can be rented for events on campus (**Figure 9**). We set up tables, seats, and a projector. On the tables, we laid out markers, paper, and Post-It notes. Our list of invitees included experts from a range of relevant backgrounds, including aerospace control rooms, autonomous car design, engineering design, ocean engineering, marine biology, interaction design, marine navigation, and human factors. These participants represented users of the NTNU Shore Control Lab: research-oriented experts with a keen interest in remote control operations.

We planned three exercises for our guests. These were intended to challenge them to think creatively, interact with other guests, have fun, and—most importantly—document their knowledge and ideas. One exercise that stood out for me was having guests get into groups and sketch two control centers, which they subsequently presented to everyone. The first represented the "best control center they could imagine," without any constraints on budget or even on technology. This exercise produced some zany features ("free food" and "ping pong" being among them), but also some insightful ideas ("AI working and training together with humans"; "connection to other control centers"). Second, the groups were asked to design a control center with the constraints we really had. Instead of a blank sheet, they were presented with a plan view of the Trondheim Maritime Center and our allotted space within it, as well as our true budgetary constraints. With this, their focus sharpened, yet their creativity and lack of inhibitions remained intact from the previous exercise. Detailed ideas emerged, like a "zone for virtual reality" and "three workstations: two for operators and one for their supervisor." Ideas were materializing and abstractions were turning into substantiations. The workshops helped to ground specifications for the NTNU Shore Control Lab in the needs of stakeholders and users. This was especially useful in the three final stages of the HCD process: "Specify user requirements," "Produce design solutions" and "Evaluate the design." Having rounded the HCD cycle several times (**Figure 6**), we were soon ready to procure equipment for the NTNU Shore Control Lab and begin construction.

Figure 9. Design workshop for NTNU Shore Control Lab (photo by Erik Veitch, 14 September 2020).

6.4. Experimental methods

"The single greatest obstacle to successful science," writes Michael Strevens, philosopher of science and author of "The Knowledge Machine: How Irrationality Created Modern Science," "is the difficulty of persuading brilliant minds to give up the intellectual pleasures of continual speculation and debate, theorizing and arguing, and to turn instead to a life consisting almost entirely of the production of experimental data."[107] Science is mostly boring and frustrating, contends Strevens. What we read about it just the 1% of scientific achievements. "Behind these achievements," he explains, "are long hours, days, months of tedious laboratory labour."[108] The driving force of modern science is not intellectual pursuit of theory punctuated by moments of brilliant clarity. The driving force of science is the production of data.

---

[107] Strevens, *The Knowledge Machine: How Irrationality Created Modern Science*. Chapter 1, 38:28-38:43.
[108] Strevens. Chapter 1, 38:19-38:27.

It wasn't just Strevens's book that inspired me to get into the lab and collect data (although it did help). To plan and execute a major experimental campaign takes an enormous amount of energy, months of planning, and big dents in budgets. The experimental effort must be willing to place an all-or-nothing bet on a best guess—a hypothesis. Part of what makes experiments risky is their expense: researchers' time, equipment, and laboratories do not come free of charge. Perhaps the biggest risk (at least from a publication standpoint) is a null result: when the data collected does not support the hypothesis.[109]

I knew early on that I wanted to do an experiment during my PhD that would contribute empirical data to the questions surrounding remote control centers. What I needed was a plan—a strategy for getting the best experimental bang for my buck. I found that strategy in "Design and Analysis of Experiments,"[110] a book originally published in 2009 and written by Douglas Montgomery, which is now in its ninth edition. In fact, it was a professor at Memorial University of Newfoundland that opened my eyes to Design of Experiments (DOE) during a course I did as a master's student in 2016.[111] Five years later, I recalled DOE as perhaps the best insurance policy available to the scientist considering investing in an experiment.

DOE is a way of planning out an experiment in such a way that it maximizes the ability to detect hypothesized effects. It also forces the researcher to carefully define their hypotheses *before* testing, thereby circumventing the temptation to "fish" for positive results in the data. Let's consider an example. Say we are hired as a consultant for a brewing company. We are tasked with finding out whether brewing temperature and type of malt affect the concentration of sugar in beer. Sugar gets transformed into alcohol during fermentation, so it's advantageous for the company to get a handle on sugars during the "mashing" stage, when malted grains are mixed with hot water to extract fermentable sugars. DOE can help us design an experiment that will satisfy our client, delivering the clearest answers for the minimum cost. Specifically, with DOE we can test the extent to which our client's hypothesized factor effects hold true. What's more, we can tell our client the extent to which *interaction effects* are present (i.e., the *combined* effect of temperature and malt on sugar concentration).

109 Of course, this treatment of scientific testing is skewed towards the publications process. Null results do, in fact, contribute very important findings to the scientific method. Indeed, Karl Popper wrote that a theory *must be falsifiable* to be a theory at all, and thus falsification of theories is the very driving force of science, leading to alternative hypotheses and new ideas. But try telling this to an editorial board of a scientific journal. Null results are notoriously under-represented in the scientific literature.

110 Douglas C Montgomery, *Design and Analysis of Experiments*, Ninth edition (John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2017).

111 I am indebted to Dr. Leonard Lye for having taught the course "Design of Experiments" in fall 2016 with enough passion to make a lasting impression.

The data analysis part of DOE uses Analysis of Variance, or ANOVA for short, so named because it uses statistical inference to check whether the factor effect (signal) can be distinguished from random error (noise) beyond a reasonable doubt. This process of hypothesis testing is essentially "innocent until proven guilty": a true example of Popperian scientific inference. DOE is used in countless applied research settings, including pharmaceutical testing, industrial process optimization, agricultural testing, food manufacturing, and many more. Whenever a scientist asks, "What is the effect of x on y?" they can turn to DOE to guide them towards an answer.

Perhaps the most powerful attribute of DOE is that is allows the experimenter to invest in their hypothesis wisely. What I mean by this is that DOE gives the experimenter the best margins to detect their hypothesized factor effects in the least possible number of runs. A suitable analogy might be taking just the right camera lens to shoot a given subject. The perennial question of "How many runs do I need in my experiment?" is brushed aside; the more appropriate question is "What are trying you accomplish with your experiment?" Once you have an answer to this question (e.g., accurate model prediction, basic factor screening), *then* can one deduce the number of runs needed. This saves the experiment time and money on wasteful trial-and-error approaches and allows experimenters to hedge their bets. To guide the experimenter on this path, DOE also provides a handful of fundamental rules that avoid common experimental pitfalls—pitfalls that, if undetected, could undermine results before testing even begins.

The most important of these fundamental DOE rules is randomization. When an experimental design is drawn up, the settings for all independent factors to test are drawn up (these are called *treatments*) along with a list of sequential treatments (called *runs*). To return to our beer brewing experiment, let's say we wish to test temperature by setting a low level of 60 degrees Celsius and a high level of 80 degrees Celsius, as well as two types of malted grain: barley and wheat. To test all variations, we must test $2^2 = 4$ treatments (called a "factorial" experiment). If we then decide to start with 60C and a wheat malt, perhaps out of preference or convenience, then we have already made a critical error. Remember: we must randomize the run order. The reason is simple. If we don't randomize, we can't be sure that run order affected our output. Maybe the barometric pressure, increasing over the course of the experiment, affected sugar content. Maybe we became more efficient at stirring the brewing mixture over time, and *this* affected the sugar content. How could you know for certain which undesired variables (called "confounding factors") didn't sneak into the experiment? Luckily, randomization levels the playing field for run order-induced confounding factors, confirming, post-hoc, that order was independent of variation.

Despite prescribing fundamental rules like run order randomization, DOE does not provide a one-size-fits-all experimental design. Like a carpenter reaching for their tools, the scientist has many different experimental designs at their disposal. The trick is selecting the right one.

In DOE, it is often the character of independent factors that determines which experimental design to choose. In my case, I wanted to investigate the effect of five independent factor effects on supervisory control performance. The five independent factors (which I introduced in Chapter 5.1), were skill, vigilance, multitasking, time pressure, and decision support. Testing each one at two levels yielded $2^5 = 32$ runs. So far, this was straightforward enough. The last step was to randomize the run order. However, this was where things got tricky. I had two groups of volunteer participants coming to the experiment (gamers and navigators) and it was impractical to schedule them according to a predefined randomized order. People cannot be adjusted like one can adjust temperature settings, nor can they be selected from a shelf like types of malt. People get sick, miss appointments, and arrive on the wrong day, inevitably foiling any plans of maintaining strict run order. Luckily, there was a solution. The most practical thing to do was to "block" the experiment according to scheduled groups—gamers in one group, navigators in another—and then to randomize the *block order* (treatment *within* the blocks would remain completely random). This method had a name: the "split-plot" design.

The split-plot design

The split-plot design gets its name from agricultural testing. Imagine you want to design an experiment testing the yield of a new crop in different soil mixtures. Easy. You design an experiment with the two crop varieties you wish to compare (old and new) and two soil mixtures (A and B): a 2^2-factorial experiment with 4 treatments. Then, you randomize the run order and find out you need to divide the field into four separate areas. Scratching your head, you realize this will require a lot of manual labour. Moving soil from one plot of land to another sounds more like forced labour in a gulag than good science. It would be a lot easier, you think, to divide the plot by just two soil mixtures and sow the seeds at random within them. This is the split-plot design with two blocks. In this case, the soil mixture is the so-called "hard-to-change" variable whose randomization we constrain in blocks.

But what about ANOVA? How can this be applied when randomization is constrained in one variable? Luckily, a special version called Restricted Maximum Likelihood (REML) ANOVA was developed by statisticians to handle just this.[112] REML ANOVA effectively placated effects analysis of factors that eschewed randomization without compromising the analysis.

In my case, the skill factor was the hard-to-change variable. For my split-plot design, I divided the 32 runs into four groups of eight and randomized the order of the groups, ending

---

112 R. R. Corbeil and S. R. Searle, "Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model," *Technometrics* 18, no. 1 (February 1, 1976): 31–38, https://doi.org/10.1080/00401706.1976.10489397.

up with Gamer, Gamer, Navigator, Navigator. The final run log, which is available as an appendix in *Article 4*, was then generated with the confidence I would not lose face by constraining my randomization. To borrow our analogy again: I had the right camera lens for the occasion. I scheduled all gamers first, followed by seafarers, restricting the randomization of treatments within four groups. The run log was followed as planned. During the period 14 March to 16 June 2022, all treatments were tested.[113] Later in this thesis, I will present some of the results from this experiment.

Handovers and takeovers

In designing the scenarios that participants in my experiment would undergo in the *Gemini* simulator, I had the virtual world as my oyster. I could design any scenario I could dream up and have participants complete them. However, in the interest of fidelity—that is, keeping the scenarios as realistic as possible—I landed on two distinct scenarios.

The method for the virtual simulation testing is well documented in *Article 4*, so expounding on it here is unnecessary. But one thing is worth highlighting: the two scenarios captured two distinct modes of intervention that I termed "handover" and "takeover." Later, in the analysis of the results, we will see that these two modes produced different outcomes: a testament to their distinctiveness. The key insight is that not all interventions are the same. In a collaboration between the machine autonomy and operator autonomy, control transfers can follow two paths: the machine can hand over control to the operator (a *handover*) or the operator can take control from the machine (a *takeover*). The handover path is reversable. This means that after a handover or takeover, the operator can transfer control back to the machine. It is important to note, however, that at no point can the machine take over from the human (**Figure 10**).

[113] There were two replacements done during this period (repeats of a treatment), owing to issues in data collection. An error in the simulator occurred during the treatment *T90* (replaced by *T90_2*); participant *E51* did not hold a valid navigator license (replaced by *E51_2*). This did not affect run order or ANOVA. See "Run Log" in Appendices of *Article 4* for list of all treatments and their run order.

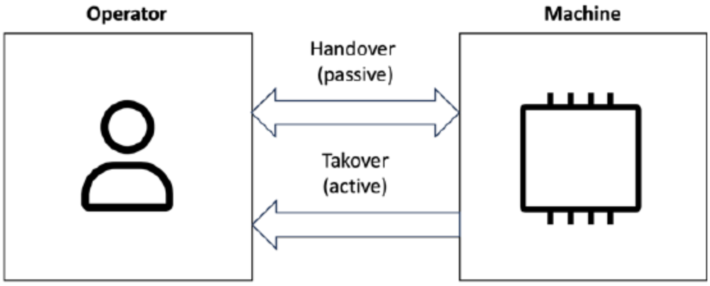


Figure 10. Handover and takeover control transfers.

The first scenario I had participants complete was a handover where the machine autonomy, upon encountering a control failure, handed over control to the operator. In the second scenario, a takeover, participants had to apply their own judgment to take over control from the machine autonomy in order to avoid a collision that the machine autonomy alone was not suited to handle. In the handover scenario, I measured performance in two ways: the first involved measuring how much time passed before the operator took control (*response time*, in seconds); the second involved computing a score based on how closely the trajectory under manual control matched with a corresponding trajectory had no handover occurred (*performance score*, in percent). The takeover scenario called for a different approach. Here, I measured performance using just a single indicator: the outcome of the scenario (*collision avoidance, near-miss,* or *collision*). As we will see later in this thesis, the question of what factors influence remote supervisory performance depends on the context of control transfer, whether it be handover or takeover.

6.5. Summary

In this chapter, I have outlined several distinct methods I adopted during my PhD research as they corresponded to the four articles I present in this thesis. Soon, we are ready to examine the results of applying these methods. But first, I want to point out one thing: the four methods I have described, for all their distinctions, actually had more in common that not. For one thing, all methods were, in essence, ways of *theorizing* (Chapter 6.2). By theorizing, I mean the action of explanatory insight, the leap from observing a chaotic world to extracting patterns and distilling order. By theorizing, I mean the process of generating hypotheses.

In this sense, my literature review method (*Article 1*) helped structure my theorizing and generate hypotheses as to what aspects of remote control design were important. Grounded Theory techniques (*Article 2*) provided the framework I needed to interpret interview and field observation data and give it narrative structure. The human-centered design techniques I presented (*Article 3*) also helped me theorize in the sense that they led me to generate tangible, although altogether impermanent, design solutions. Finally, the Design of Experiment approach (*Article 4*) forced me to refine my hypotheses about which factors were important to study. To allude to Sherlock Holmes in this chapter's epigraph, all methods in my PhD helped me suit theories to facts, resisting, as it were, twisting facts to suit theories.

7. Results

> **Trondheim: A warm spring morning in 2021.** After over a year of construction led by Egil Eide in a drafty, high-ceilinged workspace rented from Dahls Brewery, the *milliAmpere2* autonomous test ferry emerged into the daylight. Hearing the news, I walked down to the quayside near the Dora drydocks in Nyhavna, just a stone's throw from the Trondheim Maritime Center office. There it was, just like in the renderings Petter and I had made two years prior: the world's first autonomous passenger ferry, gleaming in the sunlight.

Finally—the fun part of my thesis. This is where I get to present the fruits of my labours like a proud grade school student in show-and-tell. The results I present build directly from the concepts and theories I presented in the opening chapters and from the methods I outlined in the previous chapter. As in the methods chapter, the results are presented in four parts, corresponding to the four articles appended. It's my ambition to show you here how all the results are related and follow a chronology starting from my literature review study and ending with my experimental campaign.

7.1. Literature review results

The systematic literature review indicated that the topic of my research was growing rapidly. From humble beginnings starting around 2014-15, the topic of designing remote control centers to support autonomous ship operations grew from almost nothing to almost one hundred publications in 2021 (the equivalent of about one every four days). These publications included conference papers, scientific articles, theses, and reports. These publications also encompassed many fields, including risk and safety science, ocean engineering, human factors and training, reliability engineering, and marine policy. In total, I analyzed forty-two peer-reviewed journal articles in the review, with these having met the inclusion criteria defined in the PRISMA statement and having passed abstract screening.

The analysis of the review is presented in detail in *Article 1*; here, I wish just to highlight some of the key findings and show how these findings shaped the final stages of my PhD journey.

The role of human operators

Although it wasn't clear exactly what the role of human operators would be from a regulatory perspective, one thing was demonstrably clear from the stack of papers I had read: humans were still at the center of maritime navigation. Ironically, the way most researchers ended up conveying this was by referring to a Level of Automation (LOA) taxonomy. As I mentioned before (Chapter 2.1), the LOA was originally designed to define the extent to which a machine controls a system. However, in between the lines, the opposite message was being conveyed: to what extent was the human *not* controlling a system? In the review, not a single article defined autonomous vessels using the label "full autonomy"—the highest LOA on any taxonomic scale—nor did any article indicate that full autonomy was even desirable. Humans were still somewhere in the control picture. It was just unclear where. As I wrote in *Article 1*, "Regardless of what LOA taxonomy was being used, all studies positioned themselves one or two steps below the maximum number to show that the ship was highly autonomous, albeit not autonomous enough to exclude humans."[114] The most referenced LOA taxonomy was IMO's four-level taxonomy, originally published in 2018 (see **Figure 2** in Chapter 2.1 for an overview of IMO's LOA taxonomy).

Accepting the general assertion that humans would be needed, it was time to address the ambiguity about in what capacity human control would unfold. Sifting through the articles in the review, I identified three role categories: (i) *Active*, consisting of continuous monitoring and direct or indirect remote control with decision support; (ii) *Backup*, consisting of monitoring and control intervention; and (iii) *Passive*, consisting of supervision and assistance. Interestingly, most articles positioned themselves within the *Active* category, which also indicated the lowest degree of machine autonomy. Authors described a system not unlike a modern drone (also known as Remotely Piloted Vehicles), where the operator is engaged during all phases of operation. In this mode, the operator directs local control through highly automated systems and receives decision making support by means of a digital interface. An analogy may be found in car control: the driver directs steering and stopping, but these are locally controlled by automated power-steering actuators and by an anti-lock braking system. To indicate their status, lights on the dashboard indicate when they are active (or in case they are compromised). The driver is, in other words, an *active* controller of a highly automated

[114] Veitch and Alsos, "A Systematic Review of Human-AI Interaction in Autonomous Ship Systems." Page 9.

system. The key difference with remote supervisory controllers is that operations take place remotely.

Interestingly, authors subscribing to the *Active* role of human operators often conceded that humans would be needed at least occasionally onboard the vessels. Tasks like machinery maintenance, repair and cargo loading/unloading, and mooring still elude even advanced automation. The *Backup* role, by contrast, simply defined the human role in terms of intervention. Under this format, when something goes wrong with navigation, the operator at the control center is expected to take preventative action. Likewise, if something goes wrong with a mechanical component, the operator dispatches personnel to deal with it.

This way of categorizing the role of operators had implications later in my PhD work. After my observations aboard the Bastø-Fosen ferry and after my interviews with crew aboard highly automated ferries, I recognized telltale signs of the *Backup* role unfolding in real-world operations more so that the *Active* role. As we will see later, I designed my 2022 experiment at the NTNU Shore Control Lab based on a *Backup* role, measuring human performance only as a function of successful control interventions.

Risk-based designers adopted 'best guess' approaches

Most studies I reviewed fell under the field of "safety and risk sciences." At first, this surprised me. Why weren't fields like "ocean engineering," ostensibly concerned with building things, not producing more scientific material than a field dedicated to griping about its risks? After reading the articles, however, my surprise quickly evaporated. Risk is just another word for uncertainty, and if there is anything uncertain in ocean engineering today, it's autonomous ships. At least for the time being, risk-based design (which I introduced in Chapter 2.2) is a natural candidate for grappling with how to design autonomous vessel systems. It is in this regard an appropriate antidote for addressing the uncertainty which may otherwise stall design progress.

Risk scientists are often quick to point out that risk can be computed quite simply: you take probability of failure and multiply it by the failure's consequence. This is why being a passenger in a car is riskier that being a passenger in an airplane. While the consequences of an airplane failure may be higher, its risk is effectively scaled down by a very low probability of occurrence. Driving accidents, on the other hand, have lower consequence but much higher probability of occurrence, making them riskier. What is less straightforward, however, is how to apply this computation into design of new technology systems. We have lots of statistics about cars and airplane accidents; statistics about autonomous ships, by contrast, are harder to come by—and statistics about autonomous ship *failures* are practically non-existent.

There are myriad techniques available for assessing risk in complex systems. In my literature review, I counted nine distinct methods used in fifteen different studies. The most popular methods were Systems-Theoretic Process Analysis (STPA) and Bayesian Networks (BNs). What captivated me most about these techniques was *why* so many authors favoured them. STPA is typically used for identifying hazards and revealing causal factors especially for novel and complex systems for which there is little empirical data available. BNs are also most powerful when little or even no information is available and when the source of predictive power must come from prior beliefs or limited experience. While these methods produce results, they come with a caveat: their predictions are only as good as their inputs. Working with few data is permissible, but more data is always preferred.

New risks in human-AI interaction

It is revealing that twenty-two of the forty-two articles I analyzed in my review expressed lack of research and specifically lack of empirical data as obstacles to understanding how humans and AI interact within a complex system. As we have already seen, lack of data was also the motivating factor for risk scientists adopting "best guess" approaches like Bayesian Networks (Chapter 2.2). From this, we can make a simple deduction: if risk-based design approaches are defined by lack of data and there is lack of data about human-AI interaction, it follows that risk-based design approaches are defined by human-AI interactions.

This was precisely the finding of a detailed review published in the journal "Reliability Engineering and System Safety,"[115] which reviewed all published hazards for autonomous ships to date and evaluated the most salient hazard categories. Of these, they concluded that "interaction with manned vessels and detection of objects" represented the most significant contributors to overall risk. Perhaps it is not surprising, then, that a new risk assessment method has been developed specifically for autonomous maritime systems called Human-System Integration in Autonomy (H-SIA) that places human-AI interaction at the center of modelling failures and their potential consequences.[116]

Several studies shed light on aspects of human-AI interaction where lack of data was most glaring. In this regard, the most common refrain was lack of data about how failures propagated between interactions of hardware, software, and humans. A similar dearth of data was lamented for human error probabilities during emergency response and how skills and other human factors affected remote monitoring and control. Later in this thesis, I will show how this last aspect set my course for designing an experiment using virtual simulation.

115 Chia-Hsun Chang et al., "Risk Assessment of the Operations of Maritime Autonomous Surface Ships," *Reliability Engineering & System Safety* 207 (2020): 107324.
116 Ramos et al., "Human-System Concurrent Task Analysis for Maritime Autonomous Surface Ship Operation and Safety."

7.2. Discrepancies in collaborative human-AI control

When I set out to analyze the stacks of field notes, memos, and transcribed interviews I collected in 2019-2020, some patterns began to emerge. The results of this analysis are presented in *Article 2,* but I want to highlight three findings here that stemmed primarily from the navigators' interviews. These navigators, working aboard the ferries equipped with automated crossing and docking technologies, represented a rare glimpse of what collaboration with AI currently looks like in the context of maritime navigation.

The backup role

The first thing that struck me was that the navigators referred to themselves as "backup." Some individuals used the word explicitly in their descriptions of how their role has changed since the arrival of automated navigation systems. Reflecting upon their own role as backup had the effect of assigning agency—the capacity for action—to the auto-systems. Two navigators referred to the auto-systems with the nickname "Betty" (a pseudonym)—ostensibly an homage to the female voice recording that announced stages of the journey, including alerts and warnings. Betty was, according to one navigator, "ingenious" in the fog, where zero visibly could make the delicate task of docking a 1400-ton ferry difficult. However, Betty could also be a "nag" and "do weird things," according to another. It was only a small step from assigning the machine agency to displacing themselves to a secondary role, with Betty taking charge. "We are the backup if something happens," explained one navigator. Another explained, "when what you see on the screen no longer shows the correct thing, *that's* when things get interesting." Although I didn't necessarily agree that this was the right way to align AI technologies and human operators in future iterations, this framing of their role did influence how I would eventually define the role of operators in my subsequent simulator experiments; namely, as backup.

Decision making heuristics

It became clear from navigators' accounts that interventions were the defining element in their dealings with the auto-systems on the navigation deck. The decision to take over control often followed a predictable sequence of events. For example, if small leisure boats ended up too close to the ferry's path, the navigators would take over control from the auto-system. However, at times the decision to take over control was made for less obvious reasons. To illustrate this, one navigator recounted taking over control from the auto-system to avoid rolling motions of the ship during a crossing. "Instead of rolling all the way over," they explained, "I sail a little North and then a little South to go across the waves." As any sailor knows, taking waves broadside can quickly result in uncomfortable rolling periods of the vessel—motions that can be offset by adjusting course slightly to take on waves at a smaller angle. It was unclear whether the auto-system was programmed with this knowledge. The decision to take over control in this case stemmed simply from empathizing with passengers' comfort. This showed that interventions could follow mental shortcuts (or *heuristics*, as I defined them in Chapter 2.1) to the same extent that cognitive processes could follow more logical, sequential steps. My observation of heuristic decision making influenced my thinking about how decisions will be made from the remote control room to take over control. As we will see later, I designed my experiment with a scenario impelling, not requiring, operators to take over control, and probed them in exit interviews about their decision making. The results, as I will show later, aligned more with heuristics than with the sequential cognitive processes designers seem to prefer.[117]

Skill degradation

Another finding that emerged was skill degradation associated with spending long periods of time in hands-off auto-mode. This was especially clear when interviewing navigators aboard a ferry on another route on Norway's West Coast that used auto-crossing "every day, every trip," as one officer put it. When I asked them about the last time they had driven the ferry manually, they replied, "It's been one-and-a-half years since I stopped doing it myself." Prompted about situations they might be *forced* to drive manually, they offered the case of ambulance dispatches. "With an ambulance, it's life and death," he stated, alluding to the higher speed ferries complete their crossing when carrying an ambulance in transit. What was unclear, though, was whether the officers aboard this ferry were prepared for such a case. Simply put, to what extent did they remember how to drive the ferry?

The Horten-Moss ferry, by contrast, only used their auto-systems approximately 50-70% of the time, according to their navigators. On this ferry, it emerged that new procedures had come into place to ensure operators were regularly completing crossings manually. "When I have driven a lot of auto," explained one of Horten-Moss crew, "I have to steer a couple of times myself to get the feel of it again." They continued, saying that they controlled the ship manually "at least twice per day so as not to forget how that works."

Skill degradation points to a pernicious problem in automation applications whose consequences may worsen in the transition towards remote control operations. Procedures to regularly include manual driving might effectively combat this trend, but it's unclear to what extent. I will pick up this topic again in the Discussion chapter.

117 See Chapter 8.1 for a discussion about heuristic decision making based on the interview data collected during my experimental campaign. Note that the interview data are openly available in a data archive (available for download at https://doi.org/10.18710/WYFMMP).

Design discrepancies

Upon analyzing the interviews of technology designers and navigators, discrepancies emerged between how technology designers construed collaboration with AI systems and navigators' own accounts from the field. For one, designers' referrals to "situation awareness" models of decision making was at odds with navigators' accounts of heuristic decision making. Four of the nine designers I interviewed explicitly referred to Endsley's model of situation awareness that I discussed in Chapter 2. By contrast, none of the navigators referred to it, nor did they even refer to its characteristic steps of information gathering, processing, and projection. This is illustrated in the example I highlighted previously where a navigator recalled taking over control to attend to passenger comfort. Instead of series of cognitive steps, the navigator's decision stemmed simply from imagining how passengers might experience a particular sea state.

Another discrepancy related to the role of operators. Designers preferred saying operators were "in the loop" of a larger control system—a component whose timely and decisive takeover action would, if needed, mend a compromised loop's integrity. Navigators did not talk in such terms. Instead, their reflections on their backup role somewhat begrudgingly suggested a loss of agency while at the same time not losing any of the weighty responsibilities that come with wearing a navigator's stripes. As expressed by one of the navigators, "The job hasn't change, but in auto you can sit back and let the system do it." This discrepancy influenced my thinking about the designer's role in addressing the emerging problem of AI alignment—a topic I will pick up again in the Discussion chapter.

Takeover time

An important theme that emerged in both designer and navigator interviews was that of takeover time. Takeover time can be thought of as the length of time between when a critical event occurs and when preventative action is taken—a measure of response time for an operator's intervention. Takeover time captures the operator's reaction time plus the time taken to reach a decision about what to do. As one designer expressed, "The key challenge will be to get the operator, in the shortest possible time, to get in the loop of what is going on." Navigators, too, in their reflections of being in a backup role, suggested that timely and decisive action underpinned the success of their manual actions. The importance placed on takeover time led me to take steps towards measuring it in a virtual experimentation campaign, the results of which I will present later in this chapter.

7.3. NTNU Shore Control Lab

Almost 500 days after first touring the completely empty floor at the Trondheim Maritime Center, we invited the entire Department of Design to the opening day of the NTNU Shore Control Lab (**Figure 11**). A ribbon was cut (actually, it was a wire cable), cake was served, and about 30 people crammed into the main meeting room to hear me and others explain, *what was this thing?*[118]

Figure 11. Faculty Dean Marianne Skjulhaug officially opens the NTNU Shore Control Lab with Department Head Sara Brinch (left) and Ole Andreas Alsos (right) (photo by Julian Steinke, 18 Oct 2021).

The Shore Control Lab is a test facility for designing control centers for autonomous maritime systems. It was established to meet a growing need for empirical research about human-AI interaction in the context of maritime navigation—a need made palpable from fifteen years of research about maritime autonomy at NTNU alone. The lab also synthesizes the latest thinking about remote control centers and commits to transforming an idea into something tangible—a challenging process in an academic culture predisposed to grappling with abstractions. Superficially, it consists of a room with some screens and a computer. Under the surface, though, the NTNU Shore Control Lab is a dynamic platform for designing and testing for a whole new type of work and for advancing the development of its supporting technological and organizational infrastructure (**Figure 12**).

[118] I found out later that my presentation was published on YouTube: https://youtu.be/gvns39nANxg)

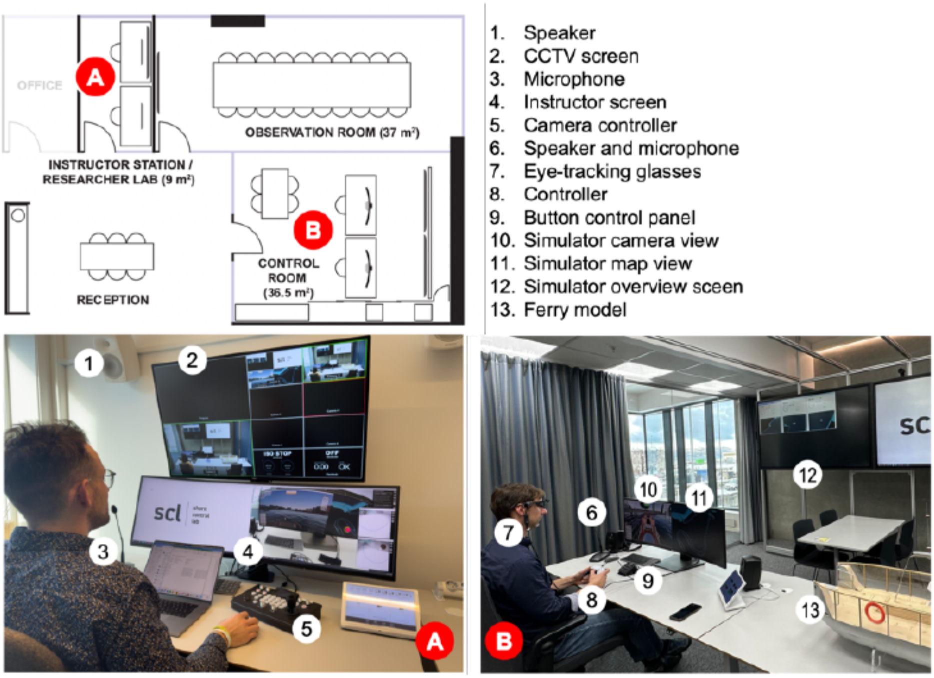


Figure 12. Overview of the NTNU Shore Control Lab (photos by Ole Andreas Alsos & Felix-Marcel Petermann, March 2023).

The Shore Control Lab is also defined by its ability to transform itself. After all, it was never intended to exist in statis or represent a one-off solution. Its reconfigurable and flexible nature has roots in the HCD process that produced it (Chapter 6.3), which commits the design to a continual cycle of improvements based on observation and feedback. Later in this thesis, I will show how the results from experiments conducted at the NTNU Shore Control Lab provided not just data to guide designers, but also suggested improvements to the lab in future design cycles.

Communications network infrastructure

One important piece of the test infrastructure that wasn't mentioned in *Article 2* is the base station atop the roof of the Trondheim Maritime Centre. It provides several network links between the *milliAmpere2* and the NTNU Shore Control Lab, including via 5G, 4G, and radio networks. During set-up and testing of the networks, which was documented by PhD candidate Ahmed Amro in fall 2021,[119] communication was established between *milliAmpere2* and the NTNU Shore Control Lab. This opened to door for the next major iteration for the Shore Control Lab's design: designing supporting systems for monitoring and control services of the real *milliAmpere2*, not just creative simulation testing featuring its digital twin.

Soon after assembling the base station, work got underway to connect the network components. Ethernet cables were run down from the roof and into the data closet on our floor in the Trondheim Maritime Center and network switches were patched to the NTNU Shore Control Lab. Soon a relatively complex network emerged that linked a mind-boggling array of components distributed across four locations. These four locations included the rooftop base-station, the *milliAmpere2*, the Trondheim Maritime Center, and an NTNU data center, where a newly purchased encrypted file server was installed to store operational data. Adding to the complexity was that Zeabuz, a spin-off company from NTNU, was using the *milliAmpere2* to test its proprietary software. New networks were added, about which we had only superficial knowledge at NTNU. Zeabuz's partner Marine Technologies, whose DNV-certified DP system was controlling the *milliAmpere2*, also connected their own proprietary remote control system called *BridgeMate* that worked on a dedicated radio link. For months I tried to keep up with the growing constellation of components in our Internet of Things, and it seemed each time I tried to map it, I discovered new sub-networks and new devices. Amro, whose PhD thesis was about network design for autonomous ferries,[120] and who used the *milliAmpere2* as his primary test case, adopted a pragmatic strategy: focus just on the networks you know about. The chaotic nature of network design must be symptomatic, I imagined, of designing for control rooms, which represented the convergence of disparate elements in the system and the teams behind them. This was a lesson to learn in itself, and a potential vulnerability to address in future work.

Preliminary results of milliAmpere2 trials

The *milliAmpere2* trials occurred over a period of three weeks in autumn 2022. During this time, the ferry completed approximately 500 crossings and transported more than 1500 passengers in total. Immediately after the trials, I interviewed three of the four "safety hosts" that worked onboard the ferry during the trials. (The fourth had a conflict of interest and was not interviewed.) At the time, I was still mulling over the question of how to define the remote operator's role, and I figured the safety hosts might shed light on this question. Since they were physically onboard the ferry during operations, safety hosts weren't remote

119 Ahmed Amro, Vasileios Gkioulos, and Sokratis Katsikas, "Communication Architecture for Autonomous Passenger Ship," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2021, 1748006X211002546.

120 Ahmed Amro, "Communication and Cybersecurity for Autonomous Passenger Ferry" (PhD Thesis, Gjøvik, Norway, NTNU, 2023), https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3064022.

operators, exactly—but they did fall under a similar category. Like remote operators, they represented the only responsible party in the human-machine collaboration, and they were expected to take preventive action in situations where it was needed. As I discovered from the interviews, they also had many additional responsibilities. Here, I assign the three safety hosts pseudonyms: Torkil, Olaf, and Mads.

All safety hosts had received training in how to operate the ferry using a joystick console and all had some form of valid navigation and passenger safety certificates that went beyond a typical small craft license. They started a shift with a short status report from the previous "team" (referring to the safety host and onboard technician who were onboard together during trials). To start a typical day, the safety host prepared the boat by disconnecting from shore power and taking in towlines, followed by starting the ferry and welcoming aboard passengers. They opened and closed the gangway and pressed the button that initiated a crossing (when they did not invite a passenger to do so). Asking what their role was, Torkil answered, "I maintained safety in the form of situation awareness and by taking over control when needed using the joystick." Torkil went on, "also, I communicate with other vessels that stop and wait and I say, OK, you can pass—we'll wait for you to pass." Torkil also said that he "answered a massive amount" of questions from the passengers. "Mostly what I do is talk to people," corroborated Olaf. "It has occurred that I had to stop or take over control, too," added Olaf, "like one time when a boat came too close—but in that case it was the other boat's fault." Another central task involved recording "any events of relevance to the voyage" into a logbook, reported Mads.

In total, the three safety hosts reported taking over control approximately twenty-five times in total during the three-week trial. Twice, manual control was taken to allow a local ferryboat to pass; the remaining takeovers occurred on one specific day when the autonomy system failed. The nature of the failure required manual control on each crossing. The ferry would stop "about twenty meters from the dock," as Mads recalled, requiring them to take over control and use the DP joysticks to complete each journey to dock. (The bug was fixed by technicians the following day).

In the end, the interviews shed light on the future role of remote supervisory controllers. I could now imagine the day when the safety host could complete their job from a remote control center, instead from onboard. From here, they would communicate verbally with passengers and count them to ensure the boat wasn't overloaded. They would also remotely communicate to other boat traffic to resolve potential traffic situations, taking over control with a joystick if necessary. Finally, they would take over control in the case of autonomy system failures, maneuvering the boat manually to dock. Such system failures would be documented, and operators would coordinate with technicians and engineers to plan repairs. The operator would also log events of interest and notable weather conditions. Locally, at least one person would be needed to moor and un-moor the vessel, as well as to connect and disconnect the charging cables.

I also recognized that the trials had come close to recreating the trials in the simulator, which I had completed earlier that year (the results of which I will present in the next part of this chapter). In the virtual scenarios, the crossing was the same right down to the location of the docks (**Figure 13**). The physical appearance of the *milliAmpere2* was also a close match, as were its inertial and hydrodynamic properties. Even the simulated boat traffic was modelled on real boats in the area. Beyond this, I had created two situations that proved realistic: an autonomy failure, requiring the operator to manually control the boat (a *handover*), and a traffic situation that required the operator to assess it, recognize its potential danger, and take over control (a *takeover*). The "backup" role was, at least in this case, proving appropriate, as was the way I recreated it in the simulator. In this next part of this chapter, I will dive into the results of this simulator testing campaign.

Figure 13. *milliAmpere2* (left; photo by Mikael Sætereid, 30 Sep 2022) and its digital twin in *Gemini* (right; screenshot taken in simulator).

7.4. Virtual simulation experiment results

Shortly after the last of thirty-two participants completed their simulated trials in my experiment, I uploaded all the data onto my computer and began the analysis process. My experiments at the NTNU Shore Control Lab, having spanned the period from March 14 to June 16, 2022, were finally over. Now, I could get to work crunching the numbers and formally test the hypotheses I had generated over three years of prior research.

The key to this process was to transform the large quantities of time series data collected from individual trials (each sampled at 5 hertz and stored in a text file), into a single number that accurately represented each operator's performance. As I described in the methods chapter (Chapter 6.4), I ended up with three such performance metrics. The first two were handover *response time* and *performance score*, which applied only to the handover scenario. The third was *collision outcome* and applied only to the takeover scenario.

The response times were straightforward to compute: simply subtract the time at which the first button press was registered from the time the critical event occurred. The performance score, however, was a little less straightforward. It was a tricky balance to strike, ensuring the score calculation was general enough to cover all thirty-two trials, yet specific enough to apply for all scenarios. Specifically, while the score formula had to be accurate for each individual, it also had to be generalized to apply to the wide variety of treatments resulting from the thirty-two factor/level combinations. In the end, I landed on an objective score out of 100% for the handover scenario that worked well for all treatments. For the takeover scenarios, I also landed on an objective scoring method. This time, the score was based on a straightforward category of collision avoidance outcomes—no formula needed. (Details on these performance metrics are available in *Article 4*.) Inputting these performance metrics into my REML ANOVA software,[121] I was ready to find out whether I could trace causal lines to factor effects at the click of a button.

The results of the analysis were like reaching into a bag of Norwegian candies: I got a variety of flavours—some expected, others unexpected. Some of my suspicions, ingrained as a list of hypotheses, were confirmed; others were upended. *Article 4* reports on the findings in detail along with visual plots of the factor effects. Rather than repeating this here, I want to present some of the highlights as they pertain to each influencing factor. Ultimately, the results encouraged closer examination and led to lively discussion of the implications as well as plans for follow-up studies. As we will see later, it also led to a publication of a risk assessment using Bayesian Networks to analyze outcomes of the collision avoidance scenario, marking a contribution to risk science for autonomous ship operations.

Vigilance decrement observed

By comparing results of the boredom questionnaires across 5- and 30-minute scenarios, I could immediately confirm that the latter reported higher levels of boredom, indicating that a vigilance decrement indeed occurred for those in the 30-minute group. Once again, it is important to note that vigilance decrement and boredom are not the same thing; however, they are *associated*. The boredom questionnaires provided a qualitative way to gauge participants' vigilance decrement by proxy of their reported experiences of boredom. The results also confirmed, beyond a reasonable doubt, that vigilance decrement set in between 5 and 30 minutes. It was encouraging to note that no other factor influenced boredom. In other words, out of the five factors including skill, multitasking, time pressure, and decision

121 "Design Expert Version 13.0.9.0" (Minneapolis, MN, USA: Stat-Ease, Inc., December 10, 2021).

support, the only factor that caused variation in boredom to any significant degree was the amount of time the scenario lasted.

There were other indicators of vigilance decrement aside from those stemming from questionnaire responses. Specifically, during the 30-minute scenarios, I noticed that participants' attention would often begin to slacken in subtle (and sometimes not so subtle) body language. **Figure 14** illustrates this vigilance decrement in three individuals by comparing body language observed in the 0- to 5-minute window to that observed in the 20- to 30-minute window. The significance of confirming the presence of vigilance decrement was that we could now take the scenario length as a proxy for vigilance. In other words, a 5-minute scenario could be taken to represent *high vigilance* and a 30-minute scenario could be taken to represent *low vigilance*.

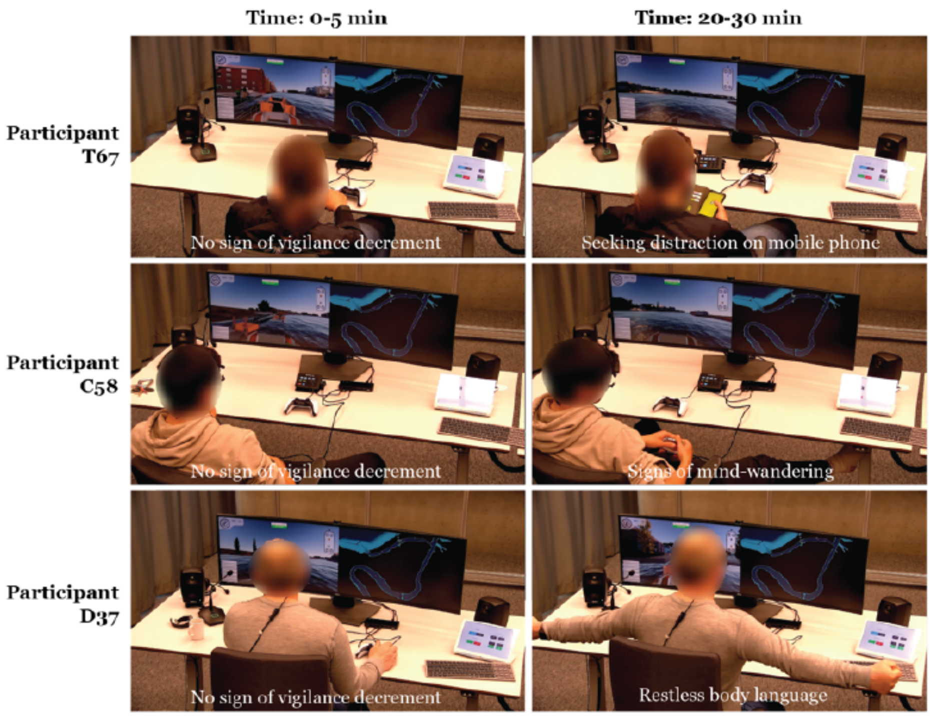


Figure 14. Vigilance decrement sets in after approximately 20 minutes.

Gamers outperformed in multitasking

One outcome I was looking forward to most was evaluating the so-called "gamer hypothesis:" would gamers really outperform expert navigators in the supervisory control role (Chapter 5.1)? The conventional thinking was that navigators had the more transferrable skills,[122] but a part of me was cheering for the underdog. Maybe the gamers, I mused, would "beat" the navigators. After all, the gamers seemed to have an advantage: the remote control room, at least superficially, had many features in common with a video game playing environment, whereas the ship's bridge environment was, in this regard, a few steps removed.

I was a little disappointed, then, when I observed a null result from the formal hypothesis testing. While gamers did score on average 6.5% better and responded to interventions 4 seconds faster and even avoided 1 more collision than their navigator counterparts, these differences were not enough to tip the balance from a mere chance result to causal effect. There was no difference, that is, until I checked the effects of factors *interacting* with the two skillset levels. That's when I discovered that gamers vastly outperformed navigators in scenarios with three ferries. Gamers were better multitaskers than navigators.

It's important to note that this only applied to performance when measured in terms of response time. Performance scores and collision outcomes indicated no such interaction effect. Regardless, the skill-multitasking interaction effect on response time was impossible to ignore. To grasp the interaction effect, consider that for scenarios with just one ferry, response time was 6 seconds for gamers and 7 seconds for navigators, on average. With three ferries, however, response was unchanged for gamers (6 seconds) but *doubled* for navigators to 14 seconds, on average (a plot of this interaction effect is presented in **Figure 15** on page 98). This was the type of result I liked—it went against conventional thinking and showed that gamers, too, had transferrable skillsets. It suggested we should collectively reconsider what skillsets are really required for supervisory controllers.

Effects of multitasking, time pressure & decision support

While skill and vigilance did not directly influence performance, as was expected, the other three factors—multitasking, time pressure and decision support—proved to be major independent influencers of remote supervisory performance. There is one obvious criticism to this outcome, which went something like this: "Okay, but I could have told you that *before* you started testing." To that criticism, I have two responses. Firstly, as I discovered during my literature review, these hypothesized effects existed only as speculations before testing, lacking formal empirical data to support them in the unique case of autonomous ship operations. Second, while it might be intuitive that multitasking, time pressure, and decision support affect a supervisory controller's performance, what isn't at all intuitive is *how important* these effects would prove to be. The experimental results gave us a transparent and repeatable way to assess just how important these factors were. In conversations about

122 For a detailed review about the state-of-the-art knowledge about skill requirements for supervisory controllers of autonomous vessels, see *Article 1*, Section 3.4.3.

remote supervisory performance, the test results were the equivalent of the colourful handholds in an indoor climbing gym: they gave us something to hold on to and make upward progress. Let's take a brief look at these three main effects.

Let's consider multitasking first. When just three vessels were present (a relatively conservative number when considering multitasking), performance scores dropped on average by 13%. In fact, in all three instances that no handover was completed at all during the allotted time, the trial in question was a three-ferry scenario. Response time also increased on average by 3 seconds—a difference consistent enough to clearly distinguish it from single-ferry operations, where response times were 7 seconds, on average. The only performance indicator that was apparently impervious to multitasking effects was collision outcome (strangely, the collision rates were *lower* for three-ferry scenarios, but the difference wasn't statistically significant).

For time pressure, when 20 seconds was available to avoid a collision (high time pressure), participants managed only 2 collision avoidances, resulting in 3 near-misses and 11 collisions. By contrast, for the group with a 60-second intervention window (low time pressure), the outcome was turned on its head: 11 collisions avoidance were recorded with only 1 near-miss and 4 collisions. This indicated that time pressure had a strong effect on performance. However, time pressure did not affect performance scores, nor did it affect response times in the handover scenario. Upon closer inspection, this apparent inconsistency had an explanation. When interventions are handed over to the operator, the decision to take over control is, in effect, made for them. For takeover scenarios, on the other hand, the timer starts ticking the moment the collision target sets its course and presents a hazard. In this time interval, the operator must identify this hazard, assess its danger, and take preventative action. The difference between 20 and 60 seconds turned out to be the equivalent of rapid versus classical chess. (See the "Time Pressure" section in Chapter 5.1 for theory background on decision making under pressure.)

Finally, decision support proved to be the most significant factor effect of all main effects. Its effect was most visible in the handover scenario. When the decision support system (DSS) was *On*, performance scores increased by 22% and response times decreased by 6 seconds, on average. Interestingly, decision support made no difference whatsoever in takeover scenarios. (I will revisit the implications of this result in the Discussion chapter.)

Table 1 presents a summary table of all main effects. The plus sign (+) indicates an increase in output as factor level increases from *Low* to *High*; a minus sign (–) indicates a drop in output. The asterisks indicate significant differences, where "significant" is defined as at most a 10% chance that the outcome was due to chance alone (known in statistics as the *alpha* or α*-level*).[123] This table was helpful for getting an indication of *how much* the factors played a role in addition to just whether they played a role. (I will leave details about the statistical basis of these inferences to the dedicated reader of *Article 4* in this thesis.)

Table 1. Summary of main factor effects on supervisory performance (n = 32).

	Level		Handover (Scenario 1)		Takeover (Scenario 2)
Factor	*Low*	*High*	Response time (seconds)	Perform-ance (/100%)	Outcome count (Avoidance/Near-Miss/Collision)
A-Skill	*Gamer*	*Navigator*	+4	-6	+3 /-4 / +1
B-Vigilance	*No decrement*	*Decrement*	-1	-8	-1 / 0 / +1
C-Multitasking	*1 ferry*	*3 ferries*	+3*	-13*	+3 / -2 / -1
D-Time pressure	*20 sec*	*60 sec*	0	+6	+9 / -2 / -7*
E-Decision support	*Off*	*On*	-6*	+22*	-1 / +2 / -1

Interaction effects

The main effects represented tests of the main hypotheses, evaluating the extent to which each expected factor influence held true. How these main effects *interacted*, though (how they combined to form effects of their own) was outside of the scope of the initial hypotheses. It came as a surprise, then, when I discovered that some interaction effects were too large to be ignored.

I have already mentioned one such interaction effect: the skill-multitasking effect that showed that gamers outperformed navigators specifically during multitasking conditions. I want to end this chapter by returning to this interaction effect and highlighting three additional interaction effects, all of which represent *new* hypotheses generated from the analysis. All the relevant interaction effects are plotted in **Figure 15**. The best way to interpret these plots is to consider that should *no* interaction effect exist, the slopes of the two lines would be parallel. The more divergent the two lines are, the stronger their combined (or interacting) effect on the response variable.

---

123 The significance level (α-level) is the subject of debate within the scientific community. However, it remains a useful tool for defining exactly what its name implies; namely, what difference is *significant*? While 10% is not uncommon, usually a 5% α-level is chosen. Generally, it is up to the researcher to define. A good guide is considering the *practical significance* for the case at hand. In my case, I wanted to allow for a more "sensitive" significance test based on the implications of my results for risk-based design and safety of autonomous ships. To illustrate: what is a significant difference in reaction time? Is 5 seconds? What about 7 seconds? Practically, I would consider both potentially significant. Setting the α-level to 10% (as a more sensitive threshold that 5%), serves to attribute a bit more practical significance to the test statistic.

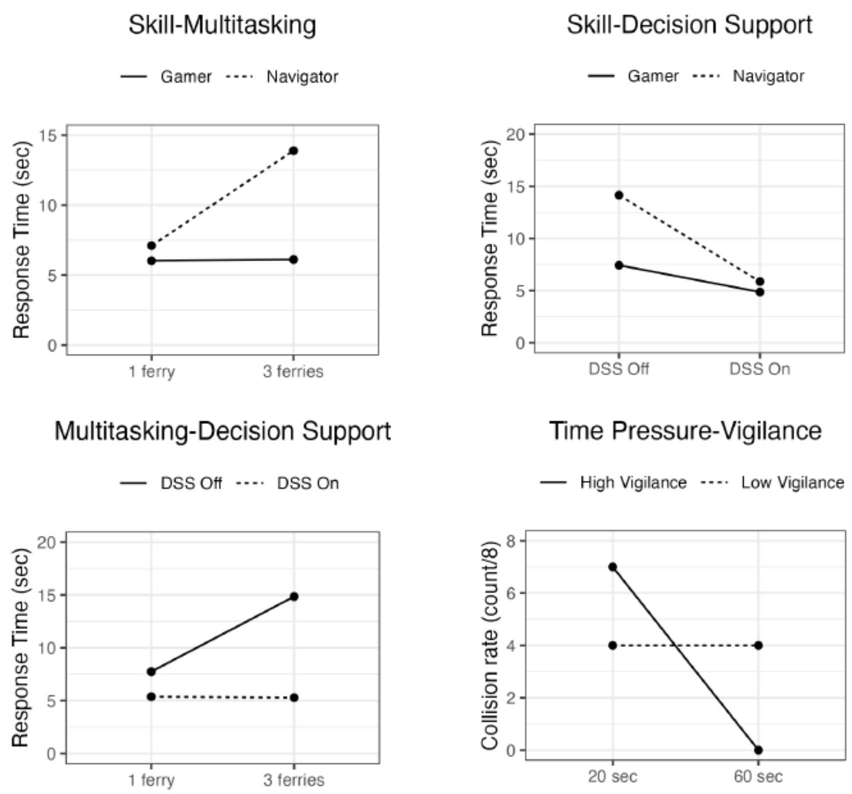


Figure 15. Interaction effect plots (points represent average values for the samples; n=8).

Below is a short summary of the interaction effects presented in **Figure 15**:

- **Skill-multitasking:** While *gamers'* response times did not change when the number of ferries changed from *one* to *three*, *navigators'* reaction times increased considerably, from 7 to 14 seconds.
- **Skill-decision support:** While *gamers'* response times were virtually unchanged when the DSS was changed from *off* to *on* (from 7 to 5 seconds), *navigators'* reaction times dropped considerably (from 14 to 6 seconds).
- **Multitasking-decision support:** With the DSS turned *on*, response times were the same when the number of ferries changed from *one* to *three*; however, with the DSS turned *off*, response times increased considerably (from 8 to 15 seconds).
- **Time pressure-vigilance:** This was the only interaction effect that applied to collision avoidance outcomes. Recall that the 5-minute scenario is a proxy for "high vigilance" and that the 30-minute scenario is a proxy for "low vigilance." When the

operator had *low vigilance*, it made no difference to the collision rate whether the available intervention time changed from *20* to *60 seconds*; however, when the operator had *high vigilance*, the collision rates dropped considerably, from 7 to 0 collisions.

The presence of such strong interaction effects showed that human factor influences did not have one-off effects on supervisory control performance, like bowling balls heading down an alley towards their target. Rather, human factor influences had knock-on effects, working more like curling, where each rock placement on the rink affected the placement of others. While investigation of interaction effects was not part of the original test scope, their strong presence in this experiment indicates that interaction effects are an important mechanism in remote supervisory control performance.

7.5. Summary

In this chapter, I have walked through a diverse range of results, including the outcome of a literature review (*Article 1*), field observations and interviews (*Article 2*), a design process of a full-scale prototype (*Article 3*), and finally an ambitious experiment with 32 participants (*Article 4*). Next, I will discuss the implications of the results and highlight patterns in the methodologies I used. Alluding to our stated definition of design (Chapter 1.2), we are beginning to glimpse the "preferred situation" of land-based control of autonomous ships. In the next chapter, I will lend some shape to this preferred situation by describing four recurring themes that define its essential features.

8. Discussion

Design is concerned with how things ought to be, with devising artifacts to attain goals.

—Herbert A. Simon, *The Sciences of the Artificial*

The chapters of this thesis have chronicled four articles that make up my thesis, starting from humble beginnings in a literature review and ending with an ambitious testing campaign. The four articles have addressed remote supervisory control from different angles, and each have contributed in their own way to a dynamic field at the crossroads of many disciplines. At this point in the thesis, however, I hope you can appreciate that the four presented articles do not stand alone.

Specifically, you might have noticed certain themes arise again and again. In tracing my PhD project from start to finish, these themes are like threads in the woven fabric of the work, visible if you know where to look. In this chapter, I zoom in on four such reoccurring themes, showing how they bridge the disparate published works together and how they form the core messages of the thesis.

8.1. Revisiting the alignment problem

During my literature review, I stumbled upon the dynamic and multidisciplinary field of "AI alignment." Like similar efforts in "Explainable AI" or "AI transparency," the field of AI alignment aims to make AI integration more user-friendly, more approachable, more accountable, and safer for a wide range of people with a wide range of abilities. Unlike other subdisciplines that have emerged, though, AI alignment contends that current technology development is discordant with its users' expectations, and that the aim is to move technology development in a preferred direction: to realign it with users' needs. Sound familiar? These concerns mirror many of designers' typical concerns (Chapter 1.2).

Before moving on, I want to reframe the question of supervisory control and the challenge of defining the role of humans in autonomous systems in terms of AI alignment. In the opening chapters of this thesis, I introduced the term "sorcerer's apprentice," a label offered by writer Brian Christian to describe the supervisory controller's role (Chapter 3.2). This term described the precarious arrangement of scrambling to the aid of a powerful yet imperfect force—the backup to a deficient AI. Being a sorcerer's apprentice implies a new partnership, a pact that must be brokered carefully to avoid potentially catastrophic divergences. Christian defines the alignment problem in terms of this emerging power struggle:

> "How to prevent such a catastrophic divergence—how to ensure that these models capture our norms and values, understand what we mean or intend, and, above all, do what we want—has emerged as one of the most central and most urgent scientific questions in the field of computer science. It has a name: *the alignment problem*."[124]

What struck me was that the alignment problem also had something of Simon's definition of design in it. It presented the type of divergence that calls for transforming situations into preferred ones, of devising artifacts concerned with how things ought to be. In this light, the discrepancies I observed in *Article 2* between technology designers' construal of AI collaboration and workers' accounts from the field could also be reframed in terms of alignment. Indeed, this idea has a substantial research pedigree, framed as a gap between "work-as-imagined" and "work-as-done," and compelling designers to address this gap.[125] Even the field of Explainable AI framed the fundamental issue with AI as one of usability discordance. The alignment community introduces us to a growing network of like-minded researchers and practitioners who dedicate their work in a context that is directly relevant for designing control rooms for autonomous operations. Next, I will draw from the experimental results of my PhD research to illustrate this case.

Human factor influences

The experimental results investigating human factor influences on remote supervisory performance served, in effect, to indicate the extent of AI misalignment. The premise that these factors influenced performance at all suggested that I, as a designer, might be

124 Christian, *The Alignment Problem: Machine Learning and Human Values.* Introduction, Page 31.

125 These terms emerged in the Resilience Engineering community and appear to stem from the distinction proposed by Erik Hollnagel and David Woods between system task descriptions and their associated cognitive tasks. See Erik Hollnagel and David D. Woods, "Cognitive Systems Engineering: New Wine in New Bottles," *International Journal of Man-Machine Studies* 18, no. 6 (June 1, 1983): 583–600, https://doi.org/10.1016/S0020-7373(83)80034-0.

misconstruing operators' capacity to use the technologies. Human performance, after all, is just the reciprocal of human error, which, as we recall, is not just a function of the operator, but also of the designer. In this sense, "human factor influences" could just as well be called "designer influences."

Consider the effects of skill, which I decomposed into two groups: gamers and navigators. The results indicated that differences between the two were not showstoppers. Initially, this surprised me. But if we were to "realign" this outcome from the vantage point of experimental hindsight, then it is not really surprising at all. Neither gamer nor navigator represented an ideal candidate for remote supervisory control any more than an experienced car driver might be qualified to drive a long-haul tractor trailer. At best, the role will demand a hybrid blend of the two skillsets: a trained operator whose competencies match the job—not the other way around. As I reported in *Article 2*, these skills are prone to degradation over time and should therefore be maintained with a work design that carefully balances remote tasks with manual tasks. By this logic, remote supervisors may benefit from occasional work aboard the vessels they are tasked to monitor; they might also benefit from periodic or even scheduled manual interventions just for the sake of practice. If one is skeptical of the effectiveness of such a work design, consider the alternative: a generation of remote operators that have never been aboard a real vessel, with many months or perhaps years having passed since the previous manual intervention.

The investigation of vigilance, too, called into question the alignment of my own expectations of the so-called "operator-in-the-loop" with the reality of how operators experienced their work. In a word, they reported that backup was "boring." In fact, as I showed in *Article 2*, boredom featured as one of the defining characteristics of the current transition to human-AI collaboration in maritime navigation. In my attempt to measure the effects of boredom on supervisory performance (which I did in terms of the quantifiably manageable "vigilance decrement"; see Chapter 5.1), I found that vigilance decrement did not visibly affect performance, as expected. However, it is likely the effects of boredom over time are more pernicious, leading to a state of diminished job satisfaction of the kind I glimpsed in my interviews with navigation crew aboard highly automated ferries.[126] The treatment of boredom underscored a difficult design problem: how does one hold operators' attention when they aren't intervening to save the day? Clearly, assigning contrived tasks only

[126] I explored the topic of job satisfaction among the navigators I interviewed in a conference article I presented in 2022. See Erik Veitch et al., "From Captain to Button-Presser: Operators' Perspectives on Navigating Highly Automated Ferries," *Journal of Physics: Conference Series* 2311 (2022): 012028, https://doi.org/10.1088/1742-6596/2311/1/012028.

serves to distract or, worse, exacerbate boredom through tasks perceived as meaningless.[127] In future studies, designers should explore alternatives to the current "backup" interpretation. This might involve, for example, having the system communicate its intentions and ask operators to approve them. It might also involve scheduled periods of remote manual control (which has the added benefit of combatting skill degradation) and meaningful secondary tasks like logging relevant events and weather conditions.

The implications of the effects of multitasking could also be framed in terms of the alignment problem. The results indicated significantly diminished performance when operating three vessels compared to one, which was concerning considering multi-vessel control is one of the driving forces of the remote supervisory model. However, we must entertain the notion that this diminished performance might have stemmed from a poor interface as well as inherently "human" factors. Take the user interface I had designed for participants in my simulator experiment as an example. The viewer functioned a bit like tabs in a web browser: the user could select a camera stream for one ferry at a time but could not view several ferries at once. Although all traffic could be monitored in an overview map viewer, the "live camera streams" were restricted to that view which had been most recently activated. In the exit interviews, gamers and navigator alike bemoaned the restrictive camera interface. Upon being asked to evaluate the interface and consider possible improvements, about a third (11 of the 32) participants recommended a viewer that displayed camera feeds for all vessels simultaneously. As participant I65 expressed, "It would've been good with a 360-camera that could take in the whole scene." The separate *Front* and *Back* camera options that were available, which required the operator to actively select one or the other, seemed misaligned with how operators oriented their attention. As participant Z99 reported, "I pressed the button to change cameras and [as a result] drove the wrong way on the joystick… you had to use an extra step to… perceive the situation as quickly as possible." This suggested that a better alignment could be achieved by leveraging what Oliva and Torralba called scene gist,[128] a concept I introduced in Chapter 3. Scene gist supports the idea that parts of a scene are recognizing semantically from the whole, rather than the other way around. In other

127 When I did my mid-term review in December 2021, one of the audience members, social anthropologist Bjarne Vandeskog, challenged me on the notion that I was creating "bullshit jobs" with designing supervisory work. Bjarne was referring to the book *Bullshit Jobs: A Theory* by David Graeber, who argues that meaningless jobs are on the rise and that those in positions of power have a moral and political prerogative to shift their focus on creating jobs that foster identity and self-respect in employees' work. When I read the book, however, I found that Graeber does not subscribe to the idea that increasing automation will take jobs or even create meaningless ones. On the contrary, he points out that increasing automation will likely lead to a better understanding of why humans are needed in certain applications in the first place. This argument fits well with my thesis and makes the case for why supervisory control is *not* a "bullshit job." See David Graeber, *Bullshit Jobs: A Theory* (New York, NY, USA: Simon & Schuster, Inc., 2018).

128 Oliva and Torralba, "Building the Gist of a Scene: The Role of Global Image Features in Recognition."

words, compartmentalizing a scene from its isolated parts and only *then* constructing its meaning is misaligned with how the human visual system works—people construct meaning and then, with the help of that meaning, identify objects.

Closer examination of operators' decision making during control takeovers was also suggestive of emergent phenomena like those described by Klein's naturalistic decision making and Suchman's situated actions. Asked in the exit interviews,[129] "Did you make a plan for what to do?" only 9 participants answered that they did (11 answered that they did not have a plan, 11 were unsure, and 1 was not applicable as they did not take over control). Of the almost two-thirds (22 out of 32) that either had no plan or were unsure, many described making a plan *after* taking over control. As D28 said, "I was about to crash, so I decided to back up, then and there—more like a reflex." Similarly, C58 replied, "No, I didn't have time, so I tried to react as fast as possible." Three participants (D37, I29, Z99) explicitly said the takeovers were "intuitive," invoking Klein's language describing decision making under pressure. Others invoked Suchman's situated actions, like M87 who described the takeover as "more on the spur of the moment," E02 who said they "had to make the plan on the spot," and B03 who reported making a plan "almost immediately when I pressed the button." These observations support the findings in *Article 2* that there is a discrepancy between how designers construe decision making as logical, sequential models and operators' own accounts of heuristic decision making. Designers may do well to align themselves more towards emergent concepts like gist, naturalistic decision making, and situated actions rather than reductionist cognitive models like Endsley's three-level situation awareness.

Other, more detail-oriented, feedback was uncovered from the interviews that hinted towards better alignment of multitasking support. Specifically, about half (17 of the 32 participants) recommended a notification-based tracking system that reported Closest Point of Approach (CPA). Among suggested improvements to the interface were colour-coding on the map viewer that displayed vessels with a low CPA in a different colour. Overall multitasking performance might have been improved had the 360-degree camera and CPA-enabled interface been available. Only additional user testing can determine how much better aligned these features are to the user.

As for time pressure, the results implied that careful consideration about expected intervention actions should be incorporated into the operational design domain. If 20 seconds proved too short a time window to avoid a collision, as the results indicated, then external factors that might in fact impose time pressure, like traffic and constrained waterways, should be carefully assessed in operational planning. When the demands on

[129] Once again, a reminder to readers that these interviews (and other experiment data) are openly available at https://doi.org/10.18710/WYFMMP.

intervention times are high, additional systems might help to mitigate the associated risks. For example, the CPA tracking envisioned by half the participants, had it been implemented, might have provided more useful advance warning than that provided by the Decision Support System. Similarly, enabling radio communication with other traffic might have proven an effective collision avoidance measure, as might simply have reducing the vessel's speed during high-risk situations like when the waterway became congested. The results were also indicative that operators should work together to solve time-demanding problems rather than working alone. In any case, circumventing the issue of time pressure on operators will ultimately be a more impactful design goal that merely understanding how time pressure hamstrings performance. Aligning the remote control center with manageable time pressure expectations, in other words, only starts with understanding the influence of time pressure on performance. The natural follow-up question is: can we design a system that prevents the identified negative influences on performance from taking hold?

Finally, the results of how decision support influenced supervisory performance could also be interpreted in terms of AI alignment. The DSS was designed, namely, to provide two decision aids: the first was a warning ten seconds prior to the critical event that alerted the operator about the impending danger; the second was bounding boxes highlighting all traffic in the vicinity (see *Article 4* for figures of the DSS). For the handover intervention, this served to alert operators about the possibility of impending control transfer, resulting in significantly improved performance. Like an alert in a calendar application reminding users of an upcoming appointment, it wasn't altogether surprising that operators heeding the alerts showed up to the interventions "on time," with reaction times reduced by a full 6 seconds compared to when it was *Off*. However, for takeovers, the DSS did not support operators, with similar collision rates occurring whether it was *On* or *Off*. One explanation stems from the fact that operators must be more self-sufficient in takeover scenarios; that they must appropriately identify and react to the hazards of their own accord, regardless of decision support prompts. An alternative "AI alignment" explanation is that the DSS used during testing was not aligned to the needs of operators. Warning prompts and bounding boxes were, in this view, not enough to support takeovers in the set-up used. In this light, the experimental results pertaining to supervisory performance with or without the DSS should be balanced with careful consideration of DSS *usability*. In future work, testing aimed at evaluating transitions of control should be considered one-and-the-same with testing to *support* transitions of control.

8.2. Human-centered AI

Once I was aware of the alignment problem, I saw it everywhere. It emerged in Level of Autonomy (LOA) taxonomies that implied a trade-off in human control for increases in machine control. It emerged in interviews with designers and navigators whose "work-as-imagined" diverged from accounts of "work-as-done." It also emerged from the experimental results in the NTNU Shore Control Lab, where diminished performance served to expose weaknesses in the design as much as in factors intrinsic to human abilities. These misalignments suggested a divergence from core user values, including safety, trustworthiness, and reliability. Like a picture hanging crooked on the wall, the design situation practically begged to be set straight. The outstanding question was *how*.

The human-centered design (HCD) framework, while not providing a fix-all, at least allowed me to assess the extent of this misalignment—like taking a level to our crooked picture. In fact, the HCD process makes misalignment its raison d'être, the continual cycle of evaluation always pointing the path towards potential improvements. From the HCD perspective, all design is misaligned in some way. The goal is this sense is not perfect alignment; rather, the goal is uncovering the steps towards realignment. Herbert Simon never promised designers would deliver the "right" solution, after all. He meant for them merely to generate a "preferred" solution.

When I stumbled across the Human-Centered AI (HCAI) framework,[130] it seemed to me like a reification of HCD principles applied to the emerging and domain-specific problems of human-AI collaboration and supervisory control. Specifically, concepts like *stages of automation* provided an alternative to LOA taxonomies that were ill-equipped to deal with collaboration; meanwhile, *supervised autonomy* offered a sense of designing for human-oriented AI in the right direction (I return to the concept of supervised autonomy later in this chapter). In *Article 1*, I presented a figure depicting my version of what the stages of automation approach might look like when compared to the common LOA taxonomies used in autonomous vessel design. This figure is recreated in **Figure 16**. Inspired by the HCAI framework, the figure also depicts the nudge that is needed towards the preferred situation of "safe, trustworthy, and reliable" collaborative control (represented by dotted lines).

130 Shneiderman, "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy."

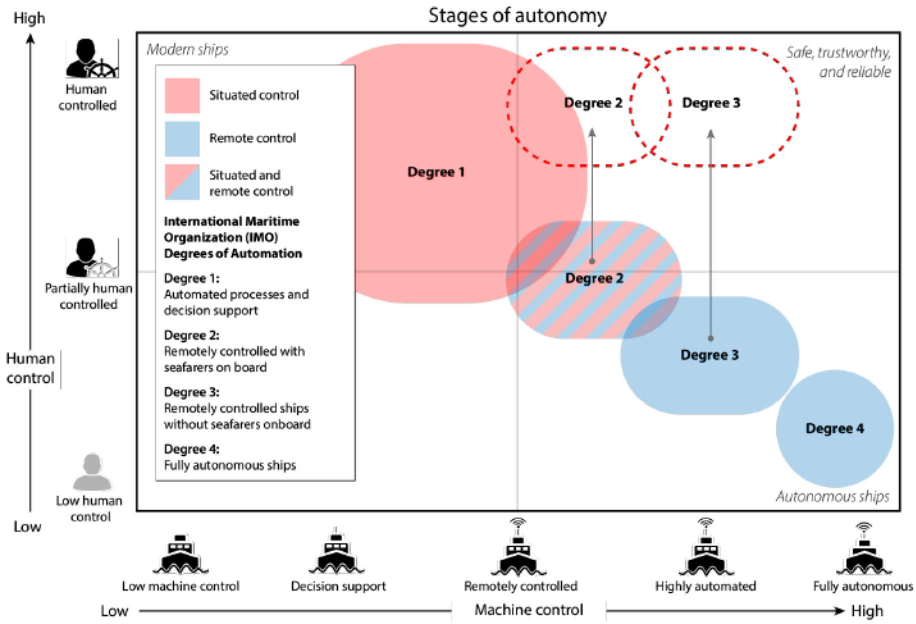


Figure 16. Stages of automation call for high human *and* machine control, after Ben Shneiderman's Human-Centered AI framework.

When I heard accounts from navigators about their perceived shift towards a backup role, it called to mind this plot. Perhaps "backup" was symptomatic of an incremental increase in machine control without a complementary increase in human control. The navigators, for their part, were generally nonplussed about the developments. "You go from being the one who performs something to just monitoring something," recalled Henrik, one of the navigators I had talked to. With Ola, another navigator I interviewed, a sense of resentment had emerged: "The engineers must have something to invent… [and] since they have the shipowners on their side, they make sure that they can get rid of people and hope that the system can be more economical." Ola continued, "The working day of the captain will get boring. You press three buttons. There's no challenge. You become an operator who monitors the system and is ready to press a button if there's a bug in the system." Taking in this feedback, I was hardly left inspired that automation was moving in a direction of improved safety, trustworthiness, and reliability. Were we to get nearer the mark, we had to address alignment in human-AI collaboration, and frameworks like HCAI provided us with the concepts to diagnose problems and get on the path to realignment.

Supervised autonomy

For decades, control rooms have been evolving in tandem with automation technology. Today, autonomous technologies and AI applications are heralding a new age in control rooms. Considering the rapid advances underway in AI development, designers of control rooms might do well to keep up.

Ben Shneiderman, in his book "Human-Centered AI," uses the term "supervised autonomy" to express the vision of supervisory control for autonomous vehicles.[131] As he explains, "'Supervised autonomous' signals that humans are monitoring performance by way of visual control panels on devices or at remote control centers, so that they can intervene in a timely manner to ensure vital outcomes."[132] Framing it this way, Shneiderman argues, "suggests that human autonomy is an equally important part of the goal."

For supervised autonomy, it is not enough that remote control centers represent a reaction to lacking human control in autonomous systems. It suggests that we must move beyond a human-machine dualism and consider remote control as *part* of the autonomous system. Supervised autonomy also casts a hopeful light on issues inherent in the backup role, shifting it, as it were, to a *supervisor* role befitting the responsibility it entails. With time, this way of framing control center design might also help to address many of the so-called ironies of automation, including the associated issues of operator vigilance, rapid transitions of control, and skill degradation. Supervised autonomy is in this sense an ally of alignment efforts whose aim is to design better human-AI collaboration.

8.3. Future work

The experiment I presented in this thesis, although ambitious in its scope, represents just a starting point in what I imagine to be a series of testing efforts. In the language of human-centered design, these testing efforts evaluate control center configurations and bring us one step closer to seamless human-AI collaboration in supervisory control. Here, I want to extend beyond the scope of the four articles in this thesis and highlight a few ways this future work might play out.

Cognitive load and stress

The experiment I presented generated *lots* of data.[133] Much of the data remains to be analyzed and presented in future publications. Among these untapped data sources were pupillometry

---

131 Shneiderman, *Human-Centered AI*. Chapter 15.
132 Shneiderman. Chapter 15, page 114.
133 Veitch, "Dataset for: Human Factor Influences on Supervisory Control of Remotely Operated and Autonomous Vessels."

measurements, which were collected with specialized eye-tracking goggles.[134] Among the hordes of pupillometry data (e.g., gaze direction, blinks, saccades), measurements of pupil diameter provided an especially promising window into an important yet elusive parameter: *cognitive load*. (I briefly introduced the theory behind how pupil diameter relates to cognitive load in Chapter 5.) Pupil diameter may also hold clues about vigilance decrement—a phenomenon I attempted to capture by means of Likert scale-based questionnaires and observation of body language.[135]

Sensor data registering Electrodermal Activity (EDA) was also collected, leveraging a well-described association between sweat response and *stress*.[136] Like cognitive load, stress is another central parameter in evaluating control center design, yet it can be difficult to measure quantitatively. Like the pupillometry data collected in the experiment, this dataset remains to be analyzed in full. Yet, preliminary analysis at the time of writing this thesis suggest that EDA holds promise as a window into operators' internal condition when facing potentially demanding and stressful interventions on the job.

Figure 17 presents a plot of pupil diameter and EDA data recorded during a handover scenario for the participant D27. From the Run Log in *Article 4*, we can see that D27 was a *Gamer* who underwent a *30-minute*-long handover scenario, had *three* vessels to supervise, and had the DSS turned *Off*. The treatment also specified that when the handover began, the participant had *20 seconds* to complete the intervention. The combination of low vigilance induced by the 30-minute monitoring stage, multitasking conditions, low intervention time, and no decision support presented a particularly challenging treatment. In the plots shown in **Figure 17**, the resulting jumps in cognitive workload and stress during intervention are visually depicted by sharp increases in both pupil diameter (left) and EDA (right) shortly after the handover. Moreover, the decreasing slope before the critical event is indicative of a vigilance decrement. Additionally, by comparing the slopes before and after the critical event, one can get an indication of the relative magnitude of the cognitive workload or stress: the larger the change or "swing" in direction, the more demanding the intervention. We observed already that vigilance decrement did not affect performance (Chapter 7.4), and that operators reported that dealing with boredom was, in effect, experienced as a chore. But the quantitative results suggest a nuance to the hypothesis that vigilance is related to

[134] I used PupilLabs's *PupilCore* product for my experiments. See their website for more information and technical specification: https://pupil-labs.com/products/core/tech-specs/

[135] My approach to analyzing video data (and especially interpreting body language) was inspired by techniques described in Christian Heath, Jon Hindmarsh, and Paul Luff, "Analysing Video: Developing Preliminary Observations," *SAGE Visual Methods*, 2010, 365–89.

[136] Cornelia Setz et al., "Discriminating Stress From Cognitive Load Using a Wearable EDA Device," *IEEE Transactions on Information Technology in Biomedicine* 14, no. 2 (March 2010): 410–17, https://doi.org/10.1109/TITB.2009.2036164.

performance: namely, that sharp jolts from a state of diminished vigilance experienced during interventions seem to require especially large demands in cognitive workload and thereby induce stress. One might rightfully wonder: over time, would a kind of intervention fatigue set in? Can we design operator work such that we shrink the jolting peaks and level out the boring troughs? Only by analyzing all the results and by completing additional testing can we explore these question in a scientific way.

Note that for easier interpretation, the pupillometry data (**Figure 17**, left) is represented as both left and right pupils and has locally estimated scatterplot smoothing (LOESS) applied. The EDA data (**Figure 17**, right) also has a LOESS smoothing applied. LOESS represents a "moving average" of noisy data and is computed based on non-parametric regression techniques. These results are indicative only and should not be interpreted as final.

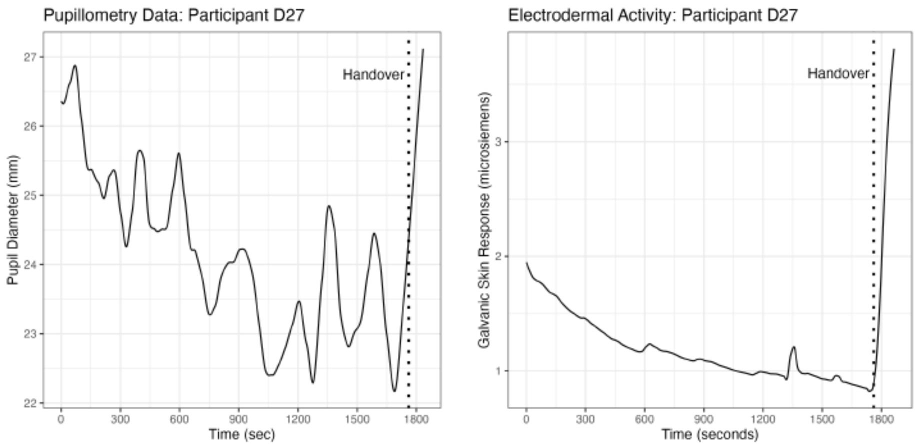


Figure 17. Preliminary pupillometry (left) and electrodermal activity (right) data.

How might pupillometry and EDA data be used in other investigations of supervisory control? One approach is to adopt an experimental blueprint similar to the one I conducted. One might, for example, imagine assessing causal lines of influence between cognitive workload (pupil diameter) and relevant factors like multitasking and time pressure. We may wish to do the same for stress (EDA). Another point of interest is the influence of skill on cognitive load and stress. Might a group specifically trained for remote supervisory control tasks display lower cognitive load and stress compared to untrained gamers or navigators? Would this lead to improved performance during critical events? Would having a teammate (the equivalent of a co-pilot) relieve stress under multitasking conditions? The research directions are practically endless.

One thing is for certain: pupillometry and EDA open investigative doors into cognitive load and stress response, respectively—two elements for which scientific assessment is traditionally difficult to achieve. More work will be needed, however, to determine the most appropriate way to analyze the data in the context of supervisory control work.

Inattentional blindness

When focused on a task, unexpected visual stimuli can go by unnoticed. I introduced the concept of inattentional blindness briefly in the opening chapters (see Chapter 2). Here, I want to discuss my own investigation into inattentional blindness in the context of remote supervisory control and outline possible directions of future work on the subject.

As you might recall, inattentional blindness describes an inability to perceive things that we do not consciously attend, even if those things are present within our visual field.[137] Inattentional blindness appears to be a natural part of being human and may even be necessary to balance conscious perception with attentional resources. However, it may be problematic if one's task is to look out for objects or events that are out-of-the-ordinary. Remote supervisory controllers seem vulnerable to inattentional blindness because their role is underpinned by just this: looking out for the out-of-the-ordinary.

My way of checking for inattentional blindness was straightforward: I added a gorilla passenger to each of the simulated ferry scenarios. The gorilla was, of course, a nod to Simons and Chabris's famous study.[138] At precisely the same time in the respective 5-minute and 30-minute scenarios, a gorilla entered the ferry along with three other (human) passengers. The gorilla stood up upon entering, thumped its chest, and stood back down. To ensure that the gorilla would enter in operators' field of view during the scenarios, the gorilla entered the ferry during the crossing in which intervention was needed. In the exit interviews,[139] I asked each participant the question: "Did you see the gorilla?"

The results indicated that half the participants (16 out of 32) did not see the gorilla. This replicated findings from previous studies. (Simons and Chabris's original study, for example, discovered that half of participants did not recall seeing the ostensibly obvious gorilla.) Also, it replicated a study by Trafton Drew and colleagues that showed that even experts experienced inattentional blindness.[140] What surprised me most, though, was when I checked

---

137 Arien Mack and Irvin Rock, *Inattentional Blindness* (Cambridge, MA, US: The MIT Press, 1998).
138 Simons and Chabris, "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events."
139 These exit interviews are available in an open database: Erik Veitch, "Dataset for: Human Factor Influences on Supervisory Control of Remotely Operated and Autonomous Vessels (In Review)" (DataverseNO, 2023), https://doi.org/10.18710/WYFMMP.
140 Drew, Võ, and Wolfe, "The Invisible Gorilla Strikes Again: Sustained Inattentional Blindness in Expert Observers."

the eye-tracking data. The eye-tracking data showed that 12 of the 16 operators who did not recall the gorilla had, in fact, gazed directly at it! (**Figure 18**). None of the five factors correlated to whether the gorilla was reported. It appeared, as predicted by theory, to be essentially random—like the outcome of a coin-flip.

Figure 18. Screenshot of eye-tracking viewer for participant O52 during testing (green/red dot shows gaze location; dotted line annotation is added).

At this stage, although the preliminary results speak for themselves, I have only started to analyze the inattentional blindness results and discuss the implications for supervisory control.[141] Like many aspects of my experiment, the discoveries raise multiple new questions for each one I had originally set out to investigate. Can we mitigate the effects inattentional blindness with training techniques? Is working in teams a good way to circumvent inattentional blindness? We can rest assured gorillas won't board autonomous ferries any time soon. However, if we consider the gorilla as a colourful metaphor for unexpected events—as *unknown unknowns*—it would be hubris to brush the inattentional blindness issue aside. The invisible gorilla poses an important question: can we support operators in their ability to detect unexpected events, leaving this vital task to more than just chance?

[141] One such discussion will be available in an upcoming publication: Erik Veitch and Ole Andreas Alsos, "Gorillas on a Boat: Inattentional Blindness during Supervisory Control of Autonomous Vessels (In Press)," 2023.

Model prediction

The results of the experiment left me with a series of five factor effects matched to a series of outputs, including response time, takeover performance, and collision outcome. It was tempting then, to use this information as a crystal ball. What I mean is, say we wish to *predict* response time. Based on our experimental results, it may seem reasonable to do so if we know the operators' skillset, the length of monitoring time, the number of vessels supervised, the time window available for intervention, and the availability of decision support. Specifically, it is tempting to apply an equation based on the coefficients and intercepts developed from the linear models produced from ANOVA to compute this prediction. However, there is one fundamental issue with this approach: assessing causation is not the same as prediction. Model prediction accuracy of the results as they currently stand will be suspect at best. There are two reasons for this. The first stems from setting up the experiment with categorical variables. This means that any factor settings outside our chosen levels (e.g., 5 vessels, 60-minutes in passive monitoring, hybrid gamer-navigator skillset) are effectively outside the scope of the experiment. The second reason stems from our lack of replicates, or multiple experimental runs with the same factor settings. Lacking replicates, we have no way of gauging variability of the observations within each treatment ("pure error"), or, conversely, the variation caused by unaccounted-for variables. Only by running replicates can we generate models worthy of prediction with a measurable degree of certainty.

The experimental work I've presented made factor screening its mission; that is, determining whether factor effects are important. Having uncovered the important factors, a reasonable strategy is to keep pulling the thread—to *augment* the experiment by adding replicates and honing the predictive power of the models for the important factors. This type of predictive power would be useful, for example, for ongoing efforts in risk-based design of autonomous ship systems. Future work may thus involve augmenting the completed experiment design to more closely study the important factor effects.

8.4. Risk-based design

Risk-based design encapsulates many of the themes we've touched on, and as such makes for a good way to end this chapter. Risk-based design, as I defined it earlier, can be considered a subset of goal-based design, as it is dedicated to drawing up functional requirements from system objectives, or goals. Risk-based design is oriented towards mapping functional requirements of safety controls to uncertainties, or risk.

Risk-based design for autonomous vessels proved to be a dynamic and growing field that matched well with the open-ended nature of problems tackled in design and with the framework of human-centered design. Risk and design both speak to human interactions with a complex world under transition. It is natural, then, that the tools for getting a handle on risk and design overlap considerably. The Bayesian Networks that proved popular among risk scientists also draw a parallel to human-centered design methods favoured by designers. Both are content with starting with a modicum of information (or even a best guess) and updating their predictions as new information arrives. Proponents of neither party suggest their results are final; on the contrary, the two approaches are built on the premise of continual improvement. This applies whether the results are risk predictions or designed configurations.

In an article led my former colleague Tingting Cheng, who contributed to my experiments in 2022 while she was a visiting PhD student at NTNU, we demonstrated how risk science and design might join forces.[142] In it, Cheng used the results of the collision outcome scenario to generate a Bayesian Network modelling causal and probabilistic relationships between Performance Shaping Factor (PSFs) and human errors identified in the empirical data. In effect, she was able to analyze the reliability of control room operators for autonomous vessels, showing, for example, that lowered available intervention time (20 seconds, in my experiment) had the most significant effects on the quality of takeovers. The work demonstrates how risk-based design ties together elements of risk assessment with design and empirical testing.

Limitations of risk-based design

Having lauded risk-based design, I should also note some of its limitations. To this aim, the history of nuclear power in the USA provides a sobering account. In his six-volume "Safe Enough? A History of Nuclear Power and Accident Risk," Thomas Wellock, historian of the U.S. Nuclear Regularity Commission, dives into the risk-based decision making underpinning the heyday of nuclear power in the 1960s and 70s. Tellingly, he writes that likelihoods of a power plant accidents were based on "expert guesswork or calculations that often produced absurd results."[143] It is also telling that the book's title is never answered. "Safe enough?" is presented as a rhetorical question. Today, we can look back on more than fifty years of nuclear power history and generate a sizeable statistical database; one that includes five major accidents (Five Mile Island, Chernobyl, and three reactor failures at Fukushima). Today, we also look back on a largely defunct and mismanaged nuclear power industry that has lost face in public perception.

The risks of autonomous shipping are significantly lower than those presented by nuclear power plants and as such do not present a one-to-one comparison. But the nuclear case is illustrative in that we are stubbornly inclined to produce risk calculations for complex

[142] Tingting Cheng et al., "Analysis of Human Errors in Human-Autonomy Collaboration in Autonomous Ships Operations through Shore Control Experimental Data (In Review)," 2023.
[143] Thomas R Wellock, *Safe Enough?: A History of Nuclear Power and Accident Risk* (University of California Press, 2021). Preface, page xv.

systems despite the "expert guesswork" it calls for. Are we biased, like Kahneman argued, towards confirmation of prior hypotheses even in the face of contradictory evidence (see Chapter 2)? Are we falling for automation bias? Or are we, as Strevens suggested, inclined not towards scientific objectivity in scientific pursuit of knowledge, but simply towards the production of more data to test our stubbornly held hypotheses (see Chapter 6)? It's hard to say. But it is illustrative that the question "safe enough?" still eludes us and perhaps always will elude us. Similarly, the question of whether a design is "good enough?" eludes designers, providing a pretext for more solutions in place of a definitive answer.

8.5. Summary

The myriad techniques used by risk scientists and designers, as well as the vast array of disciplines from which these techniques sample (e.g., human factors, cognitive science, and human-computer interaction, just to name a few), speak to richness of the problem at hand and to the cross-disciplinary insights that will be needed. In the years to come, it may very well be that the designer's role, as shepherds guiding the flock towards a preferred situation (see Chapter 1.2), will also be as liaison between institutional boundaries, as diplomats brokering harmony across disciplinary divides. My experiment alone, after all, brought together a designer, a risk scientist, and a psychologist. Overall, my PhD work also brought together sociologists, engineers, and computer scientists. Even the methodologies I used proved to have more in common than not, including human-centered design, Bayesian Networks, and qualitative analysis techniques. Control room design seems to naturally bring disciplines together. In this sense, it will be more advantageous to find similarities in approaches than to accentuate differences. Having outlined the coherency in themes and methods underlying this thesis, I will now bring it to a close in the next chapter.

9. Conclusions

This is not the end. The is not even the beginning of the end. But it is, perhaps, the end of the beginning.

—Winston Churchill

In this thesis, we have witnessed an industry in transition. Maritime transportation as we know it is changing rapidly in the wake of revolutions in computational power and techniques for simulating intelligence. However, the inescapable complexities of the world we inhabit and the need for a responsible agent in safety-critical contexts precludes such computational intelligence to act alone in the world. For autonomous vessels, the case was made in this thesis that we need supervised autonomy provided by operators working at a remote control center.

In my investigations of such a control center, which I dutifully reported in the thesis as well as in the four appended research articles, I picked and chose from a dizzying array of methodologies and approaches to help answer relevant research questions. Soon after, I made some initial discoveries. I discovered that human-AI interaction represents a rapidly growing and multidisciplinary field, even in the niche arena of maritime transportation. I discovered that risk-based design was leading the charge in the field, backed by methodologies that assiduously formalized "best guess" approaches in the face of little data, including System-Theoretic Process Analysis (STPA) and Bayesian Networks (BNs). Other discoveries shed light on the *role* of the control room operator. I found that canonical works in the field of CSCW, which ostensibly focus on automation applications, inevitably end up describing how humans supervise the automation and coordinate their actions to effectively manage a complex, largely unpredictable system. In building a full-scale prototype of a remote control room in the form of the NTNU Shore Control Lab, I discovered that control rooms are a living entity without a clear developmental start or end point, whose design needs meet at the crossroads of many disciplines. Through my experimental efforts, I made some concrete discoveries I could express in the crisp clarity of numbers. To take one such example, I found that the average response time during handover events (interventions involving a machine-to-human transfer of control) was approximately 7 seconds when supervising a single vessel. When supervising three vessels, though, this response time increased to 10 seconds. In other words, I determined the effect of multitasking on intervention response time. Besides multitasking, I also carefully investigated and formally measured the effects of other relevant factors on several performance metrics. We know now, for instance, that multitasking as well as time pressure and the availability of decision support have profound effects on response time and on performance in general during supervisory control tasks. We are also just beginning to understand these factors' influence on stress and cognitive load during control interventions.

This thesis has also awarded me the opportunity to reflect on the research process. Foremost among these reflections is my learning to theorize. That is, from my PhD I learned the art of generating hypotheses: of abduction that blurs the lines between logical deduction and subjective inference. Theorizing led me to challenge the status quo and to produce new ideas. I challenged the notion that a backup format of supervisory control is well-aligned with operators' expectations, arguing that we need a more engaging, more meaningful orientation of what it means to be "in the loop." I challenged the notion that supervisory controllers need to be licensed mariners, something that will no doubt be controversial. This idea stemmed from the experimental results, in which I showed that experienced navigators were no better, but also no worse, than video game players during simulated remote control tasks. I also challenged the notion that human-AI collaboration represents a zero-sum, as it is so often framed in conventional design approaches that presume a trade-off in human autonomy for increases in machine autonomy.

The abductive approach to hypothesis generation and testing also showed me that we interact with the world even when making ostensibly objective observations. This struck me as similar to how designers work, guided as they are by a well-trained faculty for guesswork and an unabashed willingness to put *themselves*, rather than some strict data collection method, at the center of the design process. It also struck me as similar to my observations of how my test participants made decisions under conditions of uncertainly during supervisory control tasks. Operators' interactions with the world could not be fully explained by cognitive processes that sought representations of the world outside us, nor could they be fully explained by human factors or human error that sought representations of the world inside us. Rather, operators' interactions could only be explained as a mix of the two, as situated actions or embodied cognition, whose perceptions are as much a product of interaction as the source of it. My treatment of inattentional blindness in supervisory control was a compelling example of this: we may look, but to what extent do we *see*?

The methods I adapted in this thesis, although diverse and multi-purpose, were, in the clarity of hindsight, more similar that divergent. This lent a cohesive element that ran through the work as a whole. Similarities in the diverse methods I used called attention to overarching patterns in the patchwork approach inevitable in a PhD journey. Human-centered approaches to design, for example, proved similar in their iterative and "best guess" nature to popular risk assessment methods like STPA and BNs. The concepts of naturalistic decision making, situated action, embodied cognition, and Bayesian statistics were also all suggestive of interactions that inextricably connected perception and action, rather than severing them for reductionist treatments.

I also learned that attention to one's own biases is part-and-parcel of abductive-oriented theorizing. With my overall research question as my guide, I made sure to involve a variety of research methods that complemented one another, rather than pigeonholed some particular aspect of the question. Field studies "in the wild," for instance, found their complement in laboratory studies with strictly controlled conditions. My literature review, with its lack of original, material results, found its complement in a design process with the innovative and tangible result of a full-scale prototype. To mitigate bias, I used triangulation in literature review and interview analyses, I used collaborators in design workshops and empirical experimentation. Moreover, to make my experimental work transparent, I published all the data that was used to generate the findings presented in this thesis. To make my experimental work repeatable, I made the simulator platform openly available.

This work has presented an exploration of a topic riddled with challenges and pitfalls. It seems one cannot have a conversation about remote control center design without using concepts that lack precise definitions, including concepts like AI, situation awareness, and attention. Even my chosen field of design can spark debate among its practitioners when asked to proffer a definition. What's more, seeming commonplace terms like skill, boredom, vigilance, and decision support revealed a minefield of tautological issues.

The key to it all was to accept that any multidisciplinary field dealing with complex issues is bound to meet terms that defy definition. It is the domain of the natural sciences to reduce such concepts to explanatory theory. It is the role of designer, on the other hand, to *expand* these concepts into working solutions. The best a designer can offer is a *preferred* situation; one that nudges towards an improvement on an original, one that shepherds concepts towards the tangible in the face of uncertainty, one that converges disparate disciplines under a common understanding.

In this field under transition, we can at least be confident that we are oriented in the right direction, ready to take the next steps forward.

References

"AI Ethics," May 22, 2023. https://www.ibm.com/artificial-intelligence/ethics.

Amershi, Saleema, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, et al. "Guidelines for Human-AI Interaction." In *Proc. of the 2019 CHI Conf. on Human Factors in Computing Systems (Glasgow)*, 1–13. CHI '19. New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3290605.3300233.

Amro, Ahmed. "Communication and Cybersecurity for Autonomous Passenger Ferry." PhD Thesis, NTNU, 2023. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3064022.

Amro, Ahmed, Vasileios Gkioulos, and Sokratis Katsikas. "Communication Architecture for Autonomous Passenger Ship." *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2021, 1748006X211002546.

Bainbridge, Lisanne. "Ironies of Automation." *Automatica* 19, no. 6 (1983): 775–79. https://doi.org/10.1016/0005-1098(83)90046-8.

Bentley, Richard, John A Hughes, David Randall, Tom Rodden, Peter Sawyer, Dan Shapiro, and Ian Sommerville. "Ethnographically-Informed Systems Design for Air Traffic Control," 123–29, 1992.

Brekke, Edmund F, Egil Eide, Bjørn-Olav H Eriksen, Erik F Wilthil, Morten Breivik, Even Skjellaug, Øystein K Helgesen, et al. "MilliAmpere: An Autonomous Ferry Prototype." *Journal of Physics: Conference Series* 2311, no. 1 (July 1, 2022): 012029. https://doi.org/10.1088/1742-6596/2311/1/012029.

Chang, Chia-Hsun, Christos Kontovas, Qing Yu, and Zaili Yang. "Risk Assessment of the Operations of Maritime Autonomous Surface Ships." *Reliability Engineering & System Safety* 207 (2020): 107324.

Cheng, Tingting, Erik Veitch, Ingrid B. Utne, Marilia Abilio Ramos, Ali Mosleh, Ole Andreas Alsos, and Bing Wu. "Analysis of Human Errors in Human-Autonomy Collaboration in Autonomous Ships Operations through Shore Control Experimental Data (In Review)," 2023.

Christian, Brian. *The Alignment Problem: Machine Learning and Human Values*. 1st edition. New York, NY, USA: W. W. Norton & Company, 2020.

Colby, Kenneth Mark, James B. Watt, and John P. Gilbert. "A Computer Method of Psychotherapy: Preliminary Communication." *The Journal of Nervous and Mental Disease* 142, no. 2 (1966).

https://journals.lww.com/jonmd/Fulltext/1966/02000/A_COMPUTER_METHOD_OF_PSYCHOTHERAPY__PRELIMINARY.5.aspx.

Corbeil, R. R., and S. R. Searle. "Restricted Maximum Likelihood (REML) Estimation of Variance Components in the Mixed Model." *Technometrics* 18, no. 1 (February 1, 1976): 31–38. https://doi.org/10.1080/00401706.1976.10489397.

Corbin, Juliet, and Anselm Strauss. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Fourth Edition. Thousand Oaks, CA, USA: SAGE Publications, Inc., 2015.

Danielsen, Brit-Eli. "Usability in Ship Bridge Design - A Mission Impossible? A Qualitative Study of Maritime Stakeholders' Perspectives on Usability in Ship Bridge Design." NTNU, 2023. https://hdl.handle.net/11250/3069488.

"Design Expert Version 13.0.9.0." Minneapolis, MN, USA: Stat-Ease, Inc., December 10, 2021.

Drew, Trafton, Melissa L.-H. Võ, and Jeremy M. Wolfe. "The Invisible Gorilla Strikes Again: Sustained Inattentional Blindness in Expert Observers." *Psychological Science* 24, no. 9 (September 1, 2013): 1848–53. https://doi.org/10.1177/0956797613479386.

Dybvik, Henrikke, Erik Veitch, and Martin Steinert. "Exploring Challenges with Designing and Developing Shore Control Centers (SCC) for Autonomous Ships." *Proceedings of the Design Society: DESIGN Conference* 1 (2020): 847–56. https://doi.org/10.1017/dsd.2020.131.

Ek, Sondre. "Design of Simulator for Researching Autonomous Marine Vessels." Master thesis, NTNU, 2022. https://bibsys-almaprimo.hosted.exlibrisgroup.com/permalink/f/13q4kuj/BRAGE11250/2996832.

Elo, Arpad E. *The Rating of Chessplayers, Past and Present*. Second edition. New York, N.Y., USA: Arco Publishing, Inc., 1978.

Endsley, Mica R. "Situation Awareness Misconceptions and Misunderstandings." *Journal of Cognitive Engineering and Decision Making* 9, no. 1 (2015): 4–32.

———. "Toward a Theory of Situation Awareness in Dynamic Systems." *Human Factors* 37, no. 1 (1995): 32–64. https://doi.org/10.1518/001872095779049543.

Esteva, Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks." *Nature* 542, no. 7639 (February 1, 2017): 115–18. https://doi.org/10.1038/nature21056.

Federal Aviation Administration. "FAA Safety Briefing - May/June 2021." U.S. Department of Transportation: Washington, D.C., USA, May 1, 2021. https://www.faa.gov/newsroom/safety-briefing/archived-faa-safety-briefing-magazine-issues.

Graeber, David. *Bullshit Jobs: A Theory*. New York, NY, USA: Simon & Schuster, Inc., 2018.

Grier, R.A., J.S. Warm, W.N. Dember, G. Matthews, T.L. Galinsky, J.L. Szalma, and R. Parasuraman. "The Vigilance Decrement Reflects Limitations in Effortful Attention, Not Mindlessness." *Human Factors* 45, no. 3 (2003): 349–59. https://doi.org/10.1518/hfes.45.3.349.27253.

Hansen, Mikael. "Developing a Video Game for Research and Prototyping of Unmanned Maritime Vessels." Master thesis, NTNU, 2022. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3028969.

Heath, Christian, Jon Hindmarsh, and Paul Luff. "Analysing Video: Developing Preliminary Observations." *SAGE Visual Methods*, 2010, 365–89.

Heath, Christian, and Paul Luff. "Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms." *Computer Supported Cooperative Work (CSCW)* 1, no. 1–2 (1992): 69–94. https://doi.org/10.1007/BF00752451.

Hoem, Åsa. "Risk Assessment in the Design Phase of Maritime Autonomous Ships – A Human-Centered Approach." Doctoral Thesis, NTNU, 2023. https://hdl.handle.net/11250/3063036.

Hollnagel, Erik, and David D. Woods. "Cognitive Systems Engineering: New Wine in New Bottles." *International Journal of Man-Machine Studies* 18, no. 6 (June 1, 1983): 583–600. https://doi.org/10.1016/S0020-7373(83)80034-0.

Hollnagel, Erik, David D Woods, and Nancy Leveson. *Resilience Engineering: Concepts and Precepts*. Ashgate Publishing, Ltd., 2006.

Hutchins, Edwin. *Cognition in the Wild*. MIT press, 1995.

IDEO. *The Field Guide to Human-Centered Design: Design Kit*. San Francisco, CA, USA: IDEO.org, 2015.

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

———. "Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS)." London, UK: International Maritime Organization, June 3, 2021. https://www.imo.org/en/MediaCentre/PressBriefings/pages/MASSRSE2021.aspx.

ISO. "NS-EN ISO 9241-210:2019, Ergonomics of Human-System Interaction — Part 210: Human-Centred Design for Interactive Systems." Standard Norge, September 1, 2019.

Kahneman, Daniel. *Thinking, Fast and Slow*. Farrar, 2000.

Katsivela, Marel. "Unmanned Vessels and Regulatory Concerns." *Journal of International Maritime Law* 26 (2020): 239–52.

Khan, Faisal, Rocky Taylor, Brian Veitch, Erik Veitch, and Doug Smith. "Visualizing and Understanding the Operational Dynamics of a Shipping Operation." Providence, RI, USA: Society of Naval Architects and Marine Engineers, 2018.

Klein, Gary A. *Sources of Power: How People Make Decisions*. MIT press, 2017.

Koch, Christof. *Consciousness: Confessions of a Romantic Reductionist*. MIT press, 2012.

———. *The Feeling of Life Itself: Why Consciousness Is Widespread but Can't Be Computed*. Mit Press, 2019.

Krange, Olve, and Ketil Skogen. "When the Lads Go Hunting: The 'Hammertown Mechanism' and the Conflict over Wolves in Norway." *Ethnography* 12, no. 4 (December 1, 2011): 466–89. https://doi.org/10.1177/1466138110397227.

Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Fourth edition. Chicago, IL, USA: University of Chicago Press, 2012.

Lakoff, George, and Mark Johnson. *Metaphors We Live By*. London: University of Chicago Press, 2003.

Leveson, Nancy G. *Engineering a Safer World : Systems Thinking Applied to Safety*. Cambridge, MA, USA: The MIT Press, 2016.

Lin, Jinchao, Ryan Wohleber, Gerald Matthews, Peter Chiu, Gloria Calhoun, Heath Ruff, and Gregory Funke. "Video Game Experience and Gender as Predictors of Performance and Stress during Supervisory Control of Multiple Unmanned Aerial Vehicles." In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59:746–50. SAGE Publications Sage CA: Los Angeles, CA, 2015.

Littman, Michael L., Ifeoma Ajunwa, Guy Berger, Craig Boutilier, Morgan Currie, Finale Doshi-Velez, Gillian Hadfield, et al. "Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report." Stanford, CA., USA: Stanford University, September 16, 2021. http://ai100.stanford.edu/2021-report.

Mack, Arien, and Irvin Rock. *Inattentional Blindness*. Cambridge, MA, US: The MIT Press, 1998.

Martin, Bella, and Bruce Hanington. *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions*. Beverly, MA, USA: Rockport Publishers, 2012.

Massterly. "What We Do." Making autonomy a reality, 2023. https://www.massterly.com/what-we-do.

McCarthy, John. "What Is Artificial Intelligence?" Stanford, CA. USA: Stanford University, November 12, 2017. http://jmc.stanford.edu/articles/whatisai/whatisai.pdf.

McGrayne, Sharon Bertsch. *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of Controversy*. Yale University Press, 2011.

McKinley, R. Andy, Lindsey K. McIntire, and Margaret A. Funke. "Operator Selection for Unmanned Aerial Systems: Comparing Video Game Players and Pilots." *Aviation, Space, and Environmental Medicine* 82, no. 6 (June 1, 2011): 635–42. https://doi.org/10.3357/ASEM.2958.2011.

Microsoft Research. "Guidelines for Human-AI Interaction." Accessed June 9, 2023. https://www.microsoft.com/en-us/research/project/guidelines-for-human-ai-interaction/.

Mitchell, Melanie. *Complexity: A Guided Tour*. Oxford university press, 2009.

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. "Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement." *Annals of Internal Medicine* 151, no. 4 (August 18, 2009): 264–69. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

Molloy, Robert, and Raja Parasuraman. "Monitoring an Automated System for a Single Failure: Vigilance and Task Complexity Effects." *Human Factors* 38, no. 2 (1996): 311–22. https://doi.org/10.1177/001872089606380211.

Montgomery, Douglas C. *Design and Analysis of Experiments*. Ninth edition. John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2017.

Mustvedt, Petter. "Autonom ferge designet for å frakte 12 passasjerer trygt over Nidelven." Master thesis, NTNU, 2019.

Nersessian, Nancy J. *Creating Scientific Concepts*. The MIT Press: Cambridge, MA, USA, 2010.

Norman, Don. *The Design of Everyday Things: Revised and Expanded Edition*. Basic books, 2013.

Norwegian Maritime Authority. "Forskrift om kvalifikasjoner og sertifikater for sjøfolk," 2011. https://www.sdir.no/sjofart/regelverk/rundskriv/kvalifikasjoner-og-sertifikater-for-sjofolk/.

Norwegian Safety Investigation Authority. "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." Lillestrøm, Norway: Norwegian Safety Investigation Authority, April 2021. https://www.nsia.no/Marine/Published-reports/2021-05-eng.

NTNU. "Autoferry," 2020. https://www.ntnu.edu/autoferry.

Oliva, Aude, and Antonio Torralba. "Building the Gist of a Scene: The Role of Global Image Features in Recognition." In *Progress in Brain Research*, edited by S. Martinez-Conde, S.L. Macknik, L.M. Martinez, J.-M. Alonso, and P.U. Tse, 155:23–36. Elsevier, 2006. https://doi.org/10.1016/S0079-6123(06)55002-2.

"People + AI Research." Accessed June 9, 2023. https://pair.withgoogle.com.

Ramos, Marilia A., Christoph A. Thieme, Ingrid B. Utne, and Ali Mosleh. "Human-System Concurrent Task Analysis for Maritime Autonomous Surface Ship Operation and Safety." *Reliability Engineering & System Safety* 195 (2020): 106697. https://doi.org/10.1016/j.ress.2019.106697.

Rødseth, Ørnulf Jan, Lars Andreas Lien Wennersberg, and Håvard Nordahl. "Towards Approval of Autonomous Ship Systems by Their Operational Envelope." *Journal of Marine Science and Technology* 27 (2022): 67–76. https://doi.org/10.1007/s00773-021-00815-z.

Rubinstein, Ellis, and John F. Mason. "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, no. 11 (November 1979): 33–57. https://doi.org/10.1109/MSPEC.1979.6368289.

Seth, Anil. *Being You: A New Science of Consciousness*. Penguin, 2021.

Setz, Cornelia, Bert Arnrich, John Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. "Discriminating Stress From Cognitive Load Using a Wearable EDA Device." *IEEE Transactions on Information Technology in Biomedicine* 14, no. 2 (March 2010): 410–17. https://doi.org/10.1109/TITB.2009.2036164.

Shannon, Claude E. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27, no. 3 (1948): 379–423.

Sheridan, Thomas B. *Telerobotics, Automation, and Human Supervisory Control*. MIT press, 1992.

Shneiderman, Ben. *Human-Centered AI*. Oxford University Press: Oxford, UK, 2022.

———. "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy." *International Journal of Human–Computer Interaction* 36, no. 6 (2020): 495–504. https://doi.org/10.1080/10447318.2020.1741118.

Simon, Herbert A. *The Sciences of the Artificial*. Third edition. Cambridge, MA, USA: MIT Press, 1996.

Simons, Daniel J, and Christopher F Chabris. "Gorillas in Our Midst: Sustained Inattentional Blindness for Dynamic Events." *Perception* 28, no. 9 (1999): 1059–74.

Smidts, C., S.H. Shen, and A. Mosleh. "The IDA Cognitive Model for the Analysis of Nuclear Power Plant Operator Response under Accident Conditions. Part I: Problem Solving and Decision Making Model." *Reliability Engineering & System Safety* 55, no. 1 (1997): 51–71. https://doi.org/10.1016/S0951-8320(96)00104-4.

Strevens, Michael. *The Knowledge Machine: How Irrationality Created Modern Science*. Audiobook. Prince Frederick, MD, USA: HighBridge, a division of Recorded Books, 2020.

Suchman, Lucy. "Centers of Coordination: A Case and Some Themes." In *Discourse, Tools and Reasoning: Essays on Situated Cognition*, edited by Lauren B. Resnick, Roger Säljö, Clotilde Pontecorvo, and Barbara Burge, 41–62. Berlin, Heidelberg: Springer Berlin Heidelberg, 1997. https://doi.org/10.1007/978-3-662-03362-3_3.

Suchman, Lucy A. *Human-Machine Reconfigurations: Plans and Situated Actions*. Second Edition. New York, N.Y., USA: Cambridge University Press, 2007.

Swedberg, Richard. *The Art of Social Theory*. Princeton, NJ, USA: Princeton University Press, 2014.

Tilburg, Wijnand A. P. van, and Eric R. Igou. "On Boredom: Lack of Challenge and Meaning as Distinct Boredom Experiences." *Motivation and Emotion* 36, no. 2 (June 1, 2012): 181–94. https://doi.org/10.1007/s11031-011-9234-9.

Tononi, Giulio, and Gerald M. Edelman. "Consciousness and Complexity." *Science* 282, no. 5395 (December 4, 1998): 1846–51. https://doi.org/10.1126/science.282.5395.1846.

Varela, Francisco J, Evan Thompson, and Eleanor Rosch. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT press, 2016.

Vasstein, Kjetil, Edmund Førland Brekke, Rudolf Mester, and Eide Eide. "Autoferry Gemini: A Real-Time Simulation Platform for Electromagnetic Radiation Sensors on Autonomous Ships." *IOP Conference Series: Materials Science and Engineering* 929 (November 27, 2020): 012032. https://doi.org/10.1088/1757-899x/929/1/012032.

Veitch, Erik. "Dataset for: Human Factor Influences on Supervisory Control of Remotely Operated and Autonomous Vessels." DataverseNO, March 14, 2022. https://doi.org/10.18710/WYFMMP.

———. "Dataset for: Human Factor Influences on Supervisory Control of Remotely Operated and Autonomous Vessels (In Review)." DataverseNO, 2023. https://doi.org/10.18710/WYFMMP.

Veitch, Erik, and Ole Andreas Alsos. "A Systematic Review of Human-AI Interaction in Autonomous Ship Systems." *Safety Science* 152 (August 1, 2022): 105778. https://doi.org/10.1016/j.ssci.2022.105778.

———. "Gorillas on a Boat: Inattentional Blindness during Supervisory Control of Autonomous Vessels (In Press)," 2023.

Veitch, Erik, Ole Andreas Alsos, Tingting Cheng, Kristin Senderud, and Ingrid Bouwer Utne. "Human Factor Influences on Supervisory Control of Remotely Operated and Autonomous Vessels (Preprint)." SSRN, April 17, 2023. https://dx.doi.org/10.2139/ssrn.4437731.

Veitch, Erik, Kim A Christensen, Markus Log, Erik T Valestrand, Sigurd H Lundheim, Martin Nesse, Ole A Alsos, and Martin Steinert. "From Captain to Button-Presser: Operators' Perspectives on Navigating Highly Automated Ferries." *Journal of Physics: Conference Series* 2311 (2022): 012028. https://doi.org/10.1088/1742-6596/2311/1/012028.

Veitch, Erik, Henrikke Dybvik, Martin Steinert, and Ole Andreas Alsos. "Collaborative Work with Highly Automated Marine Navigation Systems." *Computer Supported Cooperative Work (CSCW)*, October 8, 2022. https://doi.org/10.1007/s10606-022-09450-7.

Veitch, Erik, Thomas Kaland, and Ole Andreas Alsos. "Design for Resilient Human-System Interaction in Autonomy: The Case of a Shore Control Centre for Unmanned Ships." *Proceedings of the Design Society* 1 (2021): 1023–32. https://doi.org/10.1017/pds.2021.102.

Weizenbaum, Joseph. "ELIZA—a Computer Program for the Study of Natural Language Communication between Man and Machine." *Communications of the ACM* 9, no. 1 (1966): 36–45.

Wellock, Thomas R. *Safe Enough?: A History of Nuclear Power and Accident Risk*. University of California Press, 2021.

Whalen, Jack. "A Technology of Order Production: Computer-Aided Dispatch in Public Safety Communications." *Situated Order: Studies in the Social Organization of Talk and Embodied Activities*, 1995, 187–230.

Wickens, Christopher D, Justin G Hollands, Simon Banbury, and Raja Parasuraman. *Engineering Psychology and Human Performance*. Psychology Press, 2015.

Woebot Health. "Woebot Health - About Us." Accessed June 9, 2023. https://woebothealth.com/about-us/.

Wróbel, Krzysztof. "Searching for the Origins of the Myth: 80% Human Error Impact on Maritime Safety." *Reliability Engineering & System Safety*, 2021, 107942.

Part II

Research papers

Article 1

Safety Science 152 (2022) 105778

Contents lists available at ScienceDirect

Safety Science

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

A systematic review of human-AI interaction in autonomous ship systems

Erik Veitch *, Ole Andreas Alsos

NTNU, Department of Design, Kolbjørn Hejes Vei 2b, 7491 Trondheim, Norway

ARTICLE INFO

Keywords:
Automation
Artificial Intelligence
Work
Safety
Marine Navigation
Human-Computer Interaction
Safety management
Resilience Engineering
Interaction Design
Maritime Autonomous Surface Ships
STPA
Bayesian Networks

ABSTRACT

Automation is increasing in shipping. Advancements in Artificial Intelligence (AI) applications like collision avoidance and computer vision have the potential to augment or take over the roles of ship navigators. However, implementation of AI technologies may also jeopardize safety if done in a way that reduces human control. In this systematic review, we included 42 studies about human supervision and control of autonomous ships. We addressed three research questions (a) how is human control currently being adopted in autonomous ship systems? (b) what methods, approaches, and theories are being used to address safety concerns and design challenges? and (c) what research gaps, regulatory obstacles, and technical shortcomings represent the most significant barriers to their implementation? We found that (1) human operators have an active role in ensuring autonomous ship safety above and beyond a backup role, (2) System-Theoretic Process Analysis and Bayesian Networks are the most common risk assessment tools in risk-based design, and (3) the new role of shore control center operators will require new competencies and training. The field of autonomous ship research is growing quickly. New risks are emerging from increasing interaction with AI systems in safety–critical systems, underscoring new research questions. Effective human-AI interaction design is predicated on increased cross-disciplinary efforts, requiring reconciling productivity with safety (resilience), technical limitations with human abilities and expectations (interaction design), and machine task autonomy with human supervisory control (safety management).

1. Introduction

Artificial Intelligence (AI) and automation have the potential to improve safety in complex transportation systems. Unlocking that potential, though, appears to depend on a seamless integration of human and machine control, combined with well-executed strategies for managing risks in a constantly changing environment. Maritime Autonomous Surface Ships (MASSs) have in recent years emerged as a new application of vehicle automation, in turn presenting new challenges and a productive research community. In the backdrop of rapid technology development, research contributions about MASSs have come from disparate fields, including risk and safety science, human factors, policy, and engineering. Currently, there is no comprehensive review available to serve as a guidance across the multiple disciplines, despite what unites them all as a common goal: the vision of safe, effective MASS operations. The objective in this study is to systematically review the different disciplines contributing to MASS research. By identifying research themes and directions of current work, our motivation is to promote cross-pollination of research ideas and multi-disciplinary efforts as the field converges towards its common goal. Our aim in analyzing the research state-of-the-art is to contribute to a better understanding of current design frameworks for human-AI system integration. We also aim to synthesize the methods risk scientists are using to respond to the need to predict behavior of such systems in the face of considerable uncertainly and complexity.

Despite continually improving safety records, shipping is considered a dangerous industry with a high rate of fatal injuries and high consequences of maritime disasters (Hansen et al., 2002; Hetherington et al., 2006). The International Maritime Organization (IMO), the United Nations agency regulating international maritime safety, calls it “perhaps the most international of all the world’s great industries - and one of the most dangerous” (IMO, 2019). Investigations of underlying causes for marine accidents tend to point to “human error” as the single greatest contributor, by some estimates involved in 75–96% of all accidents (Rothblum, 2000). Proponents of automation, harkening its precision and untiring capabilities, have long promised an end to human errors. And yet the statistics appear unchanged: one recent safety report stated that the same “75% to 96%” of marine incidents in 2020 were at least

* Corresponding author.
E-mail addresses: erik.a.veitch@ntnu.no (E. Veitch), oleanda@ntnu.no (O. Andreas Alsos).

https://doi.org/10.1016/j.ssci.2022.105778
Received 1 May 2021; Received in revised form 29 March 2022; Accepted 4 April 2022
Available online 12 April 2022

partly due to human error (Allianz, 2020). The tendency for automation applications to introduce new modes of error, even while effectively addressing known ones, has been well documented, including in the case of ship navigation (Lützhöft and Dekker, 2002). The risk picture emerging in the wake of MASS development remains unclear, hanging in the balance as new risks latent in human-AI interactions appear only partly offset by reduced instances of human error.

While it is difficult to measure the direct effect of automation on overall safety, comparisons of yearly safety reports show that safety at sea is generally improving (Allianz, 2020). Automation in shipping has incrementally increased since the transition of sails to engines and manual rudder control to powered steering. Today, “autopilot” system controls analogous to those on cars are commonplace on ships, as is Dynamic Positioning (DP) control in vessels like ferries, offshore supply vessels, survey ships, cable layers, and drill ships that require precise maneuvering. While automation and safety seem to advance in parallel, it is unclear to what extent the two are linked, and if so by what mechanisms.

At the start of the 2010s, a shift was observed towards digital transformation of the maritime industry, captured by the moniker “Maritime 4.0” and referring especially to automated integration of real-time data into decision making (Sullivan et al., 2020). The first major project about autonomous ships was completed in the years 2012–2015, called the MUNIN project (Burmeister et al., 2014). In 2017, two Norwegian companies Yara and Kongsberg announced plans to develop the Yara Birkeland, an autonomous cargo carrier servicing three ports in Southern Norway (Yara, 2017). Today, the project is still a forerunner, with plans for “fully autonomous operation” slated for 2022 (Kongsberg, 2020a). Recent innovations have demonstrated so-called “auto-crossing” and “auto-docking” in several ferry applications (Kongsberg, 2020b; Rolls-Royce, 2018). Collectively, these technologies effectively handle entire crossings between terminals automatically, relying on bridge crew to intervene to avoid possible collisions.

Today, AI is heralding the next level of automation in transportation applications, and it is the most recent harbinger of improved system safety addressing the ubiquitous “human error” problem. Deep learning has been demonstrated in applications ranging from obstacle detection (Chen et al., 2020), obstacle avoidance (Cheng and Zhang, 2018), and mooring line fault detection (Chung et al., 2020). As a research field, AI has been characterized by periods of rapid advancement and abundant research funding (so-called “AI Springs”) followed by periods of slower progress (“AI Winters”). At the start of the 2010s, new breakthroughs in AI, made possible in part by increased computational power in Graphics Processing Units (GPUs), marked a new AI Spring. In the marine industry, it also marked first ever serious attempts at developing autonomous ships. Since 2018, MASS have been formally recognized by IMO, the agency of the United Nations promoting global maritime safety. Foremost on their agenda for MASS is meeting the growing need to reconcile perceived benefits of MASS implementation with growing concerns for safety (IMO, 2018).

1.1. The concept of shore-based control for unmanned ships

The future of MASS operations will aim at reducing the crew onboard and increasing the land-based coordination and control. The concept of the Shore Control Center (SCC) (also known as the Remote Control Center or Remote Operation Center) meets a growing need for centralized coordination with the tasks of monitoring, supervision, and intervention of MASS fleet operations. The scope of MASS operations for this paper includes open-ocean, short-sea, inland, urban, and mission-oriented operations, each of which set unique constraints for navigation, regulation, and safety management. Open-ocean MASS operations involve generally non-demanding navigation (e.g., Burmeister et al., 2014) whereas short-sea shipping routes involve trafficked shipping lanes containing aids to navigation and established Vessel Traffic Services (VTS) coordination (e.g., Lunde-Hanssen et al., 2020; NTNU, 2021). Inland applications apply to inland cargo vessels in inland waterways (e.g., Peeters et al., 2020a); urban applications apply to small vessels that transport passengers or goods in urban canals or waterways (e.g., MiT, 2020; Reddy et al., 2019; Valdez Banda et al., 2019; Wang et al., 2019), and data-collection Autonomous Surface Vessels (ASVs) are used in scientific research and exploratory field work (e.g., Dallolio et al., 2019; Dunbabin et al., 2009; Kimball et al., 2014). While we can conceptualize an SCC by its execution of land-based operational tasks like supervision, monitoring, and control intervention, the work undertaken by its operators will vary considerably across the range of MASS applications. For example, short-sea shipping will set more demands on navigation than open-ocean applications; urban passenger transport will set more demands on safety management that mission-oriented ASVs whose payload is just measurement equipment and data.

1.2. Definition of AI and applications to marine operations

AI has no formal definition. The first informal definition can be traced to a workshop on AI held in 1955 at Dartmouth College, commonly held to be the origin of the field. The concept was based on the “conjecture that every aspect of learning or any other feature of intelligence can be in principle be so precisely described that a machine can be made to simulate it (McCarthy et al., 2006).” This description largely holds true today, and many of the original topics under investigation at the first workshop on AI, like natural-language processing, neural networks, machine learning, and reasoning, still serve as guiding beacons for contemporary AI researchers and developers. The absence of a formal definition is not necessarily a hindrance. In a recent publication about the research state-of-the-art in AI, a committee of leading researchers stated that “the lack of a precise, universally accepted definition of AI probably has helped the field to grow, blossom, and advance at an ever-accelerating pace (Stone et al., 2016, pp. 12).” In this paper, we will examine AI applications for collision avoidance and computer vision used in the context of marine navigation. Examination of technology aspects are topical; the focus is on their implications for hazard identification, risk assessment, and human–computer interaction, with special attention afforded to the relevant theories and methods used across disciplines.

1.3. AI brittleness and the need for “humans-in-the-loop”

While there is significant potential for AI to improve safety by reducing “human errors” in marine operations, there are also many aspects of the technology that suggest new hazards. For example, computer vision based on deep learning is prone to so-called “adversarial attacks,” whereby strategically modified images, often imperceptible to the human eye, deceive the algorithms and produce egregious errors (Akhtar and Mian, 2018). Real-world examples of adversarial attacks have been explored for autonomous cars (Eykholt et al., 2018) but not yet for computer vision in ships. A related problem is that of so-called “tail effects,” referring to the occurrence of low-probability events that are impractical or even impossible to train as inputs into a deep learning training dataset. This introduces a subversive “tail risk;” namely, the risk arising when specific tasks like navigation are performed in an unfamiliar environment or under new conditions.

The brittleness of AI systems underscores the necessity of “humans-in-the-loop.” Human operators, especially when experienced, have the remarkable ability to integrate information in unfamiliar contexts and with apparently little information (Klein, 2017). This can be thought of as the complement of AI systems, whose decision making accuracy hinges on how similar the data is to its training set. In the much-anticipated “Outcome of the regulatory scoping exercise for the use of Maritime Autonomous Surface Ships (MASS),” the IMO listed, as one of its “high-priority issues,” the “Remote control station/centre.” This format of operating a MASS, was, they wrote, “a new concept to be implemented… and a common theme identified in several [IMO

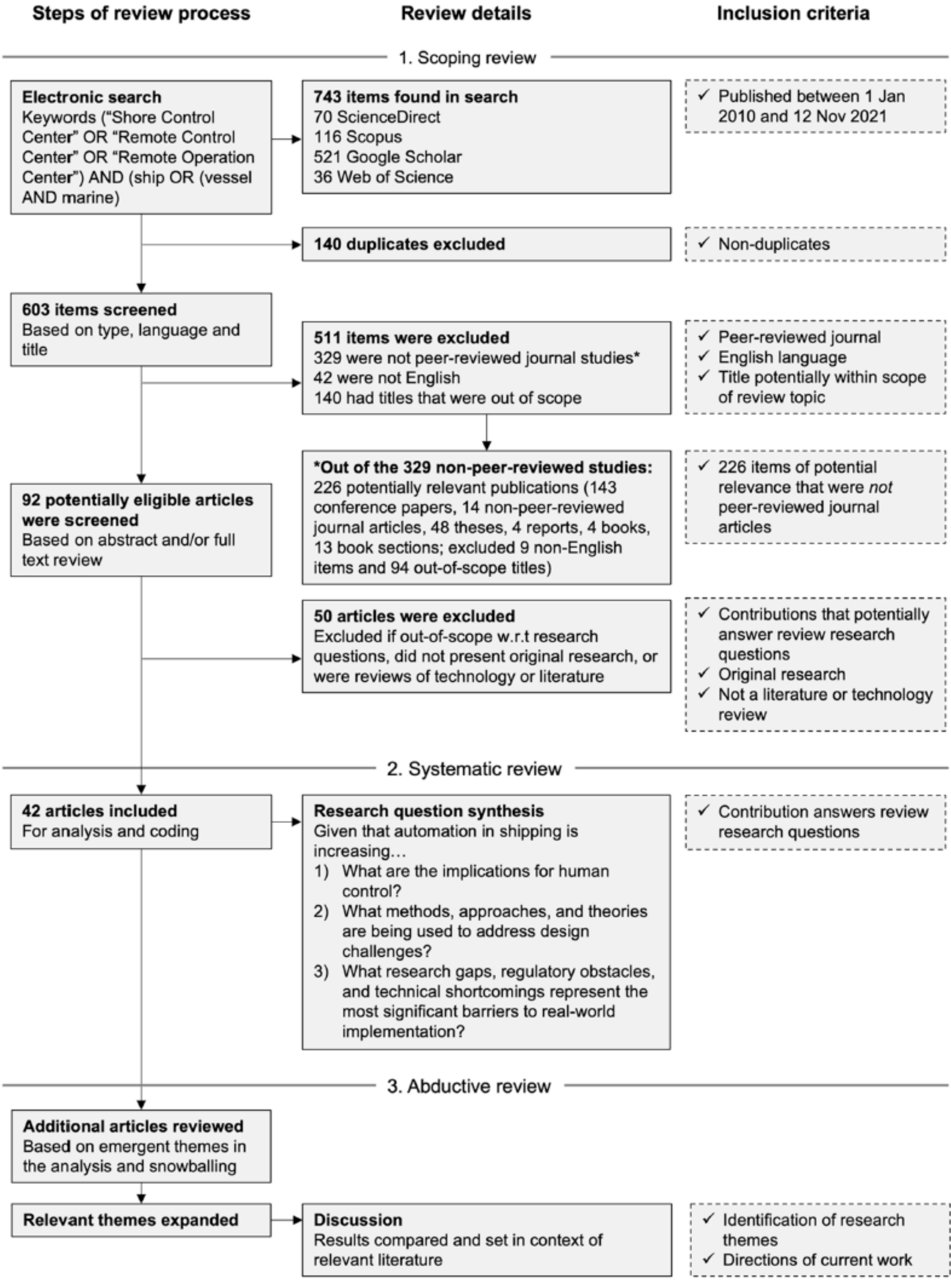


Fig. 1. Schematic showing article selection method.

Table 1
Results of systematic review.

Analysis criteria	Articles analyzed(ID can be matched to corresponding article using list in Appendix A)																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Research Question 1: How is human control being adopted?																					
Level of Autonomy taxonomies		●		●	●		●	●	●	●	●	●	●			●			●		●
Humans' roles in automated systems	●			●		●	●	●	●					●	●	●	●				●
Resilience Engineering			●														●			●	
Benefits of collaborative systems																				●	
Research Question 2: How to address design and safety?																					
Practical design approaches											●			●	●		●			●	●
Risk assessment methods																				●	
Human-computer interaction	●					●	●		●		●	●			●		●				
Empirical testing and prototypes	●			●	●		●	●	●	●	●			●			●		●		
Research Question 3: What are the outstanding knowledge gaps?																					
Risks in human-AI interaction		●		●	●	●			●		●	●		●							
Regulatory concerns		●		●	●			●			●		●		●	●					
Training for SCC operators		●	●	●	●		●	●		●			●								
Infrastructure for mixed traffic										●	●		●	●							

regulatory] instruments as a potential gap" (IMO, 2021, p. 8). Furthermore, it was outlined that the qualification, responsibility, and role of SCC operators were among the "most complex issues to be addressed." This points to the current need to articulate the concept of the SCC (called Remote control station/centre by the IMO). before researchers from diverse fields can collectively address the underlying research and design gaps.

1.4. Excessive automation and emerging risks

Recent investigations into high-profile accidents in transportation have identified interaction with highly automated systems as a root cause. Following two Boeing 737 MAX crashes in the years 2018–2019, it was revealed that the Maneuvering Characteristics Augmentation System (MCAS), designed to automatically control pitch based on sensor readings of air flow, malfunctioned without the pilots being aware that the controller existed, leading to uncontrolled dipping of the nose to counteract a non-existent stall (Nicas et al., 2019). In another accident investigation, the Tesla "Autopilot" function came under scrutiny by the National Transportation Safety Board, which surmised that "system limitations" in combination with "ineffective monitoring of driver engagement, which facilitated the driver's complacency and inattentiveness" were the probable causes of a fatal accident in California (National Transportation Safety Board, 2020, pp. 58). The task of monitoring automated systems, whether in the role of a trained pilot or passenger, is well known to be ill-suited to humans, with associated risks compounded by the need to take over control quickly (Parasuraman and Manzey, 2010; Parasuraman and Riley, 1997). Both the Boeing and Tesla examples demonstrate how poor human-AI system interaction design can lead to unacceptably dire consequences by not bringing human-in-the-loop and resilience to the fore in system design. For applications in MASS operations, which also involve control intervention in safety–critical situations, similar underlying human-system integration issues are present. Some important distinctions do exist; for example, ships have considerably slower speed, reducing the time-criticality of takeovers and simplifying interface design (Vagia and Rødseth, 2019). However, ships are also considered "under-actuated," having fewer inputs than degrees of freedom (Reyhanoglu, 1997), and may have very high inertia, requiring that navigation decisions are taken well in advance of the vessel's actual response. Controlling safety in MASS operations within acceptable risk levels will involve addressing human-AI interaction as a potential source of error, especially for time- and safety–critical interventions.

1.5. Research questions and aims

Research and development in MASS operations has increased dramatically in the last decade. In the risk sciences, significant attention has been afforded to assessing different hazard identification and risk assessment techniques for MASS operations. Designers have promoted resilience in human–computer interaction; engineers have developed and tested novel technology, and much has been written about human factors in unmanned ships. Given the emerging need for a robust human-AI interface at the SCC, as well as the contributions from multiple fields of research and the long list of obstacles slowing progression towards implementation, we organized our systematic review around three

Articles analyzed(ID can be matched to corresponding article using list in Appendix A)																					Sum
22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	
●	●	●		●	●					●	●	●	●	●	●	●	●	●	●	●	*29*
●		●	●			●	●	●				●			●		●	●	●	●	*23*
		●	●	●	●	●			●			●		●	●		●				*13*
	●				●		●								●						*5*
●	●	●	●	●	●		●		●			●	●		●	●	●	●			*20*
●	●	●	●	●	●				●			●	●			●	●	●	●	●	*15*
●		●			●							●	●		●		●			●	*16*
	●								●				●					●			*15*
		●	●	●				●	●			●	●	●	●	●	●	●	●	●	*22*
	●				●					●											*11*
				●							●		●	●							*12*
●														●		●					*7*

distinct research questions. Namely, given that automation is increasing in shipping,

1. *how is human control currently being adopted in autonomous ship operations?*
2. *what methods, approaches, and theories are being used to address safety concerns and novel design challenges?*
3. *what research gaps, regulatory obstacles, and technical shortcomings represent the most significant barriers to real-world implementation?*

The review is organized as follows. Section 2 describes the review method. Section 3 describes the outcomes of the review as guided by the three research questions. Section 4 discusses the outcomes by comparing them to relevant theory and methods outside the review. Section 5 presents conclusions of the review and traces the direction of future work.

2. Method

Understanding the state-of-art within a research topic is a complex task. There are many methods available, ranging from traditional styles that adopt an inductive analysis approach and “snowball sampling” of relevant resources, to *meta*-analysis approaches that aim to synthesize with the help of statistical tools an entire field of research, often citing hundreds of studies. In this paper, we adopted a systematic, comprehensive review approach that can be considered somewhere between traditional and *meta*-analysis review approaches. The systematic review has three virtues: it is reproduceable, introduces little bias, and it is transparent. Systematic reviews come in no universal form, although one commonly used guideline called “Preferred reporting items for systematic reviews and *meta*-analyses” (PRISMA) calls for a degree of standardization among researchers (Moher et al., 2009). We adopted PRISMA guidelines in our review.

The scope of the review was guided by the need for state-of-the-art research, not simply a state-of-the-art in implementation. We searched specifically for research about remote and shore-based control of autonomous ships because it conceptually represented the integrated MASS system, including not just elements of automated navigation but also of human control and emergent safety properties. We chose our search keywords (“Shore Control Center” and various alternatives) because these keywords represented this integrated system emerging in the wake of MASS development. The concept is also used by different fields in different ways, agnostic to any one research discipline. Moreover, the concept is relevant to all the research questions we set out to investigate. The method is divided into three phases: a scoping review, a systematic review, and an abductive review (Fig. 1).

(i) **Scoping review:** electronic search on the research topic followed by a series of screenings. The screenings occurred in two stages:
   1. Initial screening: based on title, type, and language: only titles that reflected the scope of our research questions were included, as were peer-reviewed journals, and English-language studies.
   2. Abstract screening: based on consensus of the two authors who independently analyzed the abstracts of studies passing initial screening.

(ii) **Systematic review:** full-text analysis by first author and coding in text analysis software NVivo (NVivo, 2020) according to the three research questions and relevant bibliographic details.
(iii) **Abductive review:** comparison to literature outside of the analyzed studies and expansion of the topics analyzed in the results.

2.1. Scoping review

Four databases were searched: ScienceDirect, Scopus, Web of Science, and Google Scholar. This ensured a broad indexing of academic databases, although it also resulted in many duplicates. The date range was set to 1 Jan 2010 to 12 Nov 2021. The search was restricted to the keywords "Shore Control Centre," "Remote Operation Centre," and "Remote Control Centre," plus an additional phrase to narrow the scope to marine applications ("AND Ship OR Vessel AND Marine").

Note that using double quotations in search phrases (e.g., "Shore Control Centre") in ScienceDirect and Scopus specified a search with those words together, but also included variants on punctuation and regional spelling and plural forms as well as variations on capitalization. For example, "Shore-Control Centers" and "Shore Control Centre" appeared together in the same search. For Web of Science and Google Scholar, the regional spelling variants had to be specified using the search "Shore Control Centre" AND "Shore Control Center" (but maintained all same flexibility in plurals, hyphenation, and capitals as in the other search database rules).

The initial screening was straightforward: duplicates and non-English language contributions were removed as were any publications from non-peer-reviewed journals. Note that excluding all technical reports and conferences papers potentially excluded valuable contributions, presenting a limitation in our review method. On the other hand, this exclusion criterion ensured a consistent standard of research contributions in the analysis. Titles of all remaining peer-reviewed journal articles were screened for those that were obviously out-of-scope.

The second-stage screening of articles required analyzing their abstracts and reaching a consensus on whether they should proceed to full-text review. Analysis was conducted by the two authors. Eligibility was based on whether the study potentially contributed to either one of the

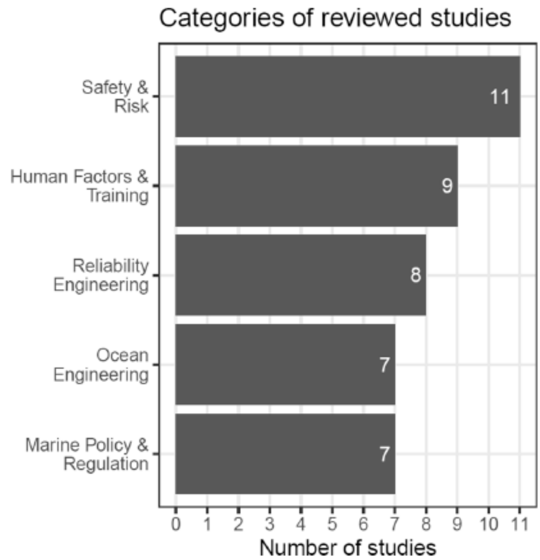


Fig. 3. Categories of all analyzed studies.

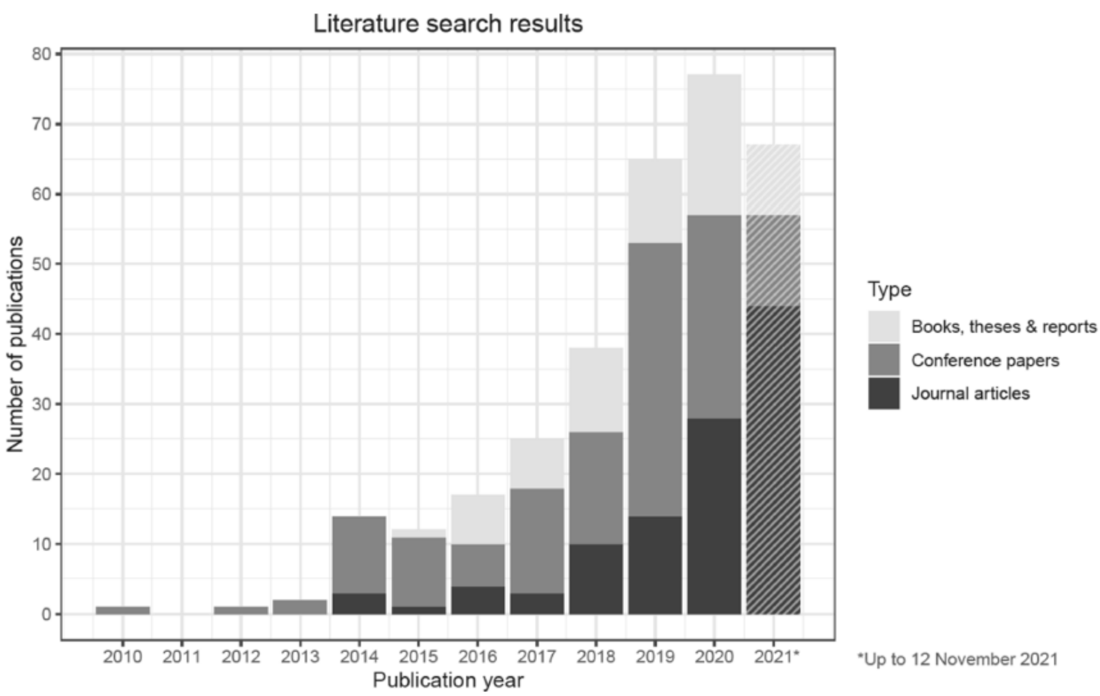


Fig. 2. Number of SCC studies published by year.

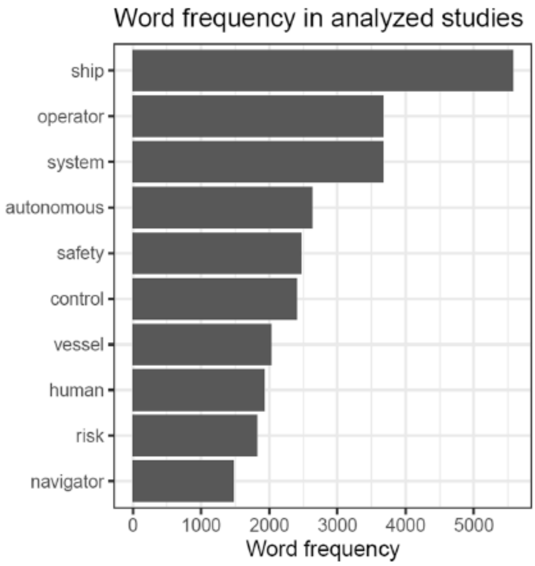


Fig. 4. Salient themes by word frequency in all analyzed studies.

three research questions defined in the Introduction. This was not always obvious from just the abstract; Grote et al., 2014 when in doubt, we included the study on the basis that the subsequent full-text analysis would confirm whether the study should be included. Also, only studies presenting original research were included, excluding opinion papers and review studies.

During the second-stage screening, we also made some additional scoping constraints to help converge our study subset. We briefly note these constraints here. We defined the scope to studies including any integrated system consisting of one or more MASSs and an SCC. From here on in we refer to such systems as just "MASS systems." The context was broad, including information needs, design approaches, systems design, complex socio-technical perspective, and elements of human–machine interaction related to remote control and intervention (from safety, risk, regulatory, technological, and maritime training perspectives) from shore-based infrastructure or from other surface vehicles. Importantly, this excluded work about UAVs and ROVs and papers focused on business cases or economics, cyber-security, law, cybernetics, insurance, or discussions about taxonomy for autonomous vessels. Additionally, we excluded works focused on just one component (e.g., communication or motion control) that only peripherally mentioned the human control and human interaction elements. The guiding principle was that the studies as a whole should reflect state-of-the-art research on MASS systems and not a state-of-the-art on its implementation.

2.2. Systematic review

Following the PRISMA guidelines, our main aim was to qualitatively synthesize the literature in the context of the defined specific research questions in a way that was reproduceable, unbiased, and transparent.

2.2.1. Review protocol, data extraction & coding

A review protocol specified how the studies were screened during full-text review against eligibility criteria and how information was extracted for answering the research questions.

Studies were screened for information that contributed to any one of the three defined research questions. This was done by coding the studies in the text analysis software NVivo (NVivo, 2020). Coding for the

Table 2
LoA taxonomies adopted in the reviewed studies.

LoA taxonomy	Description of maximum LoA for current operational concepts	Article ID
IMO (2018)	Remotely controlled ship without seafarers on board (Degree 3 out of 4): The ship is controlled and operated from another location. There are no seafarers on board.	[2, 4, 5, 8, 9, 10, 11, 13, 22, 23, 33, 35, 38, 39, 42]
NFAS (Rødseth, 2017)	Constrained autonomous (Level 7 out of 8): This is assumed to be a common mode for unmanned ships operating in relatively benign environments. The ship can operate fully automatic/autonomous in low traffic and non-restricted waters, until it sees problems that it cannot solve itself. Then it can call on shore control to resolve any problems. This reduces ship system complexity, while optimizing the benefit of having backup functions on shore.	[12, 19, 24, 37]
Utne et al. (2017)	Semi-autonomous operation or management by exception (Level 3 out of 4): The system automatically executes mission-related functions when response times are too short for human intervention. The human may override or change parameters and cancel or redirect actions within defined timelines. The operator's attention is only brought to exceptions for certain decisions (human-supervisory control).	[27, 34]
Lloyd's Register, (2016)	"Active" human in the loop (Autonomy Level 3 out of 6): Decisions and actions at the ship level are performed autonomously with human supervision. High impact decisions are implemented in a way to give human operators the opportunity to intercede and override them. Data may be provided by systems on or off the ship.	[26, 41]
Central Commission for Navigation on the Rhine (CCNR, 2021)	Conditional to High Automation (Level 3–4 out of 5): The sustained context-specific performance by a navigation automation system of all dynamic navigation tasks, including collision avoidance, with the expectation that the human helmsman will be receptive to requests to intervene and to system failures and will respond appropriately (no human intervention needed for High Automation). Applies to inland MASS only.	[16, 32, 40]
Combination / dynamic LoA	Shifts between generalized LoA levels, especially under "adaptive autonomy" as defined by Sheridan (2011) whereby "allocation of control function (to human or computer) changes with time to accommodate changes in the conditions of either the physical environment or the human" (p. 662).	[7, 21, 36]

three research questions also helped to identify likenesses among studies across more granular dimensions – dimensions that, when assigned a title, became the sub-categories that structured the presentation of the Results. There are, namely, four such sub-categories for each research question presented in the Results section.

The following information was coded during analyses of the studies:

(i) Research Question 1: How is human control currently being adopted?
   a. Level of Automation (LoA) taxonomies
   b. Humans' role in MASS operation
   c. Resilience Engineering
   d. Benefits of collaborative systems
(ii) Research Question 2: What methods, approaches, and theories are being used to address design challenges and improve safety?
   a. Practical design approaches
   b. Risk assessment methods
   c. Human-computer interaction paradigms
   d. Empirical testing and prototypes
(iii) Research Question 3: What research gaps, regulatory obstacles, and technical shortcomings represent the most significant barriers to implementation?
   a. Risk assessment methods
   b. Regulations
   c. Design for human-AI interaction
   d. Infrastructure to support mixed traffic

The coded data relevant to the research questions are summarized in Table 1.

Studies were also screened for bibliographic and descriptive details. The purpose of the former was to tabulate a list of all analyzed studies, which is included Appendix A. The purpose of the latter was to broadly characterize the contributing publications in terms of journal scope (which we sought to bin into categories) and in terms of salient themes (which could be represented by textual analysis of the most frequently used keywords). The following bibliographic and descriptive information was coded during analyses of the studies:

Table 3
Humans' roles in MASS system.

Category	Description	Article ID
Active	Continuous monitoring and decision making support: operator's role is active and engaged in all operational phases; the operator is in control of the ship, either through direct or indirect remote control. On-board roles were often specified, including emergency handling, active maintenance, lookout / watchkeeping, cargo loading and unloading.	[1, 7, 9, 15, 16, 17, 24, 29, 30, 34, 40, 42]
Backup	Monitoring and control intervention: operator's role is characterized as "backup" to the AI system; control interventions, or takeovers, can occur when the operator takes over control from the AI system, either on their own initiative or from the prompting of the AI system itself. Monitoring is mostly continuous; the operator is never far from the control position if left unattended; emphasis placed on timely emergency or contingency response.	[4, 6, 8, 21, 25, 37, 41]
Passive	Supervision and assistance: operator's role is characterized mainly by passive supervision; the operator can leave the control position and is alerted by the AI system if they are needed. Instead of handling situations as they arise, emphasis is on planning how to resolve situations before intervention is needed.	[14, 22, 28, 39]

(i) Bibliographic information (author, year, title, and journal name)
(ii) Descriptive information:
   a. Journal scope (Safety & Risk, Human Factors & Training, Reliability Engineering, Ocean Engineering, Marine Policy & Regulation)
   b. Salient themes (computed from keyword frequency).

2.3. Abductive review

The abductive review is presented in the Discussion. Here "abduction" refers to the process of forming an explanation for a phenomenon. It presupposes that the explanation we present are grounded at least partly in the works we analyzed (and partly in works outside the review) and do not simply represent our conjectures or guesswork. The term was coined by the philosopher Charles Peirce and has inspired interpretations for its use in modern science (for example, Paavola, 2006; Swedberg, 2014). We adopt the interpretation that forming an explanation is a principled process. We present our Discussion accordingly, in five sub-sections that expand upon central themes uncovered in the analysis. Inferences are drawn transparently to the Results as well as to literature outside the review's scope. This includes a diverse range of topics, including human-AI interaction for autonomous cars, ethnographic studies of control room work, the growing use of simulation in human-AI interaction studies, as well as the role of safety management in striking a balance between human and machine autonomy. In the abductive review we also reviewed studies cited in the system review in a so-called "snowballing" approach to scoping relevant literature.

3. Results

3.1. Electronic search

The electronic search resulted in 743 studies. This included, in respective order, 70, 116, 521, and 36 for ScienceDirect, Scopus, Google Scholar, and Web of Science. This included research and review articles, books, conference papers, and reports published from 1 Jan 2010 to 12 Nov 2021. After removing duplicate studies, 603 remained. Each of these was screened by title and by language, removing those studies that were clearly out-of-scope and those in a language other than English. The remainder was 318 studies, comprised of 92 peer-reviewed journal articles and 226 other publications (conference papers, books, theses, and reports). Fig. 2 presents a plot of these 318 studies organized by date and frequency of occurrence. It shows an abrupt start in the years 2013–2014, followed by steady publication rate through to 2017, after which publications increase markedly in 2018 through to 2020. The year 2021 saw a relative increase in the number of journal articles and decrease in amount of conference papers – likely a result of the COVID-19 pandemic. Note that the first publication describing the concept of a MASS system was in a 2010 paper by Im and Seo (2010), who described a "free running" or "automatically navigating" ship controlled from an SCC. The concept of "autonomous ships" with integrated control from a "shore control center" was first formally described by Rødseth et al. (2013).

Excluding all conference papers, non-peer-reviewed journal articles, theses, reports, books, and book sections, we were left with 92 potentially eligible peer-reviewed journal articles. After scanning the abstracts of these 92 articles, a further 50 were excluded. Note that it was not always obvious from the abstract whether a study should be included; in these cases, the full text was scanned, often after discussion between the co-authors until consensus was reached. Full-text analysis and coding was done for a total of 42 studies (Appendix A).

Categorizing the 92 articles by journal scope, we found that "Safety & Risk" and the related category of "Reliability Engineering" together accounted for 45% (n = 19) of the total studies analyzed (Fig. 3). A text frequency query, computed in NVivo (NVivo, 2020), showed that the word "safety" appeared among the top five most frequent words (excluding words with fewer than three letters and including stemmed words) (Fig. 4).

3.2. Research question 1: Given that automation is increasing in shipping, what are the implications for human control?

In this section, we explore the question of why human control is needed despite increasing automation. Drawing exclusively from the reviewed articles, we examine four relevant themes: (1) Level of Automation (LoA) taxonomies, (2) humans' roles in MASS systems, (3) Resilience Engineering (RE), and (4) the benefits of collaborative human-AI systems. Together, these four themes offer a cross-section of the reviewed studies that explain what at first glance seems to be a paradox: that designing for increased automation in shipping hinges upon design for human control.

3.2.1. Level of automation taxonomies

Most of the articles analyzed (twenty-nine out of forty-two), adopted a Level of Automation (LoA) taxonomy. LoAs are a way of describing how autonomous a system is by way of a linear integer scale, usually from 0 indicating full human control to some maximum number representing full machine autonomy. All articles adopted the position that full autonomy was not feasible in the scope of current research. That is, even though the terms "autonomous ship" was used, it was understood to be a "highly automated ship" involving some level of mixed human-AI control as defined by an LoA taxonomy. To illustrate this, Ramos et al. (2019, 2020b) express that full autonomy is "not expected in the near future" and Huang et al. (2020) take the stance that it is "still unrealistic." In Hannaford and Hassel (2021), the authors surveyed forty-two licensed seafarers on the matter, of whom two-thirds responded "it will not be safe to operate vessels in the future as fully autonomous" (p. 13). Regardless of what LoA taxonomy was being used, all studies positioned themselves one or two steps below the maximum number to show that the ship was highly autonomous, albeit not autonomous enough to exclude humans. Specifically, MASSs still relied on humans enough to merit the SCC in the system design.

By far the most commonly referenced LoA taxonomy in the papers analyzed was that proposed by IMO (2018). Fifteen out of the twenty-nine papers that positioned themselves within an LoA framework used this definition (Table 2). The second most commonly cited was NFAS (Rødseth, 2017), with four articles. Other LoAs used included one by Utne et al. (2017), Lloyd's Register of Shipping (Lloyd's Register, 2016), and the Central Commission for Navigation on the Rhine (CCNR, 2021) (the latter applying to inland MASS only). Three articles, while acknowledging the existence of various LoA taxonomies on offer, opted not to adhere to just one of strict definition. Instead, these studies generalized the MASS system as a combination of machine and human control, dynamically shifting with time. For example, Thieme et al., (2018) referenced several LoAs but settled on the general term "adaptive autonomy" put forward by Sheridan (2011) to avoid adherence to one particular LoA taxonomy. For a detailed review of LoA taxonomies, the reader is referred to Vagia et al. (2016).

3.2.2. Humans' roles in MASS systems

Just over half the analyzed papers (twenty-three out of forty-two) described specific roles human operators play in the MASS system. The studies were unanimous in their assertion that operators would be responsible for ensuring safety of the ship and any passengers onboard, and for ensuring protection of the surrounding environment. The ability for humans to oversee several ships at once was also described as a common feature of SCC work, as was the ability to plan voyages and take over control of the automated system.

Still, there were some notable differences in how humans' roles were described. In our review, we identified three broad categories of operator roles: (1) Active, (2) Backup, and (3) Passive. In Table 3 we characterize these three roles and list their corresponding studies. In approximately half the studies (twelve out of twenty-three), humans were identified as having an "active" role, characterized by "continuous monitoring and decision making support." The remaining role descriptions were underpinned by "backup" (seven studies) and "passive" roles (four studies). One common feature to all three roles, as noted in the studies, was their susceptibility to change, dependent as they were on technology development and operational elements (busier ports, for instance, require more active operator engagement). Still, each study could be mapped to just one of the three categories. This categorization was usually based on the case studies presented or on underlying assumptions about MASS system design. For example, Rødseth et al. (2021), while acknowledging the wide range of possibilities for human participation, narrow down the operators' primary role in the MASS system as one of backup to the automation. "It is expected that most autonomous ship systems will operate with continuous supervision from an RCC [Remote Control Center]," they write, specifying that operators need to "intervene when the automation is incapable of maintaining control" (p. 5). Wróbel et al. (2021), on the other hand, emphasize that a more active role related to diagnosing problems is needed to meet the expectation that problems can be "rectified" in a timely way. Meanwhile, others highlighted the need for planning and logistics, especially as more ships fall under command of a single SCC (e.g., Reddy et al., 2019).

Note that the categories below are independent of the studies' chosen LoA taxonomy. To illustrate this, some studies described operators' tasks as becoming more passive with higher automation (e.g., Zhou et al., 2021), while others argued that higher automation comes with the need for more active coordination among different roles (e.g., Huang et al., 2020; Relling et al., 2021).

Some studies stood out in their treatment of specific roles in the MASS system. Eriksen et al. (2021), for example, highlight that maintenance activities are critical to a ship's operability and rely on manual repair. Kooij and Hekkenberg (2021) also allude to the importance of repair activities onboard ships, pointing out that, "in contrast to [highly automated] cars and aircraft, the operators on ships do much more than

Table 4
Design approaches adopted in the reviewed studies.

Goal-based design approach	Description	Article ID
Human-computer interface design	Functional requirements are defined in terms of tasks, task goals, and operational objectives. Methods for decomposing operational objectives into distinct task goals included Task Analysis (TA) and Information-Decision-Action (IDA) (described in more detail in Section 3.3.3).	[11, 15, 17, 22, 24, 27, 29, 31, 34, 35, 37]
Safety design	Functional requirements are defined in terms of acceptable risk levels. Although this can take many forms, the over-arching purpose in generally the same: to identify hazards, estimate their occurrence, predict associated risks, and design safety controls for mitigating these risks (see Section 3.3.2 for a full list of risk analysis methods). Often referred to as "risk-based design."	[20, 22, 23, 24, 25, 26, 27, 29, 31, 34, 35, 37, 38, 39, 40]
System integration	Functional requirements stem from integrating components of a complex system. This may include diverse elements like human–computer interaction, operator tasks, and risk-based design, and emphasizes consistent modeling techniques across the board. Often referred to as "systems thinking," or "systems engineering" approaches to design.	[14, 21, 31]

'just' navigating the vehicle" (p. 2). In this sense, some human roles can be seen as holdouts to increasing automation, which include not just machinery maintenance and repair, but also cargo loading/unloading and mooring at port.

Saha (2021) draws similarities between the role of MASS operators and those of Officers of the Watch (OOW) on conventional ships, in that they are both "responsible for monitoring the ship and intervening if needed" (p. 10). Hannaford and Hassel (2021), in contrast, hesitate to describe human roles at all, on the grounds that "it is too early to identify the amount of crew and the specific roles that will be needed because the crew has not been replaced yet, and not enough research has been conducted or published" (p. 15). Yoshida et al. (2021) argue that overall responsibilities of SCC operators will increase when compared to navigators aboard conventional ships.

3.2.3. *Resilience Engineering*

Resilience Engineering (RE) is a paradigm that aims to balance productivity with safety in complex systems using safety control techniques where risks are managed over time. Instead of focusing on human errors and inferring safety from accident investigations, RE positions humans as the source of resilience in a complex system and places the emphasis on safety controls. This way of thinking sheds light on the implications for human control in the MASS system, even before we have such systems implemented. Almost one-third of the analyzed papers (thirteen out of forty-two) adopted a design framework that fit within the RE paradigm.

Here we present the relevant studies in terms of four broad principles at the core of RE. These are (1) focus on human as a source of resilience rather than as a source of error, (2) emphasis on safety controls and dynamic risk indices, (3) use of systems engineering-based risk assessment technique like System-Theoretic Process Analysis (STPA), and (4) approaches to safety science that consider human and organizational interaction elements in addition to more straightforward component failures.

Starting with the first principle, Ahvenjärvi (2016) directs attention towards the roles of software developers of the MASS system on the principle that here, too, can resilience be built in given that we design for front-line operators' flexibility and problem solving abilities. In line with the second principle, Thieme and Utne (2017) apply the "resilience-based early warning approach to development of indicators" (REWI) method to the MASS system, focusing on handling accidents, incidents, and unexpected events from an organizational perspective. Utne et al. (2020) elaborate on the "supervisory risk control" concept, showing that the system's resilience goes beyond conventional safety controls by virtue of its ability to learn, adapt and improve over time. Reddy et al. (2019) highlight the need for a "systematic, traceable, and holistic assessment" of safety controls which shares some of the organizational and managerial elements discussed in Thieme and Utne (2017) and Utne et al. (2020). Peeters et al. (2020b) designed their prototype specifically to be able to measure performance indices over time; similarly, Wu et al. (2021) designed a real-time safety index algorithm based on historical sailing voyages along the same route.

For the third RE principle, Wróbel et al. (2018) were the first to apply STPA to MASS system risk assessment. In the same year, assessing the applicability of sixty-four risk assessment methods to MASS systems, Thieme et al. (2018) promote STPA as among the most suitable candidates, in part for its ability to identify human–machine interaction risks and propose risk reduction methods. Valdez Banda et al. (2019) also used STPA, in their case to inform early-stage risk-based design of an autonomous urban passenger ferry. Zhou et al. (2021) also apply the STPA approach to MASS system design, with a special focus on identifying both safety and cyber-security hazards and their inter-related safety controls. In line with the fourth RE principle, Ramos et al. (2020a, 2020b) developed the "H-SIA method" to enable analysis of the complex system as whole, rather than modeling how component failures can propagate. Ramos et al. (2019) take a similar approach for human task analysis in MASS operations, linking the success of MASS operators' tasks to latent socio-technical factors like training and organizational culture. Overall, the studies that adopted RE principles placed humans' roles in the MASS system as central to ensuring system safety.

Table 5
Risk assessment methods adopted in the reviewed studies.

Name	Description	Article ID
BNs	**Bayesian Networks**: used to update prior beliefs about system behavior to model risk over time in an interactive network.	[27, 35, 38, 41]
FMEA	**Failure Modes and Effects Analysis**: a classic risk assessment method combining severity, likelihood of failure mode, and detection rate to compute component or system level failure probabilities.	[38, 42]
H-SIA	**Human-System Interaction in Autonomy**: "provides a framework for analyzing autonomous ship operation as an entirety, rather than each agent separately" (Ramos et al., 2020b, p. 1).	[24, 34]
REWI	**Resilience-Based Early Warning Approach to Development of Indicators**: aims at determining organizational cap- abilities to handle unexpected situations and producing quantifiable safety indicators.	[25]
STPA	**System-Theoretic Process Analysis**: used for identifying hazards and revealing causal factors especially for novel and complex systems for which there is little experience or empirical data.	[26, 27, 31, 39]
THERP	**Technique for Human Error Rate Prediction**: risk assessment focused on quantifying human errors; used in the quantitative analysis of human reliability in human factors domain.	[35]
RCM	**Reliability Centered Maintenance**: stepwise method developed in the aviation industry to optimize maintenance management and improve reliability of safety–critical systems.	[30]
FSA	**Formal Safety Assessment**: a process developed by IMO in the wake of Alpha Piper disaster; consists of systematic hazard identification, risk assessment and control, cost-benefit analysis, and safety recommendations.	[40]
CPA	**Closest Point of Approach**: risk is assessed based on proximity of other objects, or from deviation of a target from its planned path; risk is a considered as a function of the target's spatial–temporal properties.	[20, 22, 23]

3.2.4. *Benefits of collaborative human-AI systems*

Collaborative human-AI systems are a way of framing increasing automation as an effort to enhance both human and machine roles by tailoring both according to the strengths and weaknesses of the two counterparts. The purpose of increasing automation, in this sense, should not be seen simply as an effort to reduce occurrence of human errors. Rather, collaboration between human and machine should lead to greater system performance than could be achieved by either counterpart working in isolation.

Five studies out of the forty-two analyzed identified benefits of human-AI teaming that went above and beyond human error reduction. Wu et al. (2021b, 2021a), for example, develop interface designs for MASS operators that display a comparison of current navigation activities relative to those of successful historical voyages on the same route. The AI computes large amounts of data and alerts the operator only when a safety index is exceeded – when the ship deviated from its course. At this point, the operator can use the output as a heuristic for decision making (does the course deviation warrant intervention?). In this way, a synergy between the computational power of the AI system and the decision making capacity of an expert operator emerges. To consider another example, Thieme and Utne (2017) frame shore-based decision support, control, and operational logistics in terms of "dynamic safety performance monitoring." This view presumes that primary navigation tasks are computed by AI while humans attend to the higher-order tasks involved in safety management: a partnership that augments operators' role of safety–critical decision making. Similarly,

Table 6
Human-computer interaction paradigms adopted in the reviewed studies.

HCI paradigm	Description	Article ID
Cognitive modelling	Situation Awareness: a concept used in a general sense to describe an operator's semantic apprehension of their surroundings; often specifically refers to Endsley's cognitive model called of the same name (Endsley, 1995).	[1, 6, 11, 17, 22, 26, 35, 39, 42]
	Information, Decision, Action in Crew context: a framework introduced by Chang and Mosleh (2007) to model cognitive processes of crew in response to accidents in nuclear power plant operations.	[24, 34, 37]
	Task Analysis: "collective noun used in the field of ergonomics, which includes HCI [human–computer interaction], for all the methods of collecting, classifying, and interpreting data on the performance of systems that include at least one person as a system component" (Diaper, 2004).	[15, 24, 34, 37]
Embodied cognition	Sense-making: an "enactive" approach in cognitive science (distinct from "functionalist" cognitive paradigms defining the body only in terms of input and outputs); "cognition as sense-making is the exercise of skillful know-how in situated and embodied action" (Varela et al., 2016, p. xxvi). Also referred to as "ship sense" and "harmony."	[1, 6, 11, 12, 17]
	Human factors: examines latent causes of error or factors that improve system interactions in a socio-technical system. This may include cognitive functions and sense-making, but also other phenomena like vigilance, workload, trust, and fatigue.	[6, 7, 9]

Huang et al. (2020) and Ramos et al. (2019) frame AI-operator teaming in terms of enhancing practical control-related tasks like object detection and coordinating collision avoidance, rather than in terms of wholesale substitution of human control.

3.3. Given that automation is increasing in shipping, what methods, approaches, and theories are being used to address design challenges?

The forty-two studies analyzed revealed a broad range of methodological approaches. Each of these approaches represented a different investigative angle to the same underlying problem; namely, how can design challenges emerging from increased automation in shipping be addressed. Here we break down this methodological dissonance into four categories representing common strategies among the various theoretical frameworks. Each of these categories are analyzed separately below and include: (1) practical design approaches for human–computer interaction, (2) risk assessment methods used in risk-based design, (3) different human–computer interaction paradigms adopted, and (4) empirical testing or data collection efforts. These four categories emerged during the analysis, encompassing the wide array of perspectives contributing to the same topic. This "common denominator" approach to summarizing the variety of methods on offer provides a handle on the relevant theoretical foundations adopted in the contributing studies.

3.3.1. Practical design approaches for human–machine interaction

Almost half of the studies (twenty out of forty-two) adopted a practical approach for designing human–machine interaction elements of the MASS system. We summarize these approaches here, characterizing them broadly within three sub-sets of goal-based design (Table 4). Here we use the term "goal-based design" to describe the general process of mapping high-level system objectives to functional requirements. We identified three goal-based design sub-sets: (1) human–computer

Table 7
Studies that present empirical observations about MASS systems.

Source of observation	Description	Article ID
Field test	A prototype unmanned inland cargo vessel (scale model, length 4.8 m) was tested in inland waters to demonstrate its feasibility.	[19]
Field test	The SCC counterpart to [19] was tested for operator situation awareness. The researchers aimed to "stress test" the SCC prototype and investigate whether eye-tracking could provide insights into operators' SA (n = 2 participants).	[17]
Field test	Sea trials were undertaken for an "autonomous cargo test ship" in a test area outside of Wuhan, China; the ship was controlled remotely from a testing center in Netherlands.	[35]
Simulator test	First test program in a full-mission bridge simulator re-configured as an SCC, designed to investigate operators' situation awareness (n = 6 participants, all experienced seafarers).	[1]
Simulator test	Researchers designed and conducted an experiment to study VTS-MASS interactions. Testing was done in a full-mission bridge simulator (n = 24 participants; half experienced seafarers, half control group).	[10]
Simulator test	Commercial full-mission bridge simulator used to measure effect of a "guidance-support system" designed by the researchers to enhance collision avoidance performance (n = 36).	[23]
Survey	Participants were asked to evaluate the extent to which they thought that sixty-six different Knowledge, Understanding & Proficiency requirements (KUPs), set out in the STCW convention, were relevant for autonomous shipping. Responses were collected on a Likert scale (n = 109, all seafarers with a navigation license).	[4]
Survey	Participants were asked to evaluate the relevance of seventy-five human factors for SCC operators (n = 32, seafarers).	[9]
Survey and interview	Participants were asked about employment, training, safety, and feasibility of MASS systems, with results collected on a Likert scale (n = 42, all licensed navigators). Subject Matter Experts were also interviewed about similar topics (n = 7, variety of background expertise).	[5]
Workshop	Held with VTS operators in Norway; participants generated ideas about the role VTS may play in the MASS system, with special focus on regulatory challenges (n = 26).	[14]
Workshop	Series of four workshops held with experts to identify hazards and ideate safety control functions for an autonomous urban passenger ferry in early-stage design (Workshop 1–3: n = 8, Workshop 4: n = 7; participants were a variety of subject matter experts with some return participants; n = 20 total individuals).	[31]
Workshop	Series of two expert workshops to assess risks in the MASS system (Workshop 1: n = 12, variety of subject matter experts; 2: n = 8, return participants).	[40]
Interview	Semi-structured interviews with expert informants about competence requirements for SCC operators (n = 10, subject matter experts from academia, regulatory agencies, technology companies, and other relevant backgrounds).	[8]
Interviews and questionnaire	Interviewed seafarers (n = 10) and subsequently had participants watch a video of a simulator running scenarios (n = 25 trials); participants also filled out a NASA-TLX questionnaire for perceived workload.	[7]
Focus group	Interviewed experienced nautical sciences instructors (n = 3) about situation awareness in SCC operations in light of current STCW conventions on training and competency requirements for seafarers.	[11]

interface, (2) safety, and (3) system integration.

Eleven studies adopted a goal-based design approach oriented towards "human–computer interface design." Peeters et al. (2020b), for example, design and build an SCC that is integrated with a model-scale MASS prototype. Their approach starts with coming to terms with the various "activities" of operators, which take shape from assignment of operational objectives. Similarly, Yoshida et al. (2020) propose functional requirements for an interface designed to support operator situation awareness based on a "human behavioral model" of the SCC. Kooij and Hekkenberg (2021) focus on what tasks need to be replaced when moving seafarers off MASSs, leading to suggestions for a crew-reduction implementation plan. This also leads to functional requirement in terms of how many operators are onboard a given MASS and at what times. For example, the latter authors propose, "Replacing mooring, deck maintenance, bunkering, administration, and port supervision decreases the required crew to four crew members" (p. 15).

Fifteen studies adopted a goal-based design approach oriented towards "safety design." Wu et al. (2021a) propose a safety indicator system intended specifically for SCC operators and demonstrate it using data collected from the log files of sixteen voyages completed by a research vessel. The technique uses a supervised machine learning algorithm in a "black box" approach to obtain what is essentially a real-time safety indicator calculator active during a ship's voyage. Wu et al. (2021b) do something similar, based on historical data collected from VTS logs. Several other studies systematically identified hazards for the purpose of developing functional requirements for safety controls (Bolbot et al., 2021; Chang et al., 2020; Thieme and Utne, 2017; Wróbel et al., 2018; Zhou et al., 2021).

Three studies adopted a "system integration" approach to goal-based design. Relling et al. (2021) propose ways to integrate VTS with MASS traffic coordination, calling their approach a mixture of "systems thinking and participatory design thinking." Rødseth et al. (2021) argue for a systems approach to MASS design that develops functional design requirements from a range of related elements under a so-called "Operational Design Domain." This includes elements like operational complexity, human–computer interaction, and geographic location. The design work presented by Valdez Banda et al. (2019), in considering functional requirements over the entire life cycle of a MASS, is also exemplary of a systems engineering approach.

The three goal-based design approaches identified were not mutually exclusive and were often combined. Ramos et al. (2019, 2020a, 2020b), go to great lengths to describe operator tasks and link them to risks associated with MASS operational design. Utne et al. (2020), Huang et al. (2020), and Zhang et al. (2020) also combine elements of interface and safety design, in their cases mapping cognitive elements related to task handling (like perception, action, and prediction) to hazard identification and failure probabilities and their associated risks. Valdez Banda et al. (2019) combine elements of all three approaches, adopting risk-based design within the classic naval architecture "design spiral" (Evans, 1959), and adhere to systems engineering principles for life-cycle design. Yoo and Lee (2021) develop requirements for VTS and SCC operator interface designs based on collision avoidance risk assessment in a busy port.

3.3.2. Risk assessment methods used in risk-based design

In the previous section, we identified risk-based design (goal-based design oriented toward functional requirements of safety controls) as the most common approach among studies presenting practical design approaches to MASS systems. Considering the central role of risk-based design in MASS systems, we examine the studies contributing to this topic more closely. The result is a list of fifteen studies, among which we identified nine different risk assessment methods. These are listed in Table 5 along with their respective definitions and corresponding studies. Note that while each method describes a distinct modelling approach for risk assessment, they were not exclusive, often appearing in combination with other approaches.

3.3.3. Human-computer interaction paradigms

In this section, we examine distinct human–computer interaction (HCI) paradigms adopted in the reviewed studies. We observed that such paradigms yielded different hypotheses about MASS functionality and are therefore important elements when considering the breadth of relevant design approaches. Overall, we identified two broad HCI paradigms, which we define in Table 6 along with their corresponding articles. These paradigms were described in sixteen of the forty-two studies in the review.

Situation Awareness (SA) was the most common HCI paradigm, referred to in nine of the studies. Surprisingly, almost all studies adopting SA also adopted elements of sense-making in their treatment of interactions. Sense-making represents a distinct cognitive paradigm from SA in that it links cognitive processes to the body, something that SA does not do. This unlikely combination may in part be a result of MASS operators' working remotely. Operators' bodies were in the SCC while their actions involved distant MASSs, a cognitive discord that was bridged by adopting elements from both SA and sense-making paradigms.

Task Analysis (TA) was another common framework used to describe HCI. The task analyses presented in Ramos et al. (2019, 2020a, 2020b) provide detailed classification of SCC operators' roles and tasks, in their case for MASS without operators onboard. In Ramos et al. (2019), for example, the authors present a "task description" listing specific tasks goals and sub-goals, as well as descriptions of possible errors corresponding to the defined tasks (pp. 40–41). For example, the first task goal they list is "supervise safety status of the ships," which has the subtask "monitor the screens" and the corresponding possible error "not checking information on the screens" (p. 40). This way of conceptualizing the MASS system was practical for modelling risks and for guiding interface design.

The study of human factors also provided a lens on HCI. For example, Yoshida et al. (2021) frames human–computer interactions as a source of stress for SCC operators. They argue that stress "might sometimes increase due to information overload by receiving enormous amounts of visual data to compensate for the lack of the feeling of the environment inside or outside a ship" (p. 2). In another study, Wróbel et al. (2021), studying the various human factors emerging from human-computer interaction within the MASS system, adopt a systematic method called "Human Factors Analysis and Classification System for Maritime Accidents (HFACS–MA)." Both studies indicate that human factors play an important role in design of safe interactions in the MASS system.

3.3.4. Empirical testing and prototyping

Fifteen studies presented results of empirical testing and observations about MASS systems. Considering MASS systems are still in a developmental phase, such results are valuable given that they can contribute to further design iterations and extended research. The empirical material came in a variety of forms, ranging from expert workshops and interviews to simulator tests and field trials with MASS prototypes. In Table 7 we list an overview of empirical testing and data collection efforts in the reviewed studies.

While the studies in Table 7 generated new observations to inform their work, several others used historical data. These other efforts, while not contributing new data to the research community, did present useful data collection methods. For example, Thieme and Utne (2017) and Fossen and Fossen (2018) both use field data collected from Trondheim Fjord in Norway, which is a designated a test area for autonomous ships. The former use field trial data from autonomous underwater vehicles (AUVs) to make inferences about safety performance indicators for MASS operations, while the latter use live Automatic Information System (AIS) data to test ship motion prediction and visualization algorithms intended specifically for remote operators. In another study, Wu et al. (2021a) collect data from the ship log of a research vessel transiting between two locations. Taking this information to represent successfully voyages, dynamic safety indices were generated based on the extent to which route deviations were detected in real time. Wu et al. (2021b) accomplish something similar for AIS traffic data to inform collision avoidance maneuvering at a VTS or SCC.

3.4. Given that automation is increasing in shipping, what research gaps, regulatory obstacles, and technical shortcomings represent the most significant barriers to real-world implementation?

In this section, we explore the extent to which the lack of full-scale autonomous ships currently implemented in the field is due to research gaps, regulatory obstacles, and technical shortcomings. Together, the reviewed studies described MASSs along a spectrum of technology readiness levels. Some authors argue that MASS are ready for implementation so long as the appropriate operational strategies are in place (e.g., Rødseth et al., 2021; Valdez Banda et al., 2019); others argue out that MASS do not currently exist in operational scale because of unresolved design obstacles (e.g., Kooij and Hekkenberg, 2021; Relling et al., 2021). Here, we investigate four relevant themes in the effort to identify and assess what barriers currently exist on the path from concept design to real-world implementation: (1) emerging risks in human-AI interaction, (2) regulatory obstacles, (3), training for SCC operations, and (4) infrastructure needs to support mixed conventional ship and MASS traffic. These categories shed light on the directions of current research in the field and on where more work is needed.

3.4.1. Emerging risks in human-AI interaction

The MASS system introduces novel challenges related to integrating human supervisory control within a highly automated system. In the reviewed studies, a new risk picture emerged characterized by uncertainties involving how human-AI interaction will be accomplished in real-world conditions. Here we examine how human-AI interaction was positioned as a central factor in twenty-two of the analyzed studies, underpinning risk assessment and design of safety controls.

In Section 3.3.2, we presented studies that applied risk assessment methods to study MASS systems. The result of such risk assessments typically materialized as a list of potential hazards, with the hazards near the top of the list contributing the largest share of overall risk. Chang et al. (2020), for example, review all published hazard identifications for MASS operations and evaluate the most salient hazard categories. Of these, they conclude that "interaction with manned vessels and detection of objects" represent the most significant contributors to overall risk (p. 10). Wróbel et al. (2018) (which was one of the studies Chang et al. reviewed), also determined that "interaction between shore-based facilities and legal or organizational [entities]" was a significant source of hazards. This finding is corroborated in Guo et al. (2021), in which the failure probability of an autonomous urban passenger ferry is found to be "most sensitive to the failure of the remote supervisor's intervention" (p. 7). Eriksen et al. (2021), taking a different approach, frame interactions in terms of preventative maintenance and conclude that so long as conventional ship machinery is used in MASS, maintenance requirements cannot be met without personnel present on board. The overall risk picture of the MASS system, in other words, remains unclear until we design its functional requirements around human interaction. Anticipating that risks arise from interactions within a socio-technical system, the "H-SIA" method was developed specifically in Ramos et al. (2020a) to model error propagations from human-AI interactions.

Among obstacles towards understanding the mechanisms behind human-AI interaction, lack of research and especially empirical data emerged most clearly. It was expressed in twenty-two of the forty-two articles analyzed. Of these, the most cited was related to lack of data about how failures propagate between interaction of hardware, software, and humans (Bolbot et al., 2021; Fan et al., 2021; Guo et al., 2021; Ramos et al., 2019, 2020a, 2020b; Thieme et al., 2018). Also wanting was research about risk management frameworks appropriate for MASS systems (Chang et al., 2020; Ramos et al., 2020a; Thieme et al., 2018; Thieme and Utne, 2017). Some studies also pointed out that observation-based data is needed about hazards for MASS in order to shed light on mitigation strategies in the aim of eventually being able to quantify their associated risk levels (Thieme et al., 2018; Valdez Banda et al., 2019; Wróbel et al., 2018). Similarly, several studies highlighted the lack of research about human error probability quantification for human-AI interaction especially for emergency response (Ramos et al., 2019; Zhang et al., 2020; Zhou et al., 2021). Other areas where lack of empirical data was reported as a barrier to real-world implementation of MASS applied to safety management (Størkersen, 2020), reliability of MASS machinery (Eriksen et al., 2021), human-AI system interface design (Fan et al., 2021; Huang et al., 2020), and competence requirements for SCC operators (Kim and Mallam, 2020; Sharma and Kim, 2021). Interaction between MASS and conventional vessels was also an area lacking data (Relling et al., 2021; Thieme et al., 2018), as were observations about how human factors affect SCC operations (Hannaford and Hassel, 2021; Kari and Steinert, 2021; Wróbel et al., 2021; Yoshida et al., 2021).

3.4.2. Regulatory obstacles

Gaps in regulations for unmanned ships were identified as a major barrier to MASS system implementation in the analyzed studies. They were prominently featured in eleven out of the forty-two papers analyzed. Common to all studies was the finding that conventions like COLREGS, SOLAS, and STCW, which are all regulated by IMO, are predicated on seafarers being onboard the ship. Legal precedents are also based on human crew onboard the ship. Katsivela (2020), a maritime law scholar, argues for amendments to existing conventions that would allow for MASS and SCC to be regulated as conventional ships. Although the author suggests that these definition-based amendments are relatively straightforward to implement, the need for them is described as "urgent," as are efforts towards harmonization and standardization of terms used across many different regulatory agencies at both international and national levels. Katsivela (2020) and Yoshida et al. (2020) both highlight Rule 5 in COLREGS as an example of how current regulations may have limited applications to MASS. The rule states that "Every vessel shall at all times maintain a proper look-out by sight and hearing as well as by all available means appropriate in the prevailing circumstances so as to make a full appraisal of the situation and of the risk of collision." Both authors suggest that while sensor fusion is showing potential in accurate detection and classification of objects, the appraisal of the "situation and the risk of collision" appears to necessitate the presence of a human. Wu et al. (2021b) argue that COLREGS is "human-centered" and "formulated in an intuitive manner," adding that its qualitative nature imposes challenges on the development and implementation of rule- or machine-learning-based approaches to collision avoidance. Katsivela is particularly attentive to the vocabulary used in maritime conventions and how it accentuates gaps between regular ships and MASS. For example, the wording in international conventions like STCW uses phrases like "a personnel onboard," "those on board," and "seafarers on board," which have been attributed to the obligation of the shipowner in maritime law to "provide for a seaworthy vessel" (United Nations, 1982). Also, the definition of "seaworthiness" has legal precedent in maritime law (Tetley, 2008) requiring that "The vessel be capable of withstanding the ordinary perils of the sea, be fit for the proposed trip and be crewed by a competent crew." Katsivela (2020) suggests clarifications of definitions for "ship," "navigation bridge," and "seaworthiness" as they are defined in the current conventions to include the possibility of unmanned ships with crew on land. Kooij and Hekkenberg (2021), in their proposed incremental crew reduction strategy for MASS systems Wróbel, 2021, also meet barriers for allowing seafarers to be responsible for ships at a distance. This regulatory change, they write, is the last step in the process of moving seafarers to land-based work like that envisioned at the SCC.

Conventions for seafarer training and competence requirements were also lacking prescriptions for meeting the novel demands of skilled SCC work. These conventions are laid out in STCW, which Kim and Mallam (2020) and Sharma and Kim (2021) investigate for applicability to MASS. Out of the sixty-six knowledge, understanding, and proficiencies (so-called "KUPS") listed (IMO, 2017, Table A-II/1), Sharma and Kim (2021) surmise that only twenty-six are relevant for MASS operations, while many whole new ones will be needed. Moreover, it was found that STCW leadership strategies need to be re-evaluated (Kim and Mallam, 2020) and that potential pitfalls like "over-reliance on sensors" warrant special attention (Hannaford and Hassel, 2021). Saha (2021) also finds that STCW presents a clear gap when it comes to SCC operators and suggests new competency requirements in the form of "system understanding, communicational and technical knowledge, and maritime competence" (p. 6).

Bačkalov (2020) and Nzengu et al. (2021) focus on regulations for inland MASS. Nzengu et al. (2021) point out that inland MASS do not fall under IMO jurisdiction, since "their operations and their context differ significantly from short-sea and ocean-going ship applications" (p. 359). The authors conclude that current regulations do not allow for autonomous MASS because they by definition rely on human functions onboard. Bačkalov (2020) presents detailed suggested amendments to updated regulatory framework for inland MASS, including introducing the notions of the "remote operator" and the "remote control center," as well as refining the "notion of the Master of an inland vessel" (p. 11).

Rødseth et al. (2021) focus on the approval process of MASS systems, comparing six design guidelines published by flag states and classification societies. The authors find that the six guidelines are largely in agreement that approval will hinge on submitting "descriptions of the MASS (system) design, [its] intended operations, and the environment it is intended to operate in." However, since it was not immediately clear from the guidelines how best to describe the system, the authors develop the concept of the "operational design envelop." Intended to help designers describe the MASS system for the approval process, the operational design envelop encourages description of how responsibilities between human and automation are shared, independent of project-specific factors like geographic location and operation.

3.4.3. Training for SCC operators

Eleven studies addressed training needs of MASS operators working at the Shore Control Center – so-called "SCC operators" (SCCOs). The knowledge gap in what training and competency requirements were needed for these operators was framed as a barrier to MASS implementation.

Four of the studies framed the need for training in terms of risk, where training was seen as a safety control. Wróbel et al. (2018) and Fan et al. (2020), for example, argue that insufficient training of SCC operators can lead to inadequate safety control functions. By this logic, the reverse also holds true: "…a well-trained and experienced operator having adequate control over the vessel can find the best solution to the situation encountered" (Wróbel et al., 2018, p. 342). Zhang et al. (2020) go one step further, presenting the effect of "insufficient training" in quantitative terms among other human error factors present in SCC work. Thieme et al. (2018), in assessing various risk modelling techniques for applicability to MASS systems, find that training, competence, and experience feature repeatedly as factors influencing risk.

The remaining seven studies focused on what type of competencies are needed at the SCC. Ahvenjärvi (2016) was among the first to point out that special training will be needed to meet the demands of interacting with manned and unmanned ships in the same area. Subsequent studies agreed that a lack of knowledge about how to train SCC operators is hindering MASS implementation. Katsivela (2020) express this problem from a regulatory perspective: "…existing regulations relating to the training and certification of the crew and, therefore, the provisions of the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers 1978 (STCW) currently under the purview of the IMO's scoping exercise will have to be revised to accommodate the presence of a properly trained crew onshore" (p. 243). (Note that the scoping exercise is now completed and recommends further investigation into these competency needs, a topic we outlined in Section 1.3). Three studies compared competency requirements for conventional ships with those that are most likely required for MASS operations, orienting readers towards appropriate revisions to the STCW or to tailor-made SCC requirements (Kim and Mallam, 2020; Saha, 2021; Sharma and Kim, 2021). Baldauf et al. (2019), on the basis of results from a simulator experiment, argue that ship-handling skills and experience at sea are prerequisites for SCC operation. This is corroborated by Saha (2021) and Yoshida et al. (2020), who confer with subject matter experts and seafarers. Additional competency requirements also emerged, like those associated with digital communication and information transfer (Saha, 2021; Yoshida et al., 2020). The former concludes that, "the SCCO should possess a combination of maritime and technological competence to control the unmanned vessels with comprehensive knowledge of the remote vessel operational and monitoring system" (Saha, 2021, p. 11).

Sharma and Kim (2021) were also motivated to address the shortcomings in STCW to MASS operations, and systematically compare sixty-six KUPs set out in current convention (IMO, 2017, Table A-II/1). Their findings point to twenty-six KUPs that are no longer relevant and to emerging requirements for "technical and cognitive competencies." The former involves "IT skills, safety & security management skills, knowledge regarding engine room operations, electronic equipment and system integration" (p. 13), while the latter involves "non-routine problem solving, ability for self-regulation, critical thinking, mental readiness and systemic thinking" (p. 15). Sharma and Kim (2021), after accumulating survey results, report that most seafarers agree that training will change significantly with the onset of autonomous shipping and that skillsets will likely be similar to those of currently licensed seafarers.

The studies reviewed revealed that questions about training, until resolved, present a barrier to MASS implementation. This is the case whether framed as a safety control mechanism in risk science or as competency requirements human factors. Training, in this sense, encompasses not just skills, but qualifications, roles, and responsibilities that are at the core of safe MASS operations.

3.4.4. Infrastructure to support mixed traffic

Although autonomous vessels stand ready for implementation, the reality of marine traffic is that it will always be mixed, composed of both MASS and manned vessels sharing the water. This mixed traffic format will introduce unique challenges. Here we briefly summarize results from seven studies encompassing three relevant topics: (1) SCC coordination with VTS, (2) development of mixed traffic navigation aids, and (3) harmonization of rules and standards for aids to marine navigation.

It emerged from four studies that VTS will play an important role in coordinating MASS traffic. VTS is regulated globally by an organization called the International Association of Marine Aids to Navigation and Lighthouse Authorities (IALA). Katsivela (2020) refers to a 2019 report by IALA that affirms that the organization is preparing for the advent of MASS by monitoring technology development for MASS navigation services and by promoting harmonized rules and standards regarding data transferred via different services (IALA, 2019, pp. 10-11). Baldauf et al. (2019) focus on challenges associated with interaction between SCC and VTS. The authors' premise is that since that both infrastructures will have similar roles of coordinating marine traffic from a centralized control room, some degree of cooperation will be needed. This logic is also held by Relling et al. (2021), who examine how VTS procedures can be standardized to include MASS coordination. Their findings point to increased involvement: "VTS need to change their role from solving situations ad-hoc to assume a tactical responsibility in traffic planning and to resolve situations at an earlier stage" (p. 1). The authors underscore that this not does not replace SCC roles, but rather shares coordination tasks between SCC and VTS operators. Yoo and Lee (2021) were

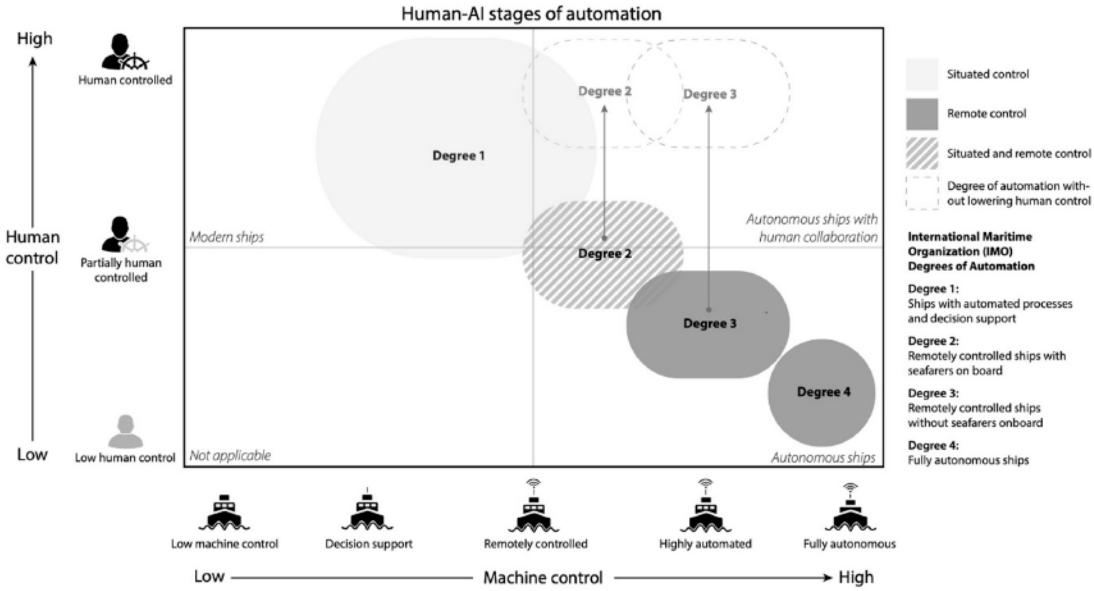


Fig. 5. IMO degrees of automation plotted onto the Human-Centered AI "Stages of Automation" framework adopted from Shneiderman (2020).

also motivated by what services VTS could provide in MASS traffic management. Their work came on the tail of IALA's announced intentions to work towards integrating "…autonomous vessels with conventional traffic, the information flow between MASS and shore authorities, and the related information exchange with conventional traffic" (IALA, 2019, p. 15). In this aim, the authors developed a "real-time collision risk assessment support system to improve the situational awareness of VTSOs [VTS operators] and MASS remote operators… regarding near-collision situations occurring in local waters" (p. 1).

An additional two studies argue that aids to navigation will have to be updated to support mixed traffic. Chang et al. (2020), after conducting a detailed risk assessment, found that "interaction with manned vessels" was among the highest contributors to overall risk of MASS operations. Thieme et al. (2018) also investigated mixed traffic from a risk science perspective, showing that traditional risk assessment methods that estimate collision frequency from conventional ship traffic will likely need to be updated for new traffic patterns that will inevitably emerge. The same authors highlight that navigation lights and buoys are used by deck officers and lookouts mainly as a type of visual ground-truth to radar and are in this way oriented uniquely to human navigation practices. Considering that MASS navigation will rely on the detection of such aids to navigation, consideration should therefore be directed towards making them more "visible" to sensor technologies.

Three studies specifically raised harmonization of rules and standards as a barrier to MASS implementation in mixed traffic. Katsivela (2020) defines harmonization as "any attempt, by whatever instrument, to minimize or eliminate discord between national commercial laws as they apply to international commercial transactions" (p. 241). Katsivela (2020) shows that the discord between MASS development, undertaken by a small number of institutions, are in discord with international standards for navigations infrastructure. The problem is especially evident for marine aids to navigation. These are regulated by IALA, whose initiatives to support digitalization of navigation (so called "e-navigation") still assumes that human operators are present on ships. Størkersen (2020) and Baldauf et al. (2019) similarly argue that harmonization of procedures across different procedures (e.g., MASS versus conventional ships, SCC versus VTS), will serve to address current obstacles to MASS implementation.

4. Discussion

In this section, we expand upon the themes uncovered in the systematic review, linking our original three research questions to farther-reaching discussions about how the introduction of AI is changing safety–critical work and about how risk scientists, designers, and researchers are addressing emerging challenges underpinning system design.

Sections 4.1 and 4.2 expand upon Research Question 1. Section 4.1 focuses on the need for "continuous monitoring and decision making support" made clear in the Results. Because this need is at odds with the traditional LoA approach, which implies a one-off focus on automation at the cost of human control, we explore alternatives to the traditional approach. Section 4.2 addresses the debate about "human error" in design, especially in the context of the de-facto risk baseline that came to light in the Results; namely, that MASS systems should be "at least as safe as" conventional ships. If automation is foremost to reduce human errors, should not MASS systems by this logic be significantly safer than conventional ships? We consider what new risks are introduced when increasing interactions with AI systems, as well as how these new risks may be addressed.

Sections 4.3–4.5 expand upon Research Question 2. Section 4.3 considers the popularity among the reviewed studies of cognitive approaches to human decision making modeling in the risk sciences. We consider limitations of such cognitive modeling approaches in light of contemporary research in the fields of computer supported collaborative work and neuroscience. Section 4.4 explores the viability of virtual simulation as a venue of systematic testing of MASS system, including a brief assessment of its strengths and weaknesses. Section 4.5 takes a second look at risk assessment methods for MASS systems. Considering the relatively wide spectrum of approaches identified in the Results, we raise the question: are some methods more appropriate that others?

Finally, Section 4.6 expands upon Research Question 3. Specifically, it addresses barriers to real-world implementation of MASS systems through the wide lens of safety management. The review results made clear that a special dexterity is needed to realize human-AI teaming flexible enough to deal with real-world demands. We examine the balance of operational flexibility with managerial control, drawing comparisons to established theories in organizational science.

4.1. The need to increase human control in navigation for AI applications

In Section 3.2.1 we reviewed forty-two studies about autonomous ships systems, guided by the question *what are the implications for human control?* Although it seemed like a paradox, the review showed that humans are needed to control autonomous ships. Part of the answer lay implicit in the way Level of Automation (LoA) taxonomies are used to describe human-AI collaboration. All taxonomies available (there were six in the studies reviewed) involved some degree of human control; no system is currently being designed for fully autonomous operations. Human control most often took the form of "Continuous monitoring and decision making support" (Section 3.2.2). Furthermore, the studies argued that this role ensured the safety of the system. Extending this idea here, we can interpret this as positioning operators as "active backup" to the automated system, where the operator is a safety control to the automated system. It seems counterintuitive, then, to categorize LoA by degree of autonomous control gained over human control lost, when in practice both are needed to ensure safety. For similar reasons, Eriksen et al. (2021) deliberately avoided using an LoA designation at all, noting they do not apply to how ship machinery systems are currently maintained. Rødseth et al. (2021) also concludes that none of the standard LoA systems do justice to explaining operators' time-critical role in the MASS system. This leaves the possibility open that other frameworks exist that may more appropriately handle the collaborative aspects of human and machine teaming as compared to the LoA frameworks adopted in this review. Here we explore alternatives to the traditional LoA approach.

One framework was recently proposed by Shneiderman (2020) who introduced a two-dimensional "Stages of Automation" in place of the more prevalent one-dimensional "Levels of Automation." This framework is what Shneiderman calls Human-Centered Artificial Intelligence (HSAI), which aims to improve AI-based systems design by jointly increasing human and machine control to achieve "Safe, Reliable, and Trustworthy" systems. Shneiderman's work focuses predominantly on the car industry, which is currently leading innovations in autonomous transportation. Specifically, the work is in response to current trends in autonomous car design that are leading to purportedly unacceptable risks by combining high autonomy with low human control. The current LoA framework used in the car industry is published by the Society of Automotive Engineers (SAE) and features six levels, ranging from Level 0 to Level 5 (SAE International, 2017). But like the frameworks designed for MASS operations encountered in the review, the SAE's focus on automation implies that human control decreases with each increment on the scale. Considering that the four-degree IMO LoA was the most common framework adopted in the reviewed studies, we mapped them onto the "Stages of Automation" framework for direct comparison (Fig. 5). IMO Degrees 1 and 2 fit roughly in the intersection of the human–machine control quadrants. Examples of these types can be seen in modern ships and early MASS designs today. IMO Degree 3 fits into the low human control-high machine control quadrant and is where most attention in MASS development is afforded today. It is surprising that the only element setting Degree 3 and Degree 2 ships apart is the complete absence of people on board, implying that MASS autonomy is improved simply by superseding human control. Finally, IMO Degree 4 appears to exist only for symbolic reasons, considering not one of the papers reviewed considers full machine autonomy technically feasible. This way of framing the popular IMO Degrees of Automation suggests that attention should be focused on "raising" Degree 2 and 3 from their current positions towards higher levels of human control, as illustrated by the arrows in Fig. 5. Fig. 5 also illustrates a trend tracing automation development in modern ships (upper left quadrant) towards the "fully autonomous" vision (lower right quadrant"). But considering the infeasibly of the fully autonomous model, one must ask why technology trends tend to orient towards such a model. The reviewed studies, after all, repeatedly emphasized the importance of improving human control, not supplanting it altogether with automation. Consider, for example, the results presented in Section 3.2.2, where we reported twenty-three studies specifically addressing humans' roles in autonomous ship systems. It follows that re-aligning design efforts toward the upper right quadrant, labelled "autonomous ships with human collaboration" in Fig. 5, may stand as a more appropriate guiding model. Of the reviewed studies, Rødseth et al. (2021) came closest to the "Stages of Automation" approach by avoiding adherence to current LoA taxonomies and instead presenting a two-dimensional grid with human control versus automation. "We believe that this is a useful approach," they write, "… as it clearly defines the responsibilities of the automation system versus humans…" (p. 5).

The review also led to a clearer understanding of specific tasks that necessitate shared human–machine control in safe MASS systems. For example, Huang et al. (2020) and Wu et al. (2021b) remind readers that interpreting regulations like COLREGS in rule-based programming code is challenging, suggesting that an AI-based collision avoidance system may require some level of collaboration with a human operator. Moreover, a recent empirical study by Rutledal et al. (2020) shows that collision avoidance is about more than just following codified rules. The study recorded instances in marine traffic where adherence to COLREGS was intentionally broken to resolve a give-way vessel in possible collision situations. Nonetheless, efforts in encoding COLREGs both in traditional symbolic AI and uncertainty-based or heuristic methods have seen considerable attention for over two decades (Statheros et al., 2008; Tam et al., 2009). Efforts in this area generally improve in accuracy over time, with recent work showing high accuracy in adopting Rapidly-Exploring Random Tree (RRT) algorithms (Chiang and Tapia, 2018; Zaccone et al., 2019). However, limitations like adverse weather conditions and non-AIS vessel targets suggest that more work is needed before these systems can be robust enough to make decisions on their own, relegating them for now to the lowest IMO Degree of "decision support" automation. The "Stages of Automation" framework may thus have a practical application for designing such collaborative collision avoidance systems, addressing the need to reconcile outputs of algorithms and the context of human control in varying conditions. Platooning, which combines human control and automation in an elegant way, represents one such reconciliation of human and machine control collaboration (e.g., Colling et al., 2021; Munim et al., 2021).

Aside from interpretation of COLREGs, several other important elements of seafaring appeared more straightforward for humans to accomplish than for AI systems. This included, for instance, managing Very High Frequency (VHF) marine radio communication and cargo loading and unloading. Checking the literature, we found that it has long been observed that tasks humans find naturally straightforward are often firmly outside the grasp of advanced AI. First articulated by Moravec (1988, pp. 15), this principle has since been coined the "Easy things are hard" paradox in AI. This has important implications especially for regulatory gaps that were identified in the review, because requirements of "seaworthiness" as well as statutory rules for "proper lookout by sight and hearing" (COLREGs) and responding to others in distress at sea (SOLAS) should be considered firmly within in the realm of human interpretation and control – firmly beyond the reach of AI methods.

4.2. New challenges emerging from increased human-AI interaction

Studies assessing risk of MASS operations often surmised that MASS should be "at least as safe as conventional ships" (Fan et al., 2020; Reddy et al., 2019; Relling et al., 2021; Saha, 2021; Thieme et al., 2018; Valdez

Banda et al., 2019; Zhou et al., 2021). Similarly, Utne et al. (2020) cite a DNV GL position paper recommending a goal-based Autonomous Ship Code (ASC) regulated by IMO stating that "autonomous and remote-controlled ships shall be as safe as conventional ships of the same type" (DNV GL, 2018). The question we raise here is whether this "at least as safe as" baseline is appropriate from a risk perspective. The predominant argument, after all, for introducing autonomous systems is that they will improve safety by reducing instances of "human errors" – sources widely held to account for 75–95% of marine casualties (Rothblum, 2000). Following this logic, should not system design produce *safer* systems – and by a significant margin? The automotive industry, to contrast with shipping, appears to follow the latter logic. Leading developers of autonomous cars proclaim significant improvements in safety when compared to their conventional counterparts (e.g., Tesla, 2019; Waymo, 2020). The logic follows from a "human error" narrative similar to the shipping case, with oft-cited statistics attributing attributed the bulk of accidents to human driver errors (e.g., U.S. National Highway Traffic Safety Administration, 2015, which sets the statistic at 92–96%). The comparison to automation in the car industry is not without precedent; Rødseth et al. (2021) adopted their central concept of "operational design domain" from design approaches of autonomous cars. Some important differences between cars and ships do exist, however, which the authors diligently highlight. "Ships," they write, "[and] especially large ones, have a much higher cost and damage potential compared to road vehicles… Låg, 2019 [also] most merchant ships move slowly and have more space to use for maneuvering" (p. 5). As a bottom line, though, safety baselines for MASS based on conventional ships should be brought to question. This is especially the case given that "human error" seems impossible to fully circumvent in ship systems, in that they will always exist by nature of human interaction, regardless of how much autonomous control is incorporated (Wróbel et al., 2017).

The review also showed that assessing the performance of autonomous systems within frameworks of conventional ship navigation occasionally led to experimental pitfalls. For example, in Man et al. (2018), the authors conducted an experiment in a full-mission bridge simulator set up such that participants operated vessels that appeared to be navigated autonomously when in fact they were be steered by researchers in another room. The authors set out to study the interface design of the SCC by measuring how they used bridge equipment to handle scenarios of remote operation. However, they concluded that "when the operators utilize commonly used navigational and collision avoidance technologies but in a different way, they would have problems in developing sufficient situation awareness for remote supervisory control tasks (pp. 241)." In other words, by attempting to study needs of remote operation by using tools found on a conventional ship's bridge, the researchers could only conclude that the tools on a conventional bridge were not sufficient. A similar problem was described by Baldauf et al. (2019), who, after running experiments with participants in a ship simulator to investigate interactions between a simulated MASS and VTS, observed that the results were limited by the equipment available. While concluding that VHF marine radio was needed and that frequent cross-checks between bridge navigation equipment and the bridge windows were important, the results generally failed to link results to the underlying research question about MASS interaction design. The primary conclusion that one can draw from the two studies is that conventional navigation equipment is probably not sufficient for novel MASS applications. They serve as examples of what can happen when making inferences about paradigm-challenging observations in this way, analogous to over-fitting a model to observations, leading to results that do not satisfy hypothesis testing. The pitfall is forewarned by Kari and Steinert (2021), who write, "the working environment in the SCC is completely different from the traditional onboard bridge" (p. 17).

Increased interaction with AI systems appears to come with increased risk. Zhou et al. (2021) make the case that with every new interaction comes a security vulnerability, increasing the risk of cyber-threats. Meanwhile, Yoshida et al. (2021) argue that increased AI interaction results in higher mental workload among seafarers, subsequently undermining their performance. The latter also show that navigators use lookouts at their "eyes," trusting them differently than they would sensors in an automated lookout system. "Human–human communication is mentally more comfortable," they write, "than using autonomous support and navigation system in this situation" (p. 19). The findings are in line with the so-called "cooperative eye hypothesis," which explains that humans evolved to have large sclera (whites of the eyes) to be able to follow the gaze of others in cooperative activities relying on joint attention and communicative interaction (Kobayashi and Kohshima, 2001, 1997; Tomasello et al., 2007).

In a similar line of thinking, the increased human-AI interaction presented by the MASS system was framed by Sharma and Kim (2021) as "increasing invisible interactions." The authors continue: "It is paramount for the crews onboard to be able to have a holistic and systemic understanding of the systems and its interactions, to be able to comprehend the complexity, to evaluate the interrelations of subsystems and to subsequently generate the best decisions and course of actions" (p. 17). Whether it was framed as misconstrued safety statistics, vulnerabilities to cyber-attack, or compromised human factors effects on work performance, it appeared that increased interaction between human work tasks and AI functions contributed to a significant extent of the added risk in MASS systems.

4.3. Limits of cognitive models of situation awareness for risk modeling

In Section 3.3.2 we showed that state-of-the-art risk assessment methods of MASS operations included the human-in-the-loop by adopting simplified models for cognitive processes for human operators. In particular, the IDAC model described by Chang and Mosleh (2007) played an important role in H-SIA. In Ramos et al. (2019), the authors identified cognitive errors stemming from tasks using the IDAC framework by re-describing task goals (in terms of sub-goals and high-level goal plans) until one element in the cognitive model (Information, Decision, Action) could be isolated and subsequently linked to errors within a network of possible events, forming a coherent risk picture for the system. This approach is exemplary of human cognition modeling, which is useful in risk modeling by virtue of its predictive power of human behavior in novel situations. However, limitations of modelling cognitive processes of operators in control applications have been a source of debate in the literature (Endsley, 2015; Flach, 2015). For at least four decades, neuroscientists have accepted that modeling human cognition is faced with the apparently insuperable challenge of infinitely many mechanisms that can generate any given observation (Anderson, 1978 Shneiderman, 2016). Recent theories in neuroscience are showing promising ways to reconcile constraints of cognitive models with peoples' underlying irrationality, including for decision making (Lieder and Griffiths, 2020). Their applicability to risk modeling, however, has not yet been explored. There also remains important social elements to cognition, which have been investigated in depth in the case of bridge navigation by Hutchins (1995). Heath and Luff (1991) also demonstrated, with the case of control room operators for the London Underground, that "awareness" existed in a collaborative sense, implying that decision making emerges in a team rather than individual context for control room work. Limitations in modelling human cognition for use in risk assessments of human-AI interactions for collaborative control room work remain under-explored.

4.4. Virtual simulation methods

In Section 3.3.2 we uncovered a dilemma particular to the problem of risk modeling for new applications like MASS operations. This inherent paradox is as follows: risk models, which are needed to predict risk in novel operations, rely on historical data that do not exist because the operations in question are novel. While risk identification methods like STPA and expert judgment-based techniques like Bayesian Networks (BNs) circumvent the need for large amounts of historical data, verification will inevitably depend on empirical observations. The strategy of building test prototypes and operating them in controlled areas sanctioned for autonomous vessels is one strategy already adopted by some contemporary research efforts. But this approach is also expensive and, as observed by researchers like Baldauf et al. (2019), Peeters et al. (2020a), and Thieme et al. (2018), is faced with uncertainty symptomatic of early design phases. One potential solution lies in simulation. Several studies we analyzed highlighted the important role simulators play in investigating emerging human-AI interaction challenges. For example, Kari and Steinert (2021) write that simulator experiments will "help to assess and evaluate the role of human–machine and human–human interactions" (p. 17). Saha (2021) also points to benefits of simulators for training and assessment of SCC operators. In simulation, virtual versions of sensors sense a virtual world, allowing researchers and developers to study how the AI works without resource-intensive field testing. Scenarios can be built and tested rapidly, including emergency scenarios that are impractical or impossible to coordinate in real life. Graphics can make the virtual worlds immersive, opening possibilities to training and testing human factors and interaction solutions. The automotive industry has adopted virtual simulation to research autonomous cars, with one notable example being the open-source platform CARLA (Dosovitskiy et al., 2017). Recent efforts aimed at developing analogous platforms for MASS systems have emerged (Vasstein et al., 2020). Given that risk scientists broadly recognize the potential value of virtual simulation, and given its successful applications in autonomous car research, we can expect that future work will rely upon simulation before operation of MASS is a reality. Still, simulation is not without its limitations. The method depends upon accurate scenario development and is faced with the so-called "long tail problem" that precludes the simulation of rare events encountered only in the wide spectrum of possible events characteristic of the real world. Also, introducing MASSs will inevitably change the behavior of conventional marine traffic once implemented, a second-order effect that that alludes simulation.

4.5. Risk assessment methods for MASS systems

Risk assessment proved to be a popular topic among studies contributing to the theme of MASS systems. In Section 3.3.2, we analyzed fifteen studies that contributed risk assessments of MASS systems, which in total adopted nine different methods. The breadth of risk assessment methods was high compared to the number of studies adopting them. This raised the question of whether some methods were more suitable than others for MASS applications. This question was investigated in Thieme et al. (2018), who pegged System-Theoretic Process Analysis (STPA) as among most promising risk assessment methods for MASS system hazard analysis. Four of the reviewed studies used STPA for this purpose (Utne et al., 2020; Valdez Banda et al., 2019; Wróbel et al., 2018; Zhou et al., 2021), making it the most popular technique along with BNs. One reason for its apparent appeal is that STPA does not rely upon large amounts of historical safety records for purposes of inference. This is a virtue shared by BN techniques, which, as expressed by Thieme et al. (2018), offer "flexibility of modeling methods and input from experts" (p. 151). Four studies demonstrated, with detailed case studies, how BNs could be implemented into risk models (Chang et al., 2020; Guo et al., 2021; Utne et al., 2020; Zhang et al., 2020). BN approaches framed input from experts as a valuable source in the absence of operational data; this was in contrast to other studies whose authors, while also relying on expert judgment, framed qualitative data as an input source with limited value in the absence of objective, statistical data (e.g., in the FMEA technique adopted by Fan et al., 2021 and FSA approach in Bolbot et al., 2021).

Examining STPA approaches more closely, Wróbel et al. (2020), who champion the technique, demonstrate that organizational and human interaction issues remain under-represented in risk assessment, while focus on the technological components of the system receive the locus of attention. Wrobel et al. (2021) argue that it may even be impossible to quantify human errors in shipping applications. STPA circumvents the needs to set percentages to human errors, and instead focuses on how human interact with a system in a real-world setting. "Why bother quantifying such a complex social phenomenon [as human error]," ask the authors rhetorically, "using various and incomparable approaches, instead of investigating how to fit the human element into a technical system so that both operate at their optimum?" (p. 10). STPA is not without its critics. Bolbot et al. (2021), for example, highlight that STPA does not offer practical recommendations for completing a hazard identification. The authors opted instead for a traditional Formal Safety Assessment (FSA) approach, which they claim is also more straightforward to approve. The first risk assessment published on MASS also used FSA, highlighting similar reasons for doing so (Rødseth and Burmeister, 2015).

Other risk assessment methods framed safety controls in terms of preventative maintenance (Eriksen et al., 2021) or in terms of safety indices based on deviations of MASSs traversing known routes (Wu et al., 2021a). Like STPA and BNs, these methods, too, rely upon expert judgment as a model input. Preventative maintenance requires the expert machinist for input; building safety indices from navigation patterns relies on training Machine Learning algorithms on "expert" data. Even the FSA approach described by Bolbot et al. (2021) relied upon expert advice for hazard identification, which is explicitly recommended in its guidelines for novel or innovative designs (IMO, 2013, pp. 10-12). In this way, all risk assessment methods in the review made use of a combination of qualitative data in the form of expert judgment and quantitative methods in the form of statistical inference.

Aside from STPA and BN methods, another risk assessment approach that showed promise was the Hybrid Causal Logic (HCL) model. Originally developed for the offshore petroleum industry by Røed et al. (2009), HCL describes event propagations and their consequences. Outcomes are then linked to BNs, allowing for probabilistic input of subjective aspects like human and organization behavior. The H-SIA method introduced by Ramos et al. (2020a) was inspired by the HCL approach but lacked the implementation of BNs. A full HCL modeling approach for MASS operations was demonstrated by Wu et al. (2020), but it focused on interactions between manned and unmanned vessels rather than on interactions of land-based operators and the unmanned vessels under their control. An HCL model for a MASS control system can lead to risk assessments that are updated based on operators' experience and continually revised over time.

Among other potentially useful approaches is the Functional Resonance Analysis Method (FRAM), which has been adopted in dynamic safety management in complex socio-technical systems like healthcare and aviation. It was assessed in Thieme et al. (2018) for MASS systems and found to hold promise. However, there are currently no FRAM studies in the literature applied to MASS systems.

4.6. Safety management holds clues for reconciling human and machine autonomies

Safety management will play a crucial role in human-AI systems for safety critical applications. Theories that underpin safety management in terms of safety and (human) autonomy may also help to frame the emerging problem reconciling safety and (machine) autonomy in terms of practical organizational mechanisms. This applies to MASS operations but also all transportation systems introducing human-AI teaming.

In this review, there was only one paper that directly addressed safety management for MASS operations. The author frames safety management as both a "source of engagement and frustration among seafarers" and suggests that remotely controlled vessels open the door to amending conventions set out in International Safety Management (ISM) Code specifically for increasing task flexibility for operators and for reducing cumbersome paperwork (Størkersen, 2020). Historically, research in work design has focused on striking a balance between human autonomy of front-line workers and centralized managerial control. In situations that demand dexterity from operators in the face of uncertainty, this balance means combining a high degree of both operational flexibility and organizational stability. The theoretical groundwork was established by Weick (1976) who framed this balance as a contradiction between centralization and decentralization and as a contradiction between safety and worker autonomy. Still, an analogy may be made between front-line human workers and the AI tools being currently developed. Drawing this parallel, we can extend the safety management frameworks to consider safety management for human-AI systems. Weick also theorized that specific mechanisms were present: so-called "loose couplings" that bridge the divide between safety (centralized, stable, managerial control) and autonomy (decentralized, flexible, task-based work). Recent work by Grote (2020) has picked up the debate, suggesting three coupling mechanisms that can be used in safety management towards High Reliability Organizations (HROs): safety rules, safety leadership, and safety culture. While specific examples like participatory rulemaking, transformational leadership, and safety climate outline practical tools in this framework, the work remains theoretical and without direct empirical verification. Regardless, many elements of the conventional work design debate can carry over to the debate in human-AI control. We outline three carry-over elements here. The first is that many aspects of personal worker safety are now moot considering that AI tasks can replace them. For example, SCC workers are not spending shifts at sea, virtually eliminating "slips, trips, falls, and vehicles accidents" traditionally associated with personal or occupational safety. New occupational hazards will present themselves at the SCC, but the hazards will be fewer and less consequential. The second is that principles we have drawn from studying HROs match the aims of human-AI systems in that the focus is on things going right, rather than counteracting things going wrong. For example, Rasmussen says, "Rather than striving to control behavior by *fighting deviations* from a particular pre-planned path, the focus should be on the control of behavior by *making the boundaries explicit and known* and by giving opportunities to develop *coping skills at boundaries* (Rasmussen, 1997, pp. 191). Rasmussen was referring to boundaries between worker control and managerial control, but they apply also to AI control and supervisory control. The idea of making "boundaries explicit and clear" echo the need to understand the AI system limitations as well as the underlying motivation of "Explainable AI" (Gunning and Aha, 2019). The "coping skills" at boundaries refer to the adaptability and contextuality of working in shared control between an AI system and human operators. The third and final element that we discuss here is the opportunity to learn from HROs for designing resilient human-AI systems. This is challenging, given the lack of empirical data on the subject. However, adopting Grote's framework and the concept of mechanisms for loose coupling between human autonomy and safety, one could imagine how strategies from HROs can carry over to loose couplings between AI and safety. Examples include participatory rulemaking, whereby the operator provides regular feedback to train AI systems, and strong cultural norms supporting switches between modes of operation.

5. Conclusion

In this article, we summarized the research state-of-the-art for autonomous ship systems, surveying a growing body of scientific literature published in peer-reviewed journals encompassing risk and safety science, human factors and training, marine policy, and ocean and reliability engineering. Given that automation is increasing in shipping, we posed three research questions: (1) how is human control currently being adopted in autonomous ship systems? (2) what methods, approaches, and theories are being used to address safety concerns and design challenges? and (3) what research gaps, regulatory obstacles, and technical shortcomings represent the most significant barriers to their real-world implementation?

In a systematic review, we collated findings from forty-two relevant peer-reviewed articles. Although the review spanned five broad disciplines (marine policy, ocean engineering, human factors, reliability engineering, and risk science), two underlying themes underpinned all findings: safety and control. The two themes were closely linked, with safety framed as a property emerging from reconciling control between machine autonomy and human autonomy. From this perspective, machine autonomy executes defined tasks while human autonomy manages supervisory goals under changing external factors and uncertainty.

Starting with the first research question, we found that human control plays a critical role even for the most advanced autonomous ship systems. Autonomous ships and their land-based counterparts, so-called Shore Control Centers (also called Remote Control Centers or Remote Operating Centers), represent a collaborative system. Most researchers envision human roles that go above and beyond simple backup to the automation, involving roles in all operational phases (e.g., watchkeeping, preventative maintenance, cargo loading and unloading, and emergency handling). During regular navigation, the AI system should enhance operators' decision making for route planning and collision avoidance, not replace it outright. In line with this perspective, new thinking is emerging about how to frame autonomous system design as advancing in stages of human-AI interaction, rather than in levels of machine control.

In line with the second research question, we synthesized current thinking about which risk assessment methods are best suited for autonomous ship systems. Our findings pointed to System-Theoretic Process Analysis (STPA) and Bayesian Networks (BNs) as the most appropriate tools to use in goal-based design of safety controls. These techniques do not rely on historical data about accident frequency, and instead leverage expert human input. A parallel can be drawn in this regard to Machine Learning algorithms and resilience thinking, two relevant fields that also frame human expertise as a source of power, not as a source of error. Current thinking raises safety management frameworks originally intended to bridge the divide between worker autonomy and managerial control as relevant for human-AI systems. Parallels were drawn in this sense to control coupling mechanisms bridging the divide between machine autonomy and human control.

Finally, addressing the third research question, we sampled the research state-of-art for the most significant obstacles standing in the way of real-world implementation off MASS systems. Notable obstacles included regulations, which hinge upon definitions ships as having seafarers physically onboard, and training requirements, for which gaps abound in current maritime conventions. The most significant obstacle involved the uncertainly surrounding human-AI interactions. Research is needed to shed light about how people work when teamed with AI, especially in safety–critical contexts.

Autonomous ships will soon sail in our oceans, coastal waters, and inland waterways. Current research shows that humans will have just as important a role in ensuring safety of autonomous ships as for conventional ships. The question remains how this role will be most appropriately carried out and best to coordinate it with AI system functions.

CRediT authorship contribution statement

Erik Veitch: Investigation, Methodology, Writing – original draft, Writing – review & editing. **Ole Andreas Alsos:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

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.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Research Council of Norway under the MAROFF-2 project "Land-based Operation of Autonomous Ships" (LOAS) (grant number 296527) directed by project leader Kongsberg Maritime AS and partnered with the Institute for Energy Technology (IFE). The authors also acknowledge the support of the Department of Design at NTNU.

Appendix A:. Dataset with ID numbers

ID	Author, year	Title	Journal	Category
1	Man et al. (2018)	Human factor issues during remote ship monitoring tasks: An ecological lesson for system design in a distributed context	International Journal of Industrial Ergonomics	Human Factors & Training
2	Kim and Mallam (2020)	A Delphi-AHP study on STCW leadership competence in the age of autonomous maritime operations	WMU Journal of Maritime Affairs	Human Factors & Training
3	Ahvenjärvi (2016)	The human element and autonomous ships	International Journal on Marine Navigation and Safety of Sea Transportation	Human Factors & Training
4	Sharma and Kim (2021)	Exploring technical and non-technical competencies of navigators for autonomous shipping	Maritime Policy & Management	Human Factors & Training
5	Hannaford and Van Hassel (2021)	Risks and benefits of crew reduction and or removal with increased automation on the ship operator	Applied Sciences	Human Factors & Training
6	Kari and Steinert (2021)	Human factor issues in remote ship operations: Lesson learned by studying different domains	Journal of Marine Science and Technology	Human Factors & Training
7	Yoshida et al. (2021)	Identification of the relationship between MASS and the operators' mental workload	Journal of Marine Science and Technology	Human Factors & Training
8	Saha (2021)	Mapping competence requirements for future shore control center operators	Maritime Policy & Management	Human Factors & Training
9	Wróbel et al. (2021)	On the influence of human factors on safety of remotely controlled merchant vessels	Applied Sciences	Human Factors & Training
10	Baldauf et al. (2019)	Merging conventionally navigating ships and MASS: Merging VTS, FOC and SCC	International Journal on Marine Navigation and Safety of Sea Transportation	Marine Policy, Regulation & Navigation
11	Yoshida et al. (2020)	Regulatory requirements on the competence of remote operator in maritime autonomous surface ship	Applied Sciences	Marine Policy, Regulation & Navigation
12	Størkersen (2020)	Safety management in remotely controlled vessel operations	Marine Policy	Marine Policy, Regulation & Navigation
13	Katsivela (2020)	Unmanned vessels and regulatory concerns	The Journal of International Maritime Law	Marine Policy, Regulation & Navigation
14	Relling et al. (2021)	The contribution of Vessel Traffic Services to safe coexistence between automated and convention	Maritime Policy & Management	Marine Policy, Regulation & Navigation
15	Kooij and Hekkenberg (2021)	Identification of a task-based implementation path for unmanned autonomous ships	Maritime Policy & Management	Marine Policy, Regulation & Navigation
16	Nzengu et al. (2021)	Regulatory framework analysis for the unmanned inland vessel	WMU Journal of Maritime Affairs	Marine Policy, Regulation & Navigation
17	Peeters et al. (2020a, 2020b)	An inland shore control centre for monitoring or controlling unmanned inland cargo vessels	Journal of Marine Science and Engineering	Ocean Engineering
18	Fossen and Fossen (2018)	eXogenous Kalman Filter for visualization and motion prediction of ships using live AIS data	Modeling, Identification and Control	Ocean Engineering
19	Peeters et al. (2020a, 2020b)	An unmanned inland cargo vessel: Design, build, and experiments	Ocean Engineering	Ocean Engineering
20	Wu et al. (2021)	Sailing status recognition to enhance safety awareness and path routing for a commuter ferry	Ships and Offshore Structures	Ocean Engineering
21	Rødseth et al. (2021)	Towards approval of autonomous ship systems by their operational envelope	Journal of Marine Science and Technology	Ocean Engineering
22	Yoo and Lee (2021)	Collision risk assessment support system for MASS, RO and VTSO support	Journal of Marine Science and Technology	Ocean Engineering
23	Wu et al. (2021)	Navigating patterns analysis for on-board guidance support in crossing collision avoidance operations	IEEE Intelligent Transportation Systems Magazine	Ocean Engineering
24	Ramos et al. (2020a, 2020b)	Human-system concurrent task analysis for Maritime Autonomous Surface Ship operation and safety	Reliability Engineering and System Safety	Reliability Engineering
25	Thieme and Utne (2017)	Safety performance monitoring of autonomous marine systems	Reliability Engineering and System Safety	Reliability Engineering
26	Wróbel et al. (2018)	System-theoretic approach to safety of remotely-controlled merchant vessel	Ocean Engineering	Reliability Engineering
27	Utne et al. (2020)	Towards supervisory risk control of autonomous ships	Reliability Engineering and System Safety	Reliability Engineering
28	Reddy et al. (2019)	Zero-emission autonomous ferries for urban water transport: cheaper, cleaner alternative to bridges and manned vessels	IEEE Electrification Magazine	Reliability Engineering
29	Huang et al. (2020)	A ship collision avoidance system for human–machine cooperation during collision avoidance	Ocean Engineering	Reliability Engineering

(continued on next page)

(*continued*)

ID	Author, year	Title	Journal	Category
30	Eriksen et al. (2021)	An RCM approach for assessing reliability challenges and maintenance needs of unmanned cargo ships	Reliability Engineering and System Safety	Reliability Engineering
31	Valdez Banda et al. (2021)	A systemic hazard analysis and management process for the concept design phase of an autonomous vessel	Reliability Engineering and System Safety	Reliability Engineering
32	Bačkalov (2020)	Safety of autonomous inland vessels: An analysis of regulatory barriers in the present technical standards in Europe	Safety Science	Safety & Risk
33	Fan et al. (2020)	A framework to identify factors influencing navigational risk for Maritime Autonomous Surface Ships	Ocean Engineering	Safety & Risk
34	Ramos et al. (2020a, 2020b)	A generic approach to analysing failures in human-system interaction in autonomy	Safety Science	Safety & Risk
35	Zhang et al. (2020)	A probabilistic model of human error assessment for autonomous cargo ships focusing on human–autonomy collaboration	Safety Science	Safety & Risk
36	Thieme et al. (2018)	Assessing ship risk model applicability to Marine Autonomous Surface Ships	Ocean Engineering	Safety & Risk
37	Ramos et al. (2019)	Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events	Safety Science	Safety & Risk
38	Chang et al. (2020)	Risk assessment of the operations of maritime autonomous surface ships	Reliability Engineering and System Safety	Safety & Risk
39	Zhou et al. (2021)	A system-theoretic approach to safety and security co-analysis of autonomous ships	Ocean Engineering	Safety & Risk
40	Bolbot et al. (2021)	A novel risk assessment process-application to an autonomous inland waterways ship	Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability	Safety & Risk
41	Guo et al. (2021)	Risk assessment of collisions of an autonomous passenger ferry	Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability	Safety & Risk
42	Fan et al. (2021)	Towards a Framework of Operational-Risk Assessment for a Maritime Autonomous Surface Ship	Energies	Safety & Risk

References

Ahvenjärvi, S., 2016. The human element and autonomous ships. TransNav: Int. J. Marine Navigation Saf. Sea Transp. 10.

Allianz, 2020. Safety and Shipping Review 2020: An annual review of trends and developments in shipping losses and safety. Allianz Global Corporate & Specialty, Munich, Germany.

Anderson, J.R., 1978. Arguments concerning representations for mental imagery. Psychol. Rev. 85, 249–277. https://doi.org/10.1037/0033-295X.85.4.249.

Bačkalov, I., 2020. Safety of autonomous inland vessels: An analysis of regulatory barriers in the present technical standards in Europe. Saf. Sci. 128, 104763 https://doi.org/10.1016/j.ssci.2020.104763.

Baldauf, M., Fischer, S., Kitada, M., Mehdi, R., Al-Quhali, M., Fiorini, M., 2019. Merging conventionally navigating ships and MASS-Merging VTS, FOC and SCC? TransNav: Int. J. Marine Navigation Saf. Sea Transp. 13.

Bolbot, V., Theotokatos, G., Andreas Wennersberg, L., Faivre, J., Vassalos, D., Boulougouris, E., Jan Rødseth, Ø., Andersen, P., Pauwelyn, A.-S., Van Coillie, A., 2021. A novel risk assessment process: Application to an autonomous inland waterways ship. Proc. Institution Mech. Eng., Part O: J. Risk Reliab. 1748006X211051829.

Burmeister, H.-C., Bruhn, W., Rødseth, Ø.J., Porathe, T., 2014. Autonomous unmanned merchant vessel and its contribution towards the e-Navigation implementation: The MUNIN perspective. Int. J. e-Navigation Maritime Economy 1, 1–13.

CCNR, 2021. Central Commission for the Navigation of the Rhine - Automation [WWW Document]. URL https://www.ccr-zkr.org/12050000-en.html (accessed 11.22.21).

Chang, C.-H., Kontovas, C., Yu, Q., Yang, Z., 2020. Risk assessment of the operations of maritime autonomous surface ships. Reliab. Eng. Syst. Saf. 207, 107324.

Chang, Y.H.J., Mosleh, A., 2007. Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: Part 1: Overview of the IDAC Model. Reliab. Eng. Syst. Saf. 92, 997–1013. https://doi.org/10.1016/j.ress.2006.05.014.

Chen, Z., Chen, D., Zhang, Y., Cheng, X., Zhang, M., Wu, C., 2020. Deep learning for autonomous ship-oriented small ship detection. Saf. Sci. 130, 104812 https://doi.org/10.1016/j.ssci.2020.104812.

Cheng, Y., Zhang, W., 2018. Concise deep reinforcement learning obstacle avoidance for underactuated unmanned marine vessels. Neurocomputing 272, 63–73. https://doi.org/10.1016/j.neucom.2017.06.066.

Chung, M., Kim, S., Lee, K., Shin, D.H., 2020. Detection of damaged mooring line based on deep neural networks. Ocean Eng. 209, 107522 https://doi.org/10.1016/j.oceaneng.2020.107522.

Colling, A., Hekkenberg, R., van Hassel, E., 2021. A Viability Study of Waterborne Platooning on the Lower Rhine. Eur. J. Transp. Infrastruct. Res. 21, 71–94.

Dallolio, A., Agdal, B., Zolich, A., Alfredsen, J.A., Johansen, T.A., 2019. Long-Endurance Green Energy Autonomous Surface Vehicle Control Architecture, in: OCEANS 2019 MTS/IEEE SEATTLE. pp. 1–10. https://doi.org/10.23919/OCEANS40490.2019.8962768.

Diaper, D., 2004. Understanding Task Analysis for Human-Computer Interaction. In: Diaper, D., Stanton, N.A. (Eds.) The Handbook of Task Analysis for Human-Computer Interaction, LEA Publishers, Mahwah, NJ.

Gl, D.N.V., 2018. Remote-controlled and autonomous ships in the maritime industry (Position Paper). Hamburg, Germany.

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V., 2017. CARLA: An open urban driving simulator, in: Proceedings of Machine Learning Research. In: Presented at the Proceedings of the 1st Annual Conference on Robot Learning, pp. 1–16.

Dunbabin, M., Grinham, A., Udy, J., 2009. In: An Autonomous Surface Vehicle for Water Quality Monitoring. Australian Robotics and Automation Association, Sydney, Australia, pp. 1–6.

Endsley, M.R., 2015. Situation awareness misconceptions and misunderstandings. J. Cogn. Eng. Decision Making 9, 4–32.

Endsley, M.R., 1995. Toward a theory of situation awareness in dynamic systems. Hum. Factors 37, 32–64.

Eriksen, S., Utne, I.B., Lützen, M., 2021. An RCM approach for assessing reliability challenges and maintenance needs of unmanned cargo ships. Reliab. Eng. Syst. Saf. 210, 107550 https://doi.org/10.1016/j.ress.2021.107550.

Evans, J.H., 1959. Basic design concepts. J. Am. Soc. Naval Engineers 71, 671–678. https://doi.org/10.1111/j.1559-3584.1959.tb01836.x.

Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., Prakash, A., Kohno, T., Song, D., 2018. Robust physical-world attacks on deep learning visual classification. In: Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1625–1634.

Fan, C., Montewka, J., Zhang, D., 2021. Towards a Framework of Operational-Risk Assessment for a Maritime Autonomous Surface Ship. Energies 14, 3879.

Fan, C., Wróbel, K., Montewka, J., Gil, M., Wan, C., Zhang, D., 2020. A framework to identify factors influencing navigational risk for Maritime Autonomous Surface Ships. Ocean Eng. 202, 107188.

Flach, J.M., 2015. Situation awareness: Context matters! A commentary on Endsley. J. Cogn. Eng. Decision Making 9, 59–72.

Grote, G., 2020. Safety and autonomy: A contradiction forever? Saf. Sci. 127, 104709 https://doi.org/10.1016/j.ssci.2020.104709.

Grote, G., Weyer, J., Stanton, N.A., 2014. Beyond human-centred automation – concepts for human–machine interaction in multi-layered networks. Ergonomics 57, 289–294. https://doi.org/10.1080/00140139.2014.890748.

Gunning, D., Aha, D., 2019. DARPA's Explainable Artificial Intelligence (XAI) Program. AIMag 40, 44–58. https://doi.org/10.1609/aimag.v40i2.2850.

Guo, C., Haugen, S., Utne, I.B., 2021. Risk assessment of collisions of an autonomous passenger ferry. Proc. Institution Mech. Engineers, Part O: J. Risk Reliability, 1748006X211050714.

Chiang, H.L., Tapia, L., 2018. COLREG-RRT: An RRT-Based COLREGS-Compliant Motion Planner for Surface Vehicle Navigation. IEEE Rob. Autom. Lett. 3, 2024–2031. https://doi.org/10.1109/LRA.2018.2801881.

Hannaford, E., Hassel, E.V., 2021. Risks and Benefits of Crew Reduction and/or Removal with Increased Automation on the Ship Operator: A Licensed Deck Officer's Perspective. Appl. Sci. 11, 3569.

Hansen, H.L., Nielsen, D., Frydenberg, M., 2002. Occupational accidents aboard merchant ships. Occup. Environ. Med. 59, 85. https://doi.org/10.1136/oem.59.2.85.

Heath, C., Luff, P., 1991. Collaborative Activity and Technological Design: Task Coordination in London Underground Control Rooms. In: Bannon, L., Robinson, M.,

Schmidt, K. (Eds.), Proceedings of the Second European Conference on Computer-Supported Cooperative Work ECSCW '91. Springer Netherlands, Dordrecht, pp. 65–80. https://doi.org/10.1007/978-94-011-3506-1_5.
Hetherington, C., Flin, R., Mearns, K., 2006. Safety in shipping: The human element. J. Saf. Res. 37, 401–411. https://doi.org/10.1016/j.jsr.2006.04.007.
Huang, Y., Chen, L., Negenborn, R.R., van Gelder, P.H.A.J.M., 2020. A ship collision avoidance system for human-machine cooperation during collision avoidance. Ocean Eng. 217, 107913 https://doi.org/10.1016/j.oceaneng.2020.107913.
Hutchins, E., 1995. Cognition in the Wild. MIT press.
IALA, 2019. Position on the development of aids to navigation services (No. C70-8.2.1.2). International Association of Marine Aids to Navigation and Lighthouse Authorities.
Im, N., Seo, J.-H., 2010. Ship manoeuvring performance experiments using a free running model ship. TransNav Int. J. Marine Navigation Saf. od Sea Transp. 4.
IMO, 2021. Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS) (No. MSC.1/Circ.1638). London, UK.
IMO, 2019. Maritime Safety [WWW Document]. URL https://www.imo.org/en/OurWork/Safety/Pages/default.aspx (accessed 4.27.21).
IMO, 2018. IMO takes first steps to address autonomous ships [WWW Document]. URL http://www.imo.org/en/MediaCentre/PressBriefings/Pages/08-MSC-99-MASS-scoping.aspx (accessed 9.9.20).
IMO, 2017. International Convention on Standards ofTraining, Certification and Watchkeeping for Seafarers (STCW) 1978, as amended in 1995/2010.
IMO, 2013. Guidelines for the approval of alternatives and equivalents as provided for in various IMO instruments (No. MSC.1/Circ.1455). IMO, London, UK.
Kari, R., Steinert, M., 2021. Human factor issues in remote ship operations: Lesson learned by studying different domains. J. Marine Sci. Eng. 9 https://doi.org/10.3390/jmse9040385.
Katsivela, M., 2020. Unmanned Vessels and Regulatory Concerns. J. Int. Maritime Law.
Kim, T., Mallam, S., 2020. A Delphi-AHP study on STCW leadership competence in the age of autonomous maritime operations. WMU J. Maritime Affairs 19, 163–181.
Kimball, P., Bailey, J., Das, S., Geyer, R., Harrison, T., Kunz, C., Manganini, K., Mankoff, K., Samuelson, K., Sayre-McCord, T., Straneo, F., Traykovski, P., Singh, H., 2014. The WHOI Jetyak: An autonomous surface vehicle for oceanographic research in shallow or dangerous waters. In: 2014 IEEE/OES Autonomous Underwater Vehicles (AUV). pp. 1–7. https://doi.org/10.1109/AUV.2014.7054430.
Klein, G.A., 2017. Sources of power: How people make decisions. MIT press.
Kobayashi, H., Kohshima, S., 2001. Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. J. Hum. Evol. 40, 419–435. https://doi.org/10.1006/jhev.2001.0468.
Kobayashi, H., Kohshima, S., 1997. Unique morphology of the human eye. Nature 387, 767–768. https://doi.org/10.1038/42842.
Kongsberg, 2020a. Autonomous ship project, key facts about YARA Birkeland [WWW Document]. URL https://www.kongsberg.com/maritime/support/themes/autonomous-ship-project-key-facts-about-yara-birkeland/ (accessed 11.8.21).
Kongsberg, 2020b. First adaptive transit on Bastøfosen VI [WWW Document]. URL https://www.kongsberg.com/maritime/about-us/news-and-media/news-archive/2020/first-adaptive-transit-on-bastofosen-vi/ (accessed 4.27.21).
Kooij, C., Hekkenberg, R., 2021. Identification of a task-based implementation path for unmanned autonomous ships. Maritime Policy Manage. https://doi.org/10.1080/03088839.2021.1914878.
Låg, S., 2019. ROMAS - Remote Operations of Machinery and Automation Systems - experiences from pilot testing.
Lieder, F., Griffiths, T.L., 2020. Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. Behav. Brain Sci. 43, e1 https://doi.org/10.1017/S0140525X1900061X.
Lloyd's Register, 2016. Cyber-enabled ships: ShipRight procedure - autonomous ships (Guidance Document). Lloyd's Register Group, Southampton, UK.
Lunde-Hanssen, L.S., Braseth, A.O., Strand, S., 2020. Identification of information requirements in ROC operations room (No. IFE/E-2020/007). IFE, Halden, Norway.
Lützhöft, M.H., Dekker, S.W.A., 2002. On Your Watch: Automation on the Bridge. J. Navig. 55, 83–96. https://doi.org/10.1017/S0373463301001588.
Man, Y., 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.
McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E., 2006. A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. AIMag 27, 12. https://doi.org/10.1609/aimag.v27i4.1904.
MiT, 2020. Roboat project [WWW Document]. roboat.org. URL http://www.roboat.org (accessed 11.19.20).
Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS med. 6, e1000097.
Moravec, H., 1988. Mind children: The future of robot and human intelligence. Harvard University Press.
Munim, Z.H., Saha, R., Schøyen, H., Ng, A.K., Notteboom, T.E., 2021. Autonomous ships for container shipping in the Arctic routes. J. Mar. Sci. Technol. 1–15.
Akhtar, N., Mian, A., 2018. Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey. IEEE Access 6, 14410–14430. https://doi.org/10.1109/ACCESS.2018.2807385.
National Transportation Safety Board, 2020. Collision Between a Sport Utility Vehicle Operating With Partial Driving Automation and a Crash Attenuator, Mountain View, California, March 23, 2018 (Accident Report No. NTSB/HAR-20/01). National Transportation Safety Board, Washington, D.C., USA.
Nicas, J., Kitroeff, N., Gelles, D., Glanz, J., 2019. Boeing built deadly assumptions into 737 Max, blind to a late design change. The New York Times.
NTNU, 2021. SFI AutoShip - NTNU [WWW Document]. URL https://www.ntnu.edu/sfi-autoship (accessed 3.21.21).
NVivo, 2020. . QSR International.
Nzengu, W., Faivre, J., Pauwelyn, A.-S., Bolbot, V., Lien Wennersberg, L.A., Theotokatos, G., 2021. Regulatory framework analysis for the unmanned inland waterway vessel. WMU J. Maritime Affairs. https://doi.org/10.1007/s13437-021-00237-z.
Paavola, S., 2006. On the origin of ideas: An abductivist approach to discovery.
Parasuraman, R., Manzey, D.H., 2010. Complacency and bias in human use of automation: An attentional integration. Hum. Factors 52, 381–410.
Parasuraman, R., Riley, V., 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. Hum. Factors 39, 230–253. https://doi.org/10.1518/001872097778543886.
Peeters, G., Kotzé, M., Afzal, M.R., Catoor, T., Van Baelen, S., Geenen, P., Vanierschot, M., Boonen, R., Slaets, P., 2020a. An unmanned inland cargo vessel: Design, build, and experiments. Ocean Eng. 201, 107056.
Peeters, G., Yayla, G., Catoor, T., Van Baelen, S., Afzal, M.R., Christofakis, C., Storms, S., Boonen, R., Slaets, P., 2020b. An Inland Shore Control Centre for Monitoring or Controlling Unmanned Inland Cargo Vessels. J. Marine Sci. Eng. 8, 758.
Zaccone, R., Martelli, M., Figari, M., 2019. A COLREG-Compliant Ship Collision Avoidance Algorithm. In: European Control Conference (ECC). Presented at the 2019 18th European Control Conference (ECC), pp. 2530–2535.
Ramos, M., Thieme, C.A., Utne, I.B., Mosleh, A., 2020a. Human-system concurrent task analysis for maritime autonomous surface ship operation and safety. Reliab. Eng. Syst. Saf. 195, 106697.
Ramos, M., Utne, I.B., Mosleh, A., 2019. Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events. Saf. Sci. 116, 33–44. https://doi.org/10.1016/j.ssci.2019.02.038.
Ramos, M.A., Thieme, C.A., Utne, I.B., Mosleh, A., 2020b. A generic approach to analysing failures in human–System interaction in autonomy. Saf. Sci. 129, 104808.
Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem. Saf. Sci. 27, 183–213. https://doi.org/10.1016/S0925-7535(97)00052-0.
Reddy, N.P., Zadeh, M.K., Thieme, C.A., Skjetne, R., Sorensen, A.J., Aanondsen, S.A., Breivik, M., Eide, E., 2019. Zero-Emission Autonomous Ferries for Urban Water Transport: Cheaper, Cleaner Alternative to Bridges and Manned Vessels. IEEE Electrif. Mag. 7, 32–45.
Relling, T., Lützhöft, M., Ostnes, R., Hildre, H.P., 2021. The contribution of Vessel Traffic Services to safe coexistence between automated and conventional vessels. Maritime Policy Manage. https://doi.org/10.1080/03088839.2021.1937739.
Reyhanoglu, M., 1997. Exponential stabilization of an underactuated autonomous surface vessel. Automatica 33, 2249–2254. https://doi.org/10.1016/S0005-1098(97)00141-6.
Rødseth, Ø.J., 2017. Definitions for Autonomous Merchant Ships 22.
Rødseth, Ø.J., Burmeister, H.-C., 2015. Risk assessment for an unmanned merchant ship. TransNav Int. J. Marine Navigation Safety Od Sea Transp. 9.
Rødseth, Ø.J., Kvamstad, B., Porathe, T., Burmeister, H.-C., 2013. Communication architecture for an unmanned merchant ship. In: Presented at the 2013 MTS/IEEE OCEANS-Bergen, pp. 1–9.
Rødseth, Ø.J., Wennersberg, L.A.L., Nordahl, H., 2021. Towards approval of autonomous ship systems by their operational envelope. J. Mar. Sci. Technol. 1–10.
Røed, W., Mosleh, A., Vinnem, J.E., Aven, T., 2009. On the use of the hybrid causal logic method in offshore risk analysis. Reliab. Eng. Syst. Saf. 94, 445–455. https://doi.org/10.1016/j.ress.2008.04.003.
Rolls-Royce, 2018. Press releases [WWW Document]. Rolls-Royce and Finferries demonstrate world's first Fully Autonomous Ferry. URL https://www.rolls-royce.com/media/press-releases.aspx (accessed 4.18.21).
Rothblum, A.M., 2000. Human error and marine safety. Presented at the National Safety Council Congress and Expo, U.S. Coast Guard Research & Development Center, Orlando, FL, USA.
Rutledal, D., Relling, T., Resnes, T., 2020. It's not all about the COLREGs: a case-based risk study for autonomous coastal ferries. In: Presented at the IOP Conference Series: Materials Science and Engineering, IOP Publishing, p. 012016.
SAE International, 2017. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems. SAE International. https://doi.org/10.4271/J3016_201401.
Saha, R., 2021. Mapping competence requirements for future shore control center operators. Maritime Policy Manage. https://doi.org/10.1080/03088839.2021.1930224.
Sharma, A., Kim, T., 2021. Exploring technical and non-technical competencies of navigators for autonomous shipping. Maritime Policy Manage. https://doi.org/10.1080/03088839.2021.1914874.
Shneiderman, B., 2020. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. null 36, 495–504. https://doi.org/10.1080/10447318.2020.1741118.
Shneiderman, B., 2016. Opinion: The dangers of faulty, biased, or malicious algorithms requires independent oversight. Proc. Natl. Acad. Sci. USA 113, 13538–13540. https://doi.org/10.1073/pnas.1618211113.
Statheros, T., Howells, G., Maier, K.M., 2008. Autonomous Ship Collision Avoidance Navigation Concepts, Technologies and Techniques. J. Navig. 61, 129–142. https://doi.org/10.1017/S037346330700447X.
Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., 2016. Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence.
Størkersen, K.V., 2020. Safety management in remotely controlled vessel operations. Marine Policy. https://doi.org/10.1016/j.marpol.2020.104349.

Sullivan, B.P., Desai, S., Sole, J., Rossi, M., Ramundo, L., Terzi, S., 2020. Maritime 4.0 – Opportunities in Digitalization and Advanced Manufacturing for Vessel Development. Procedia Manuf. 42, 246–253. https://doi.org/10.1016/j.promfg.2020.02.078.
Swedberg, R., 2014. The art of social theory. Princeton University Press.
Sheridan, T.B., 2011. Adaptive Automation, Level of Automation, Allocation Authority, Supervisory Control, and Adaptive Control: Distinctions and Modes of Adaptation. IEEE Trans. Syst. Man Cybern. - Part A: Syst. Hum. 41, 662–667. https://doi.org/10.1109/TSMCA.2010.2093888.
Tam, C., Bucknall, R., Greig, A., 2009. Review of collision avoidance and path planning methods for ships in close range encounters. J. Navigation 62, 455.
Tesla, 2019. Tesla Vehicle Safety Report [WWW Document]. Tesla Vehicle Safety Report. URL https://www.tesla.com/VehicleSafetyReport (accessed 4.22.21).
Tetley, W., 2008. Marine Cargo Claims. Yvon Blais 877–888.
Thieme, C.A., Utne, I.B., 2017. Safety performance monitoring of autonomous marine systems. Reliab. Eng. Syst. Saf. 159, 264–275.
Thieme, C.A., Utne, I.B., Haugen, S., 2018. Assessing ship risk model applicability to Marine Autonomous Surface Ships. Ocean Eng. 165, 140–154.
Tomasello, M., Hare, B., Lehmann, H., Call, J., 2007. Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. J. Hum. Evol. 52, 314–320. https://doi.org/10.1016/j.jhevol.2006.10.001.
United Nations, 1982. United Nations Convention on the Law of the Sea (UNCLOS).
U.S. National Highway Traffic Safety Administration, 2015. Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey (No. DOT HS 812 115). U.S. Department of Transportation, Washington, D.C., USA.
Utne, I.B., Rokseth, B., Sørensen, A.J., Vinnem, J.E., 2020. Towards supervisory risk control of autonomous ships. Reliab. Eng. Syst. Saf. 196, 106757 https://doi.org/10.1016/j.ress.2019.106757.
Utne, I.B., Sørensen, A.J., Schjølberg, I., 2017. Risk Management of Autonomous Marine Systems and Operations. In: Presented at the ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering, American Society of Mechanical Engineers Digital Collection. https://doi.org/10.1115/OMAE2017-61645.
Vagia, M., Rødseth, Ø.J., 2019. A taxonomy for autonomous vehicles for different transportation modes. Presented at the Journal of Physics: Conference Series, IOP Publishing, p. 012022.
Vagia, M., Transeth, A.A., Fjerdingen, S.A., 2016. A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed? Appl. Ergon. 53, 190–202. https://doi.org/10.1016/j.apergo.2015.09.013.
Valdez Banda, O.A., Kannos, S., Goerlandt, F., van Gelder, P.H.A.J.M., Bergström, M., Kujala, P., 2019. A systemic hazard analysis and management process for the concept design phase of an autonomous vessel. Reliab. Eng. Syst. Saf. 191 https://doi.org/10.1016/j.ress.2019.106584.
Varela, F.J., Thompson, E., Rosch, E., 2016. The embodied mind: Cognitive science and human experience. MIT press, Cambridge, MA.
Vasstein, K., Brekke, E.F., Mester, R., Eide, E., 2020. Autoferry Gemini: a real-time simulation platform for electromagnetic radiation sensors on autonomous ships. IOP Conf. Ser.: Mater. Sci. Eng. 929, 012032 https://doi.org/10.1088/1757-899x/929/1/012032.
Wang, W., Gheneti, B., Mateos, L.A., Duarte, F., Ratti, C., Rus, D., 2019. Roboat: An Autonomous Surface Vehicle for Urban Waterways. In: in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 6340–6347. https://doi.org/10.1109/IROS40897.2019.8968131.
Waymo, 2020. Waymo Safety Report. Mountain View, CA, USA.
Weick, K.E., 1976. Educational organizations as loosely coupled systems. Adm. Sci. Q. 1–19.
Wróbel, K., 2021. Searching for the origins of the myth: 80% human error impact on maritime safety. Reliab. Eng. Syst. Saf. 107942.
Wróbel, K., Gil, M., Chae, C.-J., 2021. On the influence of human factors on safety of remotely-controlled merchant vessels. Appl. Sci. (Switzerland) 11, 1–17. https://doi.org/10.3390/app11031145.
Wróbel, K., Gil, M., Montewka, J., 2020. Identifying research directions of a remotely-controlled merchant ship by revisiting her system-theoretic safety control structure. Saf. Sci. 129, 104797.
Wróbel, K., Montewka, J., Kujala, P., 2018. System-theoretic approach to safety of remotely-controlled merchant vessel. Ocean Eng. 152, 334–345.
Wróbel, K., Montewka, J., Kujala, P., 2017. Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. Reliab. Eng. Syst. Saf. 165, 155–169.
Wu, B., Li, G., Wang, T., Hildre, H.P., Zhang, H., 2021a. Sailing status recognition to enhance safety awareness and path routing for a commuter ferry. Ships Offshore Struct. 1–12.
Wu, B., Li, G., Zhao, L., Aandahl, H.-I.J., Hildre, H.P., Zhang, H., 2021b. Navigating patterns analysis for on-board guidance support in crossing collision avoidance operations.
Wu, Q., Wang, T., Diaconeasa, M.A., Mosleh, A., Wang, Y., 2020. A Comparative Assessment of Collision Risk of Manned and Unmanned Vessels. J. Marine Sci. Eng. 8, 852.
Yara, 2017. YARA and KONGSBERG enter into partnership to build world's first autonomous and zero emissions ship | Yara International [WWW Document]. URL https://www.yara.com/corporate-releases/yara-and-kongsberg-enter-into-partnership-to-build-worlds-first-autonomous-and-zero-emissions-ship/ (accessed 11.8.21).
Yoo, Y., Lee, J.-S., 2021. Collision Risk Assessment Support System for MASS RO and VTSO Support in Multi-Ship Environment of Vessel Traffic Service Area. J. Marine Sci. Eng. 9, 1143.
Yoshida, M., Shimizu, E., Sugomori, M., Umeda, A., 2021. Identification of the relationship between maritime autonomous surface ships and the operator's mental workload. Appl. Sci. (Switzerland) 11, 1–23. https://doi.org/10.3390/app11052331.
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, 8751.
Zhang, M., Zhang, D., Yao, H., Zhang, K., 2020. A probabilistic model of human error assessment for autonomous cargo ships focusing on human–autonomy collaboration. Saf. Sci. 130, 104838.
Zhou, X.-Y., Liu, Z.-J., Wang, F.-W., Wu, Z.-L., 2021. A system-theoretic approach to safety and security co-analysis of autonomous ships. Ocean Eng. 222, 108569 https://doi.org/10.1016/j.oceaneng.2021.108569.

Article 2

Computer Supported Cooperative Work (CSCW)
https://doi.org/10.1007/s10606-022-09450-7

RESEARCH ARTICLE

Collaborative Work with Highly Automated Marine Navigation Systems

Erik Veitch*[1], Henrikke Dybvik[2], Martin Steinert[2] & Ole Andreas Alsos[1]
*[1]*Department of Design, Faculty of Architecture and Design, Norwegian University of Science and Technology (NTNU), Produktdesign, 341, Gløshaugen, Kolbjørn Hejes Vei 2 B, 7491 Trondheim, Norway (E-mail: erik.a.veitch@ntnu.no);* [2]*Department of Mechanical and Industrial Engineering, Faculty of Engineering, Norwegian University of Science and Technology (NTNU), Verkstedteknisk, P317, Gløshaugen, Richard Birkelandsvei 2B, 7034 Trondheim, Norway*

Accepted: 8 September 2022

Abstract. In navigation applications, Artificial Intelligence (AI) can improve efficiency and decision making. It is not clear, however, how designers should account for human cooperation when integrating AI systems in navigation work. In a novel empirical study, we examine the transition in the maritime domain towards higher levels of machine autonomy. Our method involved interviewing technology designers (n = 9) and navigators aboard two partially automated ferries (n = 5), as well as collecting field observations aboard one of the ferries. The results indicated a discrepancy between how designers construed human-AI collaboration compared to navigators' own accounts in the field. Navigators reflected upon their role as one of 'backup,' defined by ad-hoc control takeovers from the automation. Designers positioned navigators 'in the loop' of a larger control system but discounted the role of in-situ skills and heuristic decision making in all but the most controlled takeover actions. The discrepancy shed light on how integration of AI systems may be better aligned to human cooperation in navigation. This included designing AI systems that render computational activities more visible and that incorporate social cues that articulate human work in its natural setting. Positioned within the field of AI alignment research, the main contribution is a formulation of human-AI interaction design insights for future navigation and control room work.

Keywords: Collaborative work, Interaction design, Navigation, Human–computer interaction, Autonomous ships, Artificial intelligence, Control rooms

1 Introduction

High levels of machine autonomy and Artificial Intelligence (AI) have the potential to improve work efficiency and improve human decision making. McCarthy (2007) defined AI as 'the science and engineering of making intelligent machines,' and intelligence as 'the computational part of the ability to achieve

Published online: 08 October 2022

goals in the world.' Since the field's inception in the 1950s, one of the frontiers of AI research has been navigation. Navigation – the process of moving a vehicle from one place to another – exemplifies the primary goal of computational intelligence: the capacity to execute planned action, as if by its own agency. In this study, we examine a transition currently underway in maritime navigation – a transition characterized by increasingly high levels of machine autonomy and incorporation of AI tools designed to collaborate with skilled navigators. Given the breakthroughs in AI technology in the past decade, we explore the extent to which a new human–machine interface is at hand and the extent to which systems design must realign to demands underlying a new order of work.

Driven by advances in computational power and the availability of hardware, examples of high levels of autonomy and AI in maritime applications are becoming more commonplace. Autonomous Surface Vehicles (ASVs) are plying the oceans for scientific data (e.g., Dallolio et al., 2019; Dunbabin et al., 2009; Kimball et al., 2014), autonomous passenger ferries are offering new alternatives to urban mobility (e.g., Reddy et al., 2019; Wang et al., 2019; MiT, 2020; Reddy et al., 2019; Wang et al., 2019), and Maritime Autonomous Surface Ships (MASSs) are introducing new ways to transport payload more efficiently across integrated ports (e.g., Burmeister et al., 2014; Peeters et al., 2020) In this study, we look at the case of partially automated Roll-On/Roll-On (Ro-Ro) ferries operating in Norway, where navigators complete crossings and dockings at the press of a few buttons (e.g., Kongsberg, 2020; Rolls-Royce, 2018). Looking ahead, we can expect implementation of machine learning tools designed to aid navigators make decisions (e.g., Martinsen & Lekkas, 2018; Gjærum et al., 2021; Wu et al., 2021), and computer vision to identify targets and automatically avoid collisions (e.g., Brekke et al., 2019; Helgesen et al., 2022).

High levels of autonomy and AI in sociotechnical applications like navigation rely upon collaboration with skilled human operators. ASVs need remote supervision (Utne et al., 2020), urban autonomous passenger ferries need human safety hosts, (Goerlandt and Pulsifer, 2022), and MASSs need supervision and remote control (Veitch and Alsos, 2022). Aboard the partially automated ferries we study in this article, operations depend upon the presence of a navigator who remains responsible for the vessels and its passengers and stands ready to take over control from the automated system. Despite more advanced systems that automate human manual control tasks and support decision making, the transition underway is not one of less human involvement, as one might expect, but of more collaboration between machines and humans. For designers, such systems present significant challenges. Recent accidents in aviation and car automation serve as dramatic examples of how the transition to human–machine collaboration can lead to accidents. In the years 2018 and 2019, two Boeing 737 MAX crashes revealed that the flight crew fatally lost control when counteracting a non-existent stall. A faulty airflow sensor feeding inputs to the Maneuvering Characteristics

Augmentation System (MCAS) was to blame: an automated pitch controller that the flight crew did not know how to override due to its hasty implementation (Nicas et al., 2019). In another instance, a fatal Tesla 'Autopilot' crash was found to be caused by 'system limitations' combined with 'ineffective monitoring of driver engagement, which facilitated the driver's complacency and inattentiveness' (National Transportation Safety Board, 2020, p. 58). As expressed by a leading autonomous car company in their safety report: 'While the benefits of automation are obvious, it can actually become a problem if people get tired or bored from having too little to do' (Waymo, 2020, p. 37). Whether it is an airplane, car, or even a ship, those individuals in control are increasingly finding themselves in a supervisory role, a role that Brian Christian has provocatively called the 'sorcerer's apprentice.' 'We conjure a force, autonomous but totally compliant, give it a set of instructions, and scramble like mad when we realize our instructions are imprecise or incomplete' (Christian, 2020, p. 31). In sociotechnical systems like that exemplified by a ship, where control is not executed by a single person but a whole team acting as one (Hutchins, 1995), this role, defined by the crossover between human and machine control, presents new challenges when considering work as fundamentally social action.

The premise for our study is that increased collaboration with computationally intelligent machines places new demands on its human counterparts, and that these demands can be discovered through observation and data collections efforts. Framing the current period of transition in maritime navigation as an opportunity to study these new demands, our aim is to incorporate perspectives of navigators experiencing this transition into further design iterations. Motivated by the potential of machine autonomy to enhance work efficiency and improve decision making, we seek to contribute to system design featuring a more seamless interface for coordinating action.

2 Related literature

Drawing on computer science, engineering, design, human–computer interaction, and sociology, we explore how current knowledge gaps and issues are compelling a new research direction positioned at these disciplines' crossroads. The background literature we present here sets the stage for our study, deepening and expanding the discussion about how technology designers are shaping human-AI collaborative work.

2.1 Levels of autonomy and artificial intelligence

AI has no formal definition. Far from presenting a problem for the field's practitioners, though, this lack of definition has, in the eyes of its leading experts, been precisely what has driven the field forward (Stone et al., 2016). The sociologist Levi Strauss used the term ‘floating signifier’ to describe phenomena like AI which, in evading definition, strengthen its suggestive power (Lechte, 1994, p. 26). The consequence of such a suggestive power, however, is captured in the so-called ‘AI effect,’ which describes the tendency for any new technology produced by the field, once accepted, to cast off its claims to AI. AI, in this sense, is precisely what is under development. In the development of autonomous vehicles, which represents the field’s idyllic mission of imbuing agency in a computational object, traces of the AI effect can be detected in the taxonomies commonly adopted to establish ‘how autonomous’ a vehicle is. These ‘Levels of Autonomy’ (LoA) taxonomies are not binary (autonomous or not) as one might expect. Rather, LoAs are more like standardized yardsticks for the extent to which a vehicle’s agency is independent of the human driver’s. These taxonomies have their origin in road transportation (SAE International, 2017) and have more recently been developed for maritime transportation (IMO, 2018; Rødseth, 2017). While LoA taxonomies vary, their basic structures remain the same, laying out an integer scale starting at zero or one, which represents full human control, and extending incrementally to some number that represents full machine control. For the vast majority of technology developers, this top number, like the field that proposed it, is a floating signifier. Only the intermediate numbers, which presume a collaborative approach to the myriad actions involved in driving a vehicle, are considered feasible.

Despite the apparently intractable goals underlying machine autonomy, the field of AI has been remarkably productive in producing technologies and techniques enabling intermediate LoAs. The theoretical underpinnings of modern computational machine learning techniques like Deep Neural Network (DNNs) have been around for decades, but only in the past decade has computational power enabled their widespread use. Advancements in machine learning techniques, too, have rapidly advanced the field, including in areas like natural language processing, image and video classification and generation, planning, decision making, and integration of vision and robotics. In the face of such advancements, however, a major new challenge has arisen. As expressed in Stanford University’s ‘*AI100 Report*,’ the field’s most influential experts recognized that, ‘Perhaps the most inspiring challenge is to build machines that can cooperate and collaborate seamlessly with humans’ (Littman et al., 2021, p. 19). In response to this challenge, an active research community has sprung up. These researchers are dedicated to ‘AI alignment,’ and include not just computer engineers and designers, but also anthropologists and sociologists, safety specialists and organizational scientists. In the context of sociotechnical systems, like that exemplified by our focus on the transition in maritime navigation, there is a growing need for such multidisciplinary efforts to understand the implications of high levels of autonomy and AI in safety–critical work. We position our work within the efforts of AI alignment research, interpreting the transition underway as one necessitating a realignment of design practices with the social actions coordinating human work.

2.2 Centres of coordination

The supervisory role taking shape in the wake of higher levels of autonomy has generated interest in centres of coordination for autonomous vehicles. For maritime navigation, this is exemplified by the concept of land-based supervisory control of highly automated ships, variously referred to in the literature as 'shore control centre,' 'remote control centre,' or 'remote operating centre.' These terms, which have surfaced in the past decade, capture a renewed interest in control rooms. Control rooms were a topic of academic interest in human factors and cognitive engineering in the 1970s and 80s especially in the context of complex, sociotechnical systems like nuclear power plants (Rasmussen, 1986; Vicente, 1999). In the 1990s, control rooms were of academic interest in the field of Computer Supported Collaborative Work. Researchers in CSCW studied the sociality of computer use in natural settings like line control rooms (Heath & Luff, 1992), airline scheduling (Goodwin & Goodwin, 1996), and emergency dispatch (Whalen, 1995). Today, in the wake of technological developments enabling higher levels of autonomy, the spotlight is once again directed towards the control room, the stage upon which supervisory control and time-critical action is orchestrated, enabling the coordination of highly automated vehicles across distributed locations. In this context, we revisit Lucy Suchman's definition of 'centre of coordination':

> 'Centres of coordination are characterized in terms of participants' ongoing orientation to problems of space and time, involving the deployment of people and equipment across distances, according to a canonical timetable or the emergent requirements of rapid response to a time-critical situation.' (Suchman, 1997, p. 42)

For autonomous ships, the 'shore control centre' as a centre of coordination presents significant challenges to designers. The International Maritime Organization (IMO), the inter-governmental agency for standardisation of safety at sea, outlined such outstanding challenges in their 'Regulatory scoping exercise for the use of maritime autonomous surface ships (MASS)' (IMO, 2021). In their report, the highest priority issues concerned the role of the navigator working in a location separate from the ship environment. While navigators' responsibility for the safety of the ship remained unchanged, the environment in which they work was substituted by an information-rich landscape necessitating new skills and competencies (IMO, 2021, p. 8). In revisiting centres of coordination, we explore what concepts and theories that emerged from seminal control room studies remain relevant today, and what gaps emerge in the light of new technological and organizational developments.

2.3 Design for AI collaborative systems

The need for improved human–machine collaboration predicated by recent technological development has led to new frameworks adopting human-centred design principles to AI systems. Shneiderman (2020), for example, proposed a design framework for 'human-centred AI' based on the principles of safe, reliable, and trustworthy system interactions. The field of Human–Computer Interaction (HCI) has put forward practical guidelines for designers adapting to such frameworks (e.g., Amershi et al., 2019; Mahadevan et al., 2018). The rapidly growing field of explainable AI (XAI), too, focuses on the interaction between humans and machines, aiming to establish human-based values of interpretability and understandability at the core of 'black box' machine learning techniques (Voosen, 2017). Expanding the audience of XAI towards users, organisations, and even non-governmental agencies, Arrieta Barredo et al. (2020) envision a 'Responsible AI' initiative, which embraces values of fairness and accountability along with the mandate of model explainability at the core of XAI. The multidisciplinary field of 'machine behaviour' has also emerged recently, which sets out as its mission the empirical treatment of the ways in which human social interactions are modified by the introduction of intelligent machines (Rahwan et al., 2019). The field of CSCW, with its interest in computationally infused environments and enacted elements of work, also stands to offer distinct contributions to this discussion. Ethnomethodological works on social interactions during navigation and control of ships (Hutchins, 1995) and airplanes (Nevile, 2001) lay the theoretical groundwork for such contributions, while more recent discussions exploring 'ethical AI issues' (Fleischmann et al., 2019) and 'challenges in human-AI collaboration' (Park et al., 2019) pave the way for current research directions. The aim of our study continues in this vein, motivated by lack of knowledge about how the transition to higher levels of autonomy affects the social underpinnings enacting work in its natural setting.

2.4 Ironies of automation

Bainbridge (1983), writing in her seminal paper 'Ironies of Automation,' described the paradoxical decrease in human abilities resulting from machines designed to improve that very ability. Among human factors specialists the effect is well-known, but despite its articulation three decades ago, its consequences persist in modern system design. For example, skill degradation associated with automation emerged as a key factor in the high-profile crash of flight Air France flight 447 in 2009, which fatally stalled over the Atlantic Ocean after the automatic flight system handed control to the flight crew shortly after detecting faulty readings from an airspeed sensor. As the accident report stated, one of the contributing causes of the stall and resulting crash was 'The absence of any training, at high altitude, in manual aeroplane handling' (Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, 2012, p. 201). Consequently, guidelines now recommend pilots practice manual flying regularly, highlighting that 'continuous use of automation does not strengthen pilots' knowledge and skills in manual flight operation and in fact could lead to degradation of the pilots' ability to quickly recover the aircraft from an undesired state' (IATA, 2020, p. 5). The consequences of skill degradation are exacerbated in systems with high LoA that require timely and decisive preventative action from a skilled operator. Taking irony to be a poor premise for design, we consider instead how design activities better aligned to the needs of collaboration can avoid the pitfalls associated with automation-induced skill degradation.

3 Methodology

Our methodology consisted of field study observations and semi-structured interviews. The research design was motivated by the practical need to inform design efforts implementing high levels of machine autonomy and AI techniques in maritime navigation applications. The aim was to describe the extent to which design practices currently shaping a transition in the maritime domain are aligned with the realities of skilled, safety–critical work in the field.

3.1 Data collection

The empirical data consisted of semi-structured interviews with individuals in the design and research communities (n = 9) and navigators working aboard two partially automated passenger ferries (n = 5). To provide context about the natural setting in which the navigation work takes place, we also report on field observations conducted at the site of one of the ferries featured in the interviews. All data were collected in Norway.

Selection of interview participants was guided by theoretical sampling commonly employed in Grounded Theory Methods (Corbin and Strauss, 2015; Glaser and Strauss, 1999). This allowed us to follow up on themes of interest and target subsequent participants as new open-ended questions presented themselves. After completing nine interviews with technology designers and researchers in autumn 2019, it became clear that the perspective of navigators in the field would be of interest. Turning to this gap, a field study was conducted aboard the navigation deck of a ferry outfitted with state-of-the-art automated navigation technology. Field notes and images were collected by the first author, and once again new questions were posed. Interviews were subsequently held in summer 2021 with three of the navigators aboard this ferry, followed by two more aboard a similar ferry.

Two researchers were present for all interviews, with one leading the conversation while the other transcribed, verbatim. The first author was present for all interviews, which were conducted either in-person or via video call and ranging in length from 45 to 60 minutes. Participants consented to data collection before and after the interviews, and all interviews that were held in Norwegian were translated into English.

3.2 Interview respondents

A total of fourteen informants were interviewed, each of whom we refer to in this study with a pseudonym (Figure 1). The group whom we refer to as 'Designers' originated from academia, applied research, and industry. This group consisted of individuals with expert domain knowledge about the development of centres of coordination for highly automated ships. Their unique contribution was insights about activities shaping the transition towards higher levels of autonomy in marine navigation work. This group captured a wide breadth of perspectives

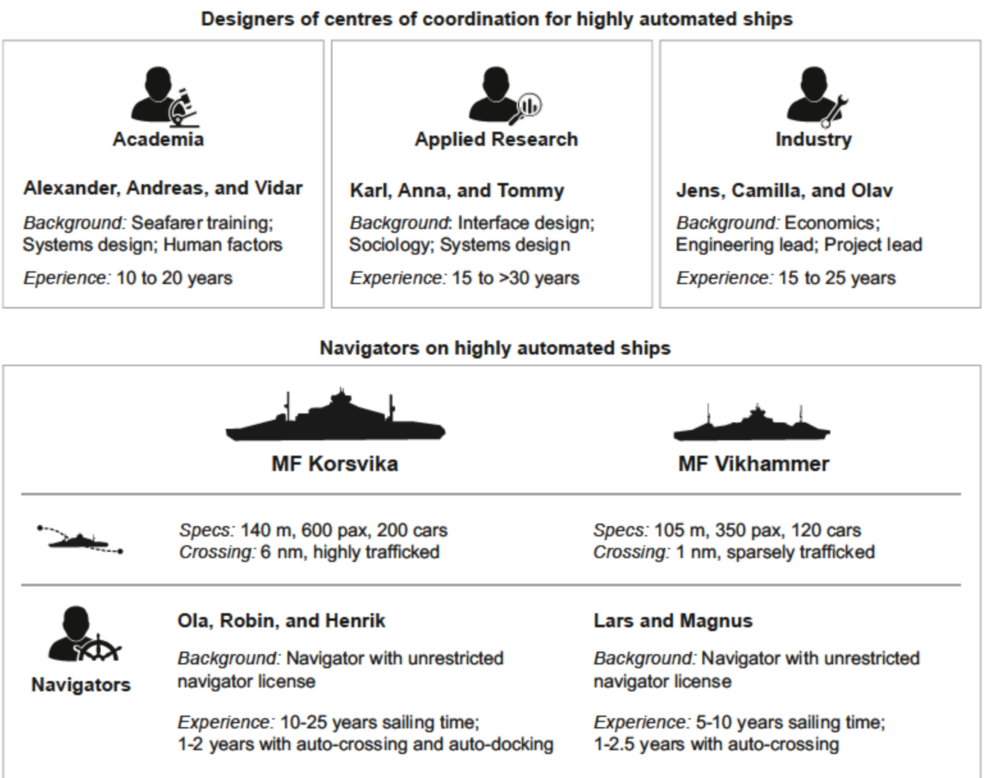


Figure 1 Description of interview informants

on technology development, holding different titles and originating from distinct professional networks separate from the networks held by the authors.

The group whom we refer to as 'Navigators' represented captains and chief mates working aboard two ferries outfitted with state-of-the-art automation technology. All navigators had a 'D1' deck officer license, the highest maritime navigation license in Norway. At the time of writing, the number of navigators working aboard partially automated ferries represented a small population. As such, we were careful to characterise them broadly to avoid de-anonymising them. The *Korsvika* (a pseudonym) was, at the time of this writing, the world's only ferry operating regularly with both auto-crossing and auto-docking, making it relatively easy to identify. It was on this ferry that field observations took place. There is a total of eight deck officers on the *Korsvika* and we interviewed three of them. The second ferry in our case study, called the *Vikhammer* (also a pseudonym), had just auto-crossing installed. The *Korsvika* and *Vikhammer* were owned and operated by different companies.

3.3 Data analysis

Our analytical approach was inspired by Grounded Theory Methods (Corbin and Strauss, 2015; Glaser and Strauss, 1999; Morse et al., 2009). Observations made on the navigation deck aboard the *Korsvika* also served an important role in the analysis, describing the context in which navigation work took place. During interviews and field observations, insights were recorded as 'memos:' dated text excerpts ranging from short notes to long, descriptive passages. No less than 101 memos were recorded in total, which served as precedents to a more structured analysis aimed at synthesizing these early insights.

In structuring the analysis, we used the software tool NVivo (NVivo, 2020). At its most fundamental level, the analysis comprised of 'codes' – units of highlighted text representing potentially relevant findings. Our analysis consisted of several hundred codes, which we assigned to categories called 'axial codes.' Special attention was afforded to retaining terms and phrases used by informants and to resisting re-interpretation in our own wording. For example, the term 'backup,' emerged as an important axial code. While only two navigators used the term expressly, the saliency of the theme was made apparent through other related codes (e.g., Ola: 'you become an operator who monitors the systems and is ready to press a button if there's a bug;' Henrik: 'When what you see on the screen no longer shows the correct thing, that's when things get interesting'). The axial coding process was iterative and was conducted by the first author and two graduate students, involving many rounds of discussion with the authors over the study period.

Eventually, we distilled our analysis into an overarching narrative structure, focused on the discrepancy between designers' construal of navigators' work and navigators' own reflective accounts. These findings are presented in Sect. 5. Before presenting this, however, it is necessary to provide some context to the findings. In Sect. 4 that follows, we outline the work activities making up a regular crossing aboard the *Korsvika*, constructed from first-hand field observations.

3.4 Methodological limitations

The empirical study consisted of both field observations and semi-structured interviews, lending our research design some distinct advantages as well as limitations. One advantage, for instance, presented itself from conducting independent, one-on-one interviews, as it led to the discovery of discrepancies between designers' and navigators' accounts of the same core activities. Similarly, this approach lent itself to making comparisons within groups. For example, when we compared accounts of navigators on different ferries, insights emerged linking their use of automation with skill degradation (Sect. 5.1). The conditions of confidentiality and anonymity, too, proved to be helpful in a way that field observation alone could not be. Informants were free to express their opinions without the potentially self-censoring effect of their colleagues' or managers' presence and reflected on their work activities as if observing them from the outside. Having interviewed the operators during a global COVID-19 pandemic, video conferencing provided a useful platform for data collection during social distancing.

4 Field observations aboard the *Korsvika*

In this section, we present field observations from the *Korsvika*. The *Korsvika* (a pseudonym) is the world's first ferry in regular service equipped with auto-crossing and auto-docking: two technologies representing a step change in the transition towards higher levels of machine autonomy in the maritime domain. For simplicity, we refer to the two technologies together as 'auto-systems.' The account that follows is a description of work during a regular crossing, as well as the environment of the navigation deck and the functionality of the auto-systems. The aim is to provide context about the roles, responsibilities, and tasks of the navigators, how these navigators interact with each other and the auto-systems, and how the adoption of higher levels of autonomy impacts their work activities. The diagram in Figure 2 can be used to orient the reader on the *Korsvika*'s navigation deck.

4.1 The *Korsvika*'s ferry service

The *Korsvika* connects vehicle traffic and foot passengers between two busy ports in Norway. The crossing takes less than 45 minutes. Operations are going smoothly when this 140-m-long roll-on/roll-off ferry, with capacity for almost 600 passengers and 200 cars, is on time with an even gap behind the other ferries that sail the same route. Because several ferries traverse the same crossing, issues can arise when one ferry is delayed, forcing the ferry behind to wait for it outside the dock. There are many factors that can affect the ferry's service, including the weather conditions and even the

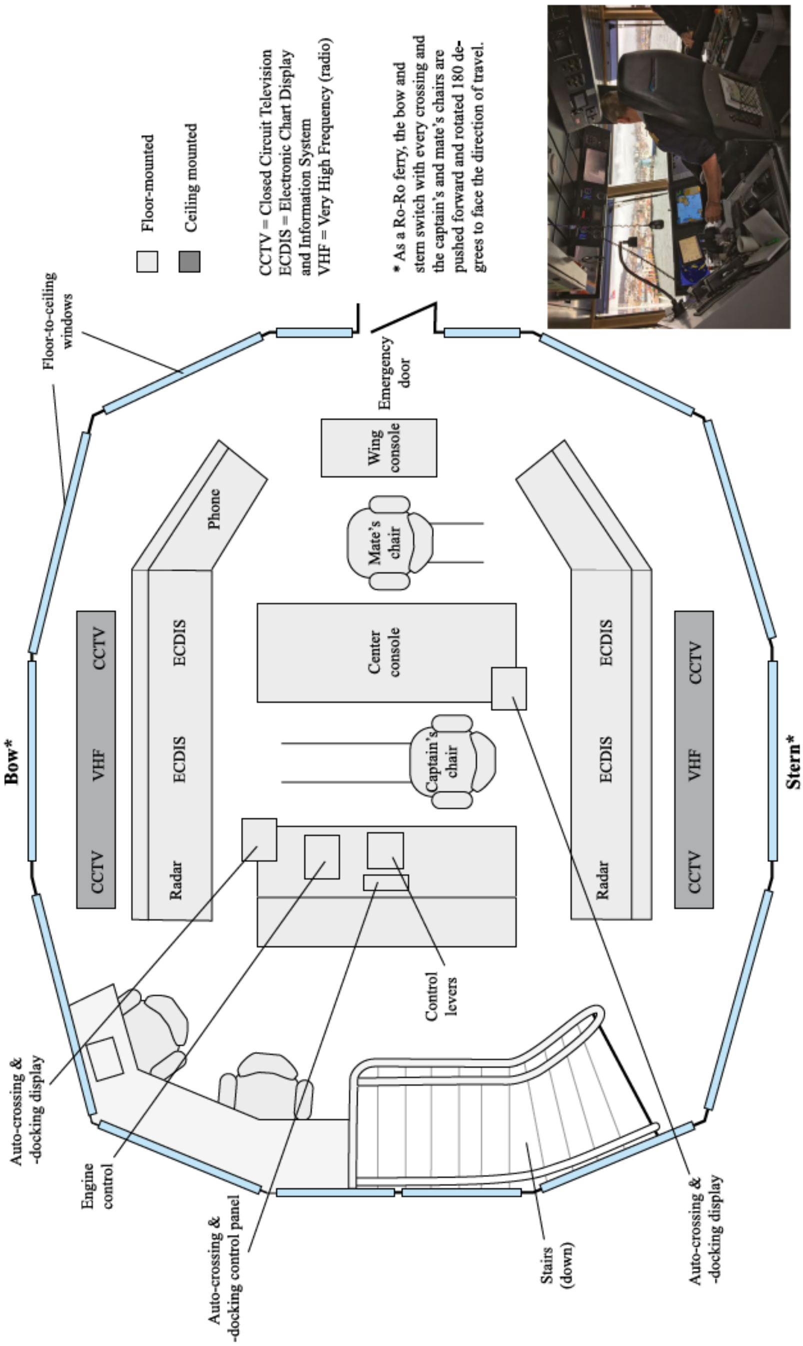


Figure 2 Layout of the bridge aboard the *Korsvika* (image taken by first author)

sailing styles of different navigators on duty. The new auto-systems installed aboard the *Korsvika* were intended to improve the efficiency of ferry service, saving fuel while providing customers with a more consistent service.

4.2 The ***Korsvika***'s crew

The captain has overall responsibility for the safety of passengers and crew. The chief mate (often shortened to 'mate') shares much of this responsibility. The captain and mate relieve one another's shifts throughout the ten-and-a-half hour working day, exchanging regular handovers in what the navigators call 'sharing a voyage.' Two bosuns handle the physical work on the main deck: loading, unloading, fitting cars, maintenance, and checks of safety equipment. One of the two navigators (captain or mate) communicates to the bosuns over a local radio and observes their actions from the bridge windows or on Closed Circuit Television (CCTV). The shipowner requires that two crew personnel must always be on the bridge, so after handovers between the captain and mate, a bosun comes up to the bridge and joins as a lookout. Other than the navigators and bosuns, the crew consists of a chief engineer, a mechanic, and cafeteria crew. Of all the crew aboard the *Korsvika*, the new auto-systems directly affect only the navigators' day-to-day work.

4.3 Loading and leaving dock

At the dock, the ferry loads vehicles and foot passengers. When loading is completed, the command 'Lift up!' is radioed to the bosuns, cueing them to close the ramp door and secure it for crossing.

Leaving the dock can be accomplished by the navigators either manually using thruster controls ('at the handles,' to use their terminology) or by pressing a button on the new auto-docking system. Currently, the auto-docking is used for 50–70% of all voyages.

Leaving the dock, the captain or mate reports their departure to the local Vessel Traffic Services (VTS) centre that they have left the dock, and VTS replies with any relevant information about traffic in the area. The navigator also keeps an eye out for small recreational boats, which are typically not detected by VTS. The new auto-systems are not yet equipped with cameras to detect possible collision targets, so the navigator must be attentive even when in auto-mode.

4.4 A regular crossing in 'auto-mode'

Shortly after the ferry is clear of the dock, auto-crossing is engaged by pressing the 'AUTO CROSS' button on the console. Nearby, on a small screen the size of a tablet computer, a touchscreen indicates that auto-crossing has been engaged and displays system information like thrust and heading. The handles on the thruster controllers move by themselves as the ship settles into its route and adjusts its speed for the crossing. The captain sits back in the chair and looks out the window, occasionally glancing at the Electronic Chart Display and Information System (ECDIS) and radar. The lookout sits in the mate's chair beside the captain, looking out the window and glancing occasionally at the captain.

Sometimes, small boats are encountered enroute. On weekends and summer holidays, there may be many such recreational boats in addition to regular commercial traffic. These small boats warrant special attention, because unlike commercial ships with trained crew, their occupants may be unfamiliar with the rules of navigation and may occasionally end up on a collision course. The auto-crossing is not yet capable of avoiding collisions. Avoiding collisions remains one of the core duties of the navigators. For larger ship traffic, whose navigators manoeuvre their vessels in accordance with Collision Regulation conventions (COLREGs), there are generally no issues avoiding collisions. Should a ship cross from either port or starboard, an agreement is usually made over the radio regarding who will adjust course or speed to pass behind the other, even if it is the give-way vessel that does so. In a give-way situation, the navigator takes over manual control by pressing the 'MANUAL' button on the auto-system console. Pulling back on the thruster, the other ship can cross ahead, whereafter the navigator can press 'AUTO CROSS,' resuming the crossing and losing little time to the timetable.

4.5 Arriving at dock and unloading

Approaching the dock, the auto-system alerts the crew with a loud beep followed by a pre-recorded voice announcing that docking is about to start. The alarm is acknowledged by the captain by pressing the 'AUTO DOCK' button that starts the auto-docking stage. Were the captain to ignore the alert, a safety measure is built in to stop the ferry in station-keeping mode, holding position some distance away from the dock.

As the ferry heads to the dock slowly under auto-docking control, the mate joins the captain (or vice-versa) in time for the docking sequence. At this point, the bosun who was on lookout duty during the crossing heads down to prepare for unloading. 'Betty's taking care of it,' announces the captain, using a nickname referring to the auto-docking system. The mate acknowledges, confirming they understood that the ferry is docking automatically.

At the dock, the captain communicates with the bosuns over radio and the ramp is lowered and unloading commences. Shortly after unloading, loading begins again. The captain's and mate's chairs are slid forward and rotated 180 degrees and the *Korsvika* sets out for its other port in the direction from where it came.

4.6 Higher levels of autonomy and centres of coordination

Currently, there are cameras installed in the *Korsvika* bow that record all marine traffic it encounters. Technology developers behind the auto-crossing and auto-docking initiatives are working towards enhancements; for example, they can use the recordings to train machine learning algorithms that can classify objects and be used in collision avoidance algorithms. As development of more advanced automation continues, there have been discussions about reducing crew aboard the ferries and controlling fleets of highly automated ferries from a land-based centre of coordination. Higher levels of autonomy have already proven successful on the *Korsvika*, improving the efficiency of fuel consumption and consistency of service in the face of highly variable external factors. Unlocking the potential benefits of higher levels of machine autonomy, though, depends on seamless integration of the AI systems with what is, at its heart, human work.

5 Interviews with designers and navigators

In this section, we present the findings of the interviews with navigators both aboard the *Korsvika* and the *Vikhammer*, as well as with technology designers and researchers shaping the transition towards higher levels of autonomy in the maritime domain. We start with the navigators, who recounted a shift to a 'backup' role subsequent to the introduction of auto-systems aboard their ships. Then, we compare this to accounts of the designers, whose construal of working with automated systems seemed misaligned with navigators' own accounts of working with automation in the field.

5.1 Navigators' perspectives: shifting to a backup role

The navigators attributed agency – a capacity for action – to the auto-systems. The influence of this agency was most evident in their descriptions of transitioning from 'hands-on' to 'backup' navigation.

The nickname assigned to the auto-systems by some of the navigators ('Betty')exemplified how machine agency could be manifested. Betty could 'take care of it,' as Robin reported, referring to the complex process of docking the 1400-ton *Korsvika* to the dock. In fog, Betty was 'ingenious' given her ability to dock in zero visibility. Betty could be a 'nag,' however, and 'do weird things,' according to Ola, who, as if by way of assuring themselves, told us that 'she has no thoughts of her own.'

> Robin: 'My captain and I, if we're auto-docking, we say that "Betty's taking care of it." Then he knows that auto mode is on. If we have normal autopilot on then I say that "Betty's not taking care of it."'

The nickname 'Betty' was used by two of the five navigators we interviewed, both aboard the *Korsvika*. Traditionally a woman's name, Betty was chosen owing to the system's female voice announcements, played at intervals to announce stages of operations or to alert navigators' attention to some procedure or sequence. Personified in this way, the navigators described interactions with the auto-system in human terms.

The agency attributed to the auto-systems underpinned the emerging 'backup' role described by the navigators. We adopted the term 'backup' from Robin, who, describing a transition in their work in recent years, said, 'We are the backup if something happens.' Other navigators described a similar role. 'You go from being the one who performs something to just monitoring something,' said Henrik of the transition, 'but when what you see on the screen no longer shows the correct thing, that's when things get interesting.'

One limitation of the auto-crossing was that it did not yet have automated collision avoidance capabilities, meaning such manoeuvres were left to navigators. Collision avoidance manoeuvres are regulated in the 1972 Convention on the International Regulations Preventing Collisions at Sea (COLREGs). The convention lays out traffic rules, like Rule 8 stating that collision avoidance actions must be 'made in ample time and with due regard to the observance of good seamanship' (IMO, 1972). Rules work best if everyone knows them, which is not always the case. 'The biggest problem is with small boats and sailboats,' Ola reported. 'They don't have the same knowledge about rules, speed, and direction,' explained Henrik. 'They think we move slower,' said Robin, 'so we have to press "MANUAL" … you don't want to run someone over.' In collision avoidance, the navigators' backup role to the auto-systems was clearly defined: take over control to adhere to the COLREGs, with special attention to small boats. Another backup role emerged, however, with less clearly defined parameters. This was illustrated by Robin who recounted an instance when they took over control to make a crossing more comfortable for passengers:

> Robin: '… these days we [the navigators] say: if it's blowing, we steer manually. Auto-crossing can be used at any time, but manual mode is more comfortable for passengers.'
> Interviewer (Erik): 'You steer the ferry [manually] so it's more comfortable for passengers?'
> Robin: 'If you have rolling, people can fall and hurt themselves. Instead of rolling all the way over, I sail a little North and then a little South to go across the waves.'

Robin's interaction with the automated system in this case is not determined by safety–critical and timely intervention, but rather on the system's inability to account for comfort of passengers. Whether in taking over control to avoid hazardous traffic situations or simply to attend to passenger comfort, the shift from a hands-on role to a backup role underscored the most salient change in navigator work after the auto-system's introduction.

One effect of shifting to a backup role was skill degradation associated with more time spent in a passive, monitoring role relative to hands-on, manual control. Skill degradation was especially apparent when comparing navigators' accounts from the *Korsvika*, who reported that 50–70% of crossings were in auto-mode, to the *Vikhammer*, who reported close to 100% automated crossings. As reported by Henrik, the crew of the *Korsvika* had taken to driving the ship manually 'at least twice per shift so as not to forget how that works.' This suggested that skill degradation set in quickly, possibly over the course of days, and that regular practice was an effective countermeasure. '…when I have driven a lot of auto,' said Henrik, 'I have to steer a couple of times myself to get the feel of it again.' Robin expanded on the subject, noting that operators' propensity for regular manual sailing practice resulting in it being incorporated into the shipowner's operating procedures:

> Robin: 'We've set it up so you'll sail it [the ferry] yourself during the day to maintain your driving. That's written in our procedures now. If you've had a holiday, you're allowed to steer the whole shift, there's no one that says you have to use auto-crossing.'

On the *Vikhammer*, in contrast, the crew had seldom sailed manually since the auto-crossing was implemented. This implied a more significant skill degradation, which might compromise safety in the eventuality of a manual takeover.

> Magnus: 'We only use auto-crossing now – every day, every trip.'
> Interviewer (Erik): 'Do you ever turn it off to take manual control?'
> Magnus: 'No.'
> Interviewer (Erik): 'When was the last time you drove manually?'
> Magnus: 'We might occasionally drive if we have an ambulance dispatched. Auto-crossing must have the lowest energy consumption, but with an ambulance it's life and health. Apart from that … it's been one-and-a-half years since I stopped doing it [driving manually] myself.'

Given how fast de-skilling was a factor among the crew of *Korsvika*, one cannot help but wonder if the crew of *Vikhammer* are prepared for an ambulance dispatch. Manual skill practice procedures, even in situations well-suited to the automation, appeared to be a useful countermeasure to skill degradation for the navigators aboard the *Korsvika*.

5.2 Designers' perspectives: prescribing action for distributed work

The interviews we held with technology designers and researchers yielded insights into how development activities are shaping the transition to increased human-AI collaboration in maritime navigation. Here we outline what this group identified as the most important design goals and what methods they are adopting to address interaction challenges between humans and machines. Then, we compare designers' construals of working with higher levels of machine autonomy with navigators' own corresponding accounts.

To begin, we outline some of the major design goals, the approaches being adopted in the industry and research communities, and what specific challenges represented outstanding gaps and issues. The main goals driving the transition towards higher levels of autonomy in the maritime domain included achieving improved 'logistics,' 'system design,' and 'centralized control.' These goals, it was envisioned, will be accomplished primarily through crew reduction relative to ship payload, as well as through centralized management of employees and ship assets from a centre of coordination. 'The whole problem statement,' said Vidar, 'can be defined as moving work farther from the pointy end to more distributed locations.' By 'pointy end,' Vidar referred to operational work in the field, a term coined by organizational scientist Rhonda Flin (Flin et al., 2008) and used often in the context of exposure to hazardous working environments. Asked to describe the vision of autonomous ships, interviewees described fleets of ships with reduced crew (or in some cases, no crew at all), whose whereabouts were tracked by trained operators in a centralized control centre. Prompted further to describe the control room, images of data-rich information displays were invoked in all interviews ('there will be large-screen displays displaying the "big picture,"' reported Karl; 'through the screen [the operators] will have access to the data they need,' said Andreas). Many of the technological artefacts located in a conventional ship were mentioned, including ECDIS, marine radio, and software for ship scheduling and voyage plans. What distinguished the control room from ship's bridge was the amount of additional data (e.g., video streams, sensor displays, and the like) and, crucially, the ability to take direct control over the ship. Here the analogy was made by five of the interviewees to VTS operators, who, tasked with monitoring traffic in busy port areas, can indirectly direct traffic by contacting navigation crew over radio. In a control room for highly automated ships, such actions could be taken directly instead of indirectly, making the control room operator effectively a remote captain in addition to traffic director.

Two interviewees described interactions at the screen interfaces in terms of 'top down' and 'bottom up' processing. As explained by Karl, this was intended to support decision making at the cognitive level, combining top-down processing ('information search') with bottom-up driven processes ('information that catches the attention of operators'). Two opposing viewpoints emerged, however: some interviewees argued that the control room should be designed to accommodate work as it takes place aboard a ship's bridge; others argued that the control room will require a ground-up approach, requiring specifications drawn up according to distinct requirements. Among the latter group, 'human-centred design,' 'prototyping,' and 'systems engineering' featured as methodologies to uncover these distinct requirements. Discussions about design strategies met a significant challenge: for highly automated ships, there were no standardised guidelines aimed for accommodating approval like those akin to conventional ships. Conventional passenger ship design, for example, is standardized according to design guidelines laid out by classification societies like DNV GL in their 'Rules for Classification' (DNV GL, 2017). For highly automated ships, adopting 'goal-based approaches' were, in place of prescriptive approaches, the most viable option towards approval of designs by regulating authorities. Characterizing this goal-based design process, five interviewees called it a 'transition,' involving testing, verification, and approval – lengthy processes typical in the highly regulated industry of shipping.

The technology for enabling high levels of machine autonomy, it appeared, was more or less available; orchestrating this technology in a real-world context, though, remained the challenge. In the boundary between human and machine, several gaps and issues were identified. The number of vessels, for instance, that each operator should control was unknown. This number was linked to the LoA of the vessel, but the LoA, too, was ambiguous, referencing various taxonomies each with its own configuration of how automated tasks and human tasks combine to navigate a ship. Specific LoA taxonomies mentioned by the interviewees included DNV GL (2018), IMO (2018), and NFAS (Rødseth, 2017). A central problem was the amount of time it takes to take over control. On the premise that such control takeovers are preventative and time-critical, the maximum allowable takeover time emerged as perhaps the single most important factor in goal-based design directed towards collaboration with the automated system.

> Tommy: 'You must quantify the person's response time. This will help a lot with the approval of a shore control centre, because then you can document, for example, that the system gives ten seconds warning and that we have done the research showing that the operators are trained for this. Today, nobody knows.'

5.3 Discrepancies between navigator and designer accounts

Comparing interviews of designers and researchers with those of navigators, certain discrepancies came to light. Two such discrepancies pointed to ways in which designers' construal of human–machine interaction diverged from those who inhabited this interface in their work. The first related to how the two groups treated decision making for control takeovers; the second related to how they reconciled their safety responsibilities while relinquishing tasks and decision making to machines.

Designers, in their efforts to build interfaces that supported navigators' work, adopted practical models for decision making based on sensory input and cognitive processing. The model of 'situation awareness,' attributed to Endsley (1995), was especially prominent, appearing independently in four of the nine interviews we held with designers. Navigators, by contrast, did not refer to situation awareness, neither directly nor by its characteristic features, which decompose decision making into distinct information processing stages. In the following excerpt, for instance, Karl, a designer, described design needs for a control room to support work for navigating highly automated ships, framed in terms of 'situation awareness' needs:

> 'What data is needed to control and monitor the [highly automated] ships: that is situation awareness need number one. Then situation awareness need number two is to display that into something understandable. Situation awareness need number three is to project that into the future. That could be a way to approach the concept [of operating highly automated ships] in a more… systematic way, perhaps.'

By contrast, navigators invoked heuristic approaches to decision making, drawing from in-situ skills informed by experience. One example of such a heuristic was illustrated by Robin who recounted taking over control to attend to passenger comfort (see Sect. 5.1). In that example, rather than following a sequence of information processing stages, the decision to take over manual control stemmed simply from imagining how passengers would experience the crossing in the given sea state.

Four of the nine designers we interviewed expressed the concept of being 'in the loop,' referring to the state of mind one must be in to take over control from automation. The navigators, by contrast, referred to this same state as 'backup.' Being backup reserved the sense of responsibility that comes with being a navigator, while losing the agency involved in manoeuvring a ship under one's own hand ('The job hasn't changed,' reported Ola, 'but in auto you can sit back and let the system do it'). Being 'in the loop,' by contrast, construed the navigators as components in a larger, cognitive system. In this 'loop,' whose terminology is rooted in control theory, the navigator was expected to passively monitor the closed loop of automated control and immediately close this loop – through timely and decisive takeover action – the moment the loop's integrity was compromised. As explained by Alexander, a designer, 'The key challenge will be to get the operator, in the shortest possible time, to get in the loop of what is going on.'

6 Discussion

In this section, we explore the implications of the field observations and interview study results, framed in terms of the knowledge gaps and issues introduced in Sect. 1 and outlined in more detail in Sect. 2. Towards this aim, we focus discussions around three relevant themes: (i) the agencies of humans and machines in collaborative navigation, (ii) the transition to centres of coordination, (iii) the social implications of AI collaboration, and (iv) control rooms of the past, present, and future.

6.1 Agencies of humans and machines in collaborative navigation

One of the most salient themes uncovered in the analysis was a transition to a 'backup' role, defined by peremptory control interventions, or 'takeovers.' For technology designers and researchers, the transition toward higher levels of autonomy in shipping culminated in centres of coordination, where operators were 'in the loop' of the system. Navigators' accounts of inhabiting this transition in their own work reflected a preoccupation with their own agency and expressed a desire to recover this agency. Backup implied two mutually exclusive activities: passive monitoring in situations for which the automation was well-suited, and active control in situations for which it was not. Backup invoked the 'sorcerer's apprentice' role (Sect. 1), necessitating timely intervention to stop the conjured force of a machine imbued with agency.

Lucy Suchman, in her 'Plans and Situated Actions' (Suchman, 2007), framed the human–machine interface in terms of co-existing intentions entrenched both in control algorithms (plans) and in-situ skills (situated actions). In the context of the backup role, navigators co-existed as passive operators when plans represented by the automation proceeded as expected, and as skilled operators when those plans were inevitably jettisoned to deal with some situation at hand.

Navigators' accounts also underscored the extent to which the canonical 'ironies of automation' applied to the present transition (Sect. 2.4). One such example emerged from the observation of skill degradation in navigators' ship-handling (Sect. 5): the auto-systems were, in effect, compromising the very thing it was designed to improve. Given the central importance of in-situ takeovers in the backup role, the manual ship handling skills seemed, paradoxically, of heightened importance in the face of increasing levels of machine autonomy.

As part of the backup role aboard the *Korsvika* and *Vikhammer*, there was a sense that in order for operations to go smoothly, navigators depended on the automation system as much as the automation system depended on the navigators. While the auto-crossing and auto-docking systems onboard represented relatively low levels of autonomy, the stakes introduced by this inter-dependence appeared to be getting higher for higher levels of machine autonomy. Demski and

Garrabrant (2019), envisioning the system requirements for an ideal cooperative AI, called this inter-dependence ‘embedded agency.’ By this design, a cooperative AI must be self-referential, capable of modelling is own impact on its environment, including how its users adapt to its presence. Dautenhahn (2007) framed this same capacity in terms of social interactivity, pointing out that activities requiring increasing degrees of interactivity require the computational system to be able to reflexively adapt to constantly changing conditions – a form of artificial ‘social’ intelligence. Navigation is exemplary of such an activity, requiring attention not just to what tasks can be automated, but how they should be automated in the context of a socially organized activity.

Whether it was framed as ‘in the loop’ by designers or ‘backup’ by navigators, being continually prepared for takeovers emerged as the defining feature in a new landscape of joint human-AI agency. The takeover, which symbolized the boundary between machine and human control, helped bring to light two specific design issues: firstly, operators’ sense of agency was upended, manifesting in skill degradation over longer periods of passive monitoring; secondly, effective collaboration between operators and highly automated navigation systems was left hanging in the balance of situated actions and computational plans in a flux of changing situations.

6.2 Transition to centres of coordination

The need for supervisory control of highly automated ships has generated renewed interest in centres of coordination for marine navigation. Referred to as ‘shore control centres,’ ‘remote operating centres,’ or ‘remote control centres’ by the informants in our study, centres of coordination of this type have emerged only in the past decade and have since grown significantly in the scientific literature (Veitch and Alsos, 2022). This renewed interest warrants a closer look at the guiding principles presented in Suchman’s original articulation of the centres of coordination concept (Suchman, 1997), which was aimed especially for designers (see Sect. 2.2). In revisiting the theoretical considerations associated with centres of coordination, we also ask whether they are still relevant given the recent technological developments in the decades following the concept’s introduction.

To begin, it is worth reiterating how centres of coordination relate to maritime navigation and to the transition to higher levels of machine autonomy. After all, the original case used to characterise them encompassed airline ground operations, a domain distinct from shipping both in sociotechnical and cultural aspects. Regardless of the differences, however, many of the core elements of centres of coordination were reified in the ‘shore control centre’ case. Specifically, the need to orient workers to the emerging requirements of safety- and time-critical situations was front and centre. The emphasis of locating technology use within socially organized activities, too, was of central interest, as were the requirements for workers to maintain competencies in reacting appropriately to emerging situations. Additionally, like in airline operations, the marine operations had to be orchestrated across different locations (e.g., port authorities, Vessel Traffic Services, other ferries and ships) and in line with a timetable.

The treatment of technology interactions as a strictly material practice, however, should be re-evaluated in the context of highly automated ships. Centres of coordination originally laid out technologies as an 'assembly of heterogeneous devices' (ibid., p. 44), placing the locus of particular actions at particular technological artifacts. Observing technology trends towards openness and interconnectedness, Monteiro et al. (2013) shifted this locus from mere 'artefacts' to 'information infrastructures,' showing the latter were distinct by virtue of networks that obscure any fixed notions of user, and even time or place of use. Recently, scholars have shifted the locus of interactivity even further from the material boundary, attributing not just agency to computational systems, but also the capacity to enact this agency in their natural environment – conditions allowing for the emergence of behavioural characteristics (Rahwan et al., 2019). Researchers in the field of 'machine behaviour' correspondingly describe as their mission 'the study of ways in which introduction of intelligent machines into social systems can alter human beliefs and behaviours' (ibid., p. 483). Experiments using games have already indicated that interacting with algorithms can increase human collaboration and may even improve group performance (Crandall et al., 2018; Shirado and Christakis, 2017). Whether the same holds true for work collaboration and navigation activities, though, remains uncertain.

Despite this shift away from the materiality of technology interaction, the core issues associated with centres of coordination raised several decades ago by-and-large still apply for today's transition in maritime navigation. Whether framed as artefacts, information infrastructures, or enacted AI agents in the CSCW sense, the interactivity of technologies in socially organized activities is still met with an inherent 'otherness' from their human collaborators. Moreover, the degree of interactivity is accentuated, rather than attenuated, for higher levels of machine autonomy.

6.3 The social implications of AI collaboration

The discrepancies we observed between designers and navigators at the human–machine interface (Sect. 5.3) reinforced the need to reorient design activities towards improved incorporation of user feedback and in-situ observation. Here, we briefly examine the role of social dynamics in this design reorientation, discussing the extent to which discrepancies can be addressed by a better understanding of social implications of AI collaboration.

Discrepancies that arose in designers' and navigators' interview accounts betrayed the ostensible straightforwardness of how decisions are reached in day-to-day work. Work, for navigators, did not unfold as a neatly distilled, stagewise process, as inferred by designers. Rather, the navigators invoked a more intuitive, heuristic decision making based on common sense and tacit knowledge gained from experience. Reflecting on their role, the navigators were more than just 'in the loop' and ready to take preventative action. They were custodians of the automated system, presiding over its operation and arbitrating in its decision making capacity in the context of real-world events. The question of how to address this gap, though, remained largely open.

Methods employed in CSCW may provide useful tools for addressing design challenges presented by developing more aligned collaborative systems. These methods, in contrast to the more prevalent cognitivist and computer science perspectives in AI systems design, consider the sociality of technology use in its natural setting. As Bødker (1991) observed with engineers immersed in computer-aided drawing, the interface between human and computer can become a site in its own right, with its own physical form and possibilities. Revisiting the seminal CSCW control room studies of the 1990s and early 2000s sheds light on how this type of site can form within a socially organised, collaborative setting (see Sect. 2.2). Extending these studies to the case of AI collaboration, what Heath and Luff (1992) described as 'mutual monitoring' in line control rooms, for instance, may be recast in the present context. Mutual monitoring originally involved instances where operators divided their attention between their own tasks and the perception of colleagues' actions through myriad cues, signals, and gestures – subtle yet essential coordinating actions in their work. A parallel can be drawn to modern 'explainable AI' (XAI) techniques, where one strategy involves generating heatmaps tracing where image recognition is 'looking' when classifying an image. In one such example, a machine learning algorithm trained to assist physicians diagnose skin cancer was designed to output details about what pixels it was analysing to reach its predictions (Esteva et al., 2017). Output in so-called 'saliency maps,' the algorithm in effect showed its collaborators 'where it was looking.' A recent review suggests that such collaborative approaches in diagnostics leads to better performance than either physician or AI working alone (Tschandl et al., 2020). Efforts like this are in line with the 'cooperative eye hypothesis,' a theory positing that humans evolved to have large sclera (whites of the eyes) to enable them to follow the gaze of others in cooperative activities, favouring selection of those able to coordinate communicative interactions (Kobayashi & Kohshima, 1997, 2001; Tomasello et al., 2007). Following this logic, enhancing explainability by 'showing where the AI is looking' may be considered among XAI efforts shifting to a more social view of computer interactivity, efforts whose merits are also recognizable in the collaborative control room setting from previous generations of ethnographic CSCW studies.

6.4 Control rooms of the past, present and future

Examining control rooms of the past and present (through literature review, expert interviews, and comparisons to other domains) has compelled us to make inferences about control rooms of the future. Here, we briefly discuss the extent to which such future explorations are rigorous and valid in the sense typically invoked by scientific research. We make the case that despite the speculative nature of our results, they constitute relevant contributions to the CSCW and design communities through their ability to articulate an under-constrained problem and generate design insights.

Our study results were speculative because, although grounded in expert interviews and field observations, they were exploratory in nature and aimed to generate rather than converge new design ideas. The starting point for the research was not a clearly defined problem calling for a clearly defined procedure; on the contrary, it was an under-specified problem calling for a correspondingly open-ended approach. In the design community, such problems often call for a 'research through design' approach (Frayling, 1993), where the goal is generating ideas through a range of pragmatic and conceptual insights. As intimated by Frayling, design is concerned with 'the new,' and as such has a close relationship with research despite the futility of its meeting the rigorous standards of a scientific research method. Inspired by Frayling's thinking, Zimmerman et al. (2007) defended 'research through design' approaches based on their propensity to produce the 'right thing' in the face of under-constrained problems. The approach's underlying contribution, they argued, was based on the strength of its potential to 'transform the world from its current state to a preferred state.' It is partly this preferred state that is so important for the researcher to articulate. In this article, we described the preferred state of future control rooms for highly automated ships through expert interviews as well as through literature review and comparisons to other domains. Set into a multi-disciplinary conceptional framework, this articulation is among the main contributions of the work, asserting that in order to effectively address a problem, it is necessary first to formulate the situation at hand. In this case, it was especially elements of social interactivity in future control rooms that was articulated (who will work in the control rooms, and what will it be like to work with increasingly automated systems?).

Similar scholarly approaches have been applied to design of centres of coordination for highly automated ships, where 'future workshops' stand out as a popular approach (e.g., Hoem et al., 2022; Lützhöft et al., 2019). In this approach, experts are invited to discuss open-ended issues under the pretext of informing design activities. What future workshops have in common with the expert interviews we utilized is that they both imagined future sociotechnical systems that fit a defined situation, shifting the focus from generating tangible solutions to eliciting insights and generating a better understanding (Lindley and Coulton, 2015). Such methods necessarily yield ambiguities, which, like Frayling's 'research through design,' reflect results that, as expressed by Gaver (2012), are 'provisional, contingent, and aspirational.' Yet, the strength of such results lies in its exploration of real issues, gaps, and opportunities, as well as in its ability to articulate the situation at hand. In this sense, the rigour and validity of such future explorations lie in its relevance, however speculative, to designers shaping that future.

6.5 Conceptual limitations and future work

Our practice-based research consisted of interviews in addition to field observations more in line with the CSCW tradition. While ethnographic methods remain indispensable in CSCW research, the addition of interview-based methodologies in our case helped to open the self-contained nature of what Monteiro et al. (2013) call the 'here and now' of field studies. Acknowledging that human–machine interaction is always in transition, such an approach contributed towards a more open-ended treatment of themes that remain constant – the 'otherness' of machine agency, the sociality of technology use – and that offer stable reference points in an inherently transitory study domain.

There were, however, some key limitations of the research design. One such limitation involved the extrapolation of work activities in a maritime navigation setting to that of a control room, and the extent to which this extrapolation provided a representative case. At the time of this study, no shore control centres for coordination of highly automated ships existed in full operational scale. The choice of studying navigation aboard highly automated ferries was used as an approximation for the control centre case. A future shore control centre will, it was argued, be organised around the same core activities – just with higher levels of autonomy and at a remote location. The choice of studying professional design activities towards building shore control centres helped to ground this extrapolation. Although this extrapolation identified several relevant themes, future work must be tuned in to the ways in which human collaboration is affected in a real control centre environment.

7 Conclusion

Maritime navigation work is in transition, marked by collaboration with increasingly high levels of machine autonomy. In this study, we framed this transition as an opportunity to study how designers are shaping work and how navigators are adjusting to the changes. Maritime navigation in this sense served as a representative case study for broader applications of safety–critical, distributed work in sociotechnical, computationally infused environments. Interviews with technology designers and navigators indicated a discrepancy between designers' construals of working with higher levels of autonomy and navigators' own reflective accounts of this work. This discrepancy was centred around the task of taking over control from the automation, a role designers called 'in the loop' and navigators called 'backup.' The discrepancy suggested a need to realign design strategies to real-world operational demands. The risks of not doing so are heightened in the face of increasing levels of autonomy and ongoing development of centres of coordination – efforts that paradoxically place more expectations on human operators, rather than less.

Considering the importance of mutual monitoring – the reflexive social articulations that coordinate work in control room environments – it was clear that collaboration with AI systems depended to a large extent on rendering computational activities more visible. Aligning with the needs of human collaborators involved displaying the AI system's actions more transparently, akin to following the gaze of a collaborator's eyes. Better alignment also pointed to designing AI systems that incorporate cues, gestures, and exclamations of their human collaborators. At least in theory, it may even require machine learning techniques incorporating embedded agency, reflexively adapting to adjustments of users influenced by the AI's presence.

The main contributions of this work are positioned within the emerging field of AI alignment research. Located at the crossroads of computer science, engineering, design, HCI, and sociology, alignment research strives to understand how people can seamlessly interact with machine autonomy. CSCW, with its preoccupation with the sociality of computer use especially in work environments, is uniquely positioned to lend perspective on the transition towards centralized control centres for highly automated maritime navigation. The contribution of this work involved the articulation of the situation at hand to help align design to the preferred real-world interplays of computational plans and human actions. Methodologically, we demonstrated the combined use of literature review, expert interviews, and field observations to ground speculative design insights for future control rooms. Conceptually, the contributions raise the relevance of multi-disciplinary theoretical frameworks and reify theory from past control rooms studies and HCI considerations.

In maritime navigation as in other applications of collaborative work, improvements to efficiency and decision making are among the potential benefits of implementing higher levels of machine autonomy and AI. The extent to which these benefits rely upon seamless coordination with human supervisors, though, remains the domain of research oriented towards the implications of collaborating with intelligent machines in work's natural settings.

Acknowledgements

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) directed by project leader Kongsberg Maritime AS and partnered with the Institute for Energy Technology (IFE). The authors also acknowledge the support of the Department of Design at NTNU and strategic funding from the Department of Mechanical and Industrial Engineering at NTNU. The authors gratefully acknowledge the support of doctoral candidates Kim Alexander Christensen and Dag Rutledal who helped with the interviews, as well as to research assistant Elise Korsmo Fjeld for help with visualisations. The authors are immensely grateful for the feedback of several people who reviewed the manuscript, including Eric Monteiro, Elena Parmiggiani, Mariann Villa, Yngve Dahl, and Hedvig Aminoff.

Funding Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital)

Declarations

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.

References

Amershi, Saleema; Weld, Dan; Vorvoreanu, Mihaela; Fourney, Adam; Nushi, Besmira; Collisson, Penny; Suh, Jina; Iqbal, Shamsi; Bennett, Paul N.; Inkpen, Kori; Teevan, Jaime; Kikin-Gil, Ruth; and Eric Horvitz (2019). Guidelines for Human-AI Interaction. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, UK, 4–9 May 2019.* New York: Association for Computing Machinery, pp. 1–13.

Arrieta Barredo, A.; Díaz-Rodríguez, N.; Del Ser, Javier; Bennetot, Adrien, Tabik; Siham, Barbado, Alberto; Garcia, Salvador; Gil-Lopez, Sergio; Molina, Daniel; Benjamins, Richard;

Chatila, Raja; and Francisco Herrera (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, vol. 58, pp. 82–115.

Bainbridge, Lisanne (1983). Ironies of Automation. *Automatica*, vol. 19, no. 6, pp. 775–779.

Bødker, Susanne (1991). *Through the Interface: A Human Activity Approach to User Interface Design*. Hillsdale, NJ: Erlbaum.

Brekke, Edmund Førland; Wilthil, Erik F.; Eriksen, Bjørn-Olav; Kufoalor, D. K. M.; Helgesen, Øystein K.; Hagen, Inger B.; Breivik, Morten; and Tor Arne Johansen (2019). The Autosea project: Developing closed-loop target tracking and collision avoidance systems. *Journal of Physics: Conference Series*, vol. 1357.

Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (2012). *Final Report: On the accident on 1st June 2009 to the Airbus A330–203 registered F-GZCP operated by Air France flight AF 447 Rio de Janerio—Paris*. Le Bourget Cedex, France: BEA.

Burmeister, Hans-Christoph; Bruhn, Wilko; Rødseth, Ørnulf Jan; and Thomas Porathe (2014). Autonomous unmanned merchant vessel and its contribution towards the e-Navigation implementation: The MUNIN perspective. *International Journal of E-Navigation and Maritime Economy*, vol. 1, pp. 1–13.

Christian, Brian. (2020). *The Alignment Problem: Machine Learning and Human Values*. New York, NY: W. W. Norton & Company.

Corbin, Juliet; and Anselm Strauss (2015). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (4th Edition). Thousand Oaks, CA: Sage Publications.

Crandall, Jacob W.; Oudah, Mayada; Tennom, Ishowo-Oloko; Fatimah, Abdallah Sherief; Bonnefon Jean-François; Cebrian Manuel; Shariff Azim; Goodrich, Michael A.; and Iyad Rahwan (2018). Cooperating with machines. *Nature Communications*, vol. 9, no. 1, p. 233.

Dallolio, Alberto; Agdal, Bendik; Zolich, Artur; Alfredsen, Jo Arve; and Tor Arne Johansen (2019). Long-Endurance Green Energy Autonomous Surface Vehicle Control Architecture. *OCEANS 2019 MTS/IEEE SEATTLE, Seattle, WA, USA, 27–31 Oct. 2019*. New York: IEEE, pp. 1–10.

Dautenhahn, Kerstin (2007). Socially intelligent robots: Dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, pp. 679–704.

Demski, Abram; and Scott Garrabrant (2019). Embedded agency. *ArXiv Preprint* ArXiv:1902.09469.

DNV GL (2017). *Rules for Classification: Ships (Part 5 Ship types, Chapter 4 Passenger ships, Edition January 2017, Amended July 2017)*. Oslo, Norway: DNV GL.

DNV GL (2018). *Remote-controlled and autonomous ships in the maritime industry.* Position paper. Oslo, Norway: DNV GL.

Dunbabin, Matthew; Grinham, Alistair; and James Udy (2009). An Autonomous Surface Vehicle for Water Quality Monitoring. In S. Scheding (ed): *Australasian Conference on Robotics and Automation (ACRA), Sydney, Australia, 2–4 December 2009*. Australia: Australian Robotics and Automation Association, pp. 1–6.

Endsley, Mica R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, vol. 37, no. 1, pp. 32–64.

Esteva, Andre; Kuprel, Brett; Novoa, Roberto A.; Ko, Justin; Swetter, Susan M.; Blau, Helen M.; and Sebastian Thrun (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, vol. 542, no. 7639, pp. 115–118.

Fleischmann, Kenneth R.; Greenberg, Sherri R.; Gurari, Danna; Stangl, Abigale; Verma, Nitin; Day, Jaxsen R.; Simons, Rachel N.; and Tom Yeh (2019). Good Systems: Ethical AI for CSCW. *CSCW '19: Conference Companion Publication of the 2019 on Computer Supported*

Cooperative Work and Social Computing, Austin, TX, USA, 9–13 November 2019. New York, NY, USA; Association for Computing Machinery, pp. 461–467.

Flin, Rhona H.; O'Connor, Paul; and Margaret Crichton (2008). *Safety at the sharp end: A guide to non-technical skills*. Farnham, UK: Ashgate Publishing, Ltd.

Frayling, Christopher (1993). Research in Art and Design. *Royal College of Art Research Papers*, vol. 1, no. 1. London, UK: Royal College of Art.

Gaver, William (2012). What Should We Expect from Research through Design? *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012*. New York, NY, USA; Association for Computing Machinery, pp. 937–946

Gjærum, Vilde B.; Strümke, Inga; Alsos, Ole A.; and Anastasios M. Lekkas (2021). Explaining a Deep Reinforcement Learning Docking Agent Using Linear Model Trees with User Adapted Visualization. *Journal of Marine Science and Engineering*, vol. 9, no. 11, p. 1178.

Glaser, Barney G.; and Anselm L. Strauss (1999). *Discovery of grounded theory: Strategies for qualitative research*. Hawthorne, NY, USA: Aldine de Gruyter.

Goerlandt, Floris; and Kenzie Pulsifer (2022). An exploratory investigation of public perceptions towards autonomous urban ferries. *Safety Science*, vol. 145, January 2022, p. 105496.

Goodwin, Charles; and Marjorie Harness Goodwin (1996). Seeing as a situated activity: Formulating planes. In Y. Engeström and D. Middleton (eds.): *Cognition and Communication at Work*. Cambridge, UK: Cambridge University Press.

Heath, Christian; and Paul Luff (1992). Collaboration and Control: Crisis Management and Multimedia Technology in London Underground Line Control Rooms. *Computer Supported Cooperative Work (CSCW)*, vol. 1, nos 1–2, March 1992, pp. 69–94.

Helgesen, Øystein Kaarstad; Vasstein, Kjetil; Brekke, Edmund Førland; and Annette Stahl (2022). Heterogeneous multi-sensor tracking for an autonomous surface vehicle in a littoral environment. *Ocean Engineering*, vol. 252, May 2022, p. 111168.

Hoem, Åsa S.; Veitch, Erik; and Kjetil Vasstein (2022). Human-centred risk assessment for a land-based control interface for an autonomous vessel. *WMU Journal of Maritime Affairs*, vol. 21, no. 2, pp. 179–211.

Hutchins, Edwin (1995). *Cognition in the Wild*. Cambridge, MA, USA: MIT Press.

IATA (2020). *Aircraft Handling and Manual Flying Skills*. Montreal, Quebec: International Air Transport Association (IATA).

IMO (1972). *International Regulations Preventing Collisions at Sea (COLREGs) (adopted 20 October 1972, entered into force 15 July 1977) 1050 UNTS 16 (COLREGs)*. London, UK: International Maritime Organization.

IMO (2018). *IMO takes first steps to address autonomous ships*. Press release, 25 May 2018. http://www.imo.org/en/MediaCentre/PressBriefings/Pages/08-MSC-99-MASS-scoping.aspx

IMO (2021). *Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS) (MSC.1/Circ.1638)*. London, UK: International Maritime Organization, 3 June 2021.

Kimball, Peter; Bailey, John; Das, Sarah; Geyer, Rocky; Harrison, Trevor; Kunz, Clay; Manganini, Kevin; Mankoff, Ken; Samuelson, Katie; Sayre-McCord, Thomas; Straneo, Fiamma; Traykovski, Peter; and Hanumant Singh (2014). The WHOI Jetyak: An autonomous surface vehicle for oceanographic research in shallow or dangerous waters. *2014 IEEE/OES Autonomous Underwater Vehicles (AUV), Oxford, MA, USA, 6–9 October 2014*. New York: IEEE, pp. 1–7.

Kobayashi, Hiromi; and Shiro Kohshima (1997). Unique morphology of the human eye. *Nature*, vol. 387, no. 6635, pp. 767–768.

Kobayashi, Hiromi; and Shiro Kohshima (2001). Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, vol. 40, no. 5, pp. 419–435.

Kongsberg (2020). *First adaptive transit on Bastøfosen VI*. Accessed 28 May 2021.

Lechte, John (1994). *Fifty key contemporary thinkers: From structuralism to postmodernity*. London, UK: Routledge.

Lindley, Joseph; and Paul Coulton (2015). Back to the Future: 10 Years of Design Fiction. *Proceedings of the 2015 British HCI Conference, Lincoln, Lincolnshire, UK, 13–17 July 2015*. New York, NY, USA; Association for Computing Machinery, pp. 210–211.

Littman, Michael L.; Ajunwa, Ifeoma; Berger, Guy; Boutilier, Craig; Currie, Morgan; Doshi-Velez, Finale; Hadfield, Gillian; Horowitz, Michael C.; Isbell, Charles; Kitano, Hiroaki; Levy, Karen; Lyons, Terah; Mitchell, Melanie; Shah, Julie; Sloman, Steven; Vallor, Shannon; and Toby Walsh (2021). *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report.* Stanford, CA, USA: Stanford University, 16 September 2016.

Lützhöft, Margareta; Hynnekleiv, Agnieszka; Earthy, Jonathan V.; and Erik S. Petersen (2019). Human-centred maritime autonomy—An ethnography of the future. *Journal of Physics: Conference Series*, vol. 1357, p. 012032.

Mahadevan, Karthik; Somanath, Sowmya; and Ehud Sharlin (2018). Communicating Awareness and Intent in Autonomous Vehicle-Pedestrian Interaction. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montréal, QC, Canada, April 21–26, 2018*. New York: Association for Computing Machinery, pp. 1–12.

Martinsen, Andreas B.; and Anastasios M. Lekkas (2018). Curved Path Following with Deep Reinforcement Learning: Results from Three Vessel Models. *OCEANS 2018 MTS/IEEE, Charleston, SC, USA, October 22–25, 2018*. Piscataway, NJ: IEEE, pp. 1–8.

McCarthy, John (2007). What is Artificial Intelligence? Stanford, CA, USA: Computer Science Department, Stanford University, 12 November 2007. http://jmc.stanford.edu/articles/whatisai/whatisai.pdf. Accessed 21 Sept 2021.

MiT (2020). Roboat. http://www.roboat.org. Accessed 20 September 2021.

Monteiro, Eric; Pollock, Neil; Hanseth, Ole; and Robin Williams (2013). From Artefacts to Infrastructures. *Computer Supported Cooperative Work (CSCW)*, vol. 22, nos 4–6, June 2021, pp. 575–607.

Morse, Janice M.; Bowers, Barbara J.; Charmaz, Kathy; Corbin, Juliet; Clarke, Adele E.; and Phyllis Noerager Stern (2009). *Developing grounded theory: The second generation*. Walnut Creek, CA, USA: Left Coast Press Inc.

National Transportation Safety Board (2020). *Collision Between a Sport Utility Vehicle Operating with Partial Driving Automation and a Crash Attenuator, Mountain View, California, March 23, 2018 (Accident Report NTSB/HAR-20/01).* Washington, DC, USA: National Transportation Safety Board, 25 February 2020.

Nevile, Maurice (2001). *Beyond the black box: Talk-in-interaction in the airline cockpit*. Ph.D. dissertation. The Australian National University, Canberra.

Nicas, Jack; Kitroeff, Natalie; Gelles, David; and James Glanz (2019). *Boeing built deadly assumptions into 737 Max, blind to a late design change*. New York, NY, USA: The New York Times.

NVivo (1.0) (2020). [Computer software]. QSR International.

Park, Sun Young; Kuo, Pei-Yi; Barbarin, Andrea; Kaziunas, Elizabeth; Chow, Astrid; Singh, Karandeep; Wilcox, Lauren; and Walter S. Lasecki (2019). Identifying Challenges and Opportunities in Human-AI Collaboration in Healthcare. CSCW '19: *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, Austin, TX, USA, 9 – 13 November 2019*. New York, NY, USA: Association for Computing Machinery, pp. 506–510.

Peeters, Gerben; Yayla, Gökay; Catoor, Tim; Van Baelen, Senne; Afzal, Muhammad Raheel; Christofakis, Christos; Storms, Stijn; Boonen, René; and Peter Slaets (2020b). An Inland Shore Control Centre for Monitoring or Controlling Unmanned Inland Cargo Vessels. *Journal of Marine Science and Engineering*, vol. 8, no. 10, September 2020.

Rahwan, Iyad; Cebrian, Manuel; Obradovich, Nick; Bongard, Josh; Bonnefon, Jean-François; Breazeal, Cynthia; Crandall, Jacob W.; Christakis, Nicholas A.; Couzin, Iain D.; Jackson, Matthew O.; Jennings, Nicholas R.; Kamar, Ece; Kloumann, Isabel M.; Larochelle, Hugo; Lazer, David; McElreath, Richard; Mislove, Alan; Parkes, David C.; Pentland, Alex 'Sandy'; … and Michael Wellman (2019). Machine behaviour. *Nature*, vol. 568, no. 7753, pp. 477–486.

Rasmussen, Jens (1986). *Information Processing and Human-machine Interaction: An Approach to Cognitive Engineering*. New York, NY, USA: Elsevier Science Publishing Co.

Reddy, Namireddy Praveen; Zadeh, Mehdi Karbalaye; Thieme, Christoph Alexander; Skjetne, Roger; Sorensen, Asgeir Johan; Aanondsen, Svein Aanond; Breivik, Morten; and Egil Eide (2019). Zero-Emission Autonomous Ferries for Urban Water Transport: Cheaper, Cleaner Alternative to Bridges and Manned Vessels. *IEEE Electrification Magazine*, vol 7, no. 4, November 2019, pp. 32–45.

Rødseth, Ørnulf Jan. (2017). *Definitions for Autonomous Merchant Ships*. Trondheim, Norway: Norwegian Forum for Autonomous Ships.

Rolls-Royce (2018). Rolls-Royce and Finferries Demonstrate World's First Fully Autonomous Ferry. Press release, 3 December 2018. https://www.rolls-royce.com/media/press-releases.aspx. Accessed 1 October 2020.

SAE International (2017). *Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems*. Warrendale, PA, USA: Society of Automotive Engineers (SAE) International.

Shirado, Hirokazu; and Nicholas A. Christakis (2017). Locally noisy autonomous agents improve global human coordination in network experiments. *Nature*, vol. 545, no. 7654, pp. 370–374.

Shneiderman, Ben (2020). Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human–Computer Interaction*, vol. 36, no. 6, March 2020, pp. 495–504.

Stone, Peter; Brooks, Rodney; Brynjolfsson, Erik; Calo Ryan; Etzioni, Oren; Hager, Greg; Hirschberg, Julia; Kalyanakrishnan, Shivaram; Kamar, Ece; and Sarit Kraus (2016). *Artificial intelligence and life in 2030: The one hundred year study on artificial intelligence*. Stanford, CA: Stanford University.

Suchman, L. (1997). Centers of Coordination: A Case and Some Themes. In L. B. Resnick, R. Säljö, C. Pontecorvo, & B. Burge (eds): *Discourse, Tools and Reasoning: Essays on Situated Cognition*. Berlin, Germany: Springer, pp. 41–62.

Suchman, Lucy A. (2007). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd Edition). Cambridge, UK: Cambridge University Press.

Tomasello, Michael; Hare, Brian; Lehmann, Hagen; Josep Call (2007). Reliance on head versus eyes in the gaze following of great apes and human infants: The cooperative eye hypothesis. *Journal of Human Evolution*, vol. 52, no. 3, pp. 314–320.

Tschandl, Philipp; Rinner, Christoph; Apalla, Zoe; Argenziano, Giuseppe; Codella, Noel; Halpern, Allan; Janda, Monika; Lallas, Aimilios; Longo, Caterina; Malvehy, Josep; Paoli, John; Puig, Susana; Rosendahl, Cliff; Soyer, H. Peter; Zalaudek, Iris; and Harald Kittler (2020). Human–computer collaboration for skin cancer recognition. *Nature Medicine*, vol. 26, no. 8, pp. 1229–1234.

Utne, Ingrid Bouwer; Rokseth, Børge; Sørensen, Asgeir J.; and Jan Erik Vinnem (2020). Towards supervisory risk control of autonomous ships. *Reliability Engineering & System Safety*, vol. 196, pp. 106757.

Veitch, Erik; and Ole Andreas Alsos (2022). A systematic review of human-AI interaction in autonomous ship systems. *Safety Science*, vol. 152, pp. 105778.

Vicente, Kim J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. Boca Raton, FL: CRC Press.

Voosen, Paul (2017). The AI detectives. *Science*, vol. 357, no. 6346, July 2017, pp. 22–27.

Wang, Wei; Gheneti, Banti; Mateos, Luis A.; Duarte, Fabio; Ratti, Carlo; and Daniela Rus (2019). Roboat: An Autonomous Surface Vehicle for Urban Waterways. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019*. New York: IEEE, pp. 6340–6347.

Waymo (2020). Waymo Safety Report. https://waymo.com/safety. Accessed 13 September 2021.

Whalen, Jack (1995). A technology of order production: Computer-aided dispatch in public safety communications. In P. Have, G. Psathas (eds): *Situated Order: Studies in the Social Organization of Talk and Embodied Activities*. Boston, MA, USA: International Institute for Ethnomethodology and Conversation Analysis, pp. 187–230.

Wu, Baiheng; Li, Guoyuan; Wang, Tongtong; Hildre, Hans Petter; and Houxiang Zhang (2021). Sailing status recognition to enhance safety awareness and path routing for a commuter ferry. *Ships and Offshore Structures*, vol. 16, no. 1, pp. 1–12.

Zimmerman, John; Forlizzi, Jodi; and Shelley Evenson (2007). Research through Design as a Method for Interaction Design Research in HCI. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, *San Jose, CA, USA, 28 April – 3 May 2007*. New York, NY, USA; Association for Computing Machinery, pp. 493–502.

Article 3

DESIGN FOR RESILIENT HUMAN-SYSTEM INTERACTION IN AUTONOMY: THE CASE OF A SHORE CONTROL CENTRE FOR UNMANNED SHIPS

Veitch, Erik Aleksander;
Kaland, Thomas;
Alsos, Ole Andreas

Norwegian University of Science and Technology

ABSTRACT

Artificial intelligence is transforming how we interact with vehicles. We examine the case of Maritime Autonomous Surface Ships (MASS), which are emerging as a safer and more effective solution for maritime transportation. Despite the focus on autonomy, humans are predicted to have a central role in MASS operations from a Shore Control Centre (SCC). Here, operators will provide back-up control in the event of system failure. There are signification design challenges with such a system. The most critical is human-system interaction in autonomy (H-SIA). We consider humans as the source of resilience in the system for adapting to unexpected events and managing safety. We ask, can Human-Centred Design (HCD) be used to create resilient interactions between MASS and SCC? Work has been done in resilience engineering for complex systems but has not been extended to H-SIA in transportation. "Resilient interaction design" is relevant as we progress from design to operational phase. We adopted the ISO 9421-210 guideline to structure our HCD approach. The result is an SCC designed for 1 Autonomy Operator (AO). The contribution is a demonstration of how resilient interaction design may lead to safer and more effective H-SIA in transportation.

Keywords: Shore Control Centre, User centred design, Design methods, Systems Engineering (SE), Autonomous Systems

Contact:
Veitch, Erik Aleksander
NTNU
Department of Design
Norway
erik.a.veitch@ntnu.no

Cite this article: Veitch, E. A., Kaland, T., Alsos, O. A. (2021) 'Design for Resilient Human-System Interaction in Autonomy: The Case of a Shore Control Centre for Unmanned Ships', in *Proceedings of the International Conference on Engineering Design (ICED21)*, Gothenburg, Sweden, 16-20 August 2021. DOI:10.1017/pds.2021.102

1 INTRODUCTION

1.1 Designing for resilience in autonomous systems: the case of unmanned ships

Artificial Intelligence (AI) is changing how we interact with vehicles. In this paper, we look at the case of unmanned ships with autonomous navigation capability. Unmanned ships are termed Maritime Autonomous Surface Ships (MASS) by the International Maritime Organization (IMO), the United Nations agency regulating international maritime safety. We use the terms 'MASS' and 'unmanned ship' interchangeably in this paper.

Historically, researchers have taken a measured approach when assessing the potential for MASS to improve safety by reducing human-related accidents, pointing out that new failures modes related to interaction with the autonomy system will present themselves (Ahvenjärvi, 2016; Chang et al., 2020; Ramos et al., 2020a). Researchers largely concede that full autonomy, whereby the AI is so complete as to remove the need for humans completely, is not a realistic operating model for which to currently plan. It emerges that humans remain central to safe operation of MASS, despite the promise of more machine autonomy. This 'irony of automation' has been observed before in automation applications (Bainbridge, 1982; Parasuraman and Wickens, 2008). We observe a similar 'irony of automation' emerging in the wake of unmanned ships, manifested in the Shore Control Centre (SCC) concept. In current operating models of MASS, the SCC serves predominantly a back-up role: when the AI fails to solve a navigation problem, the autonomy shifts to a fail-safe state until operators at the SCC intervene to resolve the problem and, once resolved, return the system to an autonomous state.

Risk researchers have predicted that the interaction between humans and the autonomy system will be the root cause for the most consequential propagation of accidents (Chang et al., 2020; Ramos et al., 2020b). The term Human-System Interaction with Autonomy (H-SIA) was introduced as way to assess the safety of this interaction (Ramos et al., 2020a). Despite the focus on risk, human interaction elements also represent an opportunity for designers to strengthen the system. Should operators at the SCC be better equipped to flexibly react and adapt to problems not only when they arise, but also before they arise, then the overall effect would be a safer and more effective system. The premise of adaptive control of safety while balancing productivity is the aim of resilience engineering (Hale and Heijer, 2006). Yet despite significant efforts in resilience engineering towards safer complex systems design and the innovations in modelling H-SIA risk, no research to date has sought to apply a Human-Centred Design (HCD) process with the aim of improving interaction resilience between SCC and MASS. This paper aims to fill that gap. The research question we ask is: can we use HCD to design for resilient interactions between MASS and SCC?

In Section 2, we describe the HCD method that we used to design a SCC. Section 3 presents findings and our SCC design. Section 4 discusses implications of this work, including for broader applications. Section 5 presents a conclusion and suggestions for future work in resilient interaction design for human-AI interaction in autonomous transportation.

1.2 Scope

In this paper, we consider SCC design applications that encompass four MASS types: urban, inland, coastal, and ocean-going (Figure 1). These are also defined in Table 1 along with some relevant reference projects.

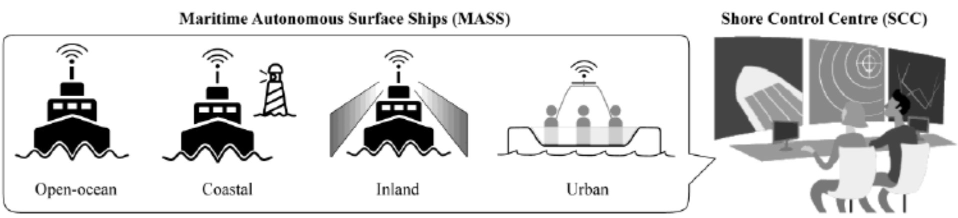


Figure 1. Autonomous maritime system comprising of MASS and SCC

Table 1. Scope of MASS types in this paper

Type	Description	Reference projects
Urban	Small vessels (length approx. 5-10 m) transporting passengers or goods in urban canals or waterways	Autoferry (Reddy et al., 2019); RoBoat (MiT, 2020)
Inland	Inland cargo vessels for European inland waterways	H2H Project (Peeters et al., 2020)
Coastal	Short-sea shipping routes in trafficked channels with aids to navigation present along the shore, Vessel Traffic Services (VTS) communication, and demanding navigation	Land-Based Operation of Autonomous Ships (LOAS); SFI AutoShip (NTNU, 2021a)
Ocean	Open-ocean transportation with non-demanding navigation	MUNIN (Burmeister et al., 2014)

1.3 Levels of autonomy

Levels of Autonomy (LoA) refers to how autonomous a system is on a spectrum from full manual control to full automatic control. For the Autonomous Maritime System (AMS) case, LoA has been shown to be an effective way to classify SCC-MASS interactions (Dybvik et al., 2020). There are many taxonomies for LoA and one comprehensive review finds that its definition is context specific (Vagia et al., 2016). For this study, we use the taxonomy in Table 2 proposed by NFAS (Rødseth, 2017) and assume the AMS is in "Constrained Autonomous" mode (Level 7).

Table 2. LOA for MASS from Rødseth (2017)

LoA	Label	Definition
1	Direct control	Direct control of ship from bridge crew, no decision support.
2	Decision support	Decision support and advice to crew on bridge. Crew decides.
3	Automatic bridge	Automated operation, under continuous supervision by crew.
4	Periodically unmanned	Continuously supervised by shore. Muster crew if necessary.
5	Remote control	Unmanned with monitoring and direct control from shore.
6	Automatic	Unmanned under automatic control, monitored from shore.
7	Constrained autonomous	Unmanned, partly autonomous, continuous shore supervision.
8	Fully autonomous	Unmanned and without supervision.

1.4 Shore control lab

The Shore Control Lab (SCL) is an SCC owned and operated by the Norwegian University of Science and Technology (NTNU) with the mission of advancing AMS research (NTNU, 2021b). Its design is the main result of this paper. In our initial scoping of similar infrastructures, we found only a handful of SCCs in existence today. Of those that do exist, none have a proven operational concept, instead ranging from early to advanced prototypes. The earliest prototypes were limited by conventional ship bridge paradigms (MacKinnon et al., 2015). Recently, a new paradigm is emerging for SCC-based navigation (Veitch et al., 2020). For example, an advanced functional prototype of an SCC for inland MASS has been built (Peeters et al., 2020) and a ship management company specializing in coastal MASS has led to innovative SCC designs (Massterly, 2021). Both projects have demonstrated efforts to base designs on specific SCC requirements rather than fitting the model of a conventional ship's bridge to the shore-based control centre mould.

As SCC prototypes advance, they are expected to unlock operational capabilities of MASS. However, before MASS can operate in marine traffic with conventional ships, there is a need for systematic and repeatable testing of human-AI interaction scenarios in the SCC to ensure safer and effective future maritime operations. The SCL is therefore a platform for addressing this research need. Its implications also have relevance to broader human-AI interaction issues in other transportation applications.

2 METHOD

We applied HCD principles to design the SCC. To do this in a way that was rigorous, tracing design decisions transparently to observations, we adopted the ISO design guideline *9241: Ergonomics of human-system interaction, Part 210: Human-centred design for interactive systems standard* (ISO, 2019). We refer to this as *ISO 9241-210* for short. This method is depicted in Figure 2. It is a relevant design guideline because it is meant for interactive systems like that typified by the SCC; although, design for autonomous interactive systems is not explicitly its focus.

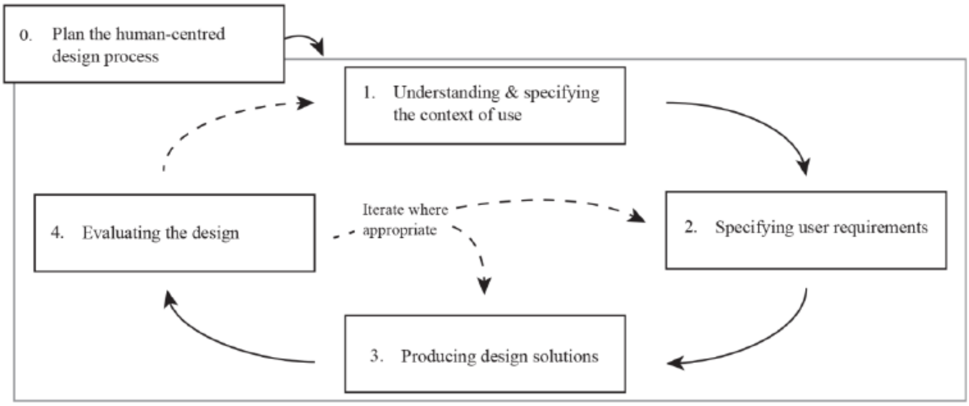


Figure 2. Human-centred design process adapted from ISO (2019)

2.1 Design activities

We used various design activities to collect information for Stages 1-4 of the HCD process in Figure 2. These design activities are depicted in Figure 3 along with their corresponding design stages. Details about each design activity are found in Sections 3.1 to 3.3.

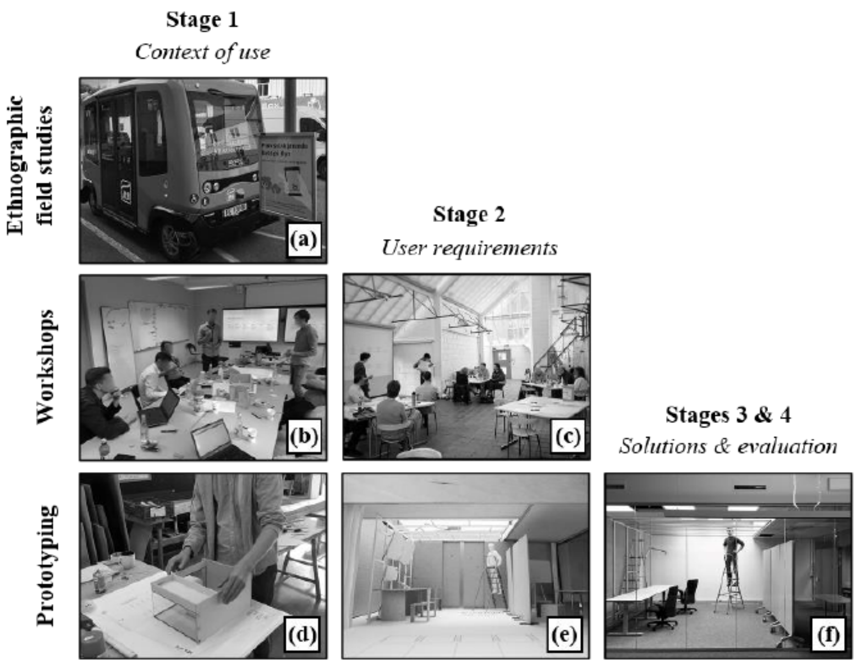


Figure 3. Design activities corresponding to stages 1-4 of the ISO 9241-210 HCD guideline. (a) Autonomous bus case study (b) Stakeholder Workshop (c) Co-Creation workshop (d) Basic prototyping (e) Detailed prototyping (f) Full-scale prototyping

3 FINDINGS

In this section we present a human-centred SCC design based on results from applying the *ISO 9241-210* guideline. It is divided into three subsections (and corresponds to the three columns in Figure 3):

1. **Understanding and specifying the context-of-use:** This phase involved information gathering and analysis with the purpose of describing the future context-of-use of SCCs. Context-of-use is broken down into four distinct elements, which are defined in Table 3. The corresponding design activities used to inform these four context-of-use elements are further listed in Table 4.
2. **Specifying user requirements:** Core user needs are listed in Figure 4 in terms of technical, physical, organizational, and cultural aspects and provide the basic functional requirements for the SCC.
3. **Producing and evaluating design solutions:** Concrete design solutions in form of physical prototypes were used to generate user-centred evaluation and design refinements. We present the full-scale SCC design (the SCL), in Figure 5.

3.1 Understanding and specifying the context of use

This was the first stage of the HCD design process (Figure 2). Our methods of data collection in this stage involved a workshop and two design ethnographic case studies. Descriptions of what these activities involved are listed in Table 4. There are two additional aspects to note about these design activities. The first is that participants in the Stakeholder Workshop (Figure 3b) were experienced university faculty members representing a variety of backgrounds, including engineering cybernetics, innovation management, interaction design, electronic systems, lab management, and sensor fusion. The second involves the formal definition of "design ethnography," which we adopt from Hanington and Martin (2012, pp. 60) as a method that "approximates the immersion methods of traditional ethnography, to… understand the user's world for design empathy and insight."

Table 3. Elements of context-of-use description (adopted from ISO, 2019)

Context-of-use element	Tag	Definition
Users and stakeholder groups	US	Groups representing a relationship with futures SCCs that can be described in terms of key goals and constraints.
Characteristics of the users	C	Relevant characteristics of the users including knowledge, skill, experience, education, training, preferences, and capabilities.
Goals and tasks of the users	GT	Overall goals of the system as well as of users and the characteristics of the tasks that can influence usability and accessibility.
Environment of the system	E	Technical environment (e.g. hardware, software, materials), physical environment (e.g. lighting, spatial layout, furniture), and relevant aspects of organizational structure and culture.

Table 4. Data collection methods informing relevant context-of-use elements

Design method	Case study	Tag (Table 3)
Stakeholder Workshop	Structured workshop to identify stakeholders and their core needs. Participants were six experienced university faculty members from various disciplines. It was led by the authors.	US, C, GT
Ethnographic case study 1	Contextual inquiry and interviews of remote air-traffic controllers at Avinor Remote Towers[1] at Bodø Airport	C, GT, E
Ethnographic case study 2	Contextual inquiry of "safety host" during normal operations and manual takeover of public autonomous city bus[2]	C, GT, E
[1] For more information, refer to Avinor (2021) [2] For more information, refer to Vy (2020)		

The results of the data collection activities in this phase are presented in the context-of-use description below. The results were integrated into a basic prototype of the SCC (Figure 3d) which was used to interactively discuss physical aspects of context-of-use like space usage and basic infrastructure (see Section 3.3 for more information about prototyping).

3.1.1 Context of use description

Who are the SCC stakeholders? We distinguished between two main stakeholder groups: developers of AMS mobilizing for MASS operational phases and the end users of the SCC operating the controls first-hand. The latter group was representative of the operators who would work at the SCC. We called these primary SCC users the "Autonomy Operators" (AOs) to illustrate their main role: as a bridge between autonomous navigation tasks handled by AI and the human navigation responsibilities stemming from oversight, back-up intervention, and safety management. After the Stakeholder Workshop, we arranged a second workshop (the "Co-Creation Workshop") with invited participants representing these two groups' interests. Although the AOs did not yet exist, the relevant backgrounds of the workshop participants allowed us to infer information about them (see Section 3.3).
What are the characteristics of SCC users? The AOs where characterized as control room operators with skills in marine navigation. Conventional Bridge Resource Management (BRM) weighed heavily on how these users were characterized, with terms like "Captain" and "Master" being used to label their roles within a team hierarchy. Considering the novelty of AMSs, we avoided applying terminology used in conventional BRM to describe AO roles and tasks and instead characterized them broadly: as operators with competence in interacting with a dynamic LoA system and in decision-making related to intervention under time pressure and elevated risk.
What are the goals and primary tasks of the users? We inferred goals and primary tasks of the AOs in terms of two operational categories: single- and multi-vessel operations. Single-vessel operation is the focus in the beginning phase of SCC development, where the primary task is safe operation of a single vessel. Tasks span planning (pre-operation stage), monitoring (operation stage), intervention (situation handling and emergencies), and assessment (post-operation stage). After single-vessel operation is verified within an acceptable risk level, the goal will be to operate multiple vessels. New primary tasks will involve fleet management (multiple vessels) and resource management (multiple AOs in a team hierarchy). Note that it is not in the scope of this paper to describe all tasks in full.
What are the characteristics of the system environment compared to similar applications? The ethnographic case studies (Table 4) led to a richer understanding of the environmental context-of-use in relation to similar applications. For example, the visit to the Avinor Remote Towers at Bodø International Airport let us observe operators turning on ambient sound to hear the in-situ airplane engines for a more immersive experience, which illustrated to us the importance of audio feedback in situation awareness when operating from a remote location.

3.2 Specifying user requirements

Here we specify functional requirements for the SCC in terms of core user requirements. The data came from the Stakeholder Workshop (Figure 3b) and the Co-Creation Workshop (Figure 3c). Results focussed on identifying users and their needs were distilled into core user requirements and presented in Figure 4. It is outside the scope of this paper to dissect the specific operator information needs. For detailed information on this topic, the reader is referred to Lunde-Hanssen et al. (2020).

3.3 Producing and evaluating design solutions

Here design solutions for the SCC are generated based on core user requirements (Section 3.2). Design solutions took the physical form of prototypes in various levels of detail and culminated in the full-scale SCC (Figure 5). During the design process, prototypes were used to generate user-based evaluation to drive further, more detailed, iterations. This paper refers to four distinct prototypes, three of which are built today (the third is still under construction). These are listed below:

1. **Basic Prototype:** assembled with Medium-Density Fibreboard (MDF), plywood, cardboard, clear plastic, and grey modelling foam with some plastic components made on a *Prusa MkII* 3D printer, constructed to 1:20 scaled and mounted on a printed floor plan. This prototype helped to interactively discuss high-level aspects like general arrangement, space and lighting, and power and data connectivity. (See Figure 3d.)
2. **Detailed Prototype:** assembled from Basic Prototype with added details. As discussions with users and experts became more detailed, so did the Detailed Prototype. (See Figure 3e).
3. **Full-scale SCC (for virtual simulator):** At the time of writing, the full-scale SCC, called the Shore Control Lab (SCL), is under construction. (See Figure 5.)
4. **Full-scale SCC (for real MASS):** Evaluation of Prototype 3 will feed into this design.

In the Co-Creation Workshop (Figure 3c), the basic prototype was interactively presented to thirteen participants representing stakeholder and expert groups to generate detailed user-oriented feedback. Participants came from a variety of backgrounds, including control room design for space applications, AI for autonomous cars, engineering design, marine navigation, and human factors. Results from this workshop fed into the detailed prototype and filtered into the full-scale SCC. The full-scale SCC, currently equipped for virtual simulation, is depicted in Figure 5 along with component descriptions.

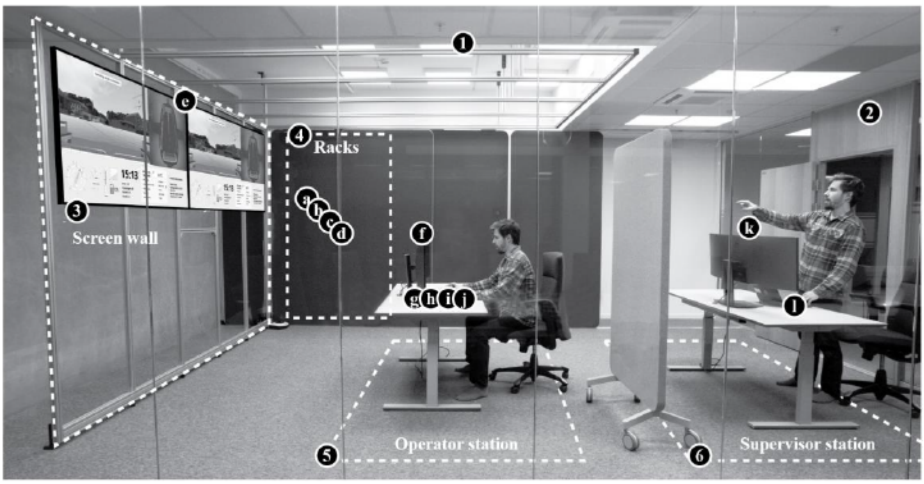
Technical
- "Mission-critical" screen positioned at Operator Station (OS). Data displayed is needed for intervention (e.g., camera view, condition monitoring). Screens are wide desktop monitors.
- "Mission support" screen(s) positioned at a distance from OS and mounted on wall. Data displayed is needed for monitoring and vessel management (e.g., map with marine traffic, weather, time-series data). Screens are large interactive touch screens.
- Push-button panel with commands for real-world interaction (e.g., vessel anchor-drop, door close, lights, emergency contacts)
- Two-way audio communication with MASS
- Audio feed from MASS (ambient sound)
- Simulator to test and train elements of H-SIA during normal operations and intervention

Cultural
- Safe and secure room
- Positive work environment

Physical
- Working environment compliant with relevant control room ergonomics standards (e.g., adequate lighting, acoustic qualities, workstation configuration)
- Minimal visual distractions in room; tidy cable management
- User-friendly user-interface (UI)
- Ergonomic sightline angles from OS to screen-wall
- Aesthetics and design of the control room contribute to a positive work environment

Organizational
- Tasks clearly defined within a hierarchy
- Operational and safety management procedures clearly defined
- Hazards identified and risks assessed
- Training for skills and decision-making competence, especially for intervention and other safety-critical scenarios
- Procedures and safety monitoring frameworks tested and verified using simulated scenarios

Figure 4. Core user requirements across four dimensions



1. Physical attributes of room: 36 m², tunable white light (4000 lm, 2700-6500 K), sliding door wall to partition racks and storage, moveable room dividers
2. Instructor Station (IS) (not shown) is a separate room with control of computers, CCTV of control room, and two-way communication with OS for directing experiments.
3. Mission-support information displays (e) mounted on floor-and-ceiling-fixed Aluflex wall frame. It is shown here depicting the open-source Gemini simulator.
4. 2 x acoustically dampened computer racks
5. Operator Station (OS) with standing desk and ergonomic chair; sightline < 20 deg to screen wall. Control of simulator and two-way communication with IS.
6. Workstation for configuration of simulator and OS A/V sources

a. Simulator running Gemini (controlled from OS & IS)
b. Digital twin of autonomy system (hardware-in-the-loop)
c. Digital twin of MASS actuators and force response
d. 4K video encoders (AVoIP) for all AV components
e 2x 75" 4K Collaborative Touchscreen Displays
f. Mission-critical display (49" curved monitor)
g. Joystick-style controller
h. Swan-neck microphone
i. Desktop speakers
j. Elgato Stream Deck with 15 programmable LCD icon buttons
k. Supervisor display (49" curved monitor)
l. Touch 10" control panel for configuring A/V sources

Figure 5. SCC physical attributes, system architecture & core technical components

4 DISCUSSION

In this section we discuss the relevance of the results in terms of future work and application to other autonomous transportation systems.

4.1 'Resilient interaction design' and applications to other fields

We have shown how HCD methods helped to uncover needs of AOs in the SCC and led to informed design of interaction solutions for the AMS. The term 'resilient interaction design' (resilient IxD) is now introduced to describe contributions of designers in the face of emerging human-AI interaction design challenges in complex systems like those present in transportation applications. The work presented in this paper can be considered as a contribution to resilient IxD for AMSs.

One of the keys to resilient IxD lies in multi-disciplinary collaboration among specialist groups working on the issues specific to the MASS and to the SCC. However, as researchers in safety engineering have discovered, compartmentalizing aspects of complex operations often lies at the root of accidents (Leveson, 2016). Resilient IxD should therefore work towards closing the gap between the autonomy system and the human interacting with it, thus taking steps towards safer operations in a system perspective

The work presented also has relevance to applications outside of maritime operations. There is a general trend for autonomous transportation systems that rely on human intervention for safety. This human intervention is often done remotely from a coordinate centre that can handle multiple vehicles simultaneously. As we have seen, human-AI interactions are vulnerable to error propagation if poorly implemented. Recent examples where poorly implemented AI-interaction design has contributed to loss of vehicle control and fatal accidents include the Boeing 737 MAX accidents (Nicas et al., 2019) and Tesla 'Autopilot' crashes (U.S National Transportation Safety Board, 2017). Takeover instances in autonomous cars are exposed to the same underlying interaction vulnerabilities as in the SC case; namely, cognitive information gathering and processing under time pressure and high risk.

4.2 Virtual simulator

It emerged during the design process that the need for training, testing, verification, and assurance were all closely linked to the ability to run controlled and repeatable scenarios: something that is only practical in a simulator. Therefore, one of the main aims of the SCL was to provide integrated virtual simulation of the AMS. This is not trivial; there are many full-mission bridge simulators, but no analogues to the SCC. Future work at the SCL will aim to fill that gap. The ability to simulate AMS operations may also contribute to verification of frameworks for monitoring operational safety indicators, such as that proposed by Thieme and Utne (2017). The literature also points to a lack of empirical data to inform risk assessment models needed to quantify probabilistic error models related to interaction between MASS and SCC (Thieme et al., 2018; Wróbel et al., 2018). Simulation in the SCC can generate empirical data in virtual scenarios to help fill this gap.

We checked the literature for autonomous vehicle simulators and found promising development in the car industry, such as that exemplified by the open platform CARLA (Dosovitskiy et al., 2017). A similar open source technology platform is being development for MASS applications, called Gemini (Vasstein et al., 2020). The SCL will use the Gemini platform for future test campaigns. Future testing will incorporate the needs of AI testing and verification with human-AI interaction needs reflecting the resilient IxD framework.

4.3 Strengths and limitations of the method

The biggest advantage of adopting an HCD approach was that it led to specification of the SCC based on user requirements, instead of abstract requirements. We have drawn clear links between user requirements and design specifications in this paper. However, there are some important limitations in the method. The most obvious was that the core users, which we termed Autonomy Operators (AOs), do not yet exist in fully operational SCCs. The role is novel, and thus we had to infer needs of AOs from various sources, including from similar applications (remote air-traffic controllers and autonomous bus safety host in Section 3.1) and from domain experts in space control rooms, engineering design, car automation, and other fields that participated in our Co-Creation Workshop (Sections 3.2 and 3.3). In the Co-Creation Workshop, there was one experienced marine navigator with a certificate for bulk carriers who joined remotely. More information from seafarers would have been helpful because seafarers represent the most similar occupation to that of the AO in the SCC and therefore the most likely to become actual AOs. A cross-check in the literature suggests the same but points to necessary amendments of training conventions to make this possible (Katsivela, 2020; Kim and Mallam, 2020). Future work should include more user-testing on seafarers, set up specifically to gauge feedback on experience and to measure performance in pre-defined tasks. Virtual simulation may provide a useful tool for this purpose.

5 CONCLUSION

We have designed an SCC for MASS in constrained autonomous mode using an HCD method. This SCC is called the 'Shore Control Lab' and will be used for research. The contribution is a novel approach to designing more resilience into the AMS by addressing safety-critical interactions between humans and AI using HCD principles. We systematically and transparently designed for AO needs, framing AOs as a central source of operational resilience in aspects such as decision-making and safety monitoring. The implications are safer and more effective operations of MASS. Other autonomous vehicles that depend on human-AI interaction for operational management and control can benefit from this approach, such as vehicle traffic with remote operator assistance. In this paper we examined the case of a single operator-single vehicle control station, although the solution is scalable for multiple vehicles managed from a single coordination centre with a team of operators. We introduce the term 'resilient interaction design' to refer to research and development contributing to solutions for safer and more effective design of human-AI interaction aspects through application of HCD principles.

ACKNOWLEDGMENTS

We acknowledge (1) NTNU for funding the *Shore Control Lab* infrastructure, (2) the Norwegian Research Council for funding the project "Land-based Operation of Autonomous Ships" (Grant Number 296527), and (3) all the workshop participants and experts for valuable information and feedback.

REFERENCES

Ahvenjärvi, S. (2016), "The human element and autonomous ships", *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, Vol. 10, No. 3, pp. 517-521. https://dx.doi.org/10.12716/1001.10.03.18.

Avinor (2021), *Remote Towers: The technology of the future at Norwegian airports* [online]. Avinor AS. Available at https://avinor.no/en/remote (accessed 3.23.21).

Bainbridge, L. (1982), "Ironies of Automation", *IFAC Proceedings*, Vol. 15, pp 129–135. https://dx.doi.org/10.1016/S1474-6670(17)62897-0

Burmeister, H.C., Bruhn, W., Rødseth, Ø.J., Porathe, T. (2014), "Autonomous Unmanned Merchant Vessel and its Contribution towards the e-Navigation Implementation: The MUNIN Perspective", *International Journal of e-Navigation and Maritime Economy*, Vol. 1, pp. 1–13. DIO: 10.1016/j.enavi.2014.12.002

Chang, C.H., Kontovas, C., Yu, Q., Yang, Z. (2020), "Risk assessment of the operations of maritime autonomous surface ships", *Reliability Engineering & System Safety*, Vol. 207, pp. 111. https://dx.doi.org/10.1016/j.ress.2020.107324

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V. (2017), "CARLA: An open urban driving simulator", *Proceedings of the 1st Annual Conference on Robot Learning*, Proceedings of Machine Learning Research, Mountain View, CA, USA, pp. 1–16.

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

Hale, A., Heijer, T. (2006), "Defining resilience", In: Hollnagel, E., Woods, D.D., Levesen, N. (Ed.), *Resilience engineering: Concepts and precepts*, CRC Press, Boca Raton, FL, USA, pp. 35–40.

Hanington, B., Martin, B. (2012), *Universal Methods of Design: 100 Ways to Research Complex Problems, Develop Innovative Ideas, and Design Effective Solutions.* Rockport Publishers, Beverly, MA, USA.

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

ISO (2019), *NS-EN ISO 9241-210:2019, Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*, European Committee for Standardization, Brussels, Belgium.

Katsivela, M. (2020), "Unmanned Vessels and Regulatory Concerns", *Journal of International Maritime Law*, Vol. 26, pp. 239252.

Kim, T., Mallam, S. (2020), "A Delphi-AHP study on STCW leadership competence in the age of autonomous maritime operations", *WMU Journal of Maritime Affairs*, Vol. 19, pp. 163–181. https://dx.doi.org/10.1007/s13437-020-00203-1

Leveson, N.G. (2016), *Engineering a Safer World: Systems Thinking Applied to Safety*. The MIT Press, Cambridge, MA, USA.

Lunde-Hanssen, L.S., Braseth, A.O., Strand, S. (2020), *Identification of information requirements in ROC operations room (No. IFE/E-2020/007)*, IFE, Halden, Norway.

MacKinnon, S., Man, Y., Baldauf, M. (2015*). Final Report: Shore Control Centre*. MUNIN Project, Gothenburg, Sweden.

Massterly (2021), *Making autonomy a reality*, [online]. Massterly AS. Available at https://www.massterly.com/news (accessed 3.21.21).

MiT (2020), *Roboat Project.* [online]. MIT. Available at: roboat.org (accessed 11.19.20).

Nicas, J., Kitroeff, N., Gelles, D., & Glanz, J. (2019). *Boeing Built Deadly Assumptions into 737 Max, Blind to a Late Design Change*. [online]. The New York Times. Available at: https://www.nytimes.com/2019/06/01/business/boeing-737-max-crash.html (accessed 3.17.21).

NTNU (2020), *Autoferry*. [online]. NTNU. Available at: ntnu.edu/autoferry (accessed 10.1.20).

NTNU (2021a), *SFI AutoShip*. [online]. NTNU. Available at: ntnu.edu/sfi-autoship (accessed 3.21.21).

NTNU (2021b), *NTNU Shore Control Lab*. [online]. NTNU. Available at: https://www.ntnu.edu/shorecontrol (accessed 4.21.21)

Parasuraman, R., Wickens, C.D. (2008), "Humans: Still Vital After All These Years of Automation", *Human Factors*, Vol. 50, pp. 511–520. https://dx.doi.org/10.1518/001872008X312198.

Peeters, G., Yayla, G., Catoor, T., Van Baelen, S., Afzal, M.R., Christofakis, C., Storms, S., Boonen, R., Slaets, P. (2020), "An Inland Shore Control Centre for Monitoring or Controlling Unmanned Inland Cargo Vessels", *Journal of Marine Science and Engineering*, Vol. 8, No. 10: 758. https://dx.doi.org/10.3390/jmse8100758

Ramos, M.A, Thieme, C.A., Utne, I.B., Mosleh, A. (2020a), "Human-system concurrent task analysis for maritime autonomous surface ship operation and safety", *Reliability Engineering & System Safety*, Vol. 195. https://dx.doi.org/10.1016/j.ress.2019.106697

Ramos, M.A., Thieme, C.A., Utne, I.B., Mosleh, A. (2020b), "A generic approach to analysing failures in human–System interaction in autonomy," *Safety Science*, Vol. 129. https://dx.doi.org/10.1016/j.ssci.2020.104808

Reddy, N.P., Zadeh, M.K., Thieme, C.A., Skjetne, R., Sorensen, A.J., Aanondsen, S.A., Breivik, M., Eide, E. (2019), "Zero-Emission Autonomous Ferries for Urban Water Transport: Cheaper, Cleaner Alternative to Bridges and Manned Vessels", *IEEE Electrification Magazine*, Vol. 7, pp. 32–45.

Rødseth, Ø.J. (2017), *Definitions for Autonomous Merchant Ships*, NFAS, Trondheim, Norway.

Salmon, P.M., Walker, G.H., Stanton, N.A. (2016), "Pilot error versus sociotechnical systems failure: a distributed situation awareness analysis of Air France 447", Vol. 17, pp. 64–79. https://dx.doi.org/10.1080/1463922X.2015.1106618

Thieme, C.A., Utne, I.B. (2017), "Safety performance monitoring of autonomous marine systems", *Reliability Engineering & System Safety*, Vol. 159, pp. 264–275. https://dx.doi.org/10.1016/j.ress.2016.11.024

Thieme, C.A., Utne, I.B., Haugen, S. (2018), "Assessing ship risk model applicability to Marine Autonomous Surface Ships", *Ocean Engineering*, Vol. 165, pp. 140–154. https://dx.doi.org/10.1016/j.oceaneng.2018.07.040

U.S. National Transportation Safety Board (2017). *Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida May 7*, 2016, Accident Report HAR1702, National Transportation Safety Board, Washington, D.C., USA.

Vagia, M., Transeth, A.A., Fjerdingen, S.A. (2016), "A literature review on the levels of automation during the years. What are the different taxonomies that have been proposed?" *Applied Ergonomics*, Vol. 53, pp. 190–202. https://dx.doi.org/10.1016/j.apergo.2015.09.013

Vasstein, K., Brekke, E.F., Mester, R., Eide, E. (2020), "Autoferry Gemini: a real-time simulation platform for electromagnetic radiation sensors on autonomous ships", *IOP Conference Series: Materials Science and Engineering*, Vol. 929, 012032. https://dx.doi.org/10.1088/1757-899x/929/1/012032

Veitch, E., Hynnekleiv, A., Lützhöft, M. (2020), "The Operator's Stake in Shore Control Centre Design: A Stakeholder Analysis for Autonomous Ships", *Proceedings of the Royal Institution of Naval Architects. Presented at the Human Factors, The Royal Institution of Naval Architects*, London, UK. https://doi.org/10.3940/hf.20

Vy (2020), *Try Europe's first self-driving reservation-only bus.* [online] Vy. Available at: https://www.vy.no/en/news/europes-first-self-driving-bus (accessed 12.4.20).

Wróbel, K., Montewka, J., Kujala, P. (2018), "System-theoretic approach to safety of remotely-controlled merchant vessel," *Ocean Engineering*, Vol. 152, pp. 334–345. https://dx.doi.org/10.1016/j.oceaneng.2018.01.020

Article 4

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

*Correspondance: erik.a.veitch@ntnu.no

Preprint date: 17 April 2023; Revised 11 May 2023

Abstract: Autonomous ships require efficient remote monitoring and human control intervention to ensure safety. However, there are knowledge gaps concerning human factors influences on remote supervisory control. We investigate the influence of five factors on remote supervisory control using a simulator experiment: (i) Skill, represented by either experienced navigators or gamers; (ii) Vigilance, manipulated by either 5 or 30 minutes in passive monitoring; (iii) Multitasking, represented by either 1 or 3 supervised vessels; (iv) Time Pressure, represented by 20- or 60-second critical time windows; (v) Decision Support, represented by presence or absence of a Decision Support System (DSS). The experiment was a randomized factorial design (n = 32) where volunteers completed a handover (automation detects a critical event and hands over control) and a takeover (operator detects a critical event and takes over control). We observed the following results: (i) Skill influenced performance when combined with Multitasking and Decision Support, favoring gamers; (ii) Vigilance influenced performance when combined with Time Pressure; (iii) Multitasking influenced performance directly, as did (iv) Time Pressure and (v) Decision Support. These outcomes contribute to the empirical basis of maritime human factors research and to safer design of autonomous vessels and their remote control centers.

Keywords: Human factors; supervisory control; autonomous ships; remote control center; simulator testing; skill; vigilance; multitasking; time pressure; decision support systems

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 applications 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 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 & Utne, 2022; J. Liu et al., 2022; Ramos et al., 2020). In this domain, charting the limits of human performance translates to Risk-Influencing Factors (RIFs): important parameters in the development of safety requirements. 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 (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 & 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 & Steinert, 2021; MacKinnon et al., 2020; Öztürk et al., 2022; Ramos et al., 2018; Wróbel et al., 2021).

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 Reinforcement Learning in path planning (e.g., Deraj et al., 2023; Rongcai et al., 2023; Wang et al., 2022), whose results, although utilizing methods like neural networks that claim to mimic the human brain, do not yet comply with the nuanced, often ambiguous stipulations of the Collision Regulations (COLREGs) (Öztürk et al., 2022). 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 control 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 and the remote control center at large were listed as the most "high priority issues" to be addressed, above seemingly more important issues like connectivity and cyber-security. 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).

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 intention is to determine how important maritime human factors influence performance of supervisory control of autonomous vessels. The outcomes 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 and operation approval.

To approach this research question, we break down maritime human supervisory control into two distinct sub-tasks: (i) *handovers*, where the automation fails and hands over control to the operator, and (ii) *takeovers*, where the operator takes over control to resolve a situation for which the automation is unsuited. Based on a literature review, five human factors are selected for the investigation and their hypothesized influences on performance are formulated. (More could have been selected; however, five factors balance experimental breadth with a manageable scope of experimental trials.) These efforts yielded what may be considered the first large-scale, systematic experiment of maritime human factor influences on supervisory control performance of autonomous vessel operations.

There are three major contributions: (i) the hypothesized effects of five important human factor influences on remote supervisory control are tested, (ii) a method is described for rigorous testing of human-machine interaction, and (iii) results are presented that can be adopted to generate more accurate models of human performance in autonomous ship systems.

1.1. Simulator test platform

We used a simulator for our experiments that re-created a real autonomous research vessel called *milliAmpere2* (Figure 1), designed and operated by the Norwegian University of Science and Technology (NTNU). Built in the Unity game platform (Version 2022.2.8; Unity, 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 was used in this study and 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).

Figure 1: Left: Autonomous urban passenger ferry milliAmpere2 during public trial operations. Right: Simulation of remote supervisory control of milliAmpere2 at NTNU Shore Control Lab

1.2. Selection of five influencing factors for investigation

In this study, five factors are selected for investigation; these are listed "A" to "E" below. Here, we explain *why* the five influencing factors were selected for the study based on a literature review. Then, we present *how* the five influencing factors were tested (Section 2, Method) and *what* these effects were (Section 3, Results), as well as their implications (Section 4, Discussion).

Factor A: Skill. The skills possessed by remote supervisors will likely influence their performance. A significant amount of research has investigated what skills – or, more specifically, what accumulated knowledge, understanding, and proficiencies (KUPs) (IMO, 2017, Table A-II/1) – are needed for remote supervisors (Veitch and Alsos, 2022). Most researchers agree that remote supervisors should possess the essential skills of conventional navigators, including ship-handing and seamanship (Baldauf et al., 2019; Saha, 2021; Yoshida et al., 2020). However, others point out that remote supervision will also require a host of new skills (Lutzhoft et al., 2019; Sharma & Kim, 2021). To date, the potentially transferrable skillsets of other groups have not been investigated. We therefore introduce 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), but never in the context of supervisory control of maritime autonomous vessels. Two levels were therefore set for the Skill factor, representing two groups of individuals with distinctly transferrable KUPs: (i) licensed maritime navigators and (ii) gamers.

Factor B: Vigilance. The amount of time in a passive monitoring role affects vigilance – the capacity to pay attention (Warm et al., 2008). Vigilance, in turn, is known to affect the performance of operators working with automated systems (Parasuraman & Riley, 1997). In psychology experiments, a drop in vigilance (known as "vigilance decrement") has been shown to set in somewhere between 20 and 30 minutes, 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 & 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). In this study, we leverage the known association between vigilance and the subjective experience of boredom to gain insights into vigilance decrement (Pattyn et al., 2008). This method is described by Senderud (2022) who, using half the dataset generated in this study and methods grounded in cognitive psychology, hypothesized that a time-related boredom experience may result in poorer performance. The work presented herein expands on Senderud's work, leveraging the full dataset and expanding the breadth of response metrics to confirm whether vigilance decrement influences supervisory performance. Two levels were selected for the Vigilance factor, representing a gap wide enough to instill a vigilance decrement in one group, but not the other: (i) 5 minutes and (ii) 30 minutes.

Factor C: Multitasking. Multitasking 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 multitasking 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 multitasking for maritime remote supervision may lead to speculative claims about how many vessels should be allotted per operator (Dybvik et al., 2020), which, in turn, may lead to inappropriate designs. Some of the uncertainly may stem from the general expectation that maritime vessels, which move relatively slowly, are natural candidates for a multi-vessel supervisory format. However, while multi-vessel formats may prove advantageous and may even address a contemporary lack of skilled mariners (Laugaland, 2022), they must be formally tested to shore up our understanding of how multitasking influences maritime supervisory control. We therefore selected two levels for the Multitasking factor that reflected the number of autonomous vessels being overseen at one time: (i) a single vessel and (ii) three vessels.

Factor D: Time Pressure. The amount of time available to attend to a critical event is, like multitasking, known to affect remote supervision performance in drone applications (Cummings & Mitchell, 2006; Gutzwiller et al., 2016; D. Liu et al., 2016). To date, however, the topic has not been formally investigated for autonomous maritime vessels. Given that all critical events impose a response window, an operator's ability to take timely and decisive preventative actions may be considered among their most important assets. 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 selected two levels for the Time Pressure factor: (i) 20 seconds and (ii) 60 seconds. These levels were set based on iterative testing in the simulation scenarios developed (see Sections 2.3.1 and 2.3.2 for descriptions of the scenarios).

Factor E: Decision Support. 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, dynamic 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., 2021). Some research has assessed navigators' user feedback qualitatively after interacting with a DSS interfaces for maritime navigation (Aylward et al., 2022; Costa et al., 2017); however, quantitative studies assessing DSS influence on performance have not been conducted to date. In the work presented herein, we designed a DSS that contained two main features: the first was object detection (see Figure 4 in Section 2.2 for an example), inspired by recent work on object tracking techniques developed by Helgesen et al. (2022); the second was textual warning prompts with a sound alert that appeared ten seconds prior to the critical event, inspired by concepts developed by Wu et al., (2021). We selected two levels for the Decision Support factor: (i) On and (ii) Off.

The relevance of Factors A through E may be illustrated in the case of the Helge Ingstad accident. The Helge Ingstad was a Norwegian frigate that collided at night with an oil tanker with 137 people onboard, leading its total loss with only minor injuries to crew (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. The findings of the accident investigation pointed to multiple contributing factors that reflect the human factor influences we have chosen to study. Namely, the bridge crew lacked sufficient skills and training (Skill); the Vehicle Traffic Services (VTS) failed to monitor Helge Ingstad on their screens (Vigilance); the navigator in charge was responsible for training two persons at the same time as they were navigating the ship (Multitasking); Helge Ingstad was sailing with a speed of 16-17 knots in a congested area leaving a short reaction time window for everybody involved (Time Pressure); and, finally, the technical systems on the bridge were inadequately used and had an inefficient design and location (Decision Support). This accident illustrates the importance of maritime human factor influences on ship safety and the relevance of the five factors we have chosen to investigate.

1.3. Hypotheses

Below is a list of five hypotheses we test formally in this study. These hypotheses correspond to the five factors we targeted and are guided by state-of-the-art knowledge on their respective expected influences presented in Section 1.2.

Hypothesis A: we expect that **skill** will influence performance; specifically, we expect licensed maritime navigators perform better than gamers.

Hypothesis B: we expect that the **vigilance** will influence performance. Demonstrating this has two parts: first, we must confirm whether a vigilance decrement exists between the 5- and

30-minute scenarios. Having confirmed this, we expect that vigilant operators in 5-minute scenarios will perform better than non-vigilant operators in 30-minute scenarios.

Hypothesis C: we expect the **multitasking** will influence performance; specifically, individuals with one vessel will perform better than those with three vessels.

Hypothesis D: we expect **time pressure** will influence performance; specifically, individuals with 60 seconds available to take preventative action will perform better that those with 20 seconds.

Hypothesis E: we expect that **decision support** will influence performance; specifically, individuals for whom the DSS is turned on will perform better than for those for who the DSS is turned off.

2. Method

In the experiment, 32 volunteer participants without any prior knowledge individually completed two scenarios. The two scenarios represent distinct supervisory intervention events for the milliAmpere2 ferry: the "own-boat" (OB). Each participant was assigned the role of a supervisory operator whose role was to "ensure the safety of the OB and its passengers" and whose primary task was to "take over remote control, if necessary." The scenarios were both initiated with the OB in autonomous mode, whereupon it crossed back and forth a 100-m-wide urban canal in Trondheim at a top speed of 3 knots, with passengers embarking and disembarking, just as in the field trials (which at the time of the experiment has not yet occurred; see Figure 1). Occasionally some pleasure crafts crossed the path of the OB. The latter successfully detected these targets and automatically stopped to let them pass. Eventually, intervention events occurred, marked by a transition to manual control by using the controllers to navigate the ferry manually to safety, whereupon the autonomous mode could once again be initiated. The key feature that allowed for systematic investigation was the experimenters' ability to adjust independent factor settings in the simulator (Section 2.3) and measure changes of dependent factors (Section 2.8).

The NTNU Shore Control Lab test facility allowed the simulator experiment to be undertaken in controlled conditions, removing effects of potentially confounding factors. The lab is based on a standard simulator set-up: the experimenter orchestrates the simulator from an Instructor Station (Figure 2a) and a volunteer participant interacts with the simulator at the Operator Station (Figure 2b). Further details about the lab, including its technical specifications and its design process, are available in Veitch et al. (2021).

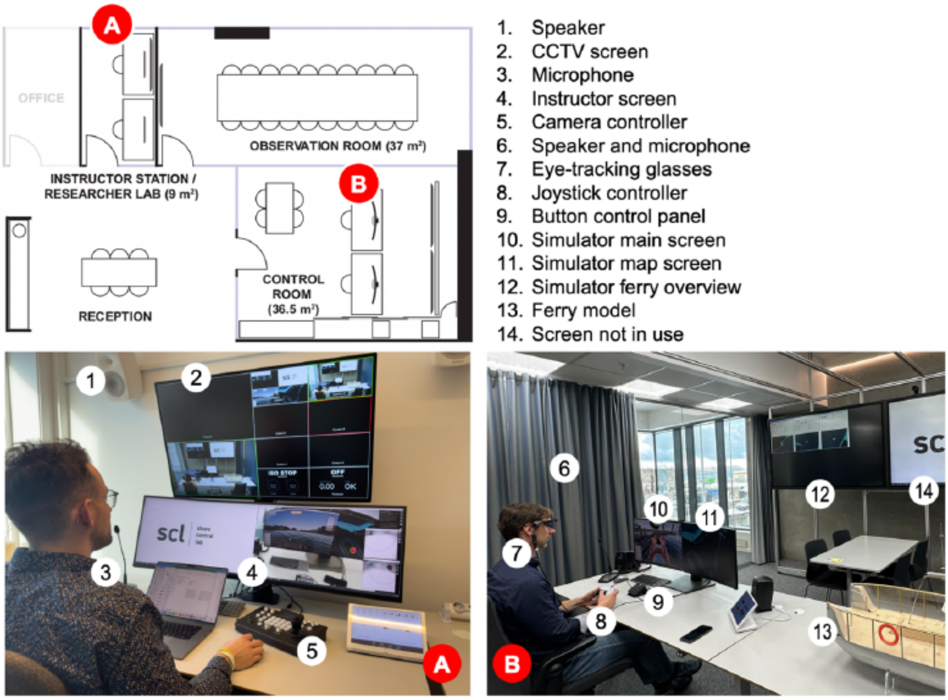


Figure 2: (a) Experimenter in Instructor Station, (b) Participant at Operation Station in Control Room (reconstruction)

For gameplay in the simulator, two controllers are available: (i) a customized Stream Deck button panel with basic controls (dock/undock, manual control on/off, switch between vessels, change camera view) and (ii) a PlayStation5 joystick controller for maneuvering the vessel and changing camera angle and source. Details about the controllers and their 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 18-65 years of age who, to permit use of eye-tracking, did not use eyeglasses (corrective lenses were permitted). Demographic information is presented in Table 1.

i. **Gamers**. The gamer group 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 hung up 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.
ii. **Navigators**. The navigator group 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.

Table 1: Demographic information about the gamer and navigator samples (n = 16 per group)

	Navigators				Gamers			
Gender	*Male*		*Female*		*Male*		*Female*	
	14		2		13		2	
Age	*Mean*	*St Dev*	*Min*	*Max*	*Mean*	*St Dev*	*Min*	*Max*
	43	9.5	28	62	25	5	20	41
Experience	*Seafaring experience (years)*				*Gaming experience (years)*			
	Mean	*St Dev*	*Min*	*Max*	*Mean*	*St Dev*	*Min*	*Max*
	14	8.1	3	28	15	5.6	7	30
Qualification	*Deck officer qualification*				*Gaming frequency (avg h/week)*			
	Class 1	11	Class 5	2	*Mean*	*St Dev*	*Min*	*Max*
	Class 2	1	Class 6	1	16.5	8.9	4.5	35
	Class 5	1	(>15 m)					
Other attributes	DP certificate		Experience with auto-crossing					
	10		2					

2.2. *Experiment protocol and data collection*

Every trial underwent the same procedure according to a prescribed experimental protocol (Figure 3). The instructor (first author) was present for all trials; the co-instructor varied (either the second or fourth author or a graduate student) but was always briefed on the protocol prior to testing. Trials took approximately 1 to 2 hours to complete, depending on the length of the scenarios.

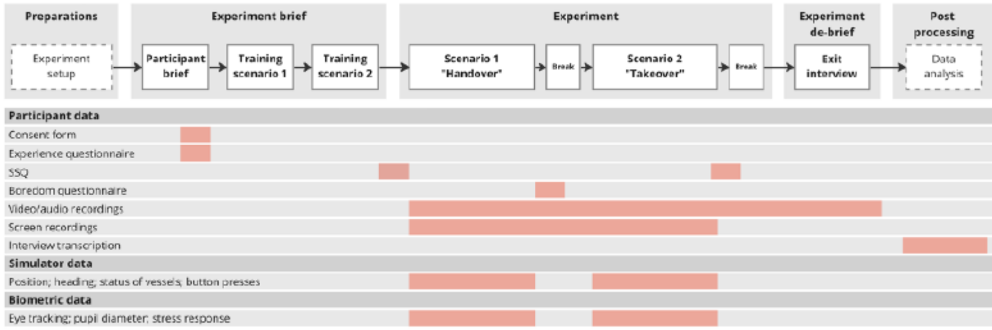


Figure 3: Experiment protocol and data collection

To begin, all volunteers consented to participation and filled out an Experience Questionnaire. There was a separate Experience Questionnaire for the gamer and navigator groups (see Appendices A and B, respectively).

Following this, participants were explained the functionality of the *milliAmpere2* with reference to a 1:10 scale model (see Figure 2b) and underwent two 5- to 10-minute training scenarios designed to familiarize them with their expected tasks, including when and how to take over in potential handover or takeover scenarios (see Figure 4). The eye-tracking glasses, biometric wristband, and portable microphone used for data collection were also placed on the participant for familiarization yet were not yet switched to the "record" position. After completing the training scenarios, the participant completed a Simulator Sickness Questionnaire (SSQ; see Appendix C) to record their "before test" condition. The SSQ, adopted from Kennedy et al. (1993), 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.

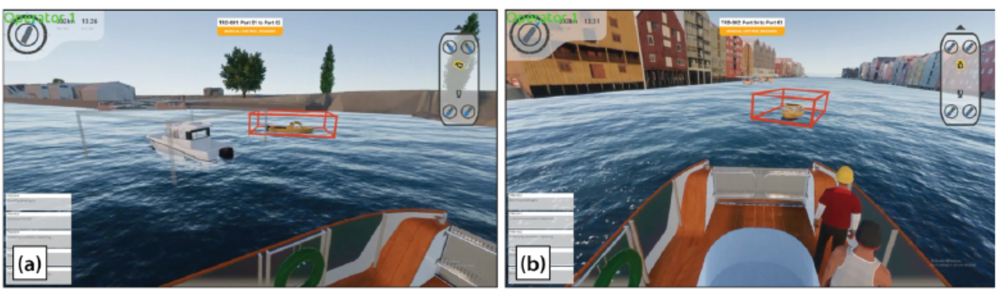


Figure 4: Two manual training scenarios: (a) maneuvering around a stationary vessel and docking (b) collision avoidance during a multi-vessel encounter. Grey boxes indicate stationary vessels; red indicates moving vessels.

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 (Appendix D). Then, after a short break, they completed Scenario 2 and filled out a second SSQ to record their "after test" condition.

Following testing, the instructor or co-instructor led an exit interview with the participant. This was a semi-structured interview consisting of 15-16 questions, which were recorded and later transcribed for analysis (see Appendix E for the Interview Guide). Note that not all interview questions are relevant to work presented herein.

Because this experiment involved collecting and recording personal data (e.g., names, demographic information, identifying features in video and audio recordings, emails) it was necessary to consider the ethical treatment of personal data. A detailed plan for conducting the experiment and managing personal data was submitted to the Norwegian Agency for Shared Services in Education and Research (Sikt), who approved the project plan (Project Number 493904) and archived the data results. Measures for confidentiality and anonymity of volunteer test participants were upheld during and after experimentation. There were no conflicts of interest with any participant, and no participants were personally known to the experimenters. All participants received a gift card as acknowledgment for their participation.

Table 2 lists the data that was collected during experimental trials. All numerical recordings were Unix time-stamped to allow for synchronization during post-processing. Note that in this article, only simulator data and participant data were used; analysis of biometric data (eye-tracking and Electrodermal Activity) is not included herein.

Table 2: Description of collected data and their sources

Type	Data	Description	Source
Simulator data	Position and heading	Position and heading for all vessels (4 Hz)	Simulator
	Status of ferry (or ferries)	Docked/un-docked, contact, boarding/de-boarding, camera view engaged (4 Hz)	Simulator
	Button-presses	All input on button-panel and handheld controller (4 Hz)	Stream Deck; Playstation5 controller
	Screen recording	Video recordings of all three screens used by participants in the simulator	
Biometric data	Eye tracking	Gaze position (50 Hz)	PupilLabs PupilCore eye-tracking glasses
	Eye tracking	Pupil diameter (both eyes, 50 Hz)	PupilLabs PupilCore eye-tracking glasses
	Stress response	Electrodermal Activity (EDA, 4 Hz)	EmpaticaE4 wristband
Participant data	Interviews	High quality interview transcriptions (annotated with body language, missing words, definitions of jargon)	Wearable microphone
	Demographics	Information about participants gauging skills and experience	Questionnaire
	Boredom	Participants' subjective feeling of boredom and related affective experiences	Questionnaire
	CCTV	Video sources showing participants during trials (front and back views)	Video camera (x2)

2.3. Scenario design & factor settings

In preparation for the experiment, two scenarios were designed in the simulator. A "scenario" in this sense can be considered a storyline that can be played and re-played in the simulator, analogous to a "level" in a video game. Both scenarios were designed to investigate performance for manual intervention in a safety-critical situation. While one scenario may have been sufficient to investigate this, there were three reasons for conducting two. Firstly, two scenarios provided a way to verify overall results through repetition; secondly, two runs provided a way to investigate how handover and takeover interventions were distinct; and thirdly, two runs minimized the risk of missed data points should a single scenario run fail to provide data.

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 more 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 parsimonious enough to allow drawing "most likely" conclusions from the results. 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 combinations yielded $2^5 = 32$ unique permutations, or "treatments" (see Appendix F for the Run Log). Each treatment was tested first on Scenario 1, then on Scenario 2, for a total of 64 sets of recorded outcomes.

Table 3: Factor settings for simulated scenarios

<table>
<tr><th></th><th>Factor</th><th>Factor description</th><th>Factor levels</th><th>Description (Scenario 1)</th><th>Description (Scenario 2)</th></tr>
<tr><td rowspan="2">A</td><td rowspan="2">Skill</td><td rowspan="2">The cumulative knowledge, understanding, and proficiencies characterizing an individual's experience.</td><td>Gamer</td><td colspan="2">Participants in the simulator are in individuals who regularly play video games</td></tr>
<tr><td>Seafarer</td><td colspan="2">Participants in the simulator are in individuals with a valid Norwegian maritime navigation certificate</td></tr>
<tr><td rowspan="2">B</td><td rowspan="2">Vigilance</td><td rowspan="2">A mental state characterized by the capacity to sustain one's attention on a task, which is associated with the experience of boredom.</td><td>5 min</td><td>Five minutes passes before the automation fails and control handover begins</td><td>Five minutes passes before a collision must be avoided by control takeover</td></tr>
<tr><td>30 min</td><td>Thirty minutes passes before automation fails and control handover begins</td><td>Thirty minutes passes before a collision must be avoided by control takeover</td></tr>
</table>

C	**Multi-tasking**	The phenomenon of dividing one's limited attention across multiple concurrent tasks.	1 ferry	Single ferry in operation at the location of the trial operation (Figure 1)	
			3 ferries	Two additional ferries routes, each separated by approximately 1 km	
D	**Time pressure**	A feature of time-critical interventions characterized by the window of time available to take preventative action.	20 sec	20 seconds are available to complete handover before scenario ends	20 seconds are available to take over control and avoid collision
			60 sec	60 seconds are available to complete handover before scenario ends	60 seconds are available to take over control and avoid collision
E	**Decision support**	A Decision Support System designed to assist navigators make safety-critical decisions.	On	Object detection; warning notification provided about the automation failure 10 seconds prior	Object detection; warning notification provided about the collision course 10 seconds prior
			Off	Default interface is displayed (e.g., video streams, speed, heading, basic alarms)	

2.3.1. Scenario 1: "Handover"

In this scenario, the autonomous mode fails and prompts the operator to take over manual control. The operator thereafter manually drives the ferry to one of the two ferry terminals using the controller (Figure 5). The prompt to take over control is a flashing alert with a beeping sound. When decision support is turned on, an additional orange alert message appears in the upper center of the screen ten seconds prior to the critical event, reading "Fatal system error! Prepare for POSSIBLE manual control takeover."

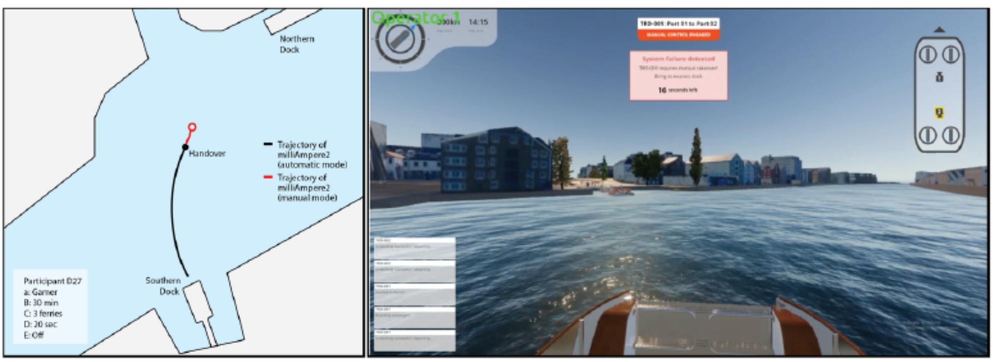


Figure 5: Scenario 1 top-down illustrative schematic (left) and screenshot main operator view in simulator (right).

2.3.2. Scenario 2: "Takeover"

In this scenario, the autonomous mode cannot resolve a multi-boat traffic situation. The operator needs to identify this situation and take evasive action to avoid the impending collision

(Figure 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 take over control and manually maneuver the ferry to safety. When decision support is turned on, an additional orange alert message appears in the upper center of the screen ten seconds 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).

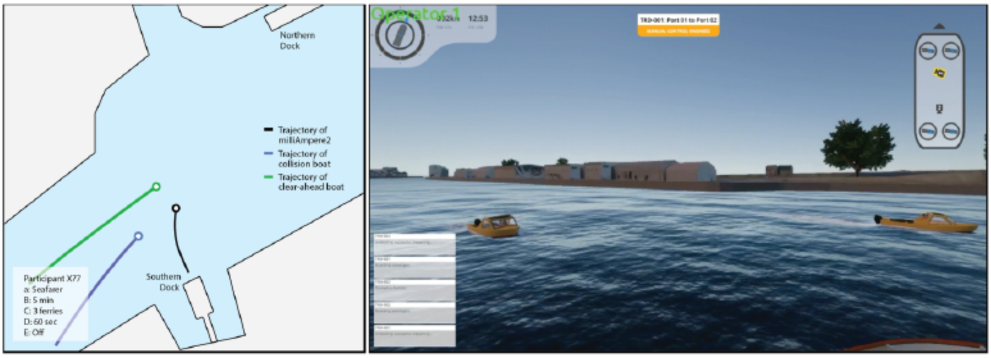


Figure 6: Scenario 2 top-down illustrative schematic (left) and screenshot main operator view in simulator (right).

Note that Scenario 2 was always run after Scenario 1. 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-minute scenarios, a short break of about 5 to 10 minutes was allotted in between.

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 (see Appendix F for the Run Log). No repetition points were conducted, which is acceptable considering the purpose of the experiment is factor screening, not model prediction.

The most important feature of the test structure was that factor treatments were run in random order. However, while the factors settings in the simulator were easy to change, the factor setting for participant’s experience was not so. The approach taken was therefore to repeat 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,” named for its propensity in agricultural testing where plots of land are sub-divided for practically testing fertilizer treatments (Montgomery, 2017, Chaper 14.5.3). The resulting restricted randomization was accounted for in the data analysis (see Section 2.8).

2.5. Handover performance and response 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%) represented an immediate control transfer with no interval of time passing between automatic control, followed by a chosen path equivalent in distance from the dock to the automation. This was computed by comparing the trajectory after the handover to its equivalent in automatic mode (Equation 1). A score of 0% thus represented a state of remaining stopped, or never taking over control. In the case where a participant reversed 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:

$$\mathrm{Score_i} = 1 - \frac{\mathrm{d_{a_i}} - \mathrm{d_{m_i}}}{\mathrm{d_{a_i}}} \times 100\% = \frac{\mathrm{d_{m_i}}}{\mathrm{d_{a_i}}} \qquad \text{Equation 1}$$

$$= \sqrt{\frac{\left(\mathrm{x_{m_i}} - \mathrm{x_0}\right)^2 + \left(\mathrm{y_{m_i}} - \mathrm{y_0}\right)^2}{\left(\mathrm{x_{a_i}} - \mathrm{x_0}\right)^2 + \left(\mathrm{y_{a_i}} - \mathrm{y_0}\right)^2}} \qquad \text{Equation 2}$$

Where d_{m_i} is the absolute distance travelled after handover (manual mode), d_{a_i} is the absolute distance travelled by baseline (automatic mode), and (x_0, y_0) is the location of the handover event. The average of all scores over the handover range was taken as the final metric of handover performance (Equation 3).

$$Score = \frac{1}{n}\sum_{i=1}^{n} Score_i \qquad \text{Equation 3}$$

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-second scenarios, the latter of which may obtain higher scores if there is more time to “catch up” to the baseline ferry trajectory at its docking position.

Figure 7 illustrates one such handover where the mean performance metric is 53%.

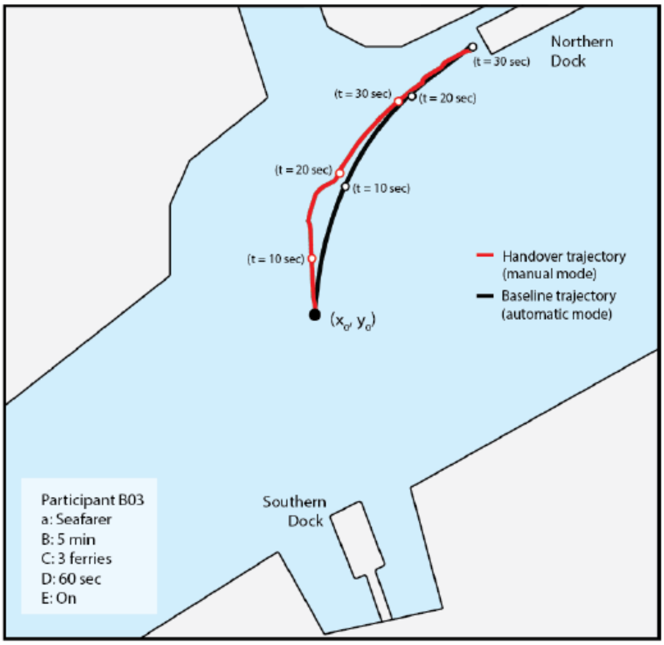


Figure 7: Top-down overview of Scenario 1: Performance metric compares handover trajectory to a baseline.

Handover response 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 seconds for a scenario where the critical event occurred after 30 minutes (1800 seconds), then the reaction time was 7 seconds.

2.6. Takeover performance (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. Figure 8 shows examples of the three performance categories.

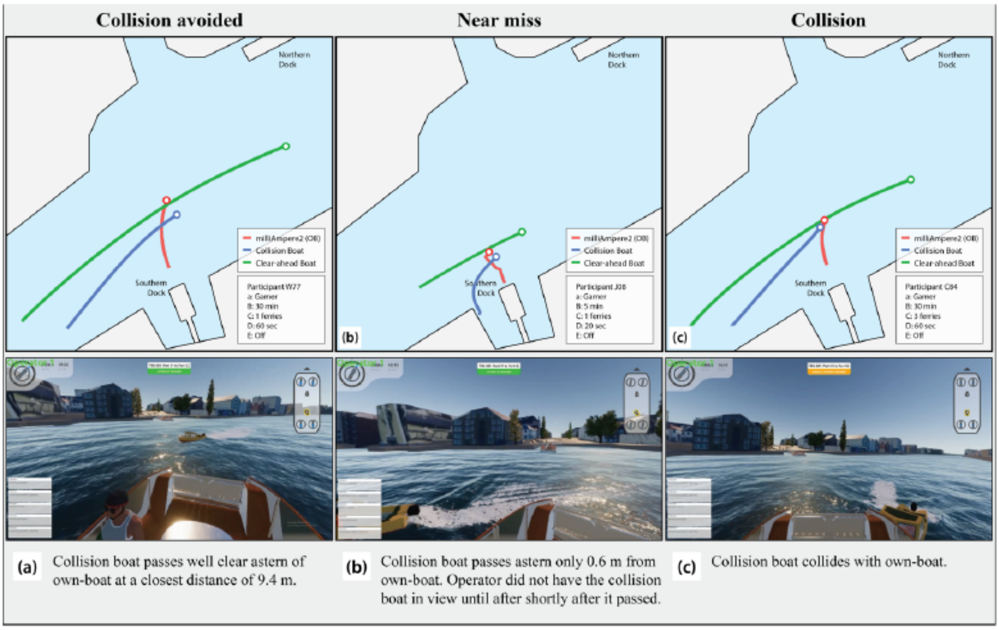


Figure 8: Examples of performance categories for Scenario 2: (a) Collision Avoided, (b) Near Miss, and (c) Collision

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 hulls (approximated to the nearest 0.1 m).

In Scenario 2, measures of response time were not recorded because there were not necessarily indicative of successful collision avoidance. This is perhaps best illustrated with an example: one participant may have detected the potential target collision and hesitated to take over until the vessel was relatively close (e.g., 2 < CPA < 5 m); another may take over almost immediately after detected the potential collision target (e.g., CPA > 5 m). Since both strategies were successful, reaction time – the interval between stimulus (target detected) to reaction (takeover) – was not indicative of overall performance.

2.7. Measuring vigilance decrement by proxy of boredom experience

It is more straightforward to ask about individuals' subjective experience of boredom than to ask about their vigilance decrement. Because subjective experience of boredom is associated with vigilance decrement (Pattyn et al., 2008), we used a two-part Boredom Questionnaire as a proxy for measuring vigilance decrement (see Appendix D). The first part asked, "To what extent did the task you just complete make you feel bored?" on a Likert scale from 1 to 7. The second was more detailed, basing its assessment of boredom across seven distinct affective experiences on a Likert scale from 1 to 5, without explicitly mentioning boredom. This two-part questionnaire method was developed and tested in van Tilburg and Igou (2012). For more details about how boredom was investigated in this experiment, see Senderud (2022).

2.8. Quantitative analysis

The aim of the quantitative analysis was to estimate the extent to which the five factors independently influenced performance and response time. 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. A significance level of $\alpha = 0.10$ was set as the prescriptive baseline of significance. 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 & 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 & 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 completely independent of run order. Should any of these diagnostics reveal signs that ANOVA assumptions were not met, then the analysis was deemed inappropriate and was not presented.

The results presented in Section 3 are those that meet the conditions listed above. Note that rather than proceeding to present model coefficients for prediction purposes, the analysis stopped short of selecting significant factors. This is in line with the aim of the experiment, which is to screen influencing factors, not develop predictive models.

3. Results

The results are divided into four sections:

i. The first section confirms that a vigilance decrement was observed across the 5- and 30-minute scenarios (Section 3.1).
ii. The second section reports on the results of handover performance, as assessed in Scenario 1 (Section 3.2).
iii. The third section reports on the results of takeover performance, as assessed in Scenario 2 (Section 3.3).
iv. The fourth section presents ANOVA test statistics for the three observed factor effects listed above (Section 3.4).

3.1. Vigilance decrement results

We begin by testing whether a vigilance decrement was observed, because confirming this will allow us to use the 5- and 30-minute-long scenarios as proxies for high and low levels of vigilance, respectively, in the rest of the analysis.

Figure 9 presents two boxplots depicting the spread and central tendencies of responses to the two-part Boredom Questionnaire (Appendix D). Likert scale responses were normalized. The results show a similar result across the two-part questionnaire, with an even distribution around 0.5, indicating that individuals' boredom experiences, while ranging considerably, centered around a neutral response.

In the boxplots, 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 and mark outliers as dots. This applies to all boxplots presented.

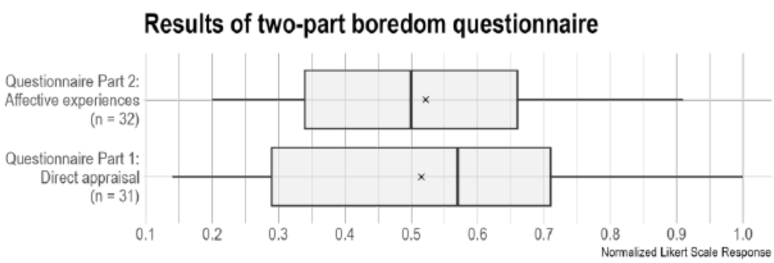


Figure 9: Boxplots of boredom questionnaire results

Figure 10 and Figure 11 present boxplots dividing the two datasets in Figure 9 across the five factors tested at low and high levels. This provides a way to visually infer factor effects and initial observations towards testing the hypothesized presence of a vigilance decrement

(Hypothesis B). Note that 31 responses were collected for Questionnaire Part 1 because one participant did not fill it out.

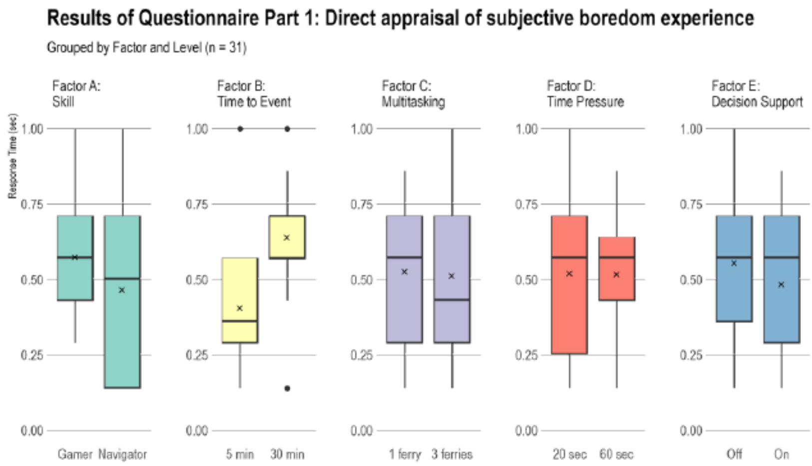


Figure 10: Boxplots of normalized responses from Boredom Questionnaire Part 1, grouped by factor and sub-divided by level

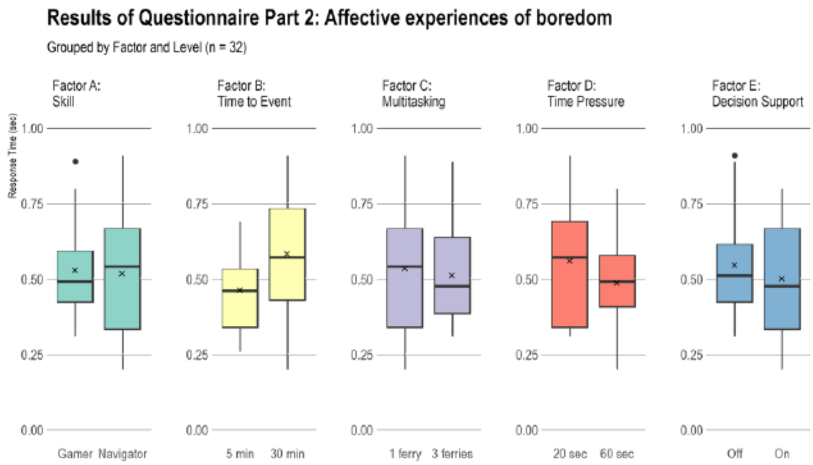


Figure 11: Boxplots of normalized responses from Boredom Questionnaire Part 2, grouped by factor and sub-divided by level

Below is a list of initial observations:

- As expected, a vigilance decrement is observed when comparing the 5- and 30-minute scenarios. This is evident as a significant rise in boredom when assessed through both direct appraisal (Part 1) and through associated affective experiences (Part 2).
- The length of the scenario appears to be the only significant contributor to boredom.

Table 4 depicts the ANOVA for each factor effect on the two boredom responses, respectively. Note that to appropriately analyze the categorical data using ANOVA, the data was normalized to a scale between 0 and 1 such that an arcsine square root transformation could be applied and uphold ANOVA assumptions (see Section 2.8 for details). All main effects are listed; however, for conciseness, only significant interaction effects are shown.

In all ANOVA tables in this article, we present three test statistics: i) Degree of freedom (DoF), (number of estimated parameters used to compute the effect's sum of squares and thereafter 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 if the null hypothesis is true; i.e., if the effect is zero).

Table 4: ANOVA tables for vigilance responses collected in the two-part Boredom Questionnaire (significant p-values *are highlighted).*

	Factor effects on vigilance *(Questionnaire Part 1: Direct appraisal of boredom)*			***Factor effects on vigilance*** *(Questionnaire Part 2: Boredom affective experiences)*		
Factor	*DoF*	*F-value*	*p-value*	*DoF*	*F-value*	*p-value*
A	24.00	1.89	0.1814	26.00	0.0269	0.8710
B	24.00	6.51	**0.0175**	26.00	3.43	**0.0753**
C	24.00	0.0028	0.9585	26.00	0.1228	0.7289
D	24.00	0.0040	0.9499	26.00	1.39	0.2494
E	24.00	1.61	0.2167	26.00	0.5375	0.4700
BC	24.00	3.16	**0.0880**			

Overall, only Factor B had a significant effect on vigilance at the 90% confidence level ($p < 0.10$) (for Part 2 results, the p-value was slightly over the confidence threshold; however, for Part 2 results, it was well below). While the interaction effect BC also appeared to influence vigilance in the Questionnaire Part 1 response, it was not repeated by the Questionnaire Part 2 results and was therefore deemed negligible.

Qualitative results support the claim that a vigilance decrement existed in the 30-minute scenarios but not in the 5-minute scenarios. In the interviews, when asked, "How would you describe your experience as a Shore Control Center Operator?" a total of 13 said they felt bored; of these, 11 had completed a 30-minute scenario. The remaining 19 described their experience in more engaged terms like "fun" or "interesting"; of these, only 5 had completed 30-minute scenario. One excerpt from participant K85 illustrates the experience of vigilance decrement in a 30-minute scenario:

> *"I soon noticed that it got a bit boring. Especially when I checked out that map, for example, and saw that there weren't going to be any dangerous situations for quite a while. Then it was a bit like, 'OK, why am I sitting here? There's not much to do here.'" – K85*

Participants' body language also showed signs of vigilance decrement, irrespective of Skill group (gamer or navigator). Figure 12 presents images taken during three participants' trials and contrasts body language for 0-5 minute and 20-30 minute time windows.

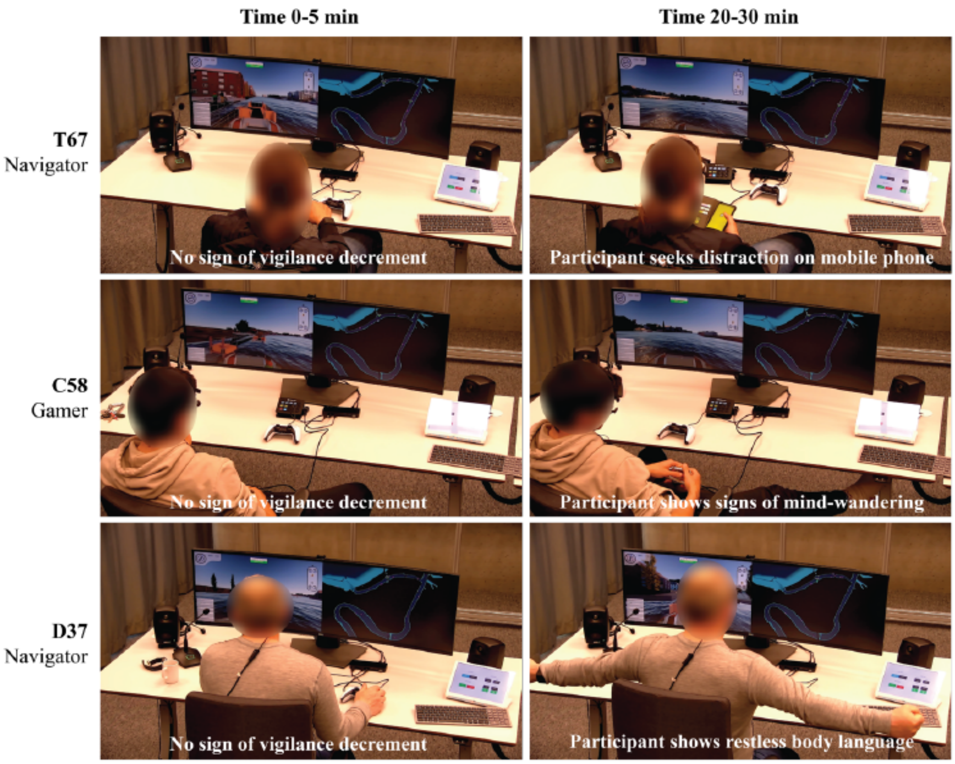


Figure 12: Qualitative indicators of vigilance decrement

To illustrate what this means in practice, consider the follow implications:

- We can confirm that a vigilance decrement sets in between 5- and 30-minutes of monitoring.
- We can therefore use vigilance levels of "Low" and "High" as proxies for 5-minutes and 30-minute scenarios, respectively, in the rest of the study.

3.2. Handover performance and response time results (Scenario 1)

Figure 13 presents a boxplot visually depicting the spread and central tendency of performance scores for all 32 trials of Scenario 1. The mean score was 29% and median was 24%; the lowest score was 0% and highest was 68%. There were three occurrences with score 0, indicating no handover at all occurred during the available time.

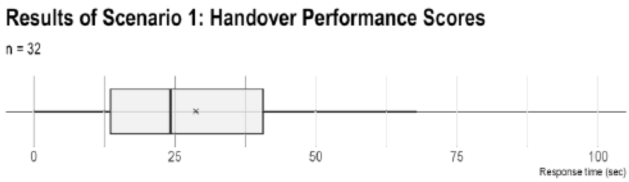


Figure 13: Boxplot of scores (/100%) in Scenario 1

Figure 14 presents boxplots dividing the dataset five times across the five factors and across low and high levels. This provides a useful way to visually assess factor effects and infer initial observations.

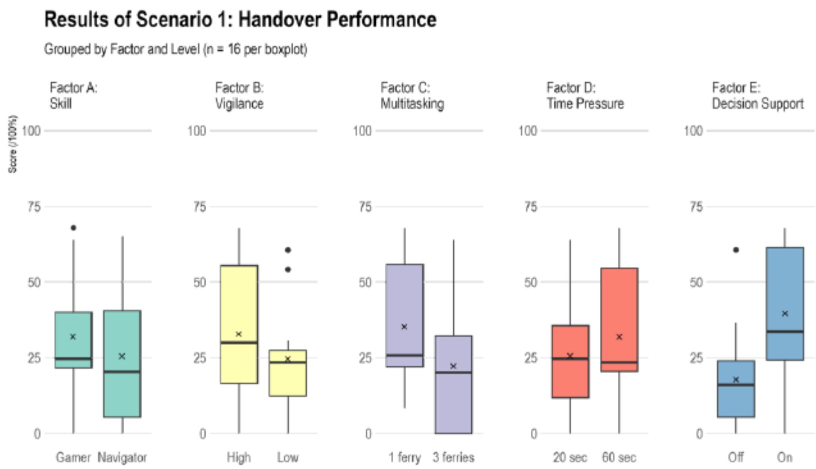


Figure 14: Boxplot of handover performance scores grouped by factor and sub-divided by level

Below is a list of initial observations inferred from Figure 14:

- Surprisingly, **gamers performed slightly better** than navigators at the task. However, the difference does not appear significant (Hypothesis A).
- As expected, **30-minute/low vigilance scenarios yielded poorer performance** than 5-minute/high vigilance scenarios; however, the difference does not appear significant (Hypothesis B).

- As expected, **scenarios with 3 ferries yielded poorer performance** than scenarios with 1 ferry. In fact, all three times operators failed the automation handover completely (score 0) was when they had responsibility for 3 ferries (Hypothesis C).
- As expected, **scenarios with 20 seconds available time yielded poorer performance** than scenarios with 60 seconds; however, the difference does not appear significant (Hypothesis D).
- As expected, **scenarios with decision support yielded better performance** than those without (Hypothesis E).

Results of ANOVA for factor effects on handover performance are presented in Table 5 (Section 3.4). Note that a square root transformation was applied to the response data to adhere to ANOVA assumptions (see Section 2.8 for details). From the ANOVA, we can confirm the initial observations: only Factors C and E had a significant effect on handover performance at the 90% confidence level ($p < 0.10$).

Furthermore, Figure 15 presents a boxplot visually depicting the spread and central tendency of handover response times for 29 trials of Scenario 1 (three data points were removed, representing those that did not complete the handover during the available time). The mean handover response time was 8 seconds, and the median was 6.5 seconds; the lowest response time was 3 seconds, and the highest was 30 seconds. The distribution of response times was heavily right tailed.

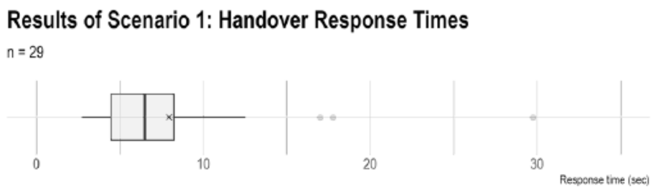


Figure 15: Boxplot of handover response times in Scenario 1

Figure 16 presents boxplots dividing the dataset five times across the five factors and across low and high levels. This provides a useful way to visually assess factor effects and infer initial observations.

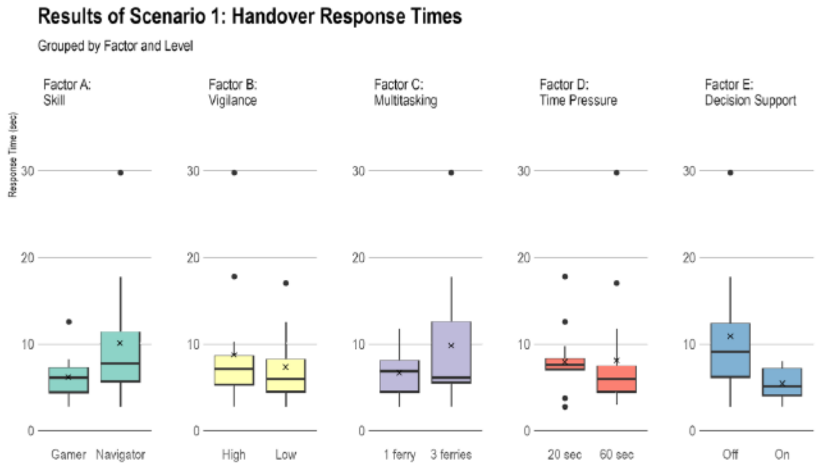


Figure 16: Boxplot of Scenario 1 handover response times grouped by factor and sub-divided by level

Below is a list of initial observations inferred from Figure 16:

- Surprisingly, **gamers had faster handover response times** than navigators; although, it is unclear whether this difference is significant (Hypothesis A).
- Surprisingly, **30-minute/low vigilance scenarios yielded slightly lower response times** than 5-minute/high vigilance scenarios; however, the difference is negligible (Hypothesis B).
- As expected, **scenarios with three ferries yielded higher response times** than scenarios with one ferry (Hypothesis C).
- Surprisingly, **time pressure** had no discernable effect (Hypothesis D).
- As expected, **scenarios with decision support yielded visibly lower response times** than those without (Hypothesis E).

Results of ANOVA for factor effects on handover response times are presented in Table 5 (Section 3.4). From the ANOVA, we can confirm the initial observations that Factors C and E had a significant effect on handover performance at the 90% confidence level ($p < 0.10$). However, while gamers did demonstrate faster response times than navigators, the difference was not statistically significant at the 90% confidence level baseline prescribed in the ANOVA assumptions. In addition, three interaction effects were found to be significant (AC, AE, and CE).

In addition to the ANOVA results, some interview excerpts qualitatively supported the claim that Multitasking negatively influenced performance. For example, B03 expressed the challenge of paying attention to multiple vessels at once; similarly, N33 described the difficulty of "keeping up":

> *"I was thinking, how many other boats should you monitor to be sure you have full control of what is going on? Situations can arise quickly… I tried the whole time to switch ferries and make up my mind about which ferry required the most attention at that time." – B03*

> *"It's a question of how many ferries there are out there: it could quickly become a bit much to keep track of. When there were three ferries, I think it was very difficult to keep up." – N33*

These results allow us to formally test the hypotheses outlined in Section 1.3:

- **Reject Hypothesis A:** Performance was similar across skillsets, whether measured by performance score or by response time.
- **Reject Hypothesis B:** Vigilance did not affect performance, whether measured by performance score or by response time.
- **Accept Hypothesis C:** Multitasking affected performance. When supervising three vessels compared to a single vessel, scores decreased by 13% and response times increased on average 3 seconds, from 7 to 10 seconds. Moreover, all three participants that never responded at all (score 0) were supervising three ferries.
- **Reject Hypothesis D:** Time pressure did not affect performance, whether measured by performance score or by response time.
- **Accept Hypothesis E:** By introducing decision support, scores increased 22% and the response times decreased by 6 seconds, from 11 to 5 seconds.
- **Interaction effect AC**: For gamers, response times were the same across single- and three-ferry scenarios (6 seconds); however, for navigators they increased 7 seconds (from 7 to 14 seconds) for three-ferry scenarios.
- **Interaction effect AE:** By introducing decision support, response time decreased by only 2 seconds (from 7 to 5 seconds) for gamers; however, it decreased by 7 seconds (from 14 to 6 seconds) for navigators.
- **Interaction effect CE:** By introducing decision support, response time decreased by only 3 seconds for single-ferry scenarios (from 8 to 5 seconds); however, it decreased by 10 seconds (from 15 to 5 seconds) for three-ferry scenarios.

3.3. Takeover performance results (Scenario 2)

Figure 17 presents a bar chart visually depicting collision avoidance outcomes for all 32 trials of Scenario 2. The most common result was "Collision" (15). The rest were "Near Misses" (4) and "Collision Avoided" (13) (see Section 2.6 for definitions of these three categories).

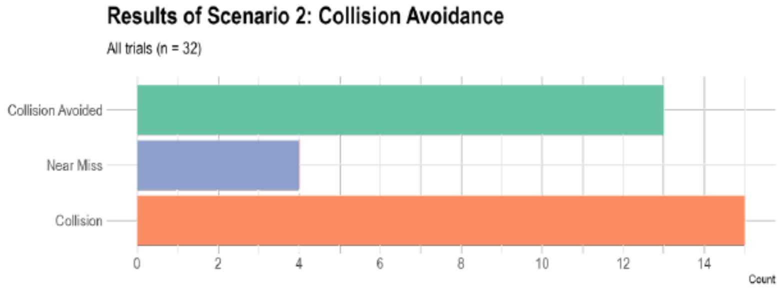


Figure 17: Bar chart of takeover performance results

Figure 18 presents a stacked bar chart dividing the dataset five times across the five factors and across low and high levels. This provides a useful way to visually assess factor effects and infer initial observations.

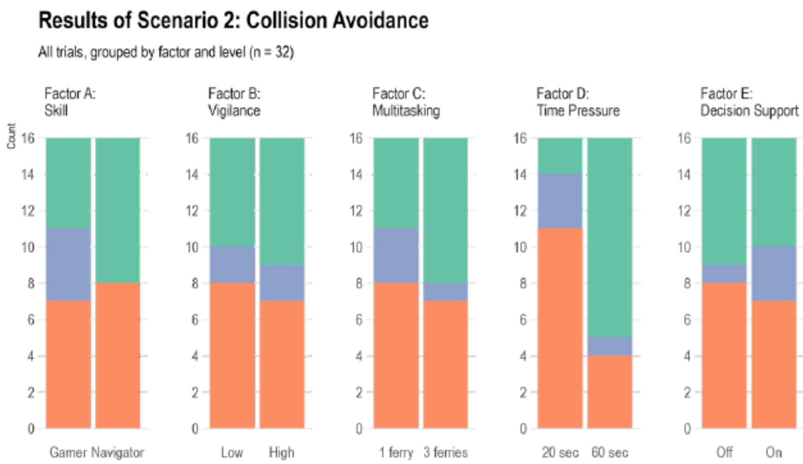


Figure 18: Stacked bar charts of takeover performance results, grouped by factor and sub-divided by level

Below is a list of initial observations inferred from Figure 18:

- Surprisingly, **gamers and navigators experienced about the same number of collisions** (Hypothesis A).
- Surprisingly, **there was about the same number of collisions for 5- and 30-minute scenarios** (Hypothesis B).
- Surprisingly, **there was about the same number of collisions for single- and three-ferry scenarios** (Hypothesis C).

- As expected, **collisions decreased markedly when the available time increased from 20 to 60 seconds** (Hypothesis D).
- Surprisingly, **there was about the same number of collisions when decision support was available**, compared to when it was not (Hypothesis E).

Results of ANOVA for factor effects on takeover performance are presented in Table 5 (Section 3.4). Note that to appropriately analyze the categorical data using ANOVA, the data was first transformed to a numerical scale between 0 and 1 such that an arcsine square root transformation could be applied and uphold ANOVA assumptions (see Section 2.8 for details). The transformation mapped "Collision Avoided" to 1, "Near Miss" to 0.5, and "Collision" to 0. From the ANOVA, we can confirm the initial observation that only Factor D had significant effects on collision results at the 90% confidence level ($p < 0.10$). In addition, the interaction effect BD was found to be statistically significant.

Qualitative observations helped to explain some of the more surprising results. Specifically, the finding that gamers and navigators experienced similar collision rates may have been due in part to an unfamiliar interface on the part of navigators. Ten out of the sixteen navigators remarked that the interface lacked standard features, including "chart plotters," "Electronic Chart Display Information System" (ECDIS), or "radar." For example, O52 complained about the lack of Closest Point of Approach (CPA) vectors on the ECDIS:

> *"It's common on ECDIS and such, where you have vectors, and where you can see if you're on a collision course… you can set an alarm, and if there's a vector that fits– if there's a CPA that is too low, then you'll get an alarm." – O52*

Furthermore, there were no indications in the interviews that participants found the DSS useful during the collision avoidance takeover sequence, helping to explain the unexpected result that performance was unchanged when it was turned off. On the contrary, there were some indications that participants found it distracting, as exemplified by T90's account:

> *"…from my career at sea, I'm used to having anti-collision alarms on almost all things, like, OK, 'You're steering towards land! Doodoo!' [mimics alarm sound] – and it was annoying because you heard it all the time, and in the end, you don't even bother to respond to it… You don't need all the gadgets in the world because that just gets annoying." – T90*

These results allow us to formally test the hypotheses outlined in Section 1.3:

- **Reject Hypothesis A:** Collision avoidance outcome was similar across skillsets.
- **Reject Hypothesis B:** Vigilance did not affect collision avoidance outcome.
- **Reject Hypothesis C:** Multitasking appeared not to affect collision avoidance outcome.
- **Accept Hypothesis D:** Time pressure affected collision avoidance outcome. When the available time was increased from 20 to 60 seconds, collisions decreased by 7, from 11 to 4. Near-misses also decreased, from 3 to 1.
- **Reject Hypothesis E:** Introducing decision support did not affect collision avoidance outcome.

- **Interaction effect BD:** Collision rates were unaffected by available time for 30 minutes scenarios; however, for 5-min scenarios, collision rates increased from 0 to 7 when the available time increased from 20 to 60 seconds.

3.4. *Results of ANOVA for performance metrics*

Table 5 presents ANOVA results of the five factor effects and relevant interaction effects for the three performance metrics presented in Sections 3.2 and 3.3 (handover performance score, handover response time, and takeover collision avoidance outcome). Note that all main effects are displayed; however, only significant interaction effects are presented for conciseness. For more details about the ANOVA methods, see Section 2.8. For a description of the three test statistics presented (DoF, F-value, and *p*-value), see Section 3.1.

Table 5: Results of ANOVA for all factor effects on performance (significant p*-values are highlighted).*

	Factor effects on handover performance scores (Scenario 1: Handover)			*Factor effects on handover response time (Scenario 1: Handover)*			*Factor effects on takeover performance results (Scenario 2: Takeover)*		
Factor	*DoF*	*F-value*	p-*value*	*DoF*	*F-value*	p-*value*	*DoF*	*F-value*	p-*value*
A	26.00	2.63	0.1170	1.94	4.64	0.1679	2.00	0.1250	0.7575
B	26.00	1.29	0.2668	17.99	0.3688	0.5513	23.00	0.2277	0.6377
C	26.00	7.57	**0.0107**	17.99	8.36	**0.0097**	23.00	0.9109	0.3498
D	26.00	0.97	0.3342	17.99	0.3010	0.5900	23.00	14.57	**0.0009**
E	26.00	11.54	**0.0022**	18.23	21.69	**0.0002**	23.00	0.0000	1.0000
AC				17.99	5.45	**0.0314**			
AE				18.18	5.51	**0.0305**			
BD							23.00	8.20	**0.0088**
CE				18.23	8.06	**0.0108**			

4. Discussion

Skillset influenced performance, but only in combination with other factors – and in a direction counter to our hypothesis. One of the unexpected findings was that background skills appeared to favor gamers over navigators in the supervisory role. Based on recent scientific literature (Baldauf et al., 2019; Lutzhoft et al., 2019; Saha, 2021; Yoshida et al., 2020), it was hypothesized that navigators would have the most transferrable skills to the role of supervisory operators and would therefore demonstrate better performance. It was surprising, therefore, when we observed the opposite: across all categories of performance, gamers displayed better performance than navigators. This was the true for handover performance (6.5% better, Figure 14), handover response time (4 seconds faster, Figure 16), and collision avoidance (1 fewer collision, Figure 18). This result indicated that gamers may have some skillsets that are more transferrable than those of licensed maritime navigators.

However, ANOVA showed that these differences had a greater than 10% probability of being due to chance – a baseline that could not allow us to conclude with confidence that gamers performed any differently than navigators (Table 5). Still, the lack of any detectable difference led us to reject the hypothesis that reflected contemporary wisdom: namely, that navigators performed better in remote supervisory tasks than gamers.

This finding comes with two important considerations. Firstly, this finding is only valid when comparing licensed maritime navigators and gamers. If we had had a third group representing a control with no transferrable skills at all, then perhaps the outcome would have more clearly favored both gamers and navigators. An alternative interpretation may be interpreted as such: both gamers and navigators may have equally transferrable skills to the role of a remote operator. The second consideration is that this result considers only the causal effect of skillsets and ignores its potential interaction effects with other performance-influencing factors.

The second of these considerations leads us to highlight two important interaction effects: (i) Skill-Multitasking (Factor AC) and (ii) Skill-Decision Support (Factor AE). The non-additive effects of both these pairs of factors are illustrated in handover response time results (Figure 19). Specifically, when compared to navigators, gamers' multitasking response times were lower (Figure 19a), and gamers' DSS-unassisted response times were lower (Figure 19b). This implies that gamers were more adept at multitasking and that they relied less on decision support than navigators. The latter implication may be explained by navigators' propensity to check ECDIS and similar tools standard in maritime navigation (but lacking in the control room set up). Gamers, by contrast, are unfamiliar with such interfaces tools and therefore did not know what they were missing.

Similar observations have been made when comparing gamers and pilots during drone operations. For example, Lin et al. (2015) found that video gaming expertise was associated with better performance during simulator 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 & Mayer, 2002), which might explain one mechanism behind their improved performance. Further afield, research focusing on human-machine teaming typically employs video games as their testing arena, (e.g., Crandall et al., 2018; Shirado and Christakis, 2017) suggesting that expert-level gameplay may harbor transferrable skills to the unique demands encountered in human-machine interaction applications.

A concern was raised that age might introduce a confounding factor with response time performance. The gamer group was, on average, 18 years younger than the navigator group – an age gap that, according to well-established literature, is sufficient to produce an age-related reaction time difference (Deary & Der, 2005). Notwithstanding, the reaction time difference was too small (on the order of tenths of seconds, according to Era et al., 1986 and Fozard et al., 1994) to make any significant difference in this experiment. Differences in reaction time measured in this experiment were hundreds of times greater than those resulting from aging;

therefore, we can conclude that the differences we observed originated from mechanisms other than age.

Additionally, neither AC nor AE interaction effects were detected for other, response-time-independent, performance metrics (Table 5), suggesting that response time and overall performance were largely independent. Indeed, response times and performance scores in Scenario 1 were only weakly correlated (Kendall's $\tau = -0.53$, $p < 0.001$). In other words, gamers may have been *faster*, but this did not mean they performed *better* than navigators. At the very least, the results indicate the extent to which response time can interact positively with multitasking and decision support technologies during critical handover events.

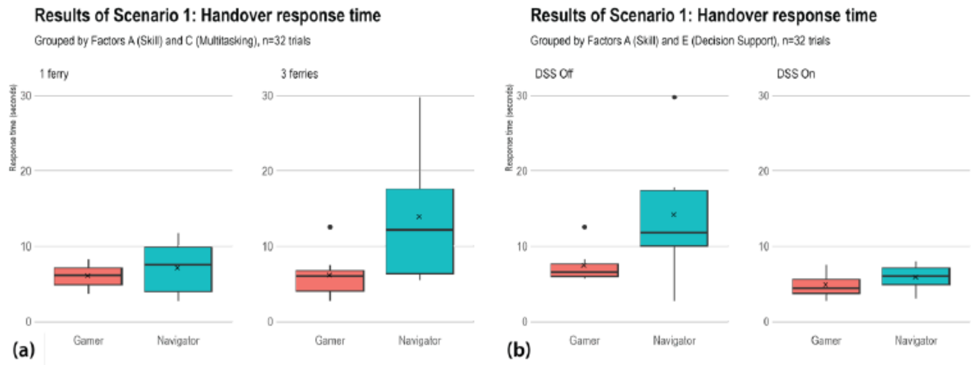


Figure 19: Boxplots illustrating interaction effects AC (a) and AE (b)

Vigilance decrement was detected and influenced performance – but only in combination with other factors and only in takeover scenarios. Although vigilance decrement was clearly observed in terms of subjective experiences of boredom (Section 3.1), we found no evidence supporting that it directly influenced remote supervision performance. This was a surprising result because the literature so clearly underscored vigilance (or the lack thereof) as an influencing factor for performance (e.g., Kari & Steinert, 2021; Ramos et al., 2018; Yoshida et al., 2021). Certainly, the observed behaviors that portrayed boredom betrayed a sense of safety (Figure 12).

Moreover, because gamers are used to high-paced video games, one could expect them to experience more boredom than navigators, the latter of whom are used to sustaining attention during monotonous days on the job. Evidence did not support this claim either, however, as shown by the lack of AB interaction effects on performance metrics (Table 5). Indeed, other research supports this findings, as illustrated by Warm et al. (2008) who found that vigilance decrement affects both experienced and inexperienced operators in equal measure. The findings serve to reify how van Tilburg and Igou (2012) concluded their in-depth studies on boredom: namely, that boredom is simply "a chore," suggesting that its potentially negative effects on performance may be avoided so long as boredom is stoically dealt with. One interview excerpt from a 30-minute scenario participant serves to illustrate this claim:

> *Interviewer [Erik]: "How would you describe your experience as an operator of the Shore Control Center?"*
>
> *Participant [I29]: "I thought it was fun. F***ing boring, too! [Laughs.]"*
>
> *Interviewer [Erik]: "It's interesting, because you have 9000 hours as a DP operator, so one would think you'd be used to this?"*
>
> *Participant [I29]: "But that's why I replied as I did instead of saying right away that it's boring. If it's my job, it doesn't matter. Okay, sure, it's boring, but when something suddenly happens, you have to be there."*

Although vigilance decrement did not influence performance directly, it does not exclude the possibility that its pernicious effects may instead be produced by interactions between factors. Indeed, the interaction effect Vigilance-Time Pressure (Factor BD) was found to significantly affect takeover performance (Table 5). Specifically, vigilance decrement appeared to produce a leveling effect on performance in 30-minute scenarios that was not present in 5-minute scenarios. Specifically, it appeared that in a vigilant state, 20 seconds was not sufficient time to avoid the collision (100% collision rate) whereas 60 seconds was sufficient (100% collision avoidance rate); in a low vigilant state, however, the outcomes appeared random across the board (both with 50% collision rate) (Figure 20). This suggests that vigilance acts as a foil for time-critical performance: its presence accentuates the effect of time pressure, and vice versa. Therefore, despite no direct effect on performance, researchers' warnings about the potential dangers of boredom during remote supervision appear warranted when considering its interaction effect with time pressure.

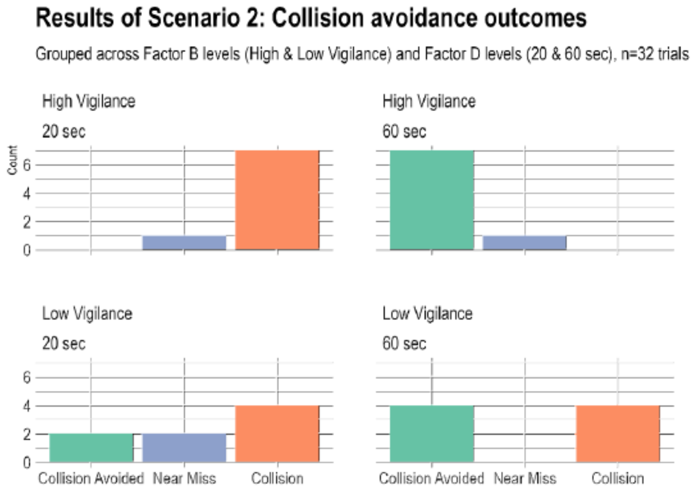


Figure 20: Bar charts of collision avoidance outcomes (Scenario 2) illustrating interaction effect BD

Over the two types of scenarios tested, only takeover scenarios demonstrated boredom's harmful effects during monitoring. A takeover, contrary to a handover, requires sustained attention to detect a potential hazard, diagnose its significance, and take the appropriate actions to deal with it. This cognitive process, which, in the context of safety-critical work is often referred to as "situation awareness" (Endsley, 1995) or Information-Decision-Action (IDA; Smidts et al., 1997), hinges in large part on the capacity to direct one's attention. This attention-directing capacity is, by definition, hampered in a state of diminished vigilance brought on by "underload" during 20-30 minutes of passive monitoring. Therefore, the claims of the potentially harmful effects of boredom may be more specifically directed towards the specific case of takeovers like that typified by appropriately identifying, reacting, and acting to avoid a collision.

Multitasking during supervisory control had a significant influence on performance. Specifically, the number of autonomous vessels proved to be a major contributor to handover performance. Specifically, operators with three vessels had a 15% drop in performance score (Figure 14) and 3-second increase in reaction time (Figure 16) compared to those with just one vessel. This result supported our hypothesis. We had expected, namely, that the increased attentional demands of monitoring several vessels would compromise performance during manual intervention.

There are several studies in the literature that corroborate these findings. For example, in one experiment, operators of drones were found to have "significantly degraded" situation awareness when tasked with handling 16 missiles compared to 8 or 12 (Cummings & Guerlain, 2007). In another experiment, air traffic controllers were found to have significantly lower situation awareness when tasked handling more than 12 aircraft (Endsley & Rodgers, 1996).

Interestingly, our confirmation that multitasking negatively affects performance runs counter to the promise that automation will improve task performance and overall operational efficiency. 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). That automation solutions should really work is predicated, it would seem, on no influence of human multitasking whatsoever. At the very least, the results suggest that the human-machine interface should be designed specifically to address multitasking to mitigate its effects on supervisory performance.

The interview responses to the questions "What can be improved?" and "What could have made the scenario easier for you?" highlighted at least three ideas for mitigating multitasking effects. The most popular (reported by 17 participants) was the inclusion of a notification-based tracking, CPA, and collision warning system. Another suggestion, reported by 11 participants, was to use 360-degree cameras onboard the vessels, thus removing the need to switch camera angles (e.g., front and back). A related suggestion, reported by 5 participants, was to display all vessels simultaneously, removing the need to switch between individual vessels. All these interface features may serve to direct users' attention more effectively, thus mitigating the potentially harmful effects of multitasking.

Multitasking also interacted with other factors to produce non-additive effects on response times in the handover scenario. Specifically, this included the interaction factor Skill-Multitasking (Factor AC; see Figure 19a) and Multitasking-Decision Support (Factor CE; see Figure 21 below). Considering the importance of time-critical response during handover, these interaction effects underscore the role of decision support in supervisory control. For example, response time tripled from 5 to 15 seconds when the DSS was turned off for three-ferry operations; although, response time changed only from 5 to 8 seconds when the DSS was turned off for single-ferry operations (Figure 21). The pernicious effects of multitasking, it would appear, can be almost entirely circumvented with a well-designed DSS.

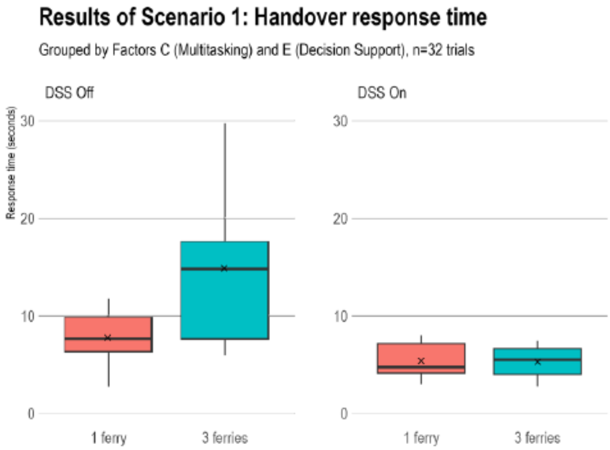


Figure 21: Boxplots illustrating interaction effect CE

The amount of time available to attend to a critical event strongly affected takeover performance. This was in line with our hypothesis (Section 1.3), with the caveat that it only affected takeovers and not handovers. The situation awareness and cognitive processing demands associated with takeovers appeared to impose a time window for completing the associated characteristic stages of information gathering, information processing, and preparation for action. This is analogous to the "maximum response time" defined by Rødseth et al. (2021), 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).

It follows intuitively that the more demanding the task, the more time is required. Surprisingly, though, this was not observed in the case of time pressure interaction effects with multitasking or decision support: the only other factor that interacted with Time Pressure was Vigilance (Factor BD). As explained before, the outcomes of collision avoidance appeared random in a state of vigilance decrement (Figure 20, bottom row), compared to causal outcomes in a state of high vigilance (Figure 20, top row). Interestingly, the results may also suggest the phenomenon of "attentional tunneling," as evidenced by poorer performance when combining high vigilance with high time pressure: a combination that might lead to a failure to consider alternative diagnostic hypotheses during a critical event (Figure 20, left column). Aside from the deleterious effects of time pressure, attentional tunneling is known to be amplified by automation complacency (Wickens & Alexander, 2009), a state in which the operator who trusts the automation is at risk of failing to notice an unexpected failure (Parasuraman & Riley, 1997).

The availability of a Decision Support System significantly influenced handover performance. This was the case both in terms of the overall performance score and response time, for which the presence of a DSS proved to be the single largest contributor to performance improvements (see Figure 14 and Figure 16, respectively). This confirmed our hypothesis and illustrated the power of DSSs for leveraging the advantages of human-machine teaming in human supervisory control applications.

Interestingly, the DSS did not affect takeover performance (Figure 18). One reason for this may be that takeovers rely on the operator's own situation awareness-related cognitive processes to a greater extent than prompts from a DSS. This is 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). The DSS, in attempting to assist the user through the high-level cognitive processes of collision target diagnosis and avoidance, may even risk distracting them (see the interview excerpts in Section 3.3 for an example). By contrast, the same DSS produced measurable performance enhancements in control handovers: a routine task that, contrary to takeovers, did not summon the full breadth of operators' cognitive powers.

Implications for risk monitoring and management. 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 to 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 also show that the operators need sufficient time to react, and that multi-tasking and the number of vessels involved influence the human response time. This underscores 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 human 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 shows no 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 operators. The results may indicate that the results of the study are impacted using a simulator instead of real-life ship operations and/or that the scenarios were not sufficiently complex enough so that the navigator experience was of limited value; however, this something that should be investigated in future studies.

4.1. *Future work and limitation of this study*

The results of this study can guide further investigations on maritime human factors. For instance, more attention should be allotted to those factors producing observable effects on performance than those not producing observable effects. Specifically, any factor producing either a direct effect or interaction effect should be studied further in the context of supervisory control of highly automated vessels. From this experiment we were only able to glean *whether* these human factor influences 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 study showed that continued investigations can be set out under the dual framework of control handovers and takeovers: two types of interventions that were shown to constitute distinct mechanisms in human supervisory control.

While our experiment employed a scientifically rigorous, transparent, and repeatable methodology, it also introduced some limitations. The most obvious of these was the nature of simulation studies themselves that, no matter how detailed or complete, will never present reality exactly. The reader can refer to many discussions on this topic (e.g., Schricker et al., 2001; Hughes and Rolek, 2003); we will not address this further. The one associated limitation we wish to highlight, however, is the interface we designed for the experiment. The DSS we designed was programmed to act like a DSS, complete with object detection and warning notifications, but it represented a “mechanical Turk” – a device merely posing as an intelligent agent. This was acceptable for the purposes of illustrating its effects in the context of a controlled experiment (the participants, after all, believed it to be a real AI-based assistant tool). However, its implementation comes with a caveat: namely, any DSS will be only as effective as its design. The presence of a DSS alone, in other words, cannot guarantee performance enhancements, and its true merit will hinge on many design factors that are not investigated in this work.

5. Conclusion

In this study, we screened influencing factors on human supervisory control of highly automated vessels and uncovered some surprising results. One such surprise involved the effects of skill and vigilance, which, despite evidence in the literature involving aircraft pilotage, our experimental observations did not detect. One explanation was how these two factors, instead of affecting performance directly, instead *combined* with other factors to influence performance. Specifically, findings indicated that performing under multitasking conditions favored gamers, and that DSS-enhanced performance was only observed for navigators. Our original hypotheses also overlooked the significant interaction effect of combining vigilance with time pressure during control takeovers. Specifically, findings indicated that effect of time pressure manifested itself only in a state of high vigilance.

The experiment also confirmed some pre-existing notions about remote supervision. This included the effects of multitasking, time pressure, and decision support, all of which had significant influences on performance. A closer examination of the results, though, revealed that factor screening alone did not tell the whole story. Just as was the case for skill and vigilance, more meaningful conclusions could be drawn from examination of interaction effects. For instance, the negative effects of Multitasking on performance were shown to be circumvented by the presence of a DSS. This was not the case for takeovers, though, which, in contrast to handovers, relied upon the cognitive demands associated with first detecting and diagnosing a hazardous situation and only thereafter committing to a decisive intervention strategy.

The factor screening outcomes of this experiment can help guide further research and design activities. In this sense, the uncovered rejections and confirmations of hypothesized factor effects do not represent full stops; rather, they should be interpreted as signposts on the path to further investigation. To illustrate this point, consider the case of decision support, which we found to produce the greatest individual influence on performance of all influencing factors studied. A designer should not be contented that simply incorporating a DSS will improve performance; on the contrary, aware of its potential impact, the designer should sharpen their focus on creating interfaces that maximize a DSS's potential (for example, by enabling mechanisms to support multitasking). 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 even very good designs may produce trade-offs in performance, efficiency, and safety.

Overall, there are several outcomes of the study. Firstly, the study contributes empirical results to research on risk assessment and modeling of human and organizational factors, as well as to risk mitigation related to human operator performance. Secondly, the study contributes to engineering and human-centered design activities aimed at developing remote supervisory infrastructure for autonomous maritime vessels. Finally, the study 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: Conceptualization, Investigation, Methodology, Formal Analysis, Writing – original draft, Writing – review & editing. **Ole Andreas Alsos:** Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Tingting Cheng:** Methodology. **Kristin Senderud:** Investigation, Methodology. **Ingrid Bouwer Utne:** Supervision, Writing – review & editing.

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.

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) directed by project leader Kongsberg Maritime AS and partnered with the Institute for Energy Technology (IFE). The authors also acknowledge the support of the Department of Design at NTNU and the entire team behind the Shore Control Lab, especially Felix Petermann for his technical support during experiments, Mikael Hanssen 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 of Dag Rutledal and Kjetil Vasstein for their valuable input on the Experience Questionnaires and on recruitment channels, and finally of Koen van de Merwe for his valuable inputs and assistance with data collection.

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. *Journal of Physics: Conference Series*, *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. *The Journal of Navigation*, *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. Scopus. https://doi.org/10.12716/1001.13.03.02

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, *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. *Journal of Physics: Conference Series*, *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. *International Journal of Human-Computer Studies*, *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. *Nature Communications*, *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. *Human 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. *Human 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. *Interacting with Computers*, *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. *Journal of Aerospace Computing, Information, and Communication*, *3*(6), 294–308. https://doi.org/10.2514/1.19599

Deary, I. J., & Der, G. (2005). Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples. *Aging, Neuropsychology, and Cognition*, *12*(2), 187–215. https://doi.org/10.1080/13825580590969235

Deraj, R., Kumar, R. S. S., Alam, M. S., & Somayajula, A. (2023). Deep reinforcement learning based controller for ship navigation. *Ocean Engineering*, *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. *Human 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. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

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. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *40*(2), 82–85. https://doi.org/10.1177/154193129604000216

Era, P., Jokela, J., & Heikkinen, E. (1986). Reaction and movement times in men of different ages: A population study. *Perceptual and Motor Skills*, *63*(1), 111–130. https://doi.org/10.2466/pms.1986.63.1.111

Fozard, J. L., Vercruyssen, M., Reynolds, S. L., Hancock, P. A., & Quilter, R. E. (1994). Age differences and changes in reaction time: The Baltimore longitudinal study of aging. *Journal of Gerontology*, *49*(4), 179–189. https://doi.org/10.1093/geronj/49.4.P179

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. *Human Factors*, *45*(3), 349–359. Scopus. https://doi.org/10.1518/hfes.45.3.349.27253

Guo, C., & Utne, I. B. (2022). Development of risk indicators for losing navigational control of autonomous ships. *Ocean Engineering*, *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 Engineering*, *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 P. M. Arezes & R. L. Boring (Eds.), *Proceedings of the AHFE 2021 Virtual Conferences on Safety Management and Human Factors, and Human Error, Reliability, Resilience, and Performance* (pp. 219–227). Springer Nature Switzerland. 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. *Reliability Engineering & System Safety*, *207*, 107347. https://doi.org/10.1016/j.ress.2020.107347

Hughes, T., & Rolek, E. (2003). Fidelity and validity: Issues of human behavioral representation requirements development. *Proceedings of the 2003 Winter Simulation Conference*, 976–982. https://doi.org/10.1109/WSC.2003.1261519

Hutchins, E. (1995). *Cognition in the wild*. MIT press.

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)* (MSC.1/Circ.1638). International Maritime Organization. https://www.imo.org/en/MediaCentre/PressBriefings/pages/MASSRSE2021.aspx

IMO. (2022). *Maritime Safety Committee (MSC 106), 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/

Kari, R., & Steinert, M. (2021). Human factor issues in remote ship operations: Lesson learned by studying different domains. *Journal of Marine Science and Engineering*, *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. *The International Journal of Aviation Psychology*, *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 Engineering*, *266*, 113120. https://doi.org/10.1016/j.oceaneng.2022.113120

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

Laugaland, J. M. (2022, August 24). *Stor mangel på sjøfolk – lovar bonus på 7000 for rekruttering*. 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. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *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 Engineering*, *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). *Journal of Marine Science and Technology*, *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. *Journal of Physics: Conference Series*, *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. *ATENA Conferences System, NAV 2015 18th International Conference on Ships and Shipping Research, Milan, Italy*, 24–25.

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

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 Manufacturing*, *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. *International Journal of Industrial Ergonomics*, *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. *Aviation, Space, and Environmental Medicine*, *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. *Human Factors*, *38*(2), 311–322. https://doi.org/10.1177/001872089606380211

Montgomery, D. C. (2017). *Design and analysis of experiments* (Ninth edition). 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 Engineering*, *251*, 111010. https://doi.org/10.1016/j.oceaneng.2022.111010

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human 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? *Physiology & Behavior*, *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 Engineering*, *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. *Reliability Engineering & System Safety*, *195*, 106697. https://doi.org/10.1016/j.ress.2019.106697

Ramos, M. A., Utne, I. B., & Mosleh, A. (2018, September 16). *On factors affecting autonomous ships operators performance in a Shore Control Center*. 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. *Journal of Marine Science and Technology*, *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 Engineering*, *272*, 113779. https://doi.org/10.1016/j.oceaneng.2023.113779

Saha, R. (2021). Mapping competence requirements for future shore control center operators. *Maritime Policy and Management*. https://doi.org/10.1080/03088839.2021.1930224

Schricker, B. C., Franceschini, R. W., & Johnson, T. C. (2001). Fidelity evaluation framework. *Proceedings of the 34th Annual Simulation Symposium*, 109–116. https://doi.org/10.1109/SIMSYM.2001.922122

Senderud, K. (2022). *Kjedsomhet og vigilans hos operatører i kontrollrom for autonome ferger* [Master thesis]. University of Bergen.

Sharma, A., & Kim, T. (2021). Exploring technical and non-technical competencies of navigators for autonomous shipping. *Maritime Policy & Management*, *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. *Applied Cognitive Psychology*, *16*(1), 97–115. https://doi.org/10.1002/acp.759

Smidts, C., Shen, S. H., & Mosleh, A. (1997). The IDA cognitive model for the analysis of nuclear power plant operator response under accident conditions. Part I: problem solving and decision making model. *Reliability Engineering & System Safety*, *55*(1), 51–71. https://doi.org/10.1016/S0951-8320(96)00104-4

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions* (Second Edition). Cambridge University Press.

Suyama, R., Miyauchi, Y., & Maki, A. (2022). Ship trajectory planning method for reproducing human operation at ports. *Ocean Engineering*, *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 Engineering*, *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).

Utne, I. B., Sørensen, A. J., & Schjølberg, I. (2017, September 25). *Risk management of autonomous marine systems and operations*. ASME 2017 36th International Conference on Ocean, Offshore and Arctic Engineering, Trondheim, Norway. 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. *Journal of Marine Science and Technology*, *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. *Motivation and Emotion*, *36*(2), 181–194. https://doi.org/10.1007/s11031-011-9234-9

Veitch, E., & Alsos, O. A. (2022). A systematic review of human-AI interaction in autonomous ship systems. *Safety Science*, *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. *Computer Supported Cooperative Work (CSCW)*. 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 Engineering*, *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. *Human 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. *The International Journal of Aviation Psychology*, *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. *Applied Sciences*, *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. *Applied Sciences*, *10*(23), 1–27. https://doi.org/10.3390/app10238751

Yoshida, M., Shimizu, E., Sugomori, M., & Umeda, A. (2021). Identification of the Relationship between Maritime Autonomous Surface Ships and the Operator's Mental Workload. *Applied Sciences*, *11*(5), 2331.

Appendix A: Gamer Experience Questionnaire

Gamer Questionnaire

Introduction	
Please answer all questions. In case something is unclear, ask the experiment coordinator. Your answers are confidential and anonymous.	
Question	**Answer**
Section 1: Basic information	
1. What is your year of birth?	______
2. What is your gender?	☐ Male ☐ Female Self-identify: ______________
Section 2: Your gaming experience	
1. How often during the week do you game (in hours)?	______
2. What specific games have you played most in the past three years? (List up to three). What is your skill level in each game? (Not in-game skill.)	

Game	**Level**
Ex. League of Legends	Ex. Expert
1.	
2.	
3.	

3. What specific games have you played most more than three years ago? (List up to three). What is your skill level in each game? (Not in-game skill.)	<table><tr><th>Game</th><th>Level</th></tr><tr><td>Ex. League of Legends</td><td>Ex. Newbie</td></tr><tr><td>1.</td><td></td></tr><tr><td>2.</td><td></td></tr><tr><td>3.</td><td></td></tr></table>
4. How many years have you been gaming?	

5. Have you ever used a ship simulator?	☐ Yes ☐ No
6. Do you have a recreational boat license (båtførerbevis)?	☐ Yes ☐ No
7. Do you have general experience in maritime activities? If yes, please describe.	
8. Why do you game?	

Appendix B: Navigator Experience Questionnaire

Navigator Questionnaire

Introduction	
Please answer all questions. In case something is unclear, ask the experiment coordinator. Your answers are confidential and anonymous.	
Question	**Answer**
Section 1: Basic information	
3. What is your year of birth?	________________
4. What is your gender?	☐ Male ☐ Female Self-identify: ______________
Section 2: Your experience at sea	
9. What level of marine qualification do you possess?	A. Deck officer class 1 B. Deck officer class 2 C. Deck officer class 3 D. Deck officer class 4 E. Deck officer class 5 F. Deck officer class 6 (vessels under 15 m) G. Deck office class 6 (vessels under 24 m) H. I am not a certified Deck Officer I. Other: ______________________________
10. How many years have you spent at sea?	________________
11. Have you operated Roll-On-Roll-Off ferries in Norway? If yes, for how many years?	☐ Yes ☐ No ________________

12. Have you operated high-speed passenger carriers (hurtigbåter) in Norway? If yes, for how many years?	☐ Yes ☐ No ____________________________
13. Which part of the Norwegian coast do you have the most experience from?	A. Oslofjorden B. Skagerrak C. Coast from Stavanger to Florø D. Coast from Florø to Smøla E. Coast from Smøla to Vega (including Trondheimsfjorden) F. Coast from Leka to Sklinna / Vestfjorden G. Coast from Støtt to Andenes (including Lofoten) H. Coast from Andenes til Lopphavet I. Coast from Lopphavet til Mehamn J. Coast from Mehamn til Grense Jakobselv K. Other: ____________________
14. What role(s) have you had on ferries?	A. Master B. Chief mate C. I have not operated a ferry D. Other: ____________________
15. When was the last time you operated a ferry?	____________________________
16. Have you ever used a ship simulator?	☐ Yes ☐ No
17. Have you used auto-crossing? If yes, how many years have you used it? How often did you / do you use it (in percentage of total trips?	☐ Yes ☐ No ____________________________
18. Have you used auto-docking? If yes, how many years have you used it? How often did you / do you use it (in percentage of total trips?	☐ Yes ☐ No ____________________________

Appendix C: Simulator Sickness Questionnaire

Simulator Sickness Questionnaire

Kennedy, R. S., Lane, N. E., Berebaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *International Journal of Aviation Psychology*, 3(3), 203-220.

Participant Number: ___________ Date:________________ Time:___________________

When: After / Before Testing

Please indicate the severity of symptoms that apply to you right now.

Symptom	0 No Symptoms	1 Minimal	2 Moderate	3 Severe
General Discomfort				
Fatigue				
Headache				
Eyestrain				
Difficulty Focusing				
Increased Salivation				
Sweating				
Nausea				
Difficulty Concentrating				
Fullness of Head				
Blurred Vision				
Dizzy (eyes open)				
Dizzy (eyes closed)				
Vertigo				
Stomach Awareness				
Burping				

Appendix D: Two-Part Boredom Questionnaire

Questionnaire

To what extent did the task you just completed make you feel bored?

1 (not at all)	2	3	4	5	6	7 (very much)
❑	❑	❑	❑	❑	❑	❑

Think about the task you just completed.

To what extent do you…

	1 (not at all)	2	3	4	5 (very much)
... feel restless and unchallenged at the same time?	❑	❑	❑	❑	❑
... think that the situation served no important purpose?	❑	❑	❑	❑	❑
... feel like doing something completely different?	❑	❑	❑	❑	❑
... feel like doing something more purposeful?	❑	❑	❑	❑	❑
... wish to turn to a more meaningful activity?	❑	❑	❑	❑	❑
... want to do something more meaningful?	❑	❑	❑	❑	❑
... want to be challenged?	❑	❑	❑	❑	❑

Boredom Experience scale (van Tilburg & Igou, 2012)

Appendix E: Interview Guide

Interview

1. Think back to the first scenario. Can you tell me briefly what you did?
2. Thick back to the second scenario. Can you tell me briefly what you did?
3. What information was useful when you were monitoring?
4. Did you notice any potentially dangerous traffic situations and hesitate whether to take over control?
5. *(Ask only if participant had decision support enabled:)* Did you see/hear the notification warnings? If so, when?
6. Think back to the control takeover. Did you make a plan for what to do? … Did you make this plan before or after pressing the Auto On/Off button?
7. Can you tell me what you did after taking over control?
8. To what extent did this follow your plan (if you had one)? (On a scale of 1 “not at all” to 5 “exactly as planned”)
9. How would you describe your experience as a Shore Control Center Operator?
10. How did it feel to take over control?
11. How much time would you estimate passed during the simulation?
12. Were you to repeat the simulation, would you have done anything differently? Why or why not?
13. Did you see the gorilla? *(If “yes”:)* Did you see it in both scenarios?
14. Think back to the display. What could be improved? Was there something you felt could help you have even better situation awareness?
15. Imagine what could have made the scenario easier for you. For example, what indicators might let you know that something is amiss with your vessel?
16. If you could have the best system in the world for remote supervision of autonomous vessels, what would that look like?

Appendix F: Run Log

Run Order (Random)	Standard Order	Group	Code ID	Factor A: Skill	Factor B: Vigilance	Factor C: Multi-tasking	Factor D: Time Pressure	Factor E: Decision Support
1	7	1	N33	Gamer	5 min	3 ferries	60 sec	Off
2	11	1	T90_2	Gamer	5 min	3 ferries	20 sec	On
3	6	1	W77	Gamer	30 min	1 ferry	60 sec	Off
4	10	1	C58	Gamer	30 min	1 ferry	20 sec	On
5	4	1	D27	Gamer	30 min	3 ferries	20 sec	Off
6	16	1	K85	Gamer	30 min	3 ferries	60 sec	On
7	13	1	D28	Gamer	5 min	1 ferry	60 sec	On
8	1	1	J08	Gamer	5 min	1 ferry	20 sec	Off
9	14	2	D33	Gamer	30 min	1 ferry	60 sec	On
10	3	2	B58	Gamer	5 min	3 ferries	20 sec	Off
11	5	2	M87	Gamer	5 min	1 ferry	60 sec	Off
12	9	2	O52	Gamer	5 min	1 ferry	20 sec	On
13	8	2	C84	Gamer	30 min	3 ferries	60 sec	Off
14	12	2	A13	Gamer	30 min	3 ferries	20 sec	On
15	15	2	E02	Gamer	5 min	3 ferries	60 sec	On
16	2	2	C15	Gamer	30 min	1 ferry	20 sec	Off
17	19	3	I65	Seafarer	5 min	3 ferries	20 sec	Off
18	18	3	E51_2	Seafarer	30 min	1 ferry	20 sec	Off
19	21	3	D71	Seafarer	5 min	1 ferry	60 sec	Off
20	30	3	D37	Seafarer	30 min	1 ferry	60 sec	On
21	25	3	P62	Seafarer	5 min	1 ferry	20 sec	On
22	24	3	T67	Seafarer	30 min	3 ferries	60 sec	Off
23	31	3	B03	Seafarer	5 min	3 ferries	60 sec	On
24	28	3	D24	Seafarer	30 min	3 ferries	20 sec	On
25	20	4	T03	Seafarer	30 min	3 ferries	20 sec	Off
26	22	4	I29	Seafarer	30 min	1 ferry	60 sec	Off
27	29	4	L95	Seafarer	5 min	1 ferry	60 sec	On
28	27	4	W78	Seafarer	5 min	3 ferries	20 sec	On
29	17	4	L64	Seafarer	5 min	1 ferry	20 sec	Off
30	26	4	H04	Seafarer	30 min	1 ferry	20 sec	On
31	32	4	Z99	Seafarer	30 min	3 ferries	60 sec	On
32	23	4	X77	Seafarer	5 min	3 ferries	60 sec	Off

Statements of co-authorship

NTNU

DECLARATION OF CO-AUTHORSHIP

Erik Aleksander Veitch is applying for the evaluation of the following thesis:

"Designing for Land-based Control of Autonomous Ships"

The declaration should describe the work process and division of labor, **specifically identifying the candidate's contribution**, as well as give consent to the article being included in the thesis.

Declaration of co-authorship on the following article: Veitch, E., & Alsos, O. A. (2022). A systematic review of human-AI interaction in autonomous ship systems. *Safety Science*, vol. 152, 105778. https://doi.org/10.1016/j.ssci.2022.105778.

Co-authors' contributions:

Alsos, O.A.:

Supervision	Oversight and leadership responsibility for the research activity planning and execution
Project administration	Management and coordination responsibility for the research activity planning and execution
Writing - Review & Editing, Original Draft, Major Review	Critical review, commentary, and revision.
Visualization	Preparation and creation of visualization in the published work
Funding acquisition	Acquisition of the financial support for the project leading to this publication
Resources	Provision of computing resources and other analysis tools

Trondheim,
....
Place, date

Ole Andreas Alsos
....
Signature co-author

PhD Candidate's (main author's) contribution:

Veitch, E.:

Conceptualization	Ideas; formulation or evolution of overarching research goals and aims
Methodology	Development or design of methodology (systematic literature review)
Formal analysis	Application of formal techniques (PRISMA statement), conducting the literature review and synthesizing results
Investigation	Conducting the research and investigation process (literature review)
Writing - Original Draft	Preparation, creation, and presentation of the published work, specifically writing the initial draft
Writing - Review & Editing, Major Review	Preparation, creation, and presentation of the published work
Visualization	Preparation, creation, and presentation of visualizations in the published work

Trondheim, 14.06.2023
………………………..
Place, date

……………………………………
Signature co-author

NTNU

DECLARATION OF CO-AUTHORSHIP

Erik Aleksander Veitch is applying for the evaluation of the following thesis:

"Designing for Land-based Control of Autonomous Ships"

The declaration should describe the work process and division of labor, **specifically identifying the candidate's contribution**, as well as give consent to the article being included in the thesis.

Declaration of co-authorship for the following article: Veitch, E., Dybvik, H., Steinert, M. et al.(2022). Collaborative Work with Highly Automated Marine Navigation Systems. *Computer Supported Cooperative Work*. https://doi.org/10.1007/s10606-022-09450-7

Co-authors' contributions:

Dybvik, H.:

• Conceptualization - Original Draft	Ideas; formulation or evolution of overarching research goals and aims
• Investigation - Original Draft	Conducting the research and investigation process (interview protocol, data collection for interviews)
• Formal analysis - Original Draft	Application of formal techniques (qualitative analysis using software tools) to analyze or synthesize study data

Stanford, 12. October. 2022.
Place, date

Henrikke Dybvik
Signature co-author

Steinert, M.:

• Supervision	Oversight and leadership responsibility for the research activity planning and execution
• Project administration	Management and coordination responsibility for the research activity planning and execution
• Writing - Review & Editing, Original Draft	Critical review and commentary

Trondheim, 22.01.2023
Place, date

Signature co-author

Alsos, O.A.:

- Supervision — Oversight and leadership responsibility for the research activity planning and execution
- Project administration — Management and coordination responsibility for the research activity planning and execution
- Writing - Review & Editing, Original Draft, Major Review, Minor Review — Critical review, commentary, and revision
- Visualization — Preparation and creation of visualizations in the published work

Trondhem,
………………………...
Place, date

…..…………………………………
Signature co-author

PhD Candidate's (main author's) contribution:

Veitch, E.:

- Conceptualization - Original Draft, Major Review, Minor Review — Ideas; formulation or evolution of overarching research goals and aims
- Methodology - Original Draft, Major Review, Minor Review — Development and design of methodology
- Formal analysis - Original Draft, Major Review, Minor Review — Application of formal techniques (qualitative analysis using software tools) to analyze or synthesize study data
- Investigation- Original Draft, Major Review, Minor Review — Conducting the research and investigation process (interview protocol, data collection for interviews, field observations)
- Data Curation- Original Draft, Major Review, Minor Review — Management activities to annotate (produce metadata), scrub data and maintain research data for initial use and later reuse
- Writing - Original Draft — Preparation, creation, and presentation of the published work, specifically writing the initial draft
- Writing - Review & Editing, Major Review, Minor Review — Preparation, creation, and presentation of the published work, specifically writing the major review and minor review
- Visualization — Preparation and creation of visualizations in the published work

Trondheim, 10 October 2022
………………………...
Place, date

…..…………………………………
Signature co-author

NTNU

DECLARATION OF CO-AUTHORSHIP

Erik Aleksander Veitch is applying for the evaluation of the following thesis:

"Designing for Land-based Control of Autonomous Ships"

The declaration should describe the work process and division of labor, **specifically identifying the candidate's contribution**, as well as give consent to the article being included in the thesis.

Declaration of co-authorship on the following article: 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*, vol. 1, pp. 1023-1032. doi:10.1017/pds.2021.102

Co-authors' contributions:

Kaland, T.:

Investigation - Original Draft	Conducting the research and investigation process (field studies, workshops, prototyping)
Formal analysis - Original Draft	Application of formal techniques (qualitative analysis using software tools) to analyze or synthesize study data

Oslo, 13.02.2023

Place, date

Thomas Kaland

Signature co-author

Alsos, O.A.:

Supervision	Oversight and leadership responsibility for the research activity planning and execution
Project administration	Management and coordination responsibility for the research activity planning and execution
Writing - Review & Editing, Original Draft, Minor Review	Critical review, commentary, and revision.
Visualization	Preparation and creation of visualization in the published work

Trondheim,

Place, date

Ole Andreas Alsos

Signature co-author

PhD Candidate's (main author's) contribution:

Veitch, E.:

Conceptualization - Original Draft, Minor Review	Ideas; formulation or evolution of overarching research goals and aims
Methodology - Original Draft, Minor Review	Development or design of methodology (Human-Centered Design methods)
Formal analysis - Original Draft, Minor Review	Application of formal techniques (qualitative analysis using software tools) to analyze or synthesize study data
Investigation- Original Draft, Minor Review	Conducting the research and investigation process (interviews, field studies, workshops, prototyping)
Writing - Original Draft	Preparation, creation, and presentation of the published work, specifically writing the initial draft
Writing - Review & Editing, Minor Review	Preparation, creation, and presentation of the published work
Visualization	Preparation, creation, and presentation of visualizations in the published work

Lisbon, 22.01.2023
....
Place, date

....
Signature co-author

NTNU

DECLARATION OF CO-AUTHORSHIP

Erik Aleksander Veitch is applying for the evaluation of the following thesis:

"Designing for Land-based Control of Autonomous Ships"

The declaration should describe the work process and division of labor, **specifically identifying the candidate's contribution**, as well as give consent to the article being included in the thesis.

Declaration of co-authorship on the following article: Veitch, E., Alsos, O.A., Cheng, T., Senderud, K., Utne, I.B. (2023). Human factor influences on supervisory control of remotely operated and autonomous vessels (In Review). https://dx.doi.org/10.2139/ssrn.4437731

Co-authors' contributions:

Alsos, O. A.:

Funding acquisition	Acquisition of the financial support for the project leading to this publication
Project administration	Responsibility for the management and coordination of research activity as Head of NTNU Shore Control Lab
Resources	Provision of experiment participants, especially through recruitment via communication channels
Supervision	Oversight and leadership for the research activity, including mentorship to the lead author
Writing (review & editing)	Critical review, commentary, and revision

Trondheim .... Ole Andreas Alsos ....

Place, date Signature co-author

Cheng, T:

Methodology	Development and design of methodology, specifically for "Scenario 2" (takeover / collision avoidance) in the simulation experiment and for interview questions pertaining to collision avoidance process

Wuhan, China, ... Tingting Cheng ....

Place, date Signature co-author

Senderud, K:

Investigation	Data collection for the "Gamer" group, including conducting exit interviews and administering questionnaires
Methodology	Development and design of methodology, specifically for the "two-part boredom questionnaire"

[redacted]
Place, date

Kristin Senderud
…………………………………
Signature co-author

Utne, I.B.:

Supervision	Oversight and leadership for the research activity, including mentorship to the lead author and third author
Writing (review & editing)	Critical review, commentary, and revision

6.6.2023
………………………...
Place, date

Ingrid B. Utne
…………………………………
Signature co-author

PhD Candidate's (main author's) contribution:

Veitch, E.:

Conceptualization	Ideas formulation and evolution of overarching research goals and aims
Methodology	Development and design of methodology (simulator experiment; Design of Experiments; "split-plot" design)
Formal analysis	Application of formal techniques to analyze and synthesize study data (model selection; REML ANOVA)
Investigation	Data collection for the "Gamer" and "Navigator" groups, including conducting exit interviews, administering questionnaires, and collecting all quantitative data (simulator data, video, audio, pupillometry, electrodermal activity)
Writing - Original Draft	Preparation, creation, and presentation of the published work, specifically writing the original draft
Writing - Review & Editing	Preparation, creation, and presentation of revisions to the work
Visualization	Preparation, creation, and presentation of visualizations

Trondheim, 31 May 2023
………………………...
Place, date

…………………………………
Signature co-author

ISBN 978-82-326-7262-2 (printed ver.)
ISBN 978-82-326-7261-5 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU
Norwegian University of
Science and Technology