2022

# Risk Assessment in the Design phase of Maritime Autonomous Ships – A Human-centred approach

DOCTORAL THESIS

ÅSA SNILSTVEIT HOEM

“Confidence is what you have until you understand the problem.”

- Woody Allen

“A lot of people in our industry haven’t had very diverce experiences.  So they don't have enough dots to connect, and they end up with very linear solutions without a broad perspective on the problem. The broader one's understanding of the human experience, the better design we will have.”

- Steve Jobs

# Preface

This thesis is submitted in partial fulfilment of the requirements for the degree of Philosophiae Doctor (PhD) at the Norwegian University of Science and Technology (NTNU) in Trondheim. The work was carried out as a part of the research project SAREPTA, founded by the Norwegian Research Council (grant number 267860) and led by Stig Ole Johnsen at SINTEF Digital.

The PhD work has been carried out at the Department of Design (ID) at NTNU. The main supervisor was Professor Thomas Porathe. During the spring of 2019, research was carried out with Professor Mary (Missy) Cummings at the Human and Autonomy Lab at Duke University.

This thesis targets readers across several fields, and the foremost are designers, risk analysts, human factor personnel and operators of Maritime Autonomous Surface Ships (MASSs) being remotely monitored and controlled. The principles and findings in this thesis may apply to other parts of autonomous transportation systems, highly automated systems and, or parts of these.

Åsa S. Hoem

Oslo, October 2022

# Summary

Maritime Autonomous Surface Ships (MASSs) are said to have a considerable impact on the shipping industry's sustainability, promising greener and safer solutions (e.g., Fan et al. (2020); Porathe et al. (2018)). Technological developments within software and hardware have led to a rapid increase in automation in many systems and applications. However, because it will change the way work is done, the chance is that it will introduce new risks. One of the biggest challenges of new technologies is the creation of new risk patterns and vulnerabilities. Technologies do not operate in a vacuum, and highly automated system operations will involve the human element as their action still represents the final and most important barrier against accident occurrence in sociotechnical systems. Hence, a considerable contribution to risks will lie in the interaction between the human element as an operator and the technological systems. In the foreseeable future, a human operator must in some way be "in the loop," supervising the operation and on stand-by to take over control from a land-based control interface referred to as a Shore Control Centre (SCC).

The overall objective of the thesis is to *provide necessary knowledge for the development of improved methods for risk assessments and mitigation in the design phase of MASS*. The objective is detailed in three sub-objectives addressed in five research articles. The articles have an interdisciplinary focus and utilise qualitative methods with research synthesis as the bearing methodology. The research areas addressed in the thesis are illustrated in Figure 1 below. The main contribution is an initial framework for a "Human-centred Risk Assessment in the design of MASS", placed at the intersection of the three research areas.

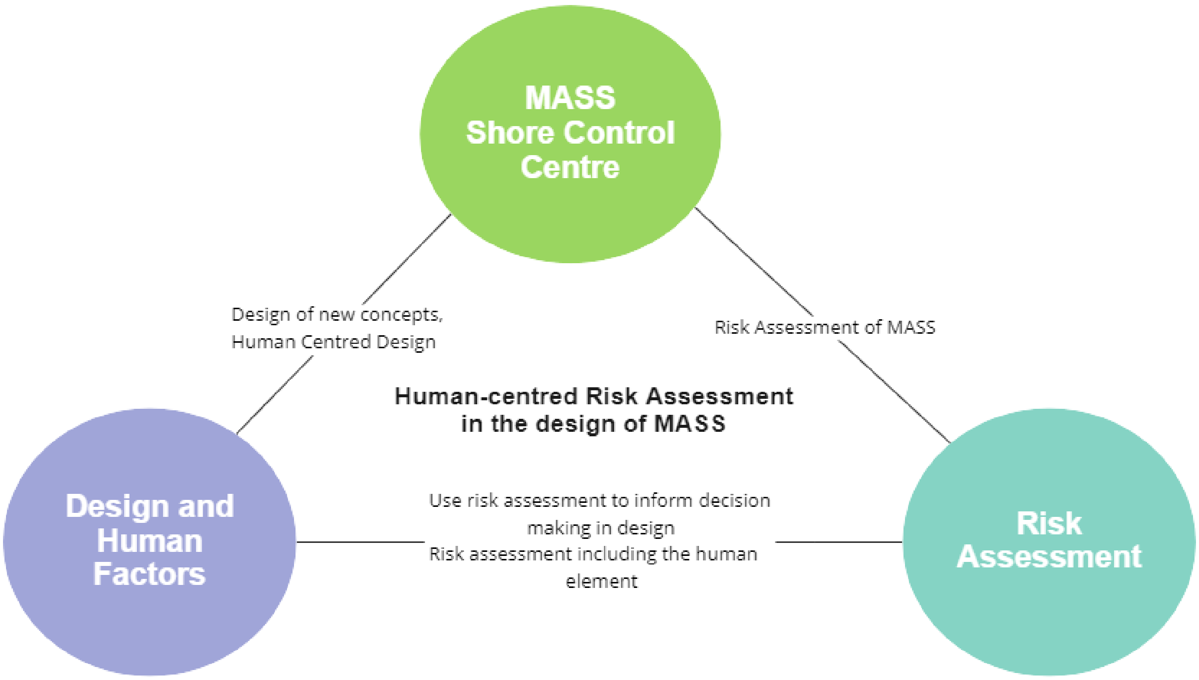

MASS Shore Control Centre
Design of new concepts, Human Centred Design
Risk Assessment of MASS
Human-centred Risk Assessment in the design of MASS
Design and Human Factors
Use risk assessment to inform decision making in design
Risk assessment including the human element
Risk Assessment

*Figure 1 Main research areas of the thesis.*

In this thesis, current risk assessment methods, tools and approaches have been reviewed to evaluate their applicability for MASS, particularly in the design of the Human Machine Interface (HMI) at the SCC. Starting up in 2017, few risk assessment approaches were published on the topic of MASS, and even fewer considered human-automation interaction associated hazards.

The thesis challenges the traditional risk concept, where risk is defined quantitatively as a product of probability and consequences. With the limited empirical data on MASS performance and the complex and software-intensive technology of MASS, accurate quantitative risk estimations are not feasible. Instead, a broader picture reflecting different views, assumptions, and ways of thinking, highlighting events, consequences, and uncertainties has been explored. This includes aspects related to Meaningful Human Control, Human-centred Design, and approaches within Safety I, II and III.

The thesis also problematises that, too often, risk assessments are isolated or separated from the design or systems engineering process. The most common consequence is that safety is treated as an after-the-fact assurance activity (Leveson & Thomas, 2018). This is typically a summative approach to risk assessments, where the focus is to evaluate if a predefined safety target (risk acceptance criteria) is met. Formative analysis, on the contrary, focuses on the process, i.e., improving the quality of the design. Risk assessments can improve the understanding of the system, safety controls and hazards of the activities under investigation. Different risk assessment methods should be applied for different purposes at different phases of the design process. By focusing on the goal of carrying out a risk assessment as a tool for designing for safety, and decision making in the design phase of MASS concepts, the main contribution of the thesis is an initial framework for an interdisciplinary risk assessment focusing on human aspects.

The Human-centred Risk Assessment in the design phase of MASS is inspired by the Scenario Analysis from the Crisis Intervention and Operability Study (CRIOP) framework. The assessment identifies safety issues by involving the end-user, i.e., including the operators' perspective, and can contribute to determining the MASS technical system's and operators' roles and responsibilities in executing different functions across various operations and situations. A stepwise approach is described in the thesis, and a qualitative case study of applying the method on an HMI prototype in a SCC for an autonomous urban passenger ferry is presented.

In conclusion, this thesis contributes to advancing theory and practice by promoting an initial framework for a formative risk assessment where the operator capabilities are considered together with the capabilities and dependencies of the MASS technical system. Further research is, however, necessary for testing and further developing the method.

# Acknowledgement

I am deeply grateful to my colleagues, friends, and family who have contributed to this PhD research and supported me throughout the exacting process; without you, this thesis would not have been completed.

First and foremost, my most profound appreciation is given to Professor Thomas Porathe. Thank you for being my main supervisor. I am grateful for your invaluable support and contributions, encouragement, positivity, and patience, and for giving me new perspectives on research and the freedom to shape my work. I would also like to thank Professor Margareta Lützhöft for being my co-supervisor, including me in your network and providing invaluable feedback on my work. Another vital person, Stig Ole Johnsen, for being my co-supervisor and mentor. Thank you for always encouraging and supporting me and trying to teach me time management and the importance of saying NO to not-so-relevant projects during my time as a PhD student and a part-time employee at SINTEF.

I would like to thank the Norwegian Research Council for providing the opportunity for this PhD through financing the SAREPTA Research project (grant number 267860). I have been blessed with support from the SAREPTA research group, which guided me through the research processes. I want to thank Jan Ørnulf Rødseth, Kay Fjørtoft, Gunnar Jenssen, Terje Moen, Tor Erik Evjemo and Trine Stene, whom I received, not only high-quality advice, detailed and timely feedback on my articles but also constant support and sincere advice to my present and future academic development. A special thanks to Prof. Missy Cummings at Duke University and the director of Duke's Humans and Autonomy Laboratory. Thank you for including me in the lab, curriculum, and student projects.

Completing this PhD thesis, a product of several years' work, has been a memorable learning experience. A great number of people at NTNU helped me with their advice, guidance, friendship, good company, and cheer. Amongst them, I would like to especially thank my co-PhD students; Brit Eli Danielsen, Erik Veitch, Dag Rutledal and Kjetil Vasstein. I am also profoundly thankful for my friends outside the university, especially Eli, Synnøve and Nora, for their warm thoughts and advice on handling academic struggles.

Finally, I want to thank my dear family, who has supported me and stayed positive. I could not have been more grateful to my mum, Sissel, for being my greatest supporter and comforter, and my grandparents and my dad for always taking pride in me. I especially want to thank my sisters, Frida and Liva, for lifting my spirits and sharing the ups and downs of scholarly exercises. During my time as a PhD candidate, I became a mom to Live and Sønnik, and I will look back at the time as a PhD candidate as the best years of my life, providing me with my greatest accomplishments. One name remains at the end, and that is Even. I am forever indebted to you for your encouragement, patience, and ability to make me laugh every day. Thank you for being you.

*Åsa*

# Contents

# Table of Figures

# Table of Tables

# List of abbreviations

*Table 1 List of abbreviations in the current PhD thesis.*

| Abbreviation | Definition |
|---|---|
| **AGCS** | Allianz Global Corporate & Specialty |
| **ALARP** | As Low As Reasonable Possible |
| **CONOPS** | Concept of Operation |
| **DHSA** | Defined Situations of Hazard and Accidents |
| **DNV** | Det Norske Veritas |
| **DOD** | US Department of Defense |
| **EMSA** | European Maritime Safety Agency |
| **EU** | European Union |
| **FAA** | Federal Aviation Administration |
| **FMEA** | Failure Modes and Effects Analysis |
| **FRAM** | Functional Resonance Accident Model |
| **FSA** | Formal Safety Assessment |
| **GUI** | Graphical User Interface |
| **HAI** | Human Automation Interaction |
| **HCRA** | Human-centred Risk Assessment |
| **HCD** | Human-Centred Design |
| **HITL** | Human-in-the-Loop |
| **HOTL** | Human-out-of-the-Loop |
| **HMI** | Human Machine Interface |
| **HRA** | Human Reliability Analysis |
| **IACG** | International Association of Classification Societies |
| **IMO** | International Maritime Organization |
| **ISO** | International Organization for Standardisation |
| **MASS** | Maritime Autonomous Surface Ships |
| **MHC** | Meaningful Human Control |
| **MSC** | Maritime Safety Committee |
| **MTO** | Man, Technology and Organisation |
| **NFAS** | Norwegian Forum for Autonomous Ships |
| **NMA** | Norwegian Maritime Authority |
| **PRA** | Probabilistic Risk Assessment |
| **RCC** | Remote Control Centre |
| **ROC** | Remote Operation Centre |
| **RQ** | Research question |
| **SCC** | Shore Control Centre |
| **SMoC** | Simple Model of Cognition |
| **STAMP** | System Theoretic Accident Model and Process |
| **STPA** | Systems-Theoretic Process Analysis |
| **WAD** | Work As Done |
| **WAI** | Work As Imagined |
| **QRA** | Quantitative Risk Assessment |

# Structure of the Thesis

The thesis is written in the form of a collection of articles. This thesis consists of two parts.

Part I – Main report: The first part introduces the theoretical background, research gaps and questions, and the theoretical background and research methodology utilised. The main results generated from each study are presented and analysed in light of the objectives, and a conclusion and possible areas for future research are indicated.

1. Introduction
2. Theoretical background
3. Research Method
4. Main Results and Discussion
5. Main Contribution
6. Conclusion and Further work
7. References

Part II – Articles: The second part is a collection of five articles that represent the main work and contribution of the PhD research. Article 1 is a background study of the thesis, while Articles 2 to 3 are empirical studies exploring recent accidents in the maritime domain and the experiences of automated technology across transportation domains. Articles 4 and 5 explore the use of an adapted method through a method and a case study. These articles are the main learning outcomes during my academic education in design.

# Publications

The following publications are included in Part II of this thesis. Articles 2 and 5 are journal articles, Article 3 is a book chapter, and the reminders (Articles 1 and 4) are conference articles.

**Article 1**

Hoem, Å. S. (2019). The present and future of risk assessment of MASS: A literature review. In Proceedings of *the 29th European Safety and Reliability Conference (ESREL), Hannover, Germany* (pp. 1666-1673). https://doi.org/10.3850/978-981-11-2724-3_0657-cd

The article is included in Part II, pp. 93-102

**Article 2**

Hoem, Å. S., Fjørtoft, K., & Rødseth, Ø. J. (2019). Addressing the accidental risks of maritime transportation: could autonomous shipping technology improve the statistics? In *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation, 13(3)* (pp.487-494). https://doi.org/10.12716/1001.13.03.01

The article is included in Part II, pp. 103-111

**Article 3**

Hoem, Å. S., Johnsen, S. O., Fjørtoft, K., Rødseth, Ø. J., Jenssen, G., & Moen, T. (2021). Improving Safety by Learning from Automation in Transport Systems with a Focus on Sensemaking and Meaningful Human Control. In *Sensemaking in Safety Critical and Complex Situations* (pp. 191-207). CRC Press. https://doi.org/10.1201/9781003003816-12

The chapter is included in Part II, pp. 113-130

**Article 4**

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 *Conference proceedings for the International Conference on Applied Human Factors and Ergonomics* (pp. 219-227). Springer. https://doi.org/10.1007/978-3-030-80288-2_26

The article is included in Part II, pp. 131-140

**Article 5**

Hoem, Å.S., Veitch, E., & Vasstein, K. (2022). Human-centred risk assessment for a land-based control interface for an autonomous vessel. *WMU Journal of Maritime Affairs* 21, (pp. 179–211). https://doi.org/10.1007/s13437-022-00278-y

The article is included in Part II, pp. 141-174

# Declaration of Authorship

Table 2 summarises each author's contribution to the publications included in this thesis. I have been the primary author in the listed articles, responsible for most of the written contributions and organizing the feedback and rewriting process. In the articles, the authors have contributed to the following tasks:

A. Initial research idea and concept
B. Data collection
C. Data analysis
D. Writing and design of the draft of the article
E. A critical review of the article
F. Organization and finalisation of the article

*Table 2 Contribution of each author to the publications enclosed in this thesis.*

| Author | Article | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Åsa Snilstveit Hoem | A-F | A-F | A-F | A-F | A-F |
| Ørnulf Rødseth | | B,C,D,E | B,C,E | B,C | |
| Kay Fjørtoft | | B,C,D,E | B,C,E | | |
| Stig Ole Johnsen | | | B,C,D,E | B,C,E | |
| Gunnar Jenssen | | | B,C,E | | |
| Terje Moen | | | B,C | | |
| Erik Veitch | | | | | B,C,E |
| Kjetil Vasstein | | | | | C,D,E |

# Publications not included in the thesis

During my PhD work, I contributed to the following four publications. These are not part of the thesis but provided data for my research and support the findings addressed.

**Article 6**
Rødseth, Ø. J., Nordahl, H., & Hoem, Å. (2018). Characterization of autonomy in merchant ships. In 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO) (pp. 1-7). IEEE.

**Article 7**
Porathe, T., Hoem, Å., Rødseth, Ø., Fjørtoft, K., & Johnsen, S. O. (2018). At least as safe as manned shipping? Autonomous shipping, safety and "human error". In Safety and Reliability–Safe Societies in a Changing World (pp. 417-425). CRC Press.

**Article 8**
Johnsen, S. O., Hoem, Å. S., Stålhane, T., Jenssen, G., & Moen, T. (2018). Risk based regulation and certification of autonomous transport systems. In Proceedings of the 28th International European Safety and Reliability Conference (ESREL 2018), Trondheim, Norway, 17–21 June 2018.

**Article 9**
Johnsen, S. O., Hoem, Å., Jenssen, G., & Moen, T. (2019, October). Experiences of main risks and mitigation in autonomous transport systems. In Journal of Physics: Conference Series (Vol. 1357, No. 1, p. 012012). IOP Publishing.

# 1 Introduction

## 1.1 Background

There is a significant interest in industry, authorities, and academia in the prospects of Maritime Autonomous Surface Ships (MASS). The International Maritime Organisation's (IMO) Maritime Safety Committee (MSC) suggested the term MASS at their 98th session in 2017. MASS is said to have a considerable impact on the shipping industry's sustainability, promising greener and safer solutions (e.g., Fan et al. (2020); Porathe et al. (2018)). However, because it will change the way work is done, the chance is that it will introduce new risks. IMO defines MASS as a ship which, to a varying degree, can operate independently of human interaction (IMO, 2019). For the regulatory scoping exercise on MASS, four degrees of autonomy have been established by IMO (2018a): 1) Crewed ship with automated processes and decision support; 2) Remotely controlled ship with seafarers on board; 3) Remotely controlled ship without seafarers on board; and 4) Fully autonomous ship. It is noted that the degree of autonomy is not necessarily intended to be linear or hierarchical; MASS can operate at one or more degrees of autonomy during a single voyage (Kim & Mallam, 2020). Human operators have different roles and interactions with ship systems and functions in each listed degree. The autonomous ship itself is only one component in a more extensive socio-technical system. When referring to MASS, the term includes the autonomous ship system with land and ship-based sensors and control systems, personnel in a Shore Control Centre (SCC) (also known as Remote Control Station/Centre or Remote Operation Centre), and other assets. Thus, it is explicit that when referring to MASS, it is not about a single autonomous ship but the autonomous ship system.

In the last decade, autonomy has been a hot topic with the development of autonomous systems, such as self-driving cars and self-controlled flying drones. The overall question for autonomy, convergent for all transport sectors, seems to be if and how technology can replace humans. Some claim that increased safety will be achieved by reducing the likelihood of "human error" when introducing more autonomy (Ramos, Thieme, et al., 2019). However, autonomy may create new types of accidents that before were averted by the humans present onboard the ship and in control. Back in 2017, at the beginning of my PhD research, we started seeing examples of fatal accidents involving cars driving in "auto-pilot" or "auto-steering" mode (e.g., NTSB (2018); Reuters (2016)). Here, the technical systems failed to detect certain objects, and the drivers or operators were not aware of or underestimated the limitation of the technical system. Hence their understanding of the situation and decision making capabilities were poor. This is a typical new accident scenario introduced by automation. Other fatal accidents involving a high level of automation are the two Boing 737 MAX crashes in 2018 and 2019, claiming nearly 350 lives (Herkert et al., 2020).

MASS introduces new technology and solutions that are said to make the maritime industry safer and more efficient (as outlined in Porathe et al. (2014)). For MASS to become a success, they must prove to be at least as safe and reliable as today's manned conventional ships. Essential questions are then: If we replace the human operator with automation, can we reduce the number of accidents? Is there a potential for new types of accidents to appear? And how can we, early in the design process of MASS, identify and mitigate possible risks in order to design and operate safe systems?

## 1.2 Motivation

The development toward autonomous operation seems to be mainly driven by a technological push. There has been less focus on risk assessment and modelling concerning the conceptualization, design, and safe operation of these systems (Parhizkar et al., 2022). The safe operation of automated and autonomous systems requires close coordination between human operators, organisations, and technical systems and components. Hence automation-related concerns regarding the "out-of-the-loop syndrome" (Endsley & Kiris, 1995) are crucial for the operational risks of MASS. As Strauch (2017) discusses, Bainbridge's (1983) main issues in "Ironies of Automation" is still valid. The key message here is the irony "that the more advanced a control system is, the more crucial may be the contribution of the human operator" (Bainbridge (1983), p. 775). For MASS, the level of autonomy will vary in a dynamic way from full human-operated control to full machine control. This dynamic autonomy brings an additional layer of complexity to the systems and operations, especially regarding the interactions and handover between human operators and autonomous technology. As for the foreseeable future (of MASS), a human operator must in some way be "in the loop," supervising the operation and on stand-by to take over control from a SCC. Hence, my PhD research hypothesises is that highly automated system operations will involve the human element, and a considerable contribution to risks will lie in the interaction between the operator and the systems. Still, as discussed by Wróbel et al. (2020), Veitch et al. (2020), and Lutzhoft et al. (2019), most of the research on the topic of MASS focuses on the high-end technical components of the system, running a risk of missing the critical human element in MASS operations.

In order to communicate the risk picture involving new technology and the human element, the following model in Figure 2 below was established in my research project and published in Porathe et al. (2018).

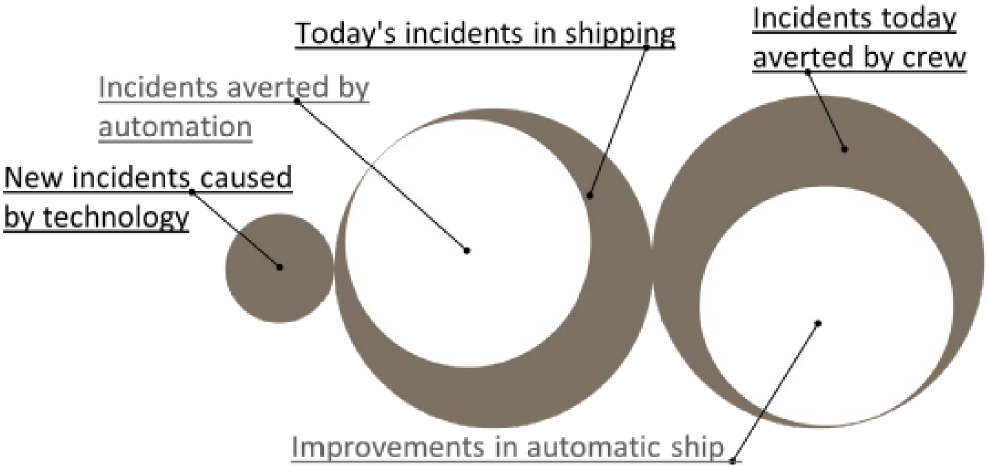

Today's incidents in shipping
Incidents today averted by crew
Incidents averted by automation
New incidents caused by technology
Improvements in automatic ship

*Figure 2 "The new risk picture", adapted from Porathe et al. (2018).*

On the one hand, we have known incidents due to "human error" that can be reduced by introducing automation (middle circle). On the other hand, we have potential incidents averted by humans today that might develop into accidents when no humans are present (right circle). Further, introducing new technology also opens up for the occurrence of new emerging incidents and accidents (far left). It is important to note that the size of the circles does not present an actual estimated risk, as this is unknown to us. Especially the right circle, where we have no actual data on incidents averted today, as it is mainly considered part of regular operation. The net result is the remaining grey areas, and the question is if this will be low enough for approval by the authorities and the general public acceptance of the new ship types. Thus, while the assumption is that the net result of automation will be fewer accidents and incidents, this remains to be shown. Trust and approval by authorities and the general public require safe design and operation of the autonomous systems. For this purpose, risk assessment considering human, software, and hardware interactions in automated and autonomous systems should be applied in the design phases of MASS.

The term design process or phases in this thesis corresponds to the design processes outlined in IMO's *Guidelines for Approval of Alternatives and Equivalents to Conventional Designs* (IMO, 2013) and is defined according to the Autonomous ship design standard provided by the AUTOSHIP project (2020). The design phase consists of a layered development process with the following phases: 1) early concept design, 2) high-level ship system design, 3) Detailed function allocation, and 4) Detailed system design. The starting point and input to the early concept design will be the general business proposal and corresponding system objectives (Rødseth et al., 2020). Compared to the traditional ship design concept (Tupper, 2013), the design of MASSs also includes the autonomous ship system where all (both physical, cyber and human) elements together ensure the sustainable operation of an autonomous ship in its intended operations or voyage (Maritime UK, 2019). Hence the design also includes the integration with a SCC.

## 1.3 Problem statement

Risk assessment uses different methods and tools to identify key contributors to risk and support the decision making regarding which safety measures to implement (Aven & Krohn, 2014). All risk assessments are of limited value if they are not used in a decision making context (Rausand, 2013). For MASS, risk assessment can be applied in the design phase (including the regulatory approval processes) and during operation. Risk assessments in the design phase are tools for decision making to assess the safety of a conceptual design. The use of risk assessment in the design process can roughly be divided into two types: formative analyses (focused on the process, e.g., to improve the quality of a design) and summative (focusing on the results of the assessment, e.g., to evaluate if a safety target is met, for validation and verification) (French et al., 2011; French & Niculae, 2005).

The maritime transportation domain is known to be conservative and heavily regulated by mandatory requirements that, in detail, prescribe equipment and procedures (Porathe, 2016). The primary safety focus for the design of conventional ships has traditionally been on prescriptive rules, which in detail define the required means for achieving a safety objective (by, e.g., redundancy requirements for technical systems and by requiring a minimum plate thickness in hulls). The prescriptive rules might act as design constraints hampering innovation and design optimization. The Risk-Based Ship Design (RBSD) framework was developed by Papanikolaou and Soares (2009) to allow for innovative ship concepts and technologies where the designers had the freedom to identify optimal and safer solutions. However, the focus in the RBSD framework is still on technical design and calculating quantitative risk estimations. The safety is quantified using a formalized quantitative risk analysis procedure and compared to predefined risk acceptance criteria (Papanikolaou & Soares, 2009), i.e., a summative approach to risk assessment.

Risk definition and perspectives in the maritime domain are strongly tied to probabilistic methods (Goerlandt & Montewka, 2015). Realist approaches dominate risk assessment applications, which consider risk as a physical attribute of a technology or system characterized by objective facts. The risk analysis heavily relies on data collected from the system or on engineering/statistical models and typically aims at accurately calculating quantitative risk measurements. The majority of risk assessment methods used today were developed some 50-70 years ago and applied to electromechanical systems with limited end-user considerations. Today, our system designs look very different and require a well-designed Human Machine Interface (HMI) for safe operation.

Most risk assessments applied in the maritime domain are technical and summative. Important questions are then if the present risk assessment methods in the maritime domain can address the emerging risks of MASS operation already at the design stage, do they include the "human in the loop" (i.e., the human element), and how can we improve the traditional methods? In the design of MASS, different risk assessment techniques and methods should be applied at different levels (concerning various aspects of the MASS operation).

Starting in September 2017, the main goal of my doctoral studies was to contribute to the field of risk analysis of highly automated ships. My PhD was part of a four-year research project named SAREPTA, an abbreviation for "Safety, autonomy, remote control and operations of industrial transport systems", funded by the Norwegian Research Council. The key objective of the project was to *provide necessary knowledge for the development of improved methods for risk assessments and mitigation in transport systems that are autonomous, remotely controlled and/or periodically unmanned.* SAREPTA is a comprehensive project covering all four modes of transportation, while my PhD project focuses on the maritime domain. The project started in the fall of 2017, and a brief literature search at that time identified few publications addressing risk analysis issues of autonomous, unmanned or remotely controlled vessels. Moreover, even fewer addressed human autonomy interactions. Hence systemizing knowledge of what can go wrong and providing a better understanding of the emerging risks involved in the operation of MASSs was necessary.

My thesis will address the challenge of the current risk assessment to improve the overall design and safety of MASS and argue for human-centred risk informed decision making in the design phase through a human-centred risk assessment. The overall objective of the thesis, the associated research questions and sub-objectives are outlined in the following section.

## 1.4 Research Questions and Objectives

The overall objective of the thesis is to:

Provide necessary knowledge for the development of improved methods for risk assessments and mitigation in the design phase of MASS.

This objective is decomposed into three research questions with corresponding sub-objectives referred to as research objectives.

**Research Question 1 (RQ1)**

1. What types of risk assessments are suggested for the design phase of autonomous, unmanned or remotely controlled ships today?
   a. What are the main issues and limitations of the risk assessment methods when identifying and addressing the accidental risks of MASS?
   b. How is the human element included in these risk assessment methods?

**Research objective 1:** Review the "state of the art" on risk analysis and assessment of MASS. Systemize and evaluate different methods according to their limitations and strengths, their applicability in the design phase, and whether they include the interaction between the autonomous system and the human operator.

The second research question addresses the “new risk picture” presented in Figure 2 by asking how MASS technology will affect today’s accidental risks in the maritime domain. In addressing the “new risk picture”, I saw the need to learn from other transportation domains to gain knowledge of experienced risks associated with introducing highly automated technologies. Another way of looking at the “new risk picture” is by identifying differentiating factors between MASS and conventional ships, and investigating how these factors will affect today’s accident statistics. The second research objective is hence divided into two parts:

**Research Question 2 (RQ2)**

2. In the design of MASS: What will the main accidental risks be, and how can they be mitigated?

a. What are the differentiating factors between MASS and conventional manned ships? How will the autonomous technology applied in MASS affect the known accidental risks in the maritime domain and what potential new risks will be introduced?
b. What are the experienced risks from operation of autonomous, unmanned or remotely controlled transportation systems today? Are these risks applicable to MASS?

**Research objective 2:** Investigate what we can learn from experiences with highly automated systems from other transportation domains and how will the autonomous technology affect the safety of MASS:

- 2a) identify the differentiating factors between MASS and conventional manned ships and investigate how the autonomous technology applied in MASS affect the “new risk picture”.
- 2b) investigate what we can learn from incidents and accidents where systems with a high degree of automation are involved; What are the related risks? What are applicable for MASS?

With the information obtained from RQ1 and RQ2, the third research question follows:

**Research Question 3 (RQ3)**

3. How can we integrate the human element in risk assessment in the design phase of MASS?

a. Are there any applicable methods for including the human element (operator) in the design of MASS?
b. What could be a good risk assessment method for identifying and assessing Human Automation Interaction-related risk in the design phase of MASS?

Consider different human factors methods used when designing control centres. How can we carry out risk assessments to describe risks and other factors that affect our ability to design and operate the socio-technical system constituting MASS? A third research objective was formulated for the last research question, using the collected information from the previous research.

**Research objective 3**: Propose a method for risk assessment in the design phase, including the human element. The technique should be human-centred, easy to use, and accessible to designers, engineers, and researchers.

The goal of the research was to develop a method that integrates expertise developed by control room-design communities, Human Factors and risk science communities, as well as prior work on risk assessment of MASS while remaining practical for industrial applications.

### 1.4.1 Limitations

My research must be viewed as one out of several possible research directions within risk assessment in the design of MASS. My research on risk assessment is based on how a design concept can and should be analysed in terms of operational risks (hazards and safety issues) concerning the human element (i.e., the operator) with the goal of reducing the operational/accidental risks by designing out these issues at an early stage. The study object of the thesis is the socio-technical system MASS. The PhD thesis discusses the development towards future maritime systems where different degrees of autonomy are realised. However, the development of autonomy is still uncertain and could take different directions (Relling, 2021).

The thesis address risk assessment in the design phase and does not cover topics like risk assessment carried out by the autonomous system during operation (e.g. dynamic risk assessment). Nevertheless, how these systems support the operator by providing real-time information is relevant in the design phase and partly addressed.

The research discusses operational/accidental risk mainly related to navigational tasks and functions and, to a limited degree, passenger handling (in Article 5). Other maritime functions, such as cargo handling, mooring, special operations etc., are not covered.

# 2 Theoretical Background

As a starting point for my PhD work, I needed to systemize and explain my understanding of the topic and form a basis for my research. This section summarizes the background of the thesis. It addresses Human Autonomy interactions in a control centre, risk assessment, and human-centred design and sets these in the context of MASS. The field of Risk assessment of MASS is an emerging topic in a growing research field. Related fields and examples of work of particular relevance are presented in this chapter. The chapter aims to cover the following three areas:

1. MASS and SCC
2. Human-centred design and HF
3. Risk assessment

## 2.1 Maritime Autonomous Surface Ship

According to *The Oxford English Dictionary*, autonomy is the right or condition of self-government (literally, “self-rule”) and the freedom from external control or influence. As Relling et al. (2018) discuss, the term is used differently in colloquial language than in the technical definition, and it is interpreted in different ways across industries. Autonomy and automation are often used interchangeably, and because machines are deterministic, algorithmic entities, the distinction between automation and autonomy lies in the eye of the human beholder (Lyons et al., 2021).

By automated (in this thesis), I mean a system that will do what it is programmed to do. Autonomy takes automation a step further and can be defined as a system’s or subsystem’s own ability of integrated sensing, perceiving, analysing, communicating, planning, decision making, and acting to achieve its goals as assigned by its human operator(s) through a designed HMI (as presented in Parhizkar et al. (2022)).

Within Autonomous Marine Systems (AMS), underwater vehicles, especially Unmanned Underwater Vehicles (UUVs), have existed for several decades and are characterised by their capability to survey the subsea environment on a larger scale than divers and submarines are able to (Yuh et al., 2011). A taxonomy for the different types of autonomous maritime vehicles is proposed by the Norwegian Forum for Autonomous Ships (NFAS), as shown in Figure 3 below (Rødseth & Nordahl, 2017).

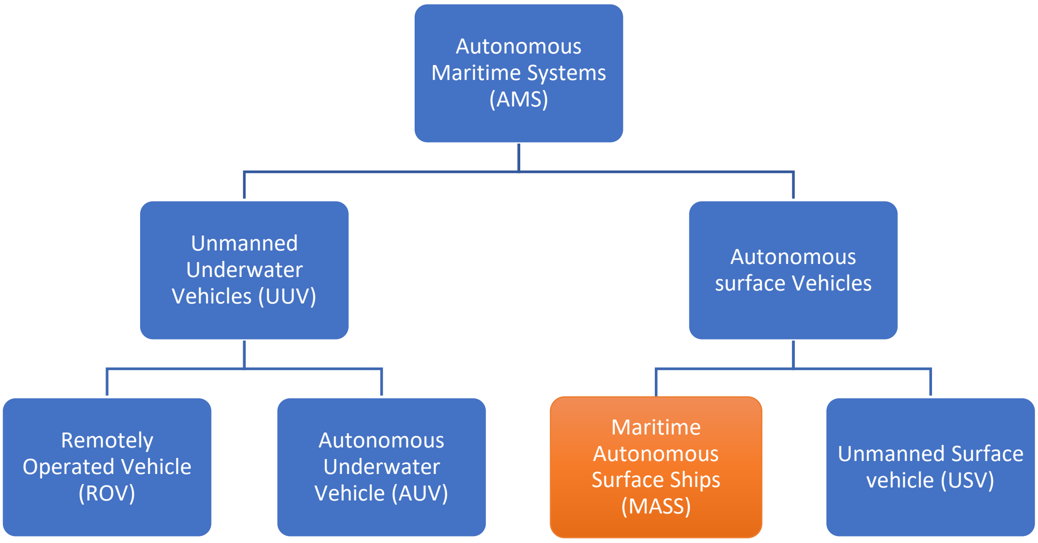

Autonomous Maritime Systems (AMS)
Unmanned Underwater Vehicles (UUV)
Autonomous surface Vehicles
Remotely Operated Vehicle (ROV)
Autonomous Underwater Vehicle (AUV)
Maritime Autonomous Surface Ships (MASS)
Unmanned Surface vehicle (USV)

*Figure 3 Types and classes of Autonomous Maritime Vehicles (AMS), derived from Rødseth & Nordahl (2017). The orange box marks the system that is investigated in this PhD project.*

IMO currently uses the term MASS for any vessel under IMO instruments' provisions, which exhibits a level of automation that is not recognized under existing instruments. In the outcome of the regulatory scoping exercise for the use of MASS (IMO, 2021), the term "MASS" is defined as a ship that can operate independently of human interaction to a varying degree. As mentioned in Chapter 1, IMO defines four degrees of autonomy, where the MASS can operate at one or more degrees of autonomy during a single voyage. The IMO framework is currently the most commonly referenced taxonomy in the literature, according to (Veitch & Alsos, 2022).

Degree One: Ship with automated processes and decision support. Seafarers are on board to operate and control shipboard systems and functions. Some operations may be automated and, at times, unsupervised but with seafarers on board ready to take control.

Degree Two: Remotely controlled ship with seafarers on board. Seafarers are available on board to take control and operate the shipboard systems and functions. The ship is controlled and operated from another location.

Degree Three: Remotely controlled ship without seafarers on board. The ship is controlled and operated from another location. There are no seafarers on board.

Degree Four: Fully autonomous ship. The operating system of the ship is able to make decisions and determine actions by itself.

In the last degree, the definition “being able to” does not necessarily mean a fully autonomous ship in the true meaning of the term autonomous, as the ship will only have the capability of making decisions and determining actions by itself but will rely on a human operator to take over control in case of an unexpected situation. This is by Rødseth et al. (2018), referred to as constrained autonomy. Here the ship has programmable limits or constraints to the actions it can take, such as a maximum deviation from planned speed or track before the crew or a remote operator must be alerted to intervene. Other authors have also understood an “autonomous ship” to be a “highly automated ship” involving some level of mixed Human Autonomy Interactions (ref. Ramos et al. (2020); Ramos, Utne, et al. (2019) and Huang et al. (2020)).

In this thesis, when referring to MASS, the focus is on the high degrees of autonomy where the ship is unmanned with the ability to be monitored and controlled from a remote control centre, referred to as a Shore Control Centre (SCC). In other words, MASS is a sociotechnical system compromising the autonomous ship, its functions, and the SCC. As argued by Hoem et al. (2021), MASS could better be an abbreviation for Maritime Autonomous Ship *System*, as they are complex sociotechnical systems consisting of equipment, machines, tools, technology, and a work organization, as shown in Figure 4 below.

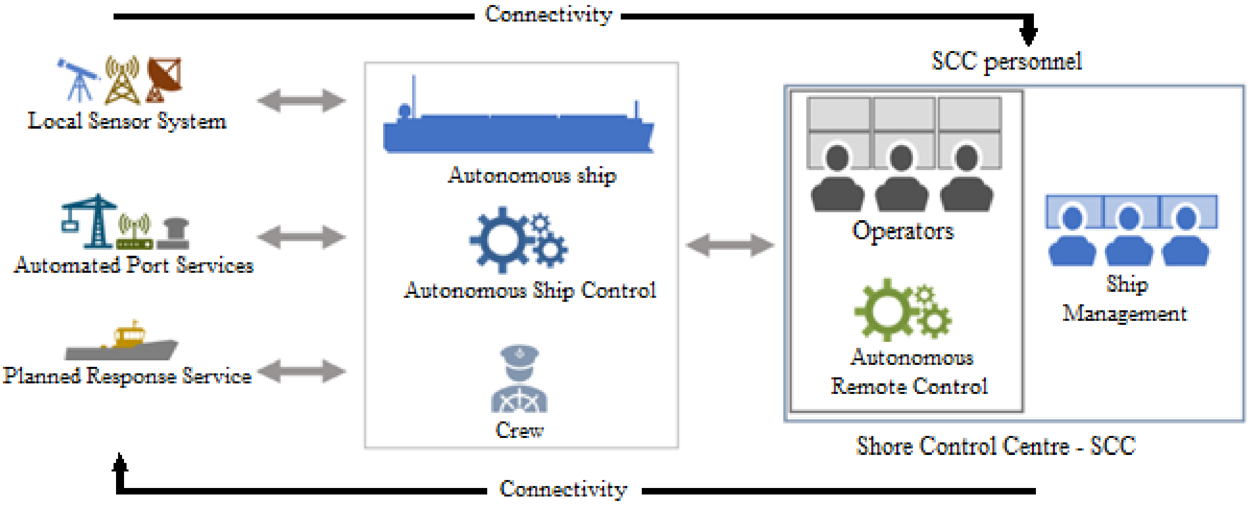


*Figure 4 Examples of components and roles in an autonomous ship system, adapted and adjusted to the content of this thesis, from Wennersberg et al. (2020).*

## 2.2 The Shore Control Centre

In the foreseeable future, it is doubtful that MASS can operate without human supervision and intervention (Porathe & Rødseth, 2019). This thesis uses the term SCC as it is the most commonly used term in the literature (Veitch & Alsos, 2022). In the literature, Shore Control Centre (SCC), Shore Operations Centre (SOC), Remote Control Centre (RCC) and Remote Operation Centre (ROC) are terms describing similar concepts.

MUNIN was the first project to develop a technical concept of a MASS back in 2015. Since then, several companies have established plans to develop MASS concepts (Infinity, 2020; Kongsberg, 2017; zeabuz, 2021). The MUNIN project made a concept study of an unmanned bulk carrier, but under the control of a shore-based remote control centre (a SCC) for the deep-sea passage between Europe and South America (MUNIN, 2016). The bulk carrier was manned during port approach and departure. The SCC facilities designed here were accomplished by using a close facsimile of a modern shipboard command structure modified to take the best advantage of the ship system automation. An operator was able to monitor and control a vessel at this workstation through HMI displays that monitored critical elements of the system. The displays were information clusters of a customized, real-time, vessel-specific dashboard, electronic sea chart, conning display, radar screen and weather chart.

Depending on the concept, one operator could monitor several vessels via a monitoring and controlling workstation. The operator could request help from a SCC Supervisor or other SCC actors like a SCC Captain (who was legally responsible for the activities of each vessel under the SCC's command) and licensed ship engineers (with expertise in the technical systems), as outlined in MacKinnon et al. (2015).

Findings in the MUNIN project related to the HMI suggest that the SCC prototype is a typical application within an exceedingly complex distributed automated system. The technology should not only be organised around the human's needs when they are not "situated" but also focus on how different parts of the system could work as a whole in the context from a genuine system perspective.

## 2.3 Design of MASS

Ship design is a complex and multifaceted process, influenced by conventions (regulations), requirements and several actors (Rumawas, 2021). Many ship design processes exist, but the process is often represented by a spiral diagram, which has been the standard way of working for decades (initially developed by Evans (1959)). However, over time ship design has become more complex with the introduction of more automated technology, more requirements to meet, more systems to optimize, and hence more analyses to be performed.

The requirements in the design spiral relate to technical aspects of the ship, such as proportions and powering, arrangements, capacities, stability and ultimately, cost estimate. The critical user interface to the human operator, the HMI, and operational structures are left out. Hence, user involvement is somewhat limited in the spiral process, resulting in situations where well-experienced seafarers must adapt to inherited poorly designed solutions (as explained in Lützhöft (2004)). Well-designed HMIs are key in reducing risk in MASS operation and in unlocking benefits from MASS.

In recent decades, Systems Theory has become a widely adopted theoretical foundation to deal with the increased complexity of engineered or designed systems. Systems Thinking is the term often used to describe what people are doing when they apply Systems Theory principles (Leveson & Thomas, 2018). In short, the main aspects of Systems Theory are, according to Leveson and Thomas (2018) (p.10):

- The system is treated as a whole, not as the sum of its parts.
- A primary concern is emergent properties, which are properties that are not in the summation of the individual components but “emerge” when the components interact. Emergent properties can only be treated adequately by taking into account all their technical and social aspects.
- Emergent properties arise from relationships among the parts of the system, that is, by how they interact and fit together.

In other words, Systems Thinking takes a sociotechnical system approach where socio-political and technological elements interact and should be oriented towards a common goal. Systems Thinking is promoted as the path to address the division between humanistic and mechanistic sciences and the subsequent technology-driven design trend that fails to answer the needs of the people who are meant to use it (Vicente, 2013). The application of Systems Thinking to create systems is called Systems Engineering. Many different Systems Engineering methods exist, but central to the methods are, according to Blanchard and Fabrycky (2013), the following common threads:

- A top-down approach to seeing the system as a whole
- A life-cycle orientation from system design and development to phase-out and disposal
- Defining system requirements and design criteria
- An interdisciplinary approach to address all design objectives

### 2.3.1 Regulations

Within the maritime domain, the general ship design rules can be divided into two main categories: (1) prescriptive rules prescribing specific design solutions and (2) goal-based rules prescribing design goals and functional requirements to meet the goals. As explained by Rødseth (2021), MASS concepts will, until new international rules are ready, need to be approved according to principals from the IMO Circular MSC.1/Circ.1455 “Approval of Alternatives and Equivalents” (IMO, 2013). This is fundamentally a risk based approach relating to goal-based rules rather than a *prescriptive rule-based* approach where operational or functional requirements must comply with the statutory rules and regulations.

The interim guidelines for MASS trials, approved by IMOs Maritime Safety Committee (MSC) in 2019, prescribe a broad range of objectives, such as risk management and compliance with mandatory instruments, amongst others. In the Interim Guidelines Subparagraph 2.2.1, it is clear that the parties to MASS trials should ensure "compliance with the intent of mandatory instruments" (IMO (2019), p. 2). In Subparagraph 2.2.2, it is left up to the flag State Administration to determine "the scope of application of mandatory instruments, […] in accordance with those instruments" for ships involved in MASS trials. Therefore, the national Administration is given the right to determine an alternative way of how this can be done. However, following the same provision, the national Administration is asked to take into account "the objectives of the trial, the anticipated capabilities and limitations of the ship and related systems and infrastructure during the trial, and the risk control measures adopted for the trial."

### 2.3.2 Norwegian Maritime Authority

In the *Guidance in connection with the construction or installation of automated functionality aimed at performing unmanned or partially unmanned operations* (NMA, 2020), the Norwegian Maritime Authority (NMA) outlines a list of design and documentation requirements to be followed based on the process described in MSC.1/Circ.1455 (IMO, 2013). The guideline further states that in the preliminary design phase, a detailed description of the entire operation of the ship should be documented in a CONOPS (Concept of Operation). Based on the CONOPS, it is required to carry out a pre-HAZID, where the entire operation is reviewed and where the focus is on the hazards that exist in the various parts of the operation (see NMA (2020), p.6). Based on the hazards identified in the HAZID, risk analyses/assessments must be carried out.

### 2.3.3 Class societies

The International Association of Classification Societies (IACS) has acknowledged that autonomy creates a need to develop new technical requirements (IACS, 2019), and several class societies have published guidelines or codes for MASS. DNV has *Guidelines for Autonomous and Remotely operated ships* (DNV, 2018), Lloyds Register has an *Unmanned Marine Systems Code* (LR, 2017), Bureau Veritas has *Guidelines for Autonomous Shipping* (Bureau Veritas, 2019), and ClassNK has *Guidelines for Concept Design of Automated Operation/Autonomous Operation of ships* (ClassNK, 2020). They all recommend applying a risk based approach to the design, verification, and validation process of MASS. Without explicitly mentioning Systems Theory, Systems Engineering or Systems Thinking, the guidelines recommend a systems engineering approach by establishing a hierarchical structure linking the technology expectations (goals) to functions and sub-functions.

## 2.4 Design of SCC for MASS operation

IMO initiated a regulatory scoping exercise for the use of MASS in 2017. In the outcome of the scoping exercise, the "Remote control station/centre" is mentioned as one of IMO's "high-priority issues" (IMO (2021), p. 8). Further stating that this is "a new concept to be implemented… and a common theme identified in several of [IMO regulatory] instruments as a potential gap" (IMO (2021), p. 8.).

Designing an SCC may have similarities with the design process of the bridge on a ship, and some industry projects have approached the design challenge by replicating the bridge onshore (Dybvik et al., 2020). However, as Dybvik et al. (2020) explain, the old seafaring model will be replaced by a completely new organisation, the SCC solution. The study further identified that designing the HMI will be the most challenging part of an SCC design. In particular, the handover from automation to human control. Knowing how to resolve this situation is a design issue and key when designing an interface. As mentioned in the previous section, the design spiral model does not address a remote control structure or the human-automation interaction. This traditional reductionist approach to ship design, where the engineering and humanistic sciences are separated, is outdated and will likely be less useful for preventing system errors (Grech et al., 2008). Reductionism has been a common heuristic in the way humans problematize things, but not always considered the best approach if we wish to design technology fit for people, especially in complex socio-technical systems like the maritime industry (Lützhöft et al., 2011).

In the literature on human-automation interaction (HAI) in MASS systems, the terms Human-Autonomy Interaction, Human-AI Interaction, Human-Computer Interaction, Human-Machine Interaction and Human-Robot Interaction are used interchangeably, all exploring the relationships between the human operator and robotic, intelligent, autonomous technology. In this thesis, the term HAI is chosen to cover the circumstances in which people interact with MASSs (including its "autonomous" capabilities) through an interface to receive information and control the task execution. For more information on the meaning of HAI and its history and future, see Sheridan and Parasuraman (2005) and Janssen et al. (2019).

## 2.5 Human-Centred Design

Human-Centred Design (HCD)[1] is a design practice (and design philosophy) where designers focus on ensuring that the design matches the needs and capabilities of the people for whom they are intended (Norman, 2013). More specifically, it can be illustrated as "an emancipatory tradition which places human needs, purpose, skill, creativity, and the human potential at the centre of activities of human organisations and the design of technological systems" (Gill, 1996).

HCD originates at the intersection of numerous fields, including engineering, psychology, anthropology, and the arts. Its origins are often traced to the founding of the Stanford University design program in 1958 by Professor John E. Arnold. He first proposed the idea that engineering design should be human-centred (Wikipedia, 2021). Since then, HCD has become one of the main design movements that govern the world of design and has been designated by the International Organization for Standardization (ISO) and the International Energy Agency (IEA) as the official approach for integrating Human Factors and usability principles, knowledge, and techniques in design practice (Giacomin, 2014). Giacomin (2014) describes HCD as "what began as the psychological study of human beings on a scientific basis for purposes of machine design" to what became "the measurement and modelling of how people interact with the world, what they perceive and experience, and what meanings they create" (p. 612).

The ISO defines HCD as an "approach to systems design and development that aims to make interactive systems more usable by focusing on the use of the system and applying human factors/ergonomics and usability knowledge and techniques" (ISO9241-210, 2019). The ISO standard

[1] Human Centred Design (HCD) and User Centred Design (UCD) are terms used interchangeably. In this thesis, the adopted term is HCD to regard for users as well as for other stakeholders affected by design practice.

defines five main activities and six key principles (including iterating the design if needed), as shown in Figure 5 below.

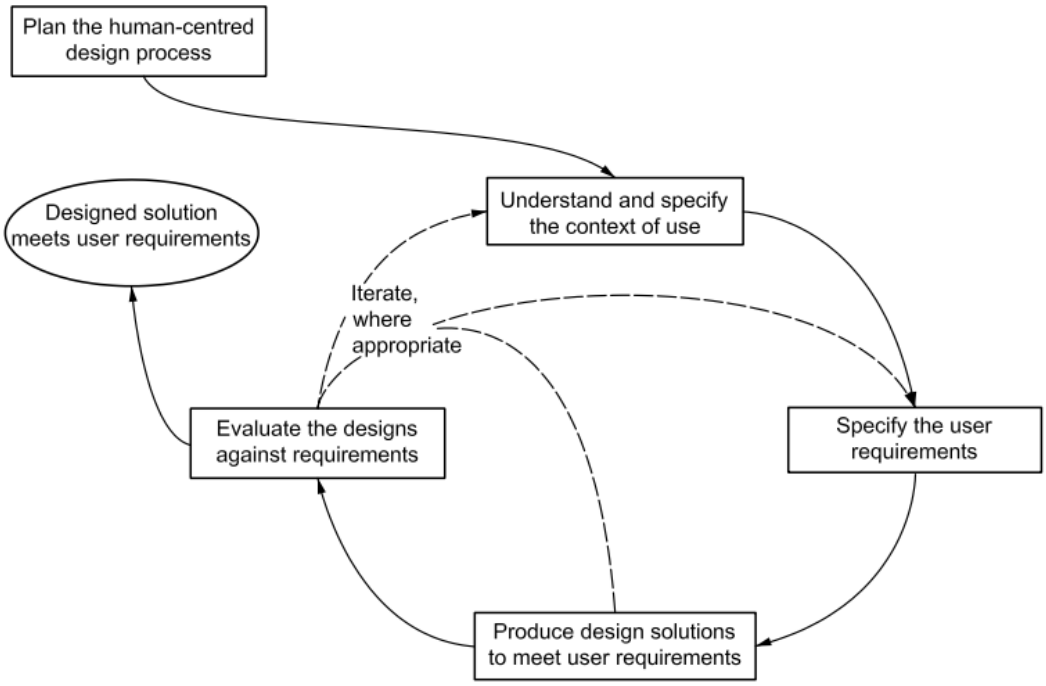

Plan the human-centred design process
Understand and specify the context of use
Designed solution meets user requirements
Iterate, where appropriate
Evaluate the designs against requirements
Specify the user requirements
Produce design solutions to meet user requirements

*Figure 5 The ISO9241-210 (2019) HCD process.*

This HCD cycle complements other design approaches employed by the designer or engineer (Costa, 2016). It is a participatory approach to design, which calls for active user involvement throughout the approach, from context and requirements to testing and evaluation.

## 2.6 Human Factors in MASS

The scientific field of Human Factors and Ergonomics (HF/E) was born after World War II, and it was mainly psychology and mostly about providing corrective ergonomics to engineered products and solutions (Helander, 2005). In the Nordic countries, the discipline is heavily influenced by sociology, while in the UK by engineering and in the US, by psychology (Relling, 2020). “Human factors” is a relatively novel concept in naval architecture and marine engineering. In the textbook for naval architects and marine engineers on Ship Design and Construction by Calhoun and Stevens (2003), “Human Factors” is defined as a broad term involving all biomedical and psychosocial considerations applying to a human in the system. The core of human factors in design is to consider humans when designing and to understand the human strengths, weaknesses, and performance variability by considering physical, cognitive, and motivation factors. Human factors challenges emerge in the boundary between humans and systems, and human factors in design is an iteration between designing and testing in a systems approach (Stanton et al., 2017). Human Factors Engineering (HFE) is one of many design aspects addressed within Human Factors. HFE involves issues of layout, equipment design, and workplace environment (Rumawas, 2016). It also addresses the HMI, including displays and controls. This thesis uses Human Factors as the term for all HF, HF/E and HFE, and adopt the general definition of “human factors” provided by the International Ergonomics (Association, 2020) and adjusts the term to the context of MASS. Thus, human factors in SCC design is hence defined as a scientific discipline concerned with understanding interactions among humans and other elements in a SCC, and the work that applies theory, principles, data, and other methods to design the HMI to optimise safety and performance, as well as the comfort, of SCC personnel.

One could think that Human Factors will be less critical for the design and development of MASS, as more of the traditional navigational tasks at the bridge is automated, and the ship might be unmanned. However, the human element at the bridge will not disappear but shift from ship to shore, where the human operator will be responsible for remote operation and supervision. Removing dependence on an operator by installing an automatic device to take over the operator's function only shifts that dependence onto the humans who design, install, test, and maintain the automated technology – who also make mistakes (Leveson, 1995). Hence, human factor considerations will be crucial in designing MASS systems with high requirements for robust and resilient hardware and software systems and the constant need for updates. Known issues at the bridge today are related to so-called "Human out of the loop" issues (Grech & Lutzhoft, 2016).

### 2.6.1 "Human in the Loop" vs "Human out of the loop"

The term "out-of-the-loop" (OOTL) performance can be linked to major issues associated with the implementation of automation (Endsley & Kiris, 1995). OOTL performance is a critical issue as it is associated with numerous negative consequences when operators are not able to identify the necessary corrective actions, respond too late, or when they have forgotten manual skills for error recovery (Kaber & Endsley, 1997). The *irony of automation* describes the challenge of keeping humans out of the loop since technology is superior to humans while asking humans to take over when technology fails (Bainbridge, 1983). Hence, many authors emphasise the importance of designing MASS with the human (operator) "in the loop" (Johnsen & Porathe, 2021; Lutzhoft et al., 2019; Relling et al., 2018; Veitch et al., 2020).

It should be noted that there is a difference between the OOTL performance issues, as referred to by Endsley, the need for "human in the loop"-designs expressed above, and the software developers' view of Human-in-the-Loop (HITL) vs Human-out-of-the-Loop (HOOTL)[2]. For software developers, the question of HITL or HOOTL in the design of AI-systems is whether or not to refer to a human at intersections or crossroads before initiating any action. In AI-Systems, a HOOTL system is a system that has sets of criteria to follow and take specific actions without deferring to a human expert (i.e., no human oversight) (Smith, 2003). While the opposite, a HITL, refers to the capability for human intervention in every decision cycle of the system. A HITL AI system will, in many cases, neither be possible nor desirable. For MASS, the importance of designing with the "human in the loop" does not mean that the human should have the capability to intervention in every decision cycle of the system. In the thesis, the terms "human in control" or "human in the loop" refers to the capability for human intervention during the design cycle of the system and monitoring of the system's operation and overseeing the overall activity of the AI system (i.e., the human receives information and can influence parts of the chain of events).

### 2.6.2 Meaningful human control

The term Meaningful Human Control (MHC) addresses the concerns of a "responsibility gap" for harm caused by these systems, i.e., that humans, not computers and their algorithms should ultimately remain in control of, and thus morally responsible for, relevant decisions about military operations. MASS, like other AI-systems, should improve individual and collective well-being. The European

[2] HOOTL should not be confused with the term Human-on-the-loop (HOTL) referred to by the EUs Ethics Guidelines for Trustworthy AI (European Commission, 2019) as the capability for human intervention during the design cycle of the system and monitoring the system's operation.

Commission has issued Ethics Guidelines for Trustworthy AI (2019) in which the following four ethical principles rooted in fundamental rights are listed: (i) Respect for human autonomy, (ii) Prevention of harm, (iii) Fairness, and (iv) Explicability. These must be respected to ensure that AI systems are developed, deployed, and used in a trustworthy manner. In order to have MHC, the allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunities for human choice (European Commission (2019), p.12). This means securing human oversight over work processes in AI systems, support humans in the working environment, and aiming for the creation of meaningful work. In other words, MHC in a SCC does not mean "direct control" but designing an interface so that the operator can decide whether their involvement in the primary task is required (van den Broek et al., 2020). Design principles such as supporting the operator's situation awareness, balancing the workload and enhancing the operator's competence and skills are essential aspects to consider when designing for the safe and resilient operation of MASS.

## 2.7 Safety of MASS

It is challenging to discuss safety and risk because of the confusing terminology, the multidisciplinary character of the topics, and the overwhelming number of books, reports, standards, and guidelines on the topic. For example, according to Blom (2016), at least 800 safety analysis tools and techniques are available across transportation domains and industries. Safety is a disciplinary term (Selvik & Signoret, 2017) that can be seen as an attribute of risk. It refers to the absence of unwanted outcomes such as incidents or accidents; hence, a reference to a condition of being safe (Hollnagel, 2014; Hollnagel et al., 2015). International organisations define "safety" as the freedom from risk which is not tolerable (ISO/IEC, 2014). Safety is commonly defined as one of the following:

1. absence of accidents and incidents (Aven, 2014)
2. freedom from unacceptable risk (Hollnagel, 2018)
3. freedom from unacceptable losses (Leveson, 2016)

These three ideas do not express the same idea. A ship navigating in the fog in an area of restricted navigation, like a narrow channel, is, per the first and last definitions, currently safe as there is no accident or incident nor any loss. But, the next moment, the ship might run aground or collide with another vessel. This aspect is covered in the risk concept – where uncertainty or probability is introduced. Safety is founded on a judgement of the magnitude and importance of the risk related to grounding or collision and its effects on the ship and the environment (Aven, 2022).

Safety research as a systematic, scientific subject is fairly young, with the pioneer works of the social science and organisational approaches to safety dating back to the seventies (e.g. Berry Turner - social science) (Haavik, 2021). Safety research often refers to three ages of Safety, each characterized by different focuses. Each perspective can be linked to the three abovementioned definitions by Aven, Hollnagel and Leveson.

#### *2.7.1.1 Safety I*

The Safety I perspective presumes that things go wrong because of identifiable malfunctions or failures of specific components of the system, such as technology, procedures, human workers and the organization in which they are embedded (Hollnagel et al., 2015). Traditional technical risk analysis methods like Failure Tree Analysis, Event Tree Analysis, and Probabilistic/Quantitative Risk Analyses (QRAs) belong to this perspective. The manifestation of Safety I is made by Hollnagel, and it is important to note that there are few or no references to Safety I among those who are given that label, as commented by Haavik (2021).

#### 2.7.1.2 Safety II

In Safety II, safety is seen as the ability to succeed under varying conditions (Hollnagel, 2018). The perspective is at odds with the Safety I perspective, where safety is defined as the absence of undesirable events and accidents and freedom of unacceptable risk. Hollnagel et al. (2015) argue to move the focus from what goes wrong to what works well. Variability is a key concept in Safety II. An activity is safe depending on the system's ability to succeed under varying conditions. Risk assessments are not highlighted in Safety II, and the traditional methods are avoided as intractable systems cannot be accurately modelled.

#### 2.7.1.3 Safety III

The split between Safety I and Safety II has been criticized by Leveson (2020a). In her work, Leveson provides a strong critique of Hollnagel's reasoning. Following her argumentation, the Safety-II approach was rejected in sophisticated engineering projects because it is not effective (Leveson (2020a), p.3). Leveson presents a third perspective based on a systems theory. However, the Safety III perspective is not "new", as it is based on System Safety that has been applied over the past 70 years in aerospace and defence to cope with increasing complexity, extensive and growing use of computers and new technology, and a changing role of humans in complex systems (Leveson, 2020a). In Safety III, safety is defined as freedom from unacceptable losses. What is considered unacceptable losses is determined by the system stakeholders. Safety III highlights the system's design process where the goal is to eliminate, mitigate or control hazards, which are the states that can lead to these losses.

This difference in view is noteworthy, but as questioned by Aven (2022), it may not be critical for the understanding, assessment, communication and management of safety but more about academic quirks of little relevance for practical safety management. Hence, the thesis will not elaborate further on the different views.

### 2.7.2 Safety Management

The disciplines of safety management and risk management are often thought to be independent when they are essentially the same discipline working towards similar goals of loss prevention or mitigation (Sloan, 2007). Banda et al. (2019) list a wide range of studies that have identified safety management gaps, challenges, and potential demands for the design of MASS. They conclude that there is a need for safety management of an autonomous ship from different angles of the entire autonomous system, including from the perspective of the human operator in the SCC.

## 2.8 Risk science

We frequently use the word *risk* in everyday life. As such, *risk* at first appears to be a relatively intuitive concept. However, when making *risk* a subject of scholarly investigation, we quickly realize that we lack a shared, let alone precise, understanding of the meaning of risk (Franzoni & Stephan, 2021). The prominent risk researcher Stan Kaplan stated the following after receiving the Distinguished Achievement Award from the Society of Risk Analysis in 1996:

> "The words of risk analysis have been and continue to be a problem. Many of you remember when our Society of Risk Analysis was brand new, one of the first things it did was to establish a committee to define the word "risk". This committee laboured for four years and then gave up, saying in its final report, that maybe it is better to not define risk. Let each author define it in their own way, only please each should explain clearly what way that is! (Kaplan, 1997), p.407)."

Aven (2016) has reviewed recent advances (the past 10-15 years) made in the risk field, focusing on fundamental ideas and thinking on which the risks fields are based upon. He concluded that many perspectives on risks exist, the scientific foundation of risk assessment is still somewhat shaky on some issues, and there are still opposing views. According to Goerlandt and Montewka (2015), these opposing views seem less known outside the theoretically oriented risk research community. Within the maritime application area, no references have been made to this. However, it is essential to present my view of the risk concept. The way we understand and describe risk strongly influences the way risk is analysed. Hence, it may have profound implications for risk management and decision making (as explained by Aven (2016)).

In the simplest, qualitative way, the risk concept is about understanding the world (in relation to risk) and how we can and should understand, assess, and manage this world. A common operational definition is given by Kaplan and Garrick (1981), who defines risk as the answer to three questions (items):

1. What can happen? (e.g. the scenarios)
2. How likely is this to happen? (e.g., the likeliness or probability associated with each scenario)
3. If it does happen, what are the consequences? (e.g. the consequences associated with each scenario)

A fourth question is, what are the uncertainties? Adding the uncertainty dimension to events and consequences has been intensively debated since the early stages of risk assessment back in the 1970s (see Aven (2012); Aven and Zio (2011); Johansen and Rausand (2015)). When addressing risk concepts of MASS, several sources of uncertainty arise, e.g., the system complexity and the lack of knowledge of and experience with MASS. Efforts are made to develop frameworks such as a semi-quantitative scale for assessing the strength of knowledge and other uncertainty dimensions (see (Bjerga & Aven, 2015; Johansen & Rausand, 2014)). However, the applicability of such frameworks is very much dependent on the available data quality, application area, level of autonomy, system complexity, and whether the technology is novel or widely used and proven (Utne et al., 2017). Hence, this thesis will not try to address any quantitative frameworks for risk analysis other than stating the limitations of QRAs that are owing to the mismatch between the nature of sociotechnical systems and how we approach them in search of prediction and control.

A distinction is made between the science of risk analysis (concerning concepts, principles, methods, and models for analysing risk) and the practice of risk analysis (concerning specific applications). There is no clear line between the two. The risk analysis and assessments addressed in the thesis are, to a varying degree, generic for the risk field but, in my case, focused on one area of application: MASS. My research looks at qualitative risk analysis, where risk analysis is carried out to make risk informed decisions during a design process, also known as risk based design or design for safety.

### 2.8.1 Risk analysis and assessment

Risk assessment is the process of finding answers to the three questions above. It entails risk identification, risk analysis and risk evaluation (ISO/IEC31010, 2019). Risk analysis is the process of comprehending the nature of risk and to determine the level of risk (ISO/IEC31010 (2019), p.11). In other words, risk assessment refers to a broader process in which, in addition to risk analysis, we evaluate risk mitigating measures.

Risk informed decision making denotes the process in which insights from risk assessments are considered together with other sources of information to make a decision that involves risk to human, environmental or material assets (Reason, 1997). Here, the risk analysis approaches and methods are typically combined with knowledge from statistics, psychology, social sciences, engineering, medicine, and many other disciplines and fields. Often, the subject of the risk assessment requires multidisciplinary and interdisciplinary activities.

## 2.9 Risk Analysis in the maritime (transportation) domain

Risk definition and perspectives in the maritime domain are strongly tied to probability. Alternative views do co-exist, but the realist approaches, rather than constructivist approaches, dominate the application area. Typically, risk assessments are well established in situations with considerable data and clearly defined boundaries for their use.  However, this is not the case for MASS, where we do not have sufficient data.

### 2.9.1 Risk-Based Ship Design

The framework for Risk-Based Ship Design (RBSD) was introduced in the SAFEDOR project[3] with the primary objective of providing evidence on the safety level of a specific design of ships (Papanikolaou & Soares, 2009). In RBSD, the risk level assessment is carried out with respect to predefined major accident categories, as shown in Figure 6 below. Another objective of the SAFEDOR project was to develop design methods and tools to assess operational extreme, accidental, and catastrophic scenarios, accounting for the human element, and integrate these into a design environment (formative approach).

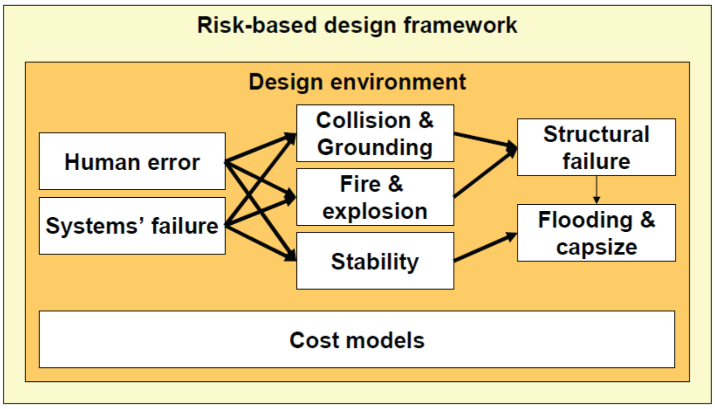


*Figure 6 Elements of the RBSD framework of SAFEDOR, adapted from Breinholt et al. (2012).*

Bayesian networks were established to evaluate probabilities and consequences for accident scenarios, and techniques like Failure Modes and Effect Analysis (FMEA) and fault trees are presented in the project. A complete list of developments can be found at Breinholt et al. (2012). One of them was an innovative bridge layout design, where interactions of the crew with the advanced equipment were identified as the focal point. However, we do not know what tools and techniques were used in this case. Still, the methods described in Breinholt et al. (2012) and the SAFEDOR Handbook (Papanikolaou & Soares, 2009) are typical engineering techniques where the “human error” is seen as

[3] SAFEDOR (Design, Operation and Regulation for Safety) was an EU project from 2005-2009 aiming to provide additional design freedom for ship and systems and an appropriate approval process that introduces safety as additional objective by proposing a regulatory framework to facilitate risk analysis as additional element of the approval process.

a cause and not as a result of poor design or organisational issues. This old view of “human error” is criticised by many as too narrow (e.g., Boring et al. (2010); Hollnagel (2000)).

The existing RBSD framework has mainly been applied for technical design (Ventikos et al., 2021). The applications that include human element considerations are relatively fewer. This is most likely because guidelines on RBSD, such as Lloyds Register’s procedures on Risk-Based Design (2016), do not provide any guidance on including human and organisational aspects of risk. Typically, Human Factors analysis in the design phase has been carried out as a separate issue.

# 3 Research Method

The purpose of a PhD project is primarily educational. A PhD indicates that the holder has obtained the necessary skills and knowledge to become a professional, independent researcher. This chapter explains "how" and "on what premises" in terms of the understanding and approach to science in the PhD research. The view adopted in this PhD is that science is a means to produce knowledge (Hansson, 2013). There exist several ideas and perspectives on what science means. Science is (in the broad sense) the practice that provides us with the most reliable (i.e., epistemically most warranted) statements that can be made, at the time being, on the subject matter covered by the community of knowledge disciplines, i.e., on nature, ourselves as human beings, our societies, our physical constructions, and our thought constructions (Hansson, 2013).

## 3.1 What is research?

Merriam-Webster defines research as "a studious inquiry or examination; especially investigation or experimentation aimed at the discovery and interpretation of facts, revision of accepted theories or laws in the light of new facts, or practical application of such new or revised theories or laws." The essence of this definition is: search for novelty, either new facts, or new theories, or new applications. Creswell (2014) state that "research is a process of steps used to collect and analyse information to increase our understanding of a topic or issue."

Research aims to enhance society by advancing knowledge through scientific theories, concepts, and ideas. A research purpose is met through forming hypotheses, collecting data by gathering evidence for theories, analysing, and contributing to developing knowledge in a field of study.

The motivation for researching a specific topic may be divided into two levels; What is the motivation for doing research in general? And why on the specific topic? For me, I started the PhD process with the desire to become an independent researcher, gain knowledge of how to carry out good research and develop and practise research skills. As for the question of why risk assessment is in the design of MASS, it is because of the exciting opportunities MASS provides and how innovative solutions allow for greener and more sustainable operations at sea. For MASS to become a success, a crucial factor is the safety of this new sociotechnical system. There are still many issues to be tackled, and I want to contribute to the theoretical considerations that need to be addressed for the practical and effective application of risk assessment in designing and developing safe MASS and have a future career as a researcher within the Maritime industry.

For defining good research, I use the current description provided by Cross (2007). Good research is:

- **Purposive:** based on the identification of an issue or problem worthy and capable of investigation
- **Inquisitive:** seeking to acquire new knowledge
- **Informed**: conducted from an awareness of previous, related research
- **Methodical:** planned and carried out in a disciplined manner
- **Communicable:** generating and reporting results that are testable and accessible to others

Science is often described as the study of the natural world, while engineering and design are devoted to creating something new in the "made world" (built environment, constructed world, the science of the artificial) (Shneiderman, 2016). Risk assessment and human-centred design are established topics that have been explored in rich literature, and it is admittedly rare that the frontiers of knowledge are pushed in new ground-breaking directions. However, a gap between engineering risk assessment practices and human-centred design exists. There might be several benefits of combining the two disciplines to reach the goal of developing a safe design considering both risks and human capabilities.

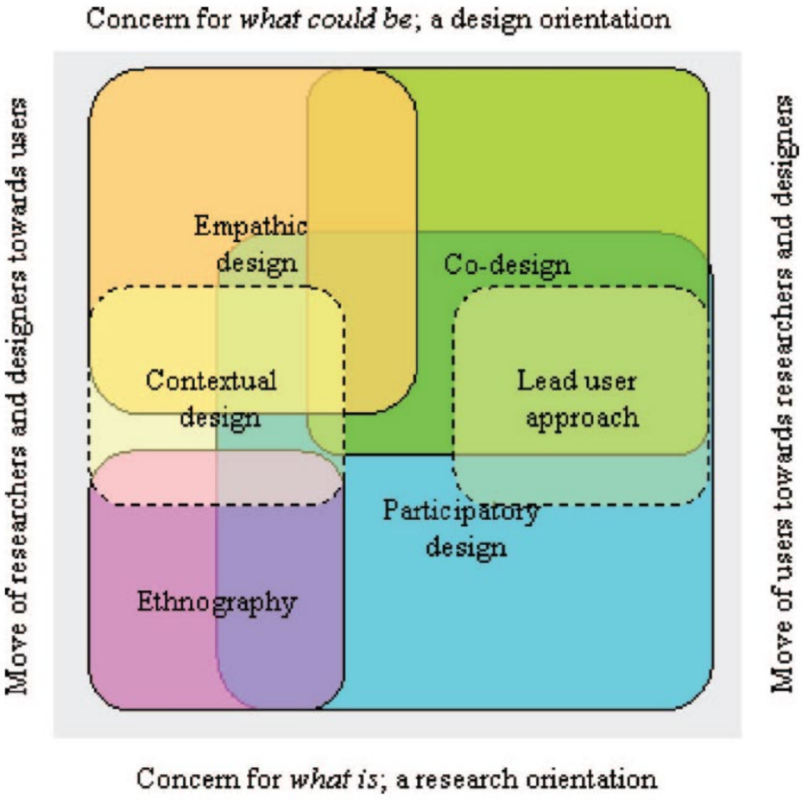
## 3.2 Design research

Design research is a relatively new field compared to traditional fields such as natural science and chemistry. Frankel and Racine (2010) have reviewed the concept of design research and how it has evolved, building on existing overviews of the field provided by Bruce Archer, Richard Buchanan, Nigel Cross, Christopher Frayling, and Ken Friedman, among others. They divide Design Research into three categories:

1. **Research for design:** research to enable design. Typically, in individual cases, by providing information, implications, and data that designers can apply to construct something. Hence associated with practice. Most practitioners and many academics associate this category of design research with the term "Design Research" (Frankel & Racine, 2010). Typical prescriptive research methods for specific design solutions include user-testing or usability testing. "There are circumstances where the best or only way to shed light on a proposition, a principle, a material, a process or a function is to attempt to construct something or to enact something, calculated to explore, embody or test it" – Archer (1995).
2. **Research through design** aims to provide an explanation or theory within a broader context (not restricted to the product in which research is conducted). Systematic design methodologies combine the practice-based research approach with elements from, e.g., social science, business, or marketing. For example, human-oriented design methodologies such as human-centred design combine human factor knowledge (from the applied social and behavioural sciences) and usability testing.
3. **Research about design** relates to basic research where the history of design, aesthetics, and design theory, as well as the analysis of design activity (Schneider, 2012), is developed.

These categories of design research are interrelated. The thesis contributes to Research for Design as the design of a SCC for MASS operation is a clinical case study subject. However, Research through Design is also relevant as I use applied research to investigate how a human-centred design (HCD) methodology can be combined with risk assessment to develop a framework to support human-centred risk assessment. As my background is in mechanical engineering, learning Research about design through courses and reading material has also been a part of my PhD curriculum.

HCD methods are not explicitly a research methodology but incorporate mainly qualitative methods and, to some degree, mixed methods (Norman et al., 2021). In HCD, researchers and designers attempt to cooperate with or learn from potential users of the products or services they are developing (Steen, 2011). HCD is in ISO9241-210 (2019), defined as an approach to systems design and development that aims to make interactive systems more usable by focusing on using the system and applying human factors/ergonomics and usability knowledge and techniques. HCD refers to a broad range of approaches, including participatory design, the lead user approach, co-design, ethnography, contextual design, and empathic design (Steen, 2011). HCD is based on four principles: 1) involving users to better understand their practices, needs, and preferences; 2) searching for an appropriate allocation of functions between people and technology; 3) organizing project iterations in conducting the research and generating and evaluating solutions; and 4) organizing multi-disciplinary teamwork (ISO9241-210, 2019). This thesis focuses on the two first principles: Involving the end-user and searching for appropriate and safe allocation of function between the human operator and the autonomous system. In the design phase of MASS, as a novel technology, the focus of HCD in this thesis will be towards a design orientation (exploring and visualizing future situations) and attempts to bring the developers and designers towards the user, but also the user towards the developers' and designers' ideas. According to Figure 7 below, the HCD approach that best fits the focus of my research is Co-design and Participatory design.



Concern for *what could be*; a design orientation
Move of researchers and designers towards users
Empathic design
Co-design
Contextual design
Lead user approach
Participatory design
Ethnography
Move of users towards researchers and designers
Concern for *what is*; a research orientation

*Figure 7 Different HCD approaches, with different starting points and emphases (adapted from Steen (2011)).*

Co-design is a contemporary form of participatory design, and they are both concerned with understanding current practices and envisioning alternative practices. Co-design can be understood as an attempt to facilitate users, researchers, designers, and others – or: diverse people with diverse backgrounds and skills – to cooperate creatively so that they can jointly explore and envision ideas, make, and discuss sketches, and tinker with mock-ups or prototypes (Steen, 2011). Steen (2011) refers to Schuler and Namioka (1993) when defining participatory design “as an approach towards computer systems design in which the people destined to use the system play a critical role in designing it”. In participatory design, one attempts to give future system users a role in its design, evaluation, and implementation.

In design research, the aim is to reach the optimal design. An optimal design involves usability considerations, and the end user is considered. Little attention is given to potential risks that may arise in the operation of the designed concept and how the design should be developed to handle safety-critical situations. A practical approach to the design is the HCD process presented in ISO9241-210 (2019). Here, understanding risk is often forgotten when performing the task “Understand and specify the context of use” and “evaluate design against requirements” (SeeFigure 1 Figure 5 in section 2.5).

The practical HCD process is an iterative and formative approach to design (part of the design process). The terms *formative* and *summative* comes from the field of education, where it is used to describe student learning: formative – providing immediate feedback to improve learning vs summative –

evaluating what was learned (Greenwich, 2022). A formative analysis in the design process does not necessarily mean that no quantitative measures are used. Albert and Tullis (2013) explains how the two approaches can be used within the product development life cycle. When running a formative study, the designer evaluates a product or design periodically while being created, identifies shortcomings, makes recommendations, and then repeats the process, until, ideally, the product comes out as perfect as possible (see Albert and Tullis (2013), p.42).

## 3.3 Research on Risk Assessments

The risk field has two main tasks, (I) to use risk assessments and risk management to study and treat the risk of specific activities (for example, the operation of an offshore installation or an investment), and (II) to perform generic risk research and development, related to concepts, theories, frameworks, approaches, principles, methods and models to understand, assess, characterise, communicate and (in a broad sense) manage/govern risk (Aven & Zio, 2014; SRA, 2021). The boundaries between the two levels (I) and (II) are not strict, and the same research methods are applied at these two levels.

In a review of recent advances on the foundation of risk assessment and risk management, Aven (2012) identifies a tension between different types of perspectives on risk analysis applications and risk conceptualisation. In recent years several attempts at integrative research have been conducted, establishing broader perspectives on the conceptualisation, assessment, and management of risk. Integrative research is linked to integrative thinking, which is the "ability to face constructively the tension of opposing ideas and, instead of choosing one at the expense of the other, generate a creative resolution of the tension in the form of a new idea that contains elements of the opposing ideas but is superior to each" (Martin (2009), p.15). Aven (2012) sees this way of thinking as essential for developing the risk field and obtaining a solid unifying scientific platform for this field.

## 3.4 Classification of research

There are several ways of classifying research. A traditional distinction is between basic research conducted to advance general knowledge and applied research undertaken to solve a practical problem (Shneiderman, 2016). This PhD project is not undertaken to solve a particular problem. Still, it is a part of the more comprehensive research program, SAREPTA (as referred to in Chapter 1), with the purpose of providing necessary knowledge for the development of improved understanding and methods for risk assessments regarding the safety, autonomy, remote control, and operations of industrial transport systems. My PhD project can be considered a combined research. Combined research is both basic and applied research. I adapt primary research findings (how to carry out risk assessments, including the human element in the design phase of MASS) and test the applicability of one method for risk assessment.

A more nuanced distinction between explorative research, testing out research, and problem-solving can be made. In this thesis, the research is primarily explorative as the remote operation of MASS is an evolving topic where little is known. I aim to answer my research questions by examining what theories and concepts are appropriate and whether existing methodologies can be used. To some degree, my PhD project can also be classified as testing out research as it aims to contribute to further development and theory building by testing out an established method for risk assessment in the design phase.

### 3.4.1 Research Methodology

Research may also be classified according to methodology, whether it uses quantitative, qualitative or mixed methods (Creswell, 2014) and whether it has an empirical or conceptual focus (Boylan et al., 2018). This PhD project may be classified as conceptual and qualitative both in focus and approach. However, mixed methods (which incorporate the elements of both quantitative and qualitative approaches) were to some degree applied to obtain a more complete understanding of the research questions.

### 3.4.2 Qualitative research methods

Qualitative research involves collecting and analysing non-numerical data to understand concepts, opinions, or experiences. It can be used to gather in-depth insights into a problem or generate new ideas for research. Quantitative research methods are suited to verifying an existing set of defined variables of an established theory, while a qualitative approach is beneficial to explore the "how" or "why" of a phenomenon rather than "how many" or "how much" (Hancock et al., 2001). Qualitative research involves purposeful use for describing, explaining, and interpreting collected data (Williams, 2007). Qualitative research can be less structured in the description because it formulates and builds new theories and has a significant sensitivity to the context in which it is implemented.

There are several different methods for conducting qualitative research. Leedy and Ormrod (2001) recommend the following five: Case studies, grounded theory, ethnography, content analysis, and phenomenological. In the PhD project, the focus has been on case studies and grounded theory research to collect and explore rich data on topics and develop theories, rather than ethnographic research, which analyses the broad cultural-sharing behaviours of individuals or groups.

### 3.4.3 Mixed methods

A mixed-methods approach (Creswell, 2014) incorporates the elements of both quantitative and qualitative approaches to obtain a more complete understanding of the research questions. The approach was chosen to address the second research question; What are the main potential (accidental) risks of MASS? Statistics (i.e., quantitative data) on known accidents in the maritime domain were used as an existing set of defined variables of an established theory (i.e., categorization of accidental events), and a qualitative approach was chosen to explore “how” and “why” autonomous technology would affect these variables.

## 3.5 The research methods of the thesis

This PhD project aims to contribute to the development of improved methods for risk assessments considering the human element by reviewing existing literature and utilising critical argumentation. Argumentation is the process of stating and reasoning from premises to conclusions in a specific context (Driver et al., 2000). The argumentation has taken place on several levels; within the mind of the candidate, between the candidate and the supervisor, within the research project group, in the department research group, with readers and reviewers in the scientific community, and through communication in the public domain.

The methodological choice for each article is summarised in Table 3 Table 3and will be introduced individually in the following sections 3.5.1 – 3.5.5.

*Table 3 Methodological choice for each research study (article).*

|  | Article Title | Main focus | Approach | Method | Data material |
|---|---|---|---|---|---|
| 1 | The present and future of risk assessment of MASS: A literature review. | State-of-the-art analysis / establishing a setting for Risk Assessment of MASS | Qualitative | Semi-systematic literature review | 8 identified articles |
| 2 | Addressing the accidental risks of maritime transportation: could autonomous shipping technology improve the statistics? |  | Qualitative<br>Mixed methods | Interviews,<br>Delphi method | Accident statistics,<br>Expert interviews,<br>Workshops |
| 3 | Improving Safety by Learning from Automation in Transport Systems with a Focus on Sensemaking and Meaningful Human Control |  | Qualitative | Integrative literature review,<br>Focus group,<br>Interviews | Articles, focus group of 9 participants who also were interviewed |
| 4 | Adopting the CRIOP framework as an Interdisciplinary Risk Analysis Method in the Design of Remote Control Centre for Maritime Autonomous Systems | Developing theory and concepts | Qualitative | Integrative literature review,<br>Interviews | Articles and CRIOP reports |
| 5 | Human-centred risk assessment for a land-based control interface for an autonomous vessel | Testing out a theory/concept (a framework for risk assessment) | Qualitative | Integrative literature review,<br>Case study | 12 participants |

In total, five research methods were applied in the PhD project, including literature review, Delphi method, case study, focus group and open-ended interviews. As an explorative applied approach is chosen for the PhD research, several methods were applied within each study/article in order to "cross-check" information and conclusions (i.e., apply triangulation). The following sections will explain these methods, analysis processes and data material.

### 3.5.1 Literature review

Building research on and relating it to existing knowledge is the building block of all academic research activities; hence literature reviews have been vital parts of the process of writing all five articles. As MASS is a relatively new and emerging concept with many ongoing projects and applications being tested out, it was essential to stay up to date on the topic of risk assessment of MASS. Hence, to which degree the literature review approach in each article followed a systematic and structured approach varied depending on the purpose of each article's research question and study. The overall goal of literature reviews was to "locate existing studies, select and evaluate contributions, analyse and synthesise data, and report the evidence in such a way that allows reasonably clean conclusions to be reached" (Denyer & Tranfield, 2009). There are several existing guidelines for literature reviews. Below are the ones that have been applied in the PhD research.

#### *3.5.1.1 Semi-systematic literature review*

While systematic reviews have strict requirements for search strategy and selecting articles for inclusion in the review, a semi-systematic literature review (or semi-structured literature review) allows for a broader selection of articles to be reviewed. This review approach is designed for topics that have been conceptualised differently and studied by various groups of researchers within diverse disciplines (Snyder, 2019). The analysis can be useful for detecting themes, theoretical perspectives, or common issues within a specific research discipline or methodology or for identifying components of a theoretical concept (Ward et al., 2009). A potential contribution could be, for example, the ability to map a field of research, synthesise the state of knowledge, and create an agenda for further research or the ability to provide a historical overview or timeline of a specific topic.

A semi-systematic literature review was applied in Article 1 to synthesise and reflect on the existing research findings on risk analysis of unmanned ships. Since the collection of literature was carried out in 2017 when IMO and other interest groups and class societies etc., had not yet published any guidelines or recommendations in terms of vocabulary for MASS, the search string involved a broad range of terms. The initial literature search was conducted to establish a picture of the most common definitions in the sense of the number of results. A second literature review was conducted in March 2018. The literature was obtained through Boolean searches in three interdisciplinary databases: Scopus, Google Scholar and Web of Science. Based on the findings in the first study, "Unmanned" was selected together with the keyword "risk identification". As an alternative way to find relevant articles, "snowballing" (i.e., tracking down references or citations in identified documents) was used to get a broader base of relevant articles.

The research articles were checked against two pre-determined criteria for their eligibility: 1) the article must be related to the maritime domain and published in a peer-reviewed journal or conference proceedings, 2) From the title or abstract of the paper, the words "risk(s)" or "accident" and a high level of automation must be present. After identifying the relevant literature, I performed an inductive coding process based on the stepwise deductive inductive approach presented in Tjora (2012) and explained in Figure 8 in section 3.6. Adjusted for the semi-analytical literature review, the process consisted of the following steps:

1) Perform an initial examination of the textual data: what is the article's main topic? Did the article present a specific risk analysis method or discuss risks more generally?
2) Identify information segments: which parts of the article in review cover the topic of interest?
3) Label the segments from categories: the type of risk analysis methods applied, what part of the MASS system is covered, and to which degree the human element is considered.
4) Reduce the overlapping categories: collect the different articles according to their risk analysis method and label them accordingly.
5) Create an overview of the categories: listed risk analysis methods.
6) Discuss the risk analysis methods by applying theory: evaluating the strengths and weaknesses of the suggested risk analyses and their applicability in the design phase of MASS.

A summary of the suggested current risk analysis categories is provided in Article 1, including a brief discussion on the consideration of human and organisational elements in these risk analyses.

#### 3.5.1.2 Integrative review

An integrative review is also known as a critical review and is closely related to the semi-structured review approach (Snyder, 2019). It aims to assess, critique and synthesise the literature on a research topic in a way that enables new theoretical frameworks and perspectives to emerge (Torraco, 2005). Most integrative literature reviews address mature topics or new, emerging topics. For mature topics, the purpose is to overview the knowledge base, critically review and potentially reconceptualise, and expand on the theoretical foundation of the specific topic as it develops. For newly emerging topics, the purpose is rather to create initial or preliminary conceptualisations and theoretical models rather than review old models.

Throughout the PhD research and especially in work with articles 3, 4 and 5, the integrative literature review has been a part of the main steps of the research approach that is close to that of research synthesis (Cooper, 2015): Research synthesis is the integration of existing knowledge and research findings pertinent to an issue. Synthesis aims to increase the generality and applicability of those findings and develop a new understanding through integration. The main steps to research synthesis in this PhD project are as follows (Cooper, 2015):

1. Define problem
2. Collection of literature: the selected sampling approach was mainly purposeful by snowballing and gathering data
3. Establishment of inclusionary and exclusionary criteria
4. Interpretation, evaluation, and synthesis of the literature
5. Development of new concepts
6. Evaluate, present, and redefine new concepts and theories.

In Article 3, the literature search topic was "meaningful human control of autonomous systems" within the four transportation domains: road, sea, air and rail. This is a new, emerging topic first identified within aviation (arising from the debate on lethal autonomous weapon systems). No references were found within rail, and only a few references within autonomous shipping and self-driving cars (Heikoop et al., 2018; van den Broek et al., 2020). Experiences from accidents and incidents and lessons learned from each domain were gathered by interviewing the domain experts in the SAREPTA project. The co-authors of the article were all members of the SAREPTA project group. The 12 members worked as a focus group and reviewed literature within their respective domains (see section 3.5.5 for more information about the focus group). They also contributed with data and experiences in the discussions during meetings in the project group where the article's topic was discussed.

In Article 4, the topic of the integrative review was the well-established framework for Crisis Intervention and Operability Study (CRIOP). The framework was evaluated based on its applicability as a risk assessment method in designing the new concept of SCC for MASS. Working as a safety engineer, I had first-hand experience applying the framework in the design process of a shore-based remote-control room for an offshore oil and gas installation. In the review process, literature and data were collected through searches in online databases and by reviewing CRIOP reports provided by one of the creators of the CRIOP method, who also co-authored the paper. The integrative review involved operational considerations for the constrained autonomy concept and use cases to critically examine and reconceptualise the framework to match the need for a risk assessment of the SCC for MASS operation. Hence, adjusting the scenarios to be evaluated and including operational envelopes (defining the human's and automation's responsibilities) was part of the new conceptual framework of an adapted CRIOP for a SCC.

Article 5 builds on the research in Article 4 but includes an integrative literature review of risk assessments of MASS focusing on formative and summative uses of risk assessment, the integration of the human element (the operator in the control centre) in risk assessments, Human factors and sociotechnical (e.g. human-centred) design principles and the requirements of IMOs Formal Safety Assessment (FSA). Article 5 consists of two main research methods: a review of the CRIOP method (an integrative review) and a case study of an applied CRIOP analysis on a prototype of a SCC. The integrative review aimed to figure out how the CRIOP framework can bridge the gaps between 1) risk based design and human-centred design, 2) the need for including the human element in risk assessment (as required in IMOs FSA) and minimise the issue of "Work as Imagined" vs "Work as done", and 3) a comparison of the CRIOP method and other current risk assessment methods suggested for the design of MASS.

### 3.5.2 Case study

A case study involves the in-depth investigation of single or multiple cases to acquire profound and detailed information related to the phenomena under investigation (Yin, 2009). A case study is an appropriate research strategy to generate a theory (Eisenhardt, 1989; Yin, 2009). It commonly includes direct observation of the event and interviews with the actors involved (Yin, 2009). A case study is a suitable approach for studying complex contemporary social events when answering research questions that start with “why” or “how”. A unique characteristic of case studies is that they allow all kinds of materials as evidence, including documents and artefacts (Yin, 2009).

In Article 5, a case study was applied to test the hypothesis that the Scenario Analysis from the CRIOP framework can be a valuable tool for risk assessment in the design phase of a SCC for MASS operation. The case study aimed to test the applicability of the Scenario Analysis by evaluating the validity, credibility, and reliability of the approach, based on the exploration of a critical scenario in a simulated SCC with experts from different disciplines. The case study was a process and outcome analysis (Yin, 2009) where we have an initial descriptive theory about the case tentative to the study and a hypothesis about the expected characteristics of the case (see Figure 4 in Article 5)

### 3.5.3 Delphi method

The Delphi method is a structured and interactive communication technique used to congregate expert opinions on a specific topic (Okoli & Pawlowski, 2004; Skinner et al., 2015). It has been widely utilised in various research fields to identify the critical issues of the subject matter from the experts' perspectives. It typically involves several iterative communication rounds in which a group of experts is asked to answer a series of questions until reaching a consensus. The responses (from experts) are typically synthesised after a first Delphi round and shared with the experts again in a second round.

The method was applied in Article 2 with some modifications to shorten the communication process and avoid non-relevant responses. As a preparation for the Delphi study, statistics on accident data were collected through reports by the European Maritime Safety Agency (EMSA, 2018) and Allianz Global Corporate & Specialty (AGCS, 2018). This was a quantitative analysis to identify today's accident picture (for conventional ships). As a mixed method approach, the quantitative data was applied to a qualitative comparison of autonomous and conventional ships. In Article 2, the main research question was: "can autonomous shipping technology improve the accidental risks of maritime transportation". This was also the topic of a workshop by the Norwegian Forum for Autonomous Ships[4] (NFAS) in the fall of 2018. In this meeting, experts from academia and industry presented projects and discussed topics related to the status and forecast of autonomous technologies and their capabilities. I gathered expert opinions during the workshop and interviewed five members during the meeting breaks. The following two open-ended questions were asked: What are the main features of autonomous technology that will improve the safety of maritime shipping (in terms of today's known accidents)? What are the risks of autonomous technology? After the first round, these opinions were returned to the co-authors, who acted as an expert group. Here, the input from the presentation, the discussions during the meeting and data from the interview were discussed and categorised into differentiating factors (between conventional and autonomous shipping) and the effects of autonomous technologies. The authors gathered for the third time to go through the categorisations and discuss how the autonomous technology would affect the new risk picture (in terms of their contribution to known accidents, accidents averted by the present crew and new accidents caused by the introduction of new technology).

### 3.5.4 Open-ended semi-structured interviews

In an open-ended interview, the interviewee is asked questions that cannot be answered with a simple yes or no. This encourages the interviewees to talk freely and extensively, thus providing information that might not be obtained otherwise (APA, 2022b). A semi-structured interview is highly flexible in terms of the questions asked, the kinds of responses sought, and how the answers are

[4] Norwegian Forum for Autonomous Ships https://nfas.autonomous-ship.org/

evaluated across interviewers or interviewees. In articles 2, 3 and 4, open-ended semi-structured interviews were carried out in person and over the phone. A few standard questions were prepared in advance, and the subsequent discussion allowed me to pursue the area of interest as it arose. The semi-structured interviews allowed for spontaneous discussion to reveal more of the applicant's expertise, opinion and argumentation compared to that of a standard predetermined question set would (see Adams and Cox (2008)). The interviewees were selected based on their competence and availability. The interviews were mostly unformal, and no recording devices were used.

### 3.5.5 Focus group

A focus group is a small set of people, typically 8 to 12 in number, which share characteristics and are selected to discuss a topic (e.g., determining typical reactions, adaptations, and solutions to any number of issues, events, or topics) of which they have personal experience (APA, 2022a). A moderator conducts the discussion and keeps it on target while encouraging free-flowing, open-ended debate.

As mentioned, the research project group in the SAREPTA project worked as a focus group for my PhD project. The group's composition was characterized by homogeneity (i.e., they had a common interest in the topic of autonomous transportation systems, and they were all senior researchers, professors, or experts within their field) but with sufficient variation to allow for contrasting opinions. The SARPTA project group consisted of 12 members and met approximately once a month during the duration of the project (from fall 2017 through spring 2021). The primary purpose of the meetings was to coordinate the project's activities and to discuss specific topics. In the meetings, I had a time slot on the agenda where I, as a PhD student, could address questions and get advice, e.g., on whom to contact for specific information (and how) or other ways of solving an issue I had run into.

## 3.6 The role of the researcher in qualitative research methods

In qualitative methodology, the researcher her/himself is the primary instrument for data collection, interpretation, and analysis. The qualitative analysis of the available literature is a demanding intellectual exercise. As an aspiring researcher, I strived/struggled to be concise and clear in my thought processes, balancing the different views and opinions on how to design a MASS system (e.g. practical considerations such as what should the role of the human operator be) and how to carry out risk assessments in the design phase. I found help in reducing the complexity of the thought processes by structuring my research in smaller steps, as advised in the stepwise deductive inductive process (as mentioned in section 3.5.1.1) presented in Figure 8 below. In this qualitative approach, I went from gathering empirical data to establishing theories and frameworks in Articles 1, 2, and 3 by going upwards through the processes in an inductive approach. While in Articles 4 and 5, I tested out a framework by going downwards through the processes in a deductive way.

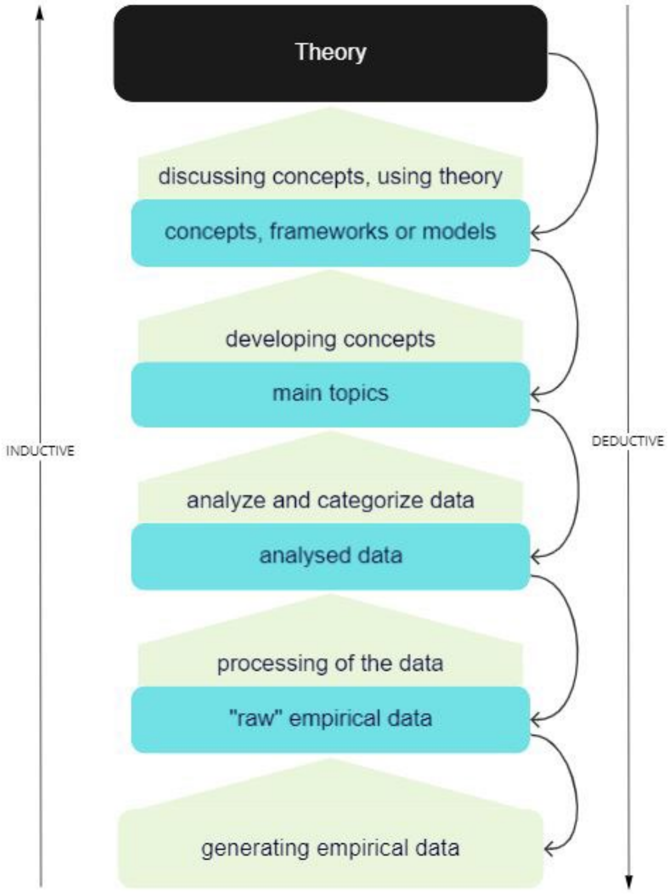

Theory
discussing concepts, using theory
concepts, frameworks or models
developing concepts
main topics
analyze and categorize data
analysed data
processing of the data
"raw" empirical data
generating empirical data
INDUCTIVE
DEDUCTIVE

*Figure 8 Stepwise deductive inductive method (adapted and adjusted from Tjora (2012)).*

This may seem like a linear process, but in reality, my research approach was much messier. Being a good researcher requires specific knowledge and skills (which must be practised) and the ability to improvise, and intuition, which grows with routine. The fact that in almost every research study, I needed to adjust the course to some degree became a source of frustration but also increased learning. At some point, I did not have control over the process when a planned case study was cancelled due to Covid-19 or when I did not get sufficient data from interviewees. I experienced both failures and successes at each step in Figure 8. In the process of carrying out the studies, I had to move up and down the different steps in the method, as I found it necessary to gather more data or evaluate other concepts (i.e., jumping several steps down in the approach).

## 3.7 Overall procedure

The overall procedure of the PhD project is outlined in Figure 9. The project consisted of three main activities:

1. Course work
2. Development of the research articles
3. Writing the thesis

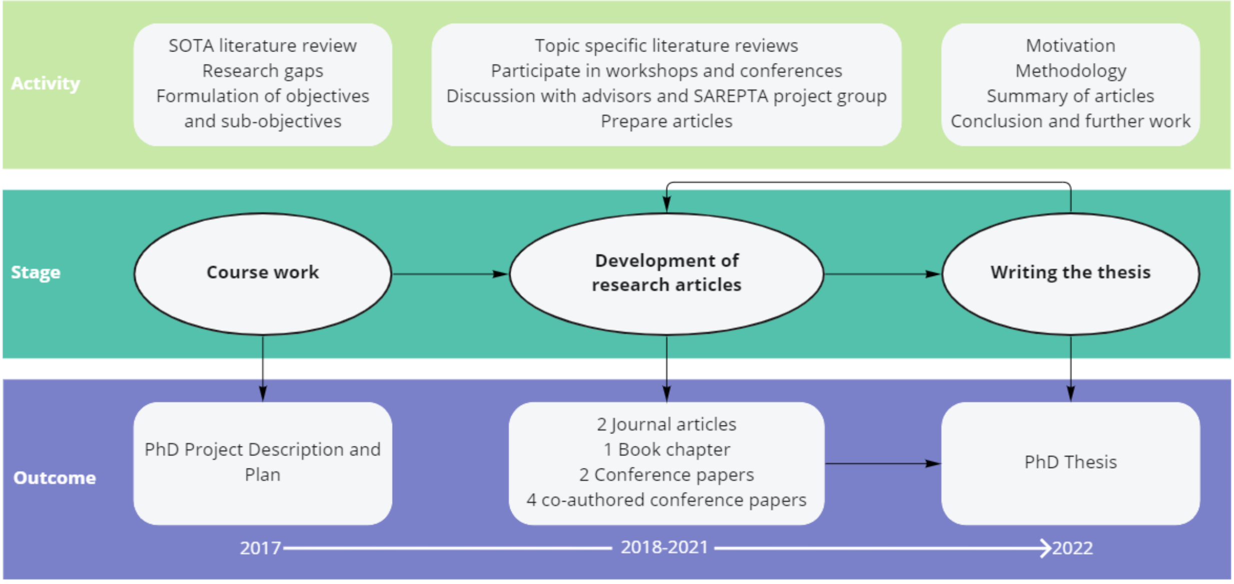

Activity
SOTA literature review
Research gaps
Formulation of objectives and sub-objectives
Topic specific literature reviews
Participate in workshops and conferences
Discussion with advisors and SAREPTA project group
Prepare articles
Motivation
Methodology
Summary of articles
Conclusion and further work
Stage
Course work
Development of research articles
Writing the thesis
Outcome
PhD Project Description and Plan
2 Journal articles
1 Book chapter
2 Conference papers
4 co-authored conference papers
PhD Thesis
2017
2018-2021
2022

*Figure 9 Overall procedure and main activities of the PhD project.*

The purpose of the coursework is to provide a broad and solid base for the PhD research. Starting up, I had expertise in both technical and practical aspects of risk analysis and assessment. Still, I needed to extend my competence in Interaction design, Human factors, HAI and research methodologies. This was achieved by a mix of specialized courses on HAI, Human Factors in the maritime domain, mixed-method research, and design research. Through my PhD project's professional network, I got in touch with Professor Missy Cummings at the Human Autonomy Lab at Duke University. This gave me the opportunity to visit their lab, follow a course in HAI and participate in and learn from their research projects during the spring of 2019 (January to mid-April).

The project plan was developed as an iterative process with a background in the project description of the SAREPTA project and identified research gaps. The initial research questions and objectives were derived from this project, but complete freedom was given to reformulate the objectives of the PhD project. After gaining more knowledge of the state of the art of risk assessment of MASS (objectives 1 and 2) and discussing the topic with scholars, the third objective and the following sub-objectives were revised and updated to narrow the initial scope. The focus was now on the risk assessment in the design phase and how to combine the need for human-centred design of SCCs and the concept of risk based design.

Reviewing the state of the art was a challenging exercise as the topic of MASS gained increased attention in terms of research interest, publications, and industry initiatives during the initial years of my PhD project. This was also a blessing because it provided new opportunities to discuss the topic with researchers from different disciplines. I was invited to share my research in meetings with the shipping industry and other actors. During the PhD project, I participated in workshops run by the research projects HUMANE[5], TRUSST[6] and Autoferry[7]. The topics of these workshops involved the safety of MASS technology and its applications in various aspects, including Human Factors, Remote

[5] The Human Maritime Autonomy Enable (HUMANE) project performed a broad, human-centred evaluation of all implications and required changes regarding MASS https://www.hvl.no/en/project/591640/.

[6] An industrial research project for Assuring Trustworthy, Safe and Sustainable Transport for All (TRUSST) https://www.zeabuz.com/trusst

[7] A cross disciplinary research project on autonomous all-electric passenger ferries for urban water transport (Autoferry) at NTNU. https://www.ntnu.edu/autoferry

Monitoring and Control, Communications and Cyber Security, Risk Management, Training and Education, Assurance, Policy, Regulation, Law, etc. These workshops provided several opportunities for data collection through interviews and open discussions among scholars, professors, and experts from different disciplines.

Except for one article (Article 1), all articles are written in collaboration with researchers in the SAREPTA project group. Especially, Senior Research Scientists Ørnulf Jan Rødseth and Kay Fjørtoft at the research institute SINTEF Ocean have contributed with their valuable insight and included me in their research on projects such as AUTOSHIP[8] and IMAT[9]. While developing all articles, I got valuable input from my main supervisor Thomas Porathe and co-supervisors, Professor Margareta Lützhöft and Senior Research Scientist Stig Ole Johnsen.

Writing the thesis involves taking a step back to reflect on my motivation, my thought processes, and the different "roads" I took in finding the "right" approach and continuously revising and updating the objectives as the research evolved throughout my studies.

## 3.8 Summary

There are limitations to the research approaches applied in the PhD research. Qualitative methodologies are generally criticised for being subjective (influenced by the researcher’s perception, opinion and judgement), not replicable or easy to derive generalisation from, and producing large quantities of data that are difficult and time-consuming to aggregate and analyse (Maxwell, 2008). Also, verifying the result of qualitative research can be challenging as the research is often open-ended, where participants can have more control over the content of the data collected compared to quantitative research. Different types of qualitative research require different levels of researchers’ control and participants’ involvement. The specific limitations of each research approach relate to the results of the articles and are best evaluated regarding the validity and reliability issues of this thesis. The terms validity and reliability are explained in the following sections. The limitation of each article is further discussed in the next chapter.

## 3.9 Validity, reliability, generalisability, and quality of research

Three requirements are often used as indicators on the quality of the research (both quantitative and qualitative): validity, reliability, and generalisability. In contrast to quantitative research, qualitative research as a whole has been constantly criticized, if not disparaged, by the lack of consensus for assessing its quality and robustness (Leung, 2015). For example, the discussion about the necessity of generalization in qualitative research and how this should be done has been going on for a long time. In this thesis, the criteria of conceptual generalization (Tjora, 2012) are of particular interest. That is, developing concepts and theories that will be of relevance to cases and applications other than the one studied. Beyond the conceptual generalizability of my research, a pragmatic approach to assessing generalizability is to adopt the same criteria as for validity (Leung, 2015). This quality criterion is further explained in the next section.

[8] The EU project, Autonomous Shipping Initiative for European Waters (AUTOSHIP), aims at speeding up the transition toward the next generation of autonomous ships https://www.autoship-project.eu/.

[9] The Integrated Maritime Autonomous Transportation Systems (iMAT) project defines, develops, and tests land-based sensors, communication, and control systems for MASS operation https://www.sintef.no/projectweb/imat/.

### 3.9.1 Validity

Validity refers to the appropriateness of the inferences made from the results —in other words, the extent to which the results accurately measure what the research intended to measure (Maxwell, 1992). The most crucial validity aspect of my research is that it is tested in dialogue with the research community, in the research project group, at conferences, in workshops and by publishing my results in scientific, peer-reviewed journals. In practice, this means I consciously relate to current theories and perspectives based on previous research within the same topic and with the same methods. I also aimed to strengthen the validity by explaining my choices and being open about how I have carried out the research. The validity of each article is established in Table 4 below.

*Table 4 The strategies used to promote research validity in the PhD research.*

| | **Method** | **Validity** |
|---|---|---|
| Article 1 | Literature review | - Systematically synthesised existing findings by explaining the process of identifying, selecting, and reviewing the literature.<br>- Categorized findings according to established risk assessment practices.<br>- Data triangulation by crosschecking collected data using multiple sources. |
| Article 2 | Literature review, Delphi method, Interviews | - Used established statistics and categories as a base for discussion.<br>- Data triangulation: used statistics from EMSA and reports from insurance carriers.<br>- Method triangulation by using different types of data collection procedures (interviews, literature review, Delphi method).<br>- Received experts' opinions to increase content validity.<br>- Experts had the opportunity to refine the researcher's understanding and findings.<br>- High representativeness in weighting the evidence, hence assessing the quality of the collected data. |
| Article 3 | Literature review, Focus group, Interviews | - Received experts' opinions to increase content validity<br>- Interactive contact with the participants in the focus group<br>- Theory triangulation by using multiple theories and perspectives<br>- Systematically synthesized the existing research findings and discussed them across the domains |
| Article 4 | Literature review, Interviews | - Theory triangulation by using multiple theories and perspectives<br>- Used various sources of evidence: reports from applied analysis, socio-technical design principles, and a design framework structure for MASS<br>- Received experts' opinions |
| Article 5 | Literature review, Case study | - Broad selection of participants to reduce biases<br>- Provided the characteristics of the participants<br>- Used an existing framework to guide the case study<br>- Looked for negative evidence in the case study |

Threats to the validity of my research are that there were only one focus group, the SAREPTA research group, that was included in my studies. However, the researchers in this group represented a variety of disciplines (i.e., Marine Engineering, Electronics, Software Engineering, Psychology, Sociology, Interaction Design, Human Factors and Computer Science).

My personal values, judgments and ideological preferences also shaped the research design and the interpretation of the results, which may also have led to biased conclusions due to information processing biases (Yin, 2009). To minimize this bias, two or more researchers were involved in each research process (except the first study) to minimize the subjectivity involved in the interpretation process.

### 3.9.2 Reliability

Reliability refers to the extent to which the results can be reproduced if the research is repeated under the same conditions (Kothari, 2004). This implies that we could ask "would the results be the same if another researcher did the same job?" And one would assume that answering yes would imply a high level of reliability. However, in qualitative research, reliability is the degree to which the finding is independent of accidental circumstances of the research (Kirk et al., 1986). In short, establishing reliability in qualitative research is about being consistent, explaining what information comes from data generation and what the researcher's own analyses are. It is important to inform about the context of the study and reflect on whether you have special knowledge and commitments that could influence access to the field, the selection/data generation, analysis, and results (Tjora, 2012).

As mentioned, the topic of MASS is relatively new and emerging, where there are still issues non explored, and the operational experience is none-existent, or at least limited. I had some preconceptions about what an ideal risk assessment would entail based on my experiences. However, I realized quickly that I needed to be open to adjusting my conceptions in the process of answering my research questions. To address the topic of MASS and risk assessment, I chose an explorative approach, hence assessing the topic from several different angels, integrating data from a variety of methods and sources of information (Maxwell, 2012). This is done by triangulation of methods, sources, and investigators/researchers. Below, in Table 5 is an overview of the strategies followed to increase the reliability of the research in each study/article.

*Table 5 The strategies to increase the reliability of the research of each study.*

| | **Method** | **Reliability** |
|---|---|---|
| Article 1 | Literature review | - Theoretical triangulation: other research in the same area is analysed<br>- Critical review carried out by a researcher with relevant background |
| Article 2 | Interviews,<br>Delphi method | - Two other researcher/experts involved in the process of analysing and co-writing the article |
| Article 3 | Literature review,<br>Focus group,<br>Interviews | - A focus group of nine experts. All reviewed the article before publication.<br>- Obtained consensus from everyone in the focus group<br>- Cross-checked findings with similar research |
| Article 4 | Literature review,<br>Interviews | - Two experts from two different disciplines involved in the development and evaluation of the suggested framework<br>- Compared the results with previous research findings |
| Article 5 | Literature review,<br>Case study | - Explained the data analysis process and limitations of the case study<br>- Compared the results with previous findings and findings with analysis of similar concepts<br>- An actual prototype of a SCC was applied in the case study<br>- Employed a clear description of a scenario analysis that can be repeated<br>- Two scholars involved in the analysis and discussion, and agreed on the findings |

### 3.9.3 Scientific quality

This PhD thesis follows the criteria for scientific quality laid out by the Norwegian Research Council NRC (2000). The research has been conducted to the best of the author's ability to emulate the criteria of originality, solidity, and relevance (as per NRC (2000)).

Originality relates to the contribution of new knowledge to the existing academic literature. New ideas are often created at the intersection between disciplines and perspectives, and this PhD preliminary research shows originality by interdisciplinary synthesising. This resulted in developing a new human-centred risk assessment method (based on already known material), providing improvements to current concepts, and applying existing knowledge to new problem areas.

Solidity means that the results are well sustained and stand up to scrutiny. New ideas should be presented in a clear and honest way that makes it possible to trace and question their basis and conclusions. In many ways, this is related to the validity, as explained in the previous section. The use of references, scientific methods, synthesis, and peer evaluation of results has been employed to satisfy solidity criteria.

The research in this work is relevant both academically and practically/domain-wise. Through my profile at the online research community *Researchgate.net*, where I have uploaded all published articles, I have received notifications when the articles are downloaded, recommended, and cited by others. This endorses that other academics have found the topic fascinating and that the findings contribute to advancing theory within or across domains. My research findings can apply to different highly automated and autonomous transportation systems, focusing on filling the gap between engineering risk assessment, human factors, and human-centred design practices. The practical utility of the research has been a focus of the explorative applied research. The findings in the case study applied in Article 5 reflect that my research is also useful from a practical point of view.

# 4 Main Results and Discussion

The PhD results are documented in the five articles attached in part II. They are presented, analysed, and discussed with reference to the theories presented in Chapter 2. Each article's main findings and contributions are summarised in the following sections related to each objective. The main results generated in each article are presented and analysed in light of its limitations. Possible areas for further research are indicated. This presentation gives only a brief overview, and the reader is referred to the articles for concrete descriptions of the detailed results.

Two articles are published in relevant international journals, and one article is published as a book chapter. The other two have been presented at peer-reviewed international conferences and published in conference proceedings.

Objectives:

- O1: Review the SOTA on risk analysis and risk assessment in the design of MASS.
- O2: Investigate what we learn from other transportation domains and how autonomous technology will affect the potential accidental risks of MASS.
- O3: Propose a method for risk assessment in the design phase, including the human element.

## 4.1 Objective 1

The first objective was to review the "state of the art" on risk analysis in the design of MASS. The motivation was to investigate what risk analysis methods were suggested by considering the most current research in the area of risk analysis and assessment of MASS. This objective is mainly addressed through the semi-systematic literature review in Article 1. However, essential contributions to the objective were identified alongside the whole PhD research process as the topic of the safety of MASS gained more attention both within the industry, regulations, and research society.

### 4.1.1 Article 1: The present and future of risk assessment of MASS – A literature review

The article addresses the following two research questions:

1. *What risk identification analysis and methods for MASS can be found in the literature today?*
2. *What are the main limitations and challenges of these risk assessments?*

The applied research method is the semi-systematic literature review, as explained in section 3.5.1.1, where the research synthesis followed the inductive stepwise approach presented in the same section. From the identified literature, eight papers were directly concerned with risk models or risk identification. Risk models assess the risk arising from ship traffic or during ship operation by providing a graphical representation of real-world phenomena. Examples of risk models are fault trees, event trees, or Bayesian belief networks (BBN). The eight papers presented different approaches to risk identification, risk analysis, and risk management and are listed according to the method in Table 6 below.

*Table 6 Identified risk methods for MASS.*

| No. | Author(s) (year) | Topic/Title | Risk methods |
|---|---|---|---|
| 1 | Rødseth, Ø, & Tjora, A (2014) | A system architecture for an unmanned ship | HazId |
| 2 | Rødseth, Ø.<br>& Burmeister, H.C. (2015) | Risk assessment for an unmanned merchant ship | |
| 3 | Rødseth, Ø. & Tjora, A (2015) | A risk based approach to the design of unmanned ship control systems | |
| 4 | Thieme, C.A. & Utne, I.B. (2017) | A risk model for autonomous marine systems and operation focusing on human–autonomy collaboration | BBN, HAC |
| 5 | Wróbel, K et al. (2017) | Towards the Development of a Risk Model for Unmanned Vessels Design and Operation | BBN, ETA |
| 6 | Utne, I.B. et al. (2017) | Risk Management of Autonomous Marine Systems and Operations | Risk management |
| 7 | Wróbel, K et al. (2017) | Towards the assessment of potential impact of unmanned vessels on maritime transportation safety | What If, HFACS |
| 8 | Wróbel, K et al. (2018) | System-theoretic approach to the safety of remotely-controlled merchant vessel | STPA |

The paper aimed to systemise and evaluate the different methods according to their limitations and challenges in terms of their applicability in the design phase and whether they include the human element or not.The review was primarily an assessment of models of risk identifications, and the article presented and discussed each model or method for risk analyses separately.

The main findings are:

- Five risk analysis methods were identified, as listed in Table 6.
- There has been a substantial progress from 2013 toward risk analysis that could be useful in the design of MASS.
- From the eight papers reviewed, it is difficult to conclude one recommended practice for risk assessment of MASS. They all cover different topics, and a few can be seen as overlapping and, to some extent, supplement each other.
- The papers highlight only parts of a socio-technical system and a few scenarios. The main focus is on the technical aspects.
- The methods in the papers (except paper no. 6) do not state the risk definition or risk measure.
- As insufficient data are available for MASSs, quantification of risk models is difficult, and the risk models in the papers are qualitative. However, most of the papers aim to use qualitative models as a basis for quantification by applying conditional probability tables.
- None of the papers have a human-centred approach, i.e., identifying, analysing and assessing operational risks from the perspective of the human operator. One paper focuses on the Human-Autonomy Collaboration (HAC) by modelling the relationship between human operator performance and the technical performance of the autonomous system. Another paper carries out a what-if analysis augmented by the Human Factors Analysis and Classification System for Marine Accidents (HFACS-MA) in order to analyse whether the introduction of MASS will increase the overall safety of maritime transportation.
- The STPA method seems to be the most theoretically documented framework suitable for the socio-technological system; however, it requires a high level of knowledge of the system architecture.
- The risk models do not present a high level of detail in the model description or structure, making it difficult to assess them.
- The literature search did not include the term “human factors” or “human error”; however, Human Factors and situation awareness issues are mentioned in five of the eight papers. It is a consensus in the majority of the papers that the contribution of Human Factors is essential.
- All eight papers acknowledge the lack of data on design solutions and system architectures and recognise that more work is necessary to develop risk analysis and assessment approaches.

**Discussion:** This was the first draft at establishing the state of the art on which type of risk identification methods were published in relation to MASS. As mentioned, there has been a drastic increase in the number of publications in the last few years. Searching for literature using the exact keywords, predefined criteria, and research questions would, if carried out now in 2022, provide a higher amount of relevant and extensive literature on the topic. Henceforth, providing a completely different review. However, the review gave me an overview of the efforts toward risk assessment in the design of MASS in 2017.

The methods are mainly different types of hazard analyses that present the MASS systems' elements, operations, and safety features. The risks are related to different kinds of known accidents, and the MASS is considered similar to an unmanned conventional ship that is remotely monitored and controlled by a SCC. There are no empirical data, so the analyses are mainly qualitative and on a high level. Thieme and Utne (2017) is the only paper to quantify the risk model (by applying conditional probability tables). All risk methods and models belong to the Safety I perspective. Only a few papers addressed the human-automation interaction, namely Wróbel et al. (2017) and Thieme and Utne (2017). When addressing the human element, the concept of Human Reliability Analysis (HRA) is often referred to in the literature (French et al., 2011).

HRA is a collection of different methods for identifying potential human failure events, qualitatively evaluating factors that influence human errors, and applying human error probabilities for each human failure event (Mosleh & Chang, 2004). However, none of the identified papers explicitly presents an HRA. In Thieme and Utne (2017), a Human-Autonomy Collaboration (HAC) risk model is presented using data from an HRA-based method in the case study. The risk model in this paper is a BBN where the HAC performance depends on the performance of the human operator and the autonomous function (technical system), respectively. These are influenced by a range of operational aspects (level of autonomy, mission duration, number of vehicles, etc.) and performance shaping factors (such as fatigue, task load, experience, training, etc.). A case study where the model is applied and quantified in a case study of an autonomous underwater vehicle (AUV, which is not a MASS) with a high Level of Autonomy (LoA). The model is not directly adaptable for MASS as it would be more complex due to the operational aspects, including interaction with other vessels and probably a higher reliance on the human operator to intervene. The model lists interface design as a factor where the quality of the HMI highly influences the way information is perceived. The HMI design would also affect the quality of the execution of tasks as the interface also entails feedback from the operator; however, this is not addressed in the paper. The HAC model might identify potential problems and issues that arise under an AUV mission by highlighting essential relationships between the human operators and the technical (autonomous) systems. In this way, it could aid the design of MASS. However, an overall risk model of all interactions would be highly complex due to the operational aspects. Making different models depending on the operational design domain/phase and the LoA (i.e., use cases/scenarios) to express and evaluate the HAC might be an improved approach.

Some of the proposed methods identified in the review were suggested for summative use, that is, to estimate the overall risk level, the likelihood of accidents for MASS based on data from conventional shipping, and the probability of adequate HAI. The summative use or risk assessments are typical for the engineering view of Risk Analysis, focusing on representing and quantifying the risks involved in a situation to facilitate making decisions (Franzoni & Stephan, 2021). Some methods may also be of formative use: recognising and roughly ranking the potential for different accidental events, hazards, and risk influencing factors can help improve the design proactively and systematically.

## 4.2 Objective 2

In the first study, I looked at the current status of RA of MASS and found some initial ideas and suggested risk modelling techniques and frameworks. In the literature review, the hypothesis of increased safety was often brought forward, and the request of MASS to be at least as safe as conventional ships. One may then ask, how safe is manned shipping today, and what are the main accidents and risks?

The second objective is two-folded. Objective 2a is to consider the characteristics of MASS and examine the possible contribution the autonomous technology will have on the known risks in merchant/conventional shipping today. The sub-objective 2b is formulated because I have chosen an explorative research approach. Hence investigate how other industries approach the risk of autonomous technology and their implications on HAI and risk assessment.

Article 1 identified a few initial ideas for risk assessments. The application to the design phase of MASS was limited; hence, the topic of risk and safety of autonomous and highly automated systems within other transportation domains should be explored. The sub-objective 2b explores how other transportation domains approach the safety of autonomous systems, their experiences, and what is applicable to MASS operation.

Objective 2b is to investigate what we learn from transportation domains where a high degree of automation is present/involved; what are the experienced accidents and related risks?; what can we learn from the experience here?; what is applicable for MASSs?

### 4.2.1 Article 2: Addressing the accidental risks of maritime transportation: could autonomous shipping technology improve the statistics?

In parallel with carrying out research for my first article, I contributed to the article/paper "At least as safe as manned shipping? Autonomous shipping, safety and "human error" (Porathe et al., 2018), where we addressed the hypothesis of increased safety of MASS due to replacing the human with automation. The paper discusses how autonomation can make shipping safer and why automation can make shipping less safe. The paper further highlights some challenges with QRA of MASS and recommends looking into new methods for risk assessment of socio-technical systems in the early design phase in order to address the new risk picture. The paper states that if autonomous unmanned ships are to become a success, they must prove successful in several areas, and safety is one of them. Thus, we might ask: how safe is then manned shipping today? Moreover, could autonomous shipping technology improve accident statistics? This is the background for the second research objective (2a) and is closely linked to research question 2 (see section 1.4).

Article 2 aims to fulfil objective 2a by investigating accident statistics for merchant ships and examining how autonomous shipping technology will affect these statistics. The chosen research method was the Delphi method, as described in Section 3.5.3, and interviews of experts to get other opinions as input to evaluating the autonomous technologies' contribution to accidental risks. The study first identified the main factors distinguishing an autonomous ship from a conventional manned vessel.

We made some initial assumptions (stated in the article), such as that an autonomous ship is a ship that is entirely unmanned with constrained shipboard autonomy and a SCC that will handle events that the automation cannot handle. The main differentiating factors were identified as: fully unmanned cargo ship, constrained autonomy, a SCC, higher technical resilience, and improved voyage planning. From accident statistics and a list of hazards identified by Bureau Veritas (2019), we identified causal and contributing factors, conditions, activities, systems, components, etc., that are critical concerning accidental risk. The relevance of these was evaluated for each category (differentiating factor) and their effect on the operation of autonomous ships (sub-category). In this process, open interviews of experts gave additional input on how autonomous technologies and their characteristics would affect the risks in terms of known accidents and possibly new types of accidents. Other questions of “how” and “why” were asked for each statement made. By bringing this back to the Delphi method group, analysis and discussion on how each differentiating factor, their characteristics and possible effects would affect the “new risk picture” was initiated by asking the following questions. Will the main differentiating factor, including their effects:

A. contribute to new types of incidents? Yes, or neutral (as it is not possible to have a positive impact on unknown risks)
B. contribute to what is most characterised regards today’s incidents in shipping? (Positive impact on risk (reduced risk of known accidents), negative impact (increased risk of known accidents) or neutral.
C. contribute to the risks (in terms of incidents and accidents) that are today handled and averted by the presence of crew onboard? A positive impact (reduced risk) implies that the effect of the differentiating factors reduces risk. While a negative impact implies that the effect is increasing the risks.

A detailed discussion and analysis of these questions are presented in Article 2. The main results are shown in Table 7 below. The effects are listed in the last columns. The colour red (R) indicates an increased contribution to risk, yellow (Y) indicates a neutral impact, and green (G) indicates a lower impact (i.e., lower risk).

*Table 7 Qualitative comparison between autonomous (unmanned) and conventional (manned) shipping, adapted from Article 2, Hoem et al. (2019).*

| **Main differentiating factors** | | **Brief description of the effects** | **A (New)** | **B (Todays)** | **C (Averted)** |
|---|---|---|---|---|---|
| ***Fully unmanned*** | | | | | |
| 1 | Higher demand on sensors, automation and shore control as one lack some of the “personal touch”, both on the environment, ship and technical system’s performance. | More technology means more complexity and possibility of technological failure, but it will also improve some of today’s operators’ erroneous actions (known as “human error”). | R | G | Y |
| 2 | Less exposure to danger for the crew. | 40% of deaths at sea are occupational hazards. | Y | G | G |
| 3 | May be unable to inspect equipment or systems that report errors or problems. | This may cause problems, especially if sufficient backup systems are not in place. | R | Y | Y |
| 4 | Slightly lower risk of fires in accommodation, galleys, laundry, and waste systems. | Improvement on today’s accident events, but more difficult fire handling and control. | R | G | Y |
| ***Constrained autonomy*** | | | | | |
| 5 | More limited, but also more deterministic response from sensors and automation. | Better HAI, due to time to get situational awareness before action. | Y | G | Y |
| 6 | Dependence on shore control operators’ performance and situational awareness. | Always rested, but not directly in the loop. | R | Y | Y |

| **Main differentiating factors** | | **Brief description of the effects** | **A (New)** | **B (Todays)** | **C (Averted)** |
|---|---|---|---|---|---|
| 7 | Dependence on the communication link to shore. | Loss of communication may cause new accident types, but high integrity req. and clear operational design domains will help. | R | Y | Y |
| 8 | Dependence on high-quality implementation of fallback solutions and definition of minimum risk conditions for the ship. | More conservative and hence safer operational procedures. | Y | G | G |
| ***Shore control centre*** | | | | | |
| 9 | Dependence on good cooperation in the shore control centre. | Training and resource management is critical. | Y | G | R |
| 10 | The intervention crew do not have to worry about personal risk and adverse conditions on board. | May be likely to find solutions to critical problems that would otherwise be lost. | Y | G | Y |
| ***Higher technical resilience*** | | | | | |
| 11 | More technical barriers against technical faults. | In case of trouble, backup systems shall be in place. | Y | G | Y |
| 12 | Much improved technical systems with built-in predictive maintenance functionality. | Less chance of trouble. | Y | G | Y |
| 13 | Dependent on maintenance at shore. | Something may be forgotten. | R | G | Y |
| ***Improved voyage planning*** | | | | | |
| 14 | Less chance of surprises during voyage. | Better planned voyage. | Y | G | G |
| 15 | More support from other functions on shore. | Improved traffic regulation. | Y | G | G |

**Discussion:** The analysis should be seen as a cursory and qualitative analysis of the risk issues from the perspectives of researchers and experts working with the development of autonomous technology at the beginning of 2019. The participants and interviewees were experts developing autonomous technologies and researchers at the Centre for Autonomous Marine Operations and Systems (AMOS)[10] at NTNU. Many interviewees had a strong faith in the capabilities of autonomous technology (referred to as technology optimists). They also stated that a substantial effort would be made to design for safe MASS operation and that the principle of "least as safe as safe as manned shipping" would be a precondition for implementing MASS. The authors of the article tried to weigh the optimistic opinions with pessimistic ones (negative evidence of the statements) when analysing and discussing the input from the experts. Hence, the article points to both positive effects on risk by limiting human intervention and challenges this effect as a human presence within the system not only allows "human-error"-induced accidents to happen but also helps prevent others from happening (Besnard & Hollnagel, 2014). Anyhow, a bias towards the positive technical contributions from autonomous ship designs was apparent. This study's limitation became even more prominent when the results were presented at a conference on Human Factors. The presentation sparked a debate on the human

[10] https://www.ntnu.edu/amos

role in MASS (the value, relevance, and purpose) and whether the autonomous technology actually will reduce "human errors" as it not only applies to operators of the system but also to its designers and manufacturers, etc. The fact that the article used the term "human error" without quotation marks became an issue. From the discussion at the conference, it became clear that the root of much of the frustration from the opponents (mainly Human Factor experts) lies in the problems of the "what-you-look-for-is-what-you-find"-principle (Lundberg et al., 2009), the "blame culture" in accident investigations (Whittingham, 2004) and (consequently) statistics implying that "human errors cause around 80 % of maritime accidents". See Wróbel (2021) for more on why this value is a poorly documented myth on a phenomenon that is far more complex than to be solved by a straightforward analysis and answer (i.e., a number).

Undoubtedly, the "human error" and human factors will shift from the ship to the SCC for an unmanned ship (Man et al., 2015; Ramos et al., 2018). However, the high complexity and uncertainty regarding the operation of MASS favour the emergence of ambiguity around the norms and criteria to interpret or judge the accidental risks involved. Rather than trying to quantify and calculate the effect autonomous technology may have on the occurrence of "human error", the question should rather be on how to fit the human element into MASS so that this socio-technological system operates at its optimum, as postulated in the Safety-III perspective and recommended by Wróbel (2021). In this context, Article 2 provides an overview of some of the areas that need special attention in the design of MASS. Hence, the topic of future risk assessments of MASS should give particular focus to the "red" indicators from Table 7 above:

- More technology means more complexity and possibilities of technological failure
- Dependency on shore control operators' performance and situational awareness: "out of the loop" performance issues
- Dependency on sufficient cooperation, competence, and resource management in the SCC.
- The critical communication link to shore
- The SCC lacks a "personal touch" or feel of the environment, ship, and technical systems' performance
- Difficult to handle and control fire onboard
- New maintenance challenges when no operator is present onboard for inspection
- And other issues are when the operator is unable to inspect equipment or systems that report errors or problems

### 4.2.2 Article 3: Improving Safety by Learning from Automation in Transport Systems with a Focus on Sensemaking and Meaningful Human Control.

In Article 3, the aim was to investigate what we can learn from automation in transportation systems across the four transportation domains. From Article 2, we learned that the safe operation of MASS would be highly dependent on the SCC operators' performance and situation awareness. *Sensemaking* was identified as a better term for the SCC operator as it also incorporates decision making (Danielsen, 2021). In addition, the concept of Meaningful Human Control was of interest as it provides some design principles and requirements for including considerations of the "human in the loop" both during design and operation. That is, humans (supported by computers and algorithms) should ultimately remain in control and responsible for relevant decisions. The responsibility may also be on the designer and producer of the autonomous systems as well as the operator. The article summarises safety challenges and lessons learned in each transportation domain gathered through the research conducted in the SAREPTA project. The applied research method was literature reviews, interviews, and discussions within the SAREPTA project group that constituted the focus group (see section 3.5).

The main findings (across the domains) were:

- The primary safety issues are technical reliability and maturity, the need for automation transparency (including awareness of the decision made by automation), the need to define what conditions the system can operate under and assigning responsibilities to human operators and the automation.
- Regarding the human element:
    - An operator is still needed, especially when there is a disruption and sensors fail to recognise an obstacle or determine the following actions.
    - Most of the projects lack early incorporation of human factors in analysis, design, testing, and certification processes. The motivation seems to be to automate as much as possible and assume that humans will monitor it.
    - HAI and how to keep the human in the loop is often considered a challenge to be solved late in the project after knowing the limitations of the technology and by considering the humans as the adapting backup
    - If human intervention is needed to handle a scenario, sensemaking must be supported within the existing limitation of human abilities.

Regarding the design phase:

- Developing autonomous or remotely controlled transportation systems (especially for AVs and MASS) appears to primarily be a technology push rather than considering and providing socio-technical solutions, including redesigning work, capturing knowledge, and addressing human factors.
- Aviation safety is considered exceptionally high. The success can be traced to systematically automating simple tasks and reducing demands on the pilot. The development is based on the science of Human Factors, building infrastructure to control and support flights, a strong focus on learning from minor incidents and accidents, and having support from control centres that have strict control of the operational domain. The Boeing 737 MAX fatal crashes are examples of automation accidents (Cruz and de Oliveira Dias, 2020) where recommended guidelines during design and certification were not followed.
- The article briefly introduces the concept of “operation envelopes” based on the concept of “Operational Design Domain” introduced for autonomous cars by SAE J2016 (2016). The operational envelope is defined by answering which functions and roles to assign to the automation versus the human operator.
- Design principles from meaningful human control should be used to verify if the interaction between automation and the human operator is safe. This can be used as an input to operational envelopes and assist in designing a good HAI supporting sensemaking
- Well-defined operational envelopes may reduce complexity and analyse the need for cues and information to support sensemaking when needed.

**Discussion:** Article 3 covers a broad topic and all four transportation domains. There is indeed much to learn across the domains. However, each domain has different operational solutions and implications, and the results must be evaluated based on their applicability. The paper did not focus on lessons learned from a maritime perspective alone but summarised the results across the “Man Technology and Organization” (MTO) perspectives. The article mentions a few automation accidents attributed to software-related failures and overlooked dependencies among systems' technical, operational, and organisational components. This, again, can be attributed to poor design or system design requirements. As discussed in Article 3, one example is the “Human out of the loop” subsystem, MCAS-system in the Boeing 737 MAX crashes. Here, the human operators (pilots) were used to being in the loop but experienced a chaotic situation where the technical system provided little information about the failed sensor and its actions to intervene without pilot input. There are (and should, of course, be some “human out of the loop systems“ in operation, but the MCAS is safety-critical and should not have been approved without proper validation of the system design based on Human Factors Design Standards such as FAA (2003), US DOD (2012), and SAE6906 (2019), as suggested by Endsley (2019).

Drawing on the experiences from the other transportation domains, the main conclusion is that it will be extremely challenging to apply a probabilistic (quantitative) approach to risk assessments of MASS due to a number of reasons. Risk analyses such as quantitative risk analyses (QRAs) are well established in situations with considerable available performance data and clearly defined boundaries for their use. But this is not the case for MASS for several reasons. Firstly, with the current short or non-existing history of MASS, we lack experience and knowledge of the failure mechanisms and the undesirable consequences that might occur. Hence, we do not have sufficient empirical data to address the probabilities (likelihood or frequencies), and the evidence base is weak. Secondly, the complex and software-intensive technology of MASS, composed of hardware components, logic control devices and a high number of sensors, introduces invisible dynamic interactions that are challenging to model. Adding interactions with a human operator on top of the system architecture increases the complexity. Thirdly, few current risk assessment models within the maritime domain are applicable for MASS as they lack the consideration of the communication connection with a SCC, impacts from software failures on system risk, and interaction between conventional and autonomous systems (Thieme, 2018).

Faced with the above-listed challenges, accurate quantitative risk estimation is difficult to achieve, and if estimated values are achieved, the uncertainty related to these numbers will be high. Nevertheless, addressing the risk to design out potential hazards and safety issues is important (and should be prioritised) as the ability to influence the safety at this stage is high compared to later development phases. When faced with large uncertainties regarding the system behaviour due to complex interactions and little available empirical data, other ways of identifying and evaluating risk should be considered (ref. Aven (2009)).

We have experienced that the interaction between the human element and automation will be a vital issue for the safety of highly automated and autonomous systems. The HAI can be addressed by considering qualitative risk assessments in the design of MASS, looking both at the hazards related to the operation of MASS and the mitigation of hazardous events by the operator through the HMI. Putting the human operator at the centre of the risk assessment by asking what can go wrong in assessing a situation (scenario) and how undesirable events can be avoided or the consequences minimised can be considered another way of identifying and evaluating risks.

The risk assessment should then focus on the human capabilities, the HMI, and the defined tasks and responsibilities envisioned for the operator vs the autonomous system, typically presented in a CONOPS or by developing operational envelopes as a part of the conceptual design of MASS.

## 4.3 Objective 3

Recent literature on Human Automation Interaction (HAI) and System safety (Leveson, 2020b) stresses that the role of humans in systems is changing. The typical assumption that operator error is the cause of most incidents and accidents is replaced with the view that operator (human) error is a symptom of a system that needs to be redesigned (and not a cause). Within the field of MASS and risk assessments, several papers mention human or operator error and organisational weaknesses as possible or actual contributing factors to accidents (Utne et al., 2020; Wróbel et al., 2017, 2018; Zhou et al., 2020), but fewer recognise the importance of focusing on human factors to improve safety. To do something about "human error", one must look at the system in which people work: the design of equipment, the usefulness of procedures, goal conflicts and high workloads. Hence risk assessments in the design phase of SCC should also address this and consider the risks from a joint human-automation collaboration viewpoint.

Experience from other domains where remote control is applied indicates the need to design according to socio-technological principles (co-active design by including the end-user, sensemaking and meaningful human control). The identified risk could hence be addressed by taking a human-centred approach and assessing human factors perspectives on what safety barriers are needed to reduce the risk related to the remote operation of MASS.

The third objective is to propose a method for risk assessment in the design phase, including the human element. The Crisis Intervention and Operability Analysis (CRIOP) method is mentioned in DNV's (2018) Guidelines on Autonomous and remotely operated ships. The method focuses on Human Factors and has proven useful in the oil and gas industry for the design of control centres. Therefore, the method was selected to be evaluated as a risk assessment method for the design of a SCC for MASSs.

### 4.3.1 Article 4: Adopting the CRIOP Framework as an Interdisciplinary Risk Analysis Method in the Design of Remote Control Centre for Maritime Autonomous Systems

The article presents an adapted scenario analysis method based on the Scenario Analysis in the CRIOP framework developed by Johnsen et al. (2011). The article highlights the need to include end-users and carry out risk based design considering the operational quality of a Remote Operation Centre for MASS. The term was chosen for this article as one of the contributing authors emphasised that the Operation or Control Centre do not necessarily need to be located onshore (as indicated in the term *SCC*). The goal of risk based design is to use information from risk analysis to design out accidents before they occur. Having a risk based design simply means carrying out risk analysis and considering potential risk in the different phases of design and hence treating safety as a life cycle issue. A risk based approach is recommended in IMOs "Guidelines for the Approval of Alternatives and Equivalents" (2013), which is currently the principle that MASSs will be approved according to. A risk based approach is (as mentioned) also recommended by Lloyd's Register (2017) and DNV (2018) in their guidelines on the design of MASS.

In the conceptual design phase of MASS, a CONOPS, with functional requirements and operational envelopes, is defined. The CONOPS gives us some idea of the responsibilities of the automation vs the human operator (and conditions for when the responsibility changes) and makes it possible to provide an initial concept of a SCC. Article 4 evaluates the CRIOP framework as an interdisciplinary scenario-based risk analysis method in the design of a SCC for MASSs and proposes an adapted version of the Scenario Analysis. The goal of the adapted version is to identify the critical actions to be carried out by the operator, the potential hazards that may occur, evaluate the operator's ability to handle critical situations, and provide design recommendations. The analysis group must include competent participants from different disciplines (involved in the design of MASS) and the end-user, the operator. The Scenario Analysis assesses the operator's actions in response to possible scenarios. Based on the scenario, a dynamic assessment is made of the interaction between essential factors in the control centre, e.g., presentation of information and time available. The methodology suggests using

Sequentially Timed Events Plotting (STEP) diagrams for a graphic representation of the scenario events. The analysis should utilise guidewords, checklists of questions and performance shaping factors. The analysis can then identify potential error sources in the information systems, the operator's ability to achieve an adequate level of situation awareness, and whether sufficient information is available to allow the operator to make decisions when required. The result of a Scenario Analysis is a list of hazards (sources of potential harm, not limited to human errors) and design issues (both technical, operational, and organisational) related to a current prototype/concept. These are so-called weak points in the design, and the Scenario Analysis's final step is to identify measures that should be taken to improve the identified weak points.

Prior experiences suggest that CRIOP helps identify significant challenges between human operator(s) and automation, as Human Factors and "best practice" guidelines are used. Often mentioned issues in CRIOP analysis of control centres in the oil and gas industry are the ability to grasp the situation "at a glance" and simplifying automation steps to let the operator understand the action taken by the automation (i.e., explainable AI). An essential feature of MASS(s) is the dynamic levels of autonomation that may change during a voyage depending on certain conditions. Hence, the handover situations and change of responsibility between the automation and the human operator will be an essential source of risk. The adapted Scenario Analysis may provide a qualitative assessment of these risks from an interdisciplinary and human-centred perspective.

Article 4 presented the idea, but it remains to evaluate the usefulness and applicability of the adapted method. Based on the findings in/from research projects such as Autoship (with the AURA framework) and Autoferry (with the urban passenger ferry milliAmpere2), a system architecture for MASSs and a prototype of an HMI at a SCC for an unmanned passenger ferry were available. Hence it became possible to draft a case study of the adapted framework, as presented in the following section.

### 4.3.2 Article 5: Human-centred risk assessment for a land-based control interface for an autonomous vessel

Building on the research in Article 4 and following up on the recommendations for further work, Article 5 presents a case study of the adapted risk assessment method inspired by the Scenario Analysis in the CRIOP framework. Additional reviews of the method were carried out to further investigate the Scenario Analysis in light of its contributions to risk analysis and design research. In the paper, the Scenario Analysis method is evaluated as a design tool in relation to the HCD design process in ISO9241-210 (2019) and compared to the STPA method, which is suggested by many researchers as a promising risk assessment to be applied to MASS concepts (Banda et al., 2019; Thieme et al., 2018; Utne et al., 2020; Wróbel et al., 2017, 2018; Zhou et al., 2020).

The article divides the use of risk assessment in the design process into two types: formative analyses (focused on the process, e.g., to improve the quality of a design) and summative (focusing on the results of the assessment, e.g., to evaluate if a safety target is met, for validation and verification) (French et al., 2011; French & Niculae, 2005). Most risk assessments applied in the maritime domain are technical and of summative use. For instance, the regulatory framework for Risk-Based Ship Design (RBSD) was introduced with the primary objective of providing evidence on the safety level of a specific design of ships (Papanikolaou & Soares, 2009), i.e., a summative approach to risk assessment, where safety is quantified using a formalized quantitative risk analysis procedure and compared to a predefined risk acceptance criterion. A challenge to this strong focus on technical risk assessments is that it is insufficient in addressing human-automation interactions (HAIs). This is partly due to the presence of other risk dimensions (e.g., typically socio-technical ones, organisational capacity, security, etc.), involvement of various actors (such as Vessel Traffic Service centre, Emergency response, etc.), and concerns related to operational issues such as how to perform decision making under uncertainty (Aven, 2016; Goerlandt, 2020). For risk based design of MASS, the summative classical approach will also be challenging for practical use as the background knowledge - the basis for the probability models and assignments – is weak, i.e., uncertain. In the face of uncertainties, the risk assessment of MASS may be better addressed by constructing scenarios that are validated according to logical consistency, psychological empathy with the main players involved, congruence with past trends, and narrative plausibility (see Aven and Renn (2009)). This is where the Scenario Analysis can be a valuable tool as a human-centred, participatory and multidisciplinary risk assessment. However, the result of a Scenario Analysis is a list of weak points and suggestions for improvements (i.e., mitigating barriers) and not a complete characterization of a risk or a risk picture. The overall goal of a Scenario Analysis is to improve the design by enabling human-centred risk informed decision making. The adapted Scenario Analysis as a Human-centred Risk Assessment (HCRA) process is presented in the article and replicated in Figure 10 below. The HCRA comprises steps 0, 1 and 2 in Figure 10. Suggesting a new design (step 3) and evaluating the design against relevant decision making criteria in step 4 should consider criteria beyond merely safety related once. Hence, the HCRA is covering parts of a larger Human-centred risk informed design process.

*Figure 10 The iterative human-centred risk assessment approach based on the HCD process (ISO9241-210, 2019).*

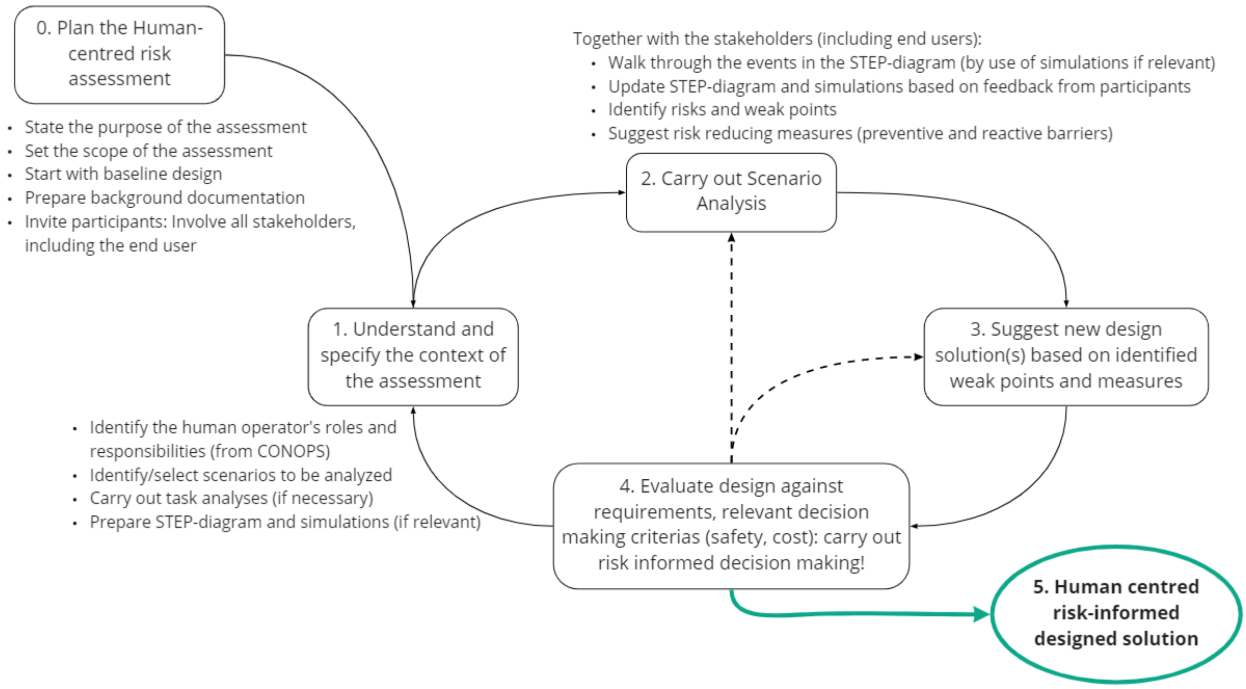

0. Plan the Human-centred risk assessment
• State the purpose of the assessment
• Set the scope of the assessment
• Start with baseline design
• Prepare background documentation
• Invite participants: Involve all stakeholders, including the end user
Together with the stakeholders (including end users):
• Walk through the events in the STEP-diagram (by use of simulations if relevant)
• Update STEP-diagram and simulations based on feedback from participants
• Identify risks and weak points
• Suggest risk reducing measures (preventive and reactive barriers)
2. Carry out Scenario Analysis
1. Understand and specify the context of the assessment
3. Suggest new design solution(s) based on identified weak points and measures
• Identify the human operator's roles and responsibilities (from CONOPS)
• Identify/select scenarios to be analyzed
• Carry out task analyses (if necessary)
• Prepare STEP-diagram and simulations (if relevant)
4. Evaluate design against requirements, relevant decision making criterias (safety, cost): carry out risk informed decision making!
5. Human centred risk-informed designed solution

*Figure 10 The iterative human-centred risk assessment approach based on the HCD process (ISO9241-210, 2019).*

A summary of the identified strengths and benefits of the adapted Scenario Analysis method:

- It is simple and easy to apply from an early design phase, combining socio-technical design principles and including risk informed decision making in the design process. The participants do not need extensive expert knowledge to facilitate the analysis, nor do they need to go through many complicated steps.
- The analysis may work as both an analytic (in analysing the human-machine interactions) and an evaluative tool (evaluating the design against requirements).
- It is cross-disciplinary and can be an arena for learning and sharing experiences.

- The method provides a common platform for understanding the operations and how the SCC operator can handle different situations. By visualizing the scenario in a simulation of the HMI and structuring the discussion to events in a STEP diagram, scenarios involving different subsystems and actors can become comprehensive and easy to understand. Which further facilitates an open discussion and brainstorming around possible risks.
- It can be a valuable tool to address the human element in risk assessment by focusing on the operators' ability to handle the situation by utilising Human Factors knowledge. Unlike traditional hazard analysis tools, the method is especially useful in identifying HAI-associated hazards.
- By involving people with experience from similar systems and including the end-users, the analysis aims to minimize the gap between Work as Imagined (WAI) and Work as Done (WAD).
- The method can be combined with the STPA and provide input to more advanced safety analysis, like the *Human System Interaction in Autonomy*-method proposed by Ramos et al. (2020).
- It can be seen as one framework to support the requirements of incorporating the "human element" in risk assessment, associating them directly with the occurrence of possible accidents, underlying causes, or influences (ref. the guidelines for FSA by IMO (2018b)).

In the case study in Article 5, we applied the method on a prototype of the HMI during the early preliminary design phase of a SCC interface for an autonomous passenger ferry. The prototype of the HMI was the first version of the initial design; hence the complexity and fidelity of the analysis were consistent with the data and information available. In the case study workshop, a scenario of a handover situation where the simulated autonomous system asks for assistance from the SCC Operator was presented. The case study was carried out on a digital platform. Twelve people, including the design and engineering team of four, attended the workshop.

A limitation of the case study was that we did not have the actual end-user present, and only one scenario was analysed. However, the validity of the method is considered high as it is acknowledged as a valuable tool in the design process and for validation and verification of Control Centres in the oil and gas industry. The validity of testing the applicability in the case study was evaluated in terms of participants' feedback and the method's ability to identify hazards, risks, and design issues (i.e., weak points). The analysis identified an extensive list of possible hazards and weak points in the case study. Moreover, the participants recognised and supported the plausibility and truthfulness of the analysis. Threats to the validity, credibility, reliability, and usefulness were also recorded and presented in Article 5.

The Scenario Analysis gave the case study workshop a necessary and efficient structure to analyse and discuss risks and mitigating measures for a SCC of autonomous ferries. The main finding in Article 5 is that this method can be a valuable tool to address the human element in risk assessment by focusing on the operators' ability to handle the situation. Hence, the analysis supports risk based design for the human control element in autonomous ferries, allowing for human in the loop-capabilities. After all, the design of an HMI supporting a safe and dynamic transition between autonomous and manual mode is a critical prerequisite for their implementation. Recommendations were made for further work on improved method guidance and the use of simulations.

**Discussion:** Article 4 proposes a method for qualitative risk assessment in the design phase, including the human element. The adapted risk assessment method is inspired by the Scenario Analysis in the CRIOP framework, and Article 5 presents a case study which evaluates the applicability of this method.

The CRIOP framework was developed primarily as a tool for Human Factor verification and validation and not primarily for risk identification and evaluation. The main focus is the ability of a control centre to safely and efficiently handle all modes of operation. It is an established, standardized scenario method primarily developed for the oil and gas industry, with its latest version dating back to 2011. It includes principals from ISO11064 "Ergonomic design of control centres" (2013) and ISO9241-210 "Ergonomics of human-system interaction - Human-centred design for interactive system" standard (2019), which is also referred to for designing layout and interfaces for a SCC for MASSs (Bureau Veritas (2019), Veitch and Alsos (2021)). These standards emphasise the need to consider the combination of humans and machines (or technology) as an overall system to be optimised within its organisational and environmental context. Human-centred design principles could improve effectiveness, efficiency, and human working conditions and reduce possible adverse effects on the safety and performance of the overall system. In a risk based design setting, taking a human-centred view on "what may go wrong" (i.e., risks) in assessing a situation/scenario in the SCC may contribute to the same improvements. A common misunderstanding of the CRIOP study is that the method only measures the risk in terms of consequences of human errors. In my research, an adapted version of the Scenario Analysis is to be considered as a human-centred risk assessment. The identified risks relate to how the overall system assesses different situations (i.e., scenarios) from the operator's perspective. The analysis involves:

- how the system is designed to handle the scenario/situation based on the defined operation envelope and CONOPS,
- how information is presented to the operator through, e.g. Explainable Artificial Intelligence (Veitch & Alsos, 2021),
- how the operator can get an overview and grasp the situation, what information is presented, what options the operator has at hand, and similar.

Hence, it is not necessarily limited to how the human may fail to observe and interpret the situation, decide on an action, and execute it, which are typically considered "human errors".

Articles 4 and 5 use the experiences of applying the CRIOP Scenario Analysis as a risk analysis tool for control centres for offshore oil and gas installation. There are essential differences between a control centre for an offshore installation and MASSs, like the navigational aspects and dynamic levels of autonomy, that bring an extra layer of complexity to the analysis. However, there is still much to learn from the experience from the oil and gas industry, especially regarding aspects of Human Factors and considerations in HMI designs. The checklists (which function as a structured interview guide to different steps of a human information processing model), the questions relating to performance shaping factors and guidewords used in the traditional CRIOP must be updated and customised for the application on SCC for MASSs. For CRIOP studies on control centres in Norway's oil and gas industry, the main source of scenarios is the Defined Situations of Hazards and Accidents (DSHAs) developed by the Petroleum Safety Authority Norway (PSA) and listed in their annual report on the "Trends in risk level (RNNP)[11]." A similar database for the maritime domain does not exist, and it should be questioned whether existing categories from accident statistics (like EMSA and AGCS applied in Article 2) are applicable. Nevertheless, it can be assumed that by systematically recording accidents and incidents involving different types of MASS, a database of DSHA for MASS could and should be established.

[11] The RNNP process has been used since 1999 to measure how the level of risk is developing in Norway's oil and gas industry. https://www.ptil.no/en/technical-competence/rnnp/

In the case study in Article 5, a semi-structured approach to the Scenario Analysis was chosen due to the early preliminary design phase of the SCC HMI. What distinguishes the adapted Scenario Analysis from the Analysis in the traditional CRIOP framework, presented in Johnsen et al. (2011), is the focus on the analysis as a risk assessment identifying hazards (especially HMI-associated ones) and mitigating measures. The scenario checklists and questions related to performance shaping factors were not systemically applied. Instead, guidewords based on these, such as “available time”, “goal conflicts”, “task allocation”, and “prioritisations”, were adopted and applied informally to help guide the brainstorming discussions on safety issues, HAI-hazards, and possible mitigating measures.

Scenario-based assessments are criticised for focusing too much on individual scenarios and lacking a systemic view. However, both bottom-up and top-down methods are necessary to address a broad risk picture. A CONOPS defines the operational aspects in the conceptual design phase and is established before a detailed system architecture. Hence, the Scenario Analysis can identify potential risks by analysing the suggested CONOPS and initial system architecture in a specific context and situation (scenario or use-cases). In the analysis, participants from different disciplines can brainstorm how the SCC should be designed from a socio-technological perspective utilising meaningful human control. The identified weak points and mitigating actions can then provide valuable input to further developing operational envelopes, CONOPS, and system architecture.

The Scenario Analysis is not systemic as it does not address an entire system with all its intended functions and applications. However, the Scenario Analysis is structured in steps of different activities. Hence, offering a systematic analysis of scenarios related to the socio-technological system that entails the autonomous ship(s), the SCC(s), and various involved actors. In order to have a systemic evaluation of the whole system, more recently developed system-theoretic methods such as STPA and FRAM are recommended (Relling et al., 2018). Nevertheless, these methods are not straightforward or easy to use and interpret (Hirata & Nadjm-Tehrani, 2019; Tian & Caponecchia, 2020).

As discussed in Article 5, we can avoid defining design needs based solely on abstraction by carrying out scenario analyses at different stages of the design process. Still, as indicated in studies by Leveson (2020b), Hollnagel (2017), Lützhöft (2004), and Sarter et al. (1997), the design process is filled with assumptions that often turn out to be unfounded in practice:

- Designers assume that users will naturally make rational and optimum decisions about automation use without needing specific training about how automation should be used.
- Users assume that the automation is more capable than it really is, do not understand its limitations, or believe that it is unnecessary and that they can do the job better than it can.
- Managers assume that automation will always produce the intended benefits, the designers have covered all the use cases and operating conditions, and mandatory use of the automation will always maximise efficiency and safety.

Discussions around these assumptions should naturally be a part of the CRIOP process. The facilitator is responsible for enlightening the participants on the possible HAI-associated risks these assumptions may contribute to.

The adapted Scenario Analysis should be considered one of several risk assessments that could be utilised to have a risk based design of MASS. It is a creative and analytic way of carrying out a qualitative approach to risk assessments of the HAI from the perspective of the human operator. The analysis should be carried out in workshops at different design stages, from preliminary/conceptual design to more detailed design and final verification. Even after a period of system operation, when operational experience and feedback are achieved, the method can be applied to assess the need for modification and investigate if the WAI vs WAD gap is sufficiently minimised.

# 5 Main Contribution – An initial framework for a Human-centred Risk Assessment

This chapter outlines the main contribution, a framework for a Human-centred Risk Assessment (HCRA) of SCCs for MASS operation. This framework can be seen as my main contribution to the SAREPTA project and aim to fulfil the main objective of the thesis, to contribute with necessary knowledge for the development of improved methods for risk assessments and mitigation in the design phase of MASS. I have chosen to call it a framework of a method, as the HCRA can be adjusted to be applied at different phases of the design and development of MASS. In other words, the HCRA is a skeletal or underlying structure of an approach to a risk assessment. The HCRA can also be called a method as it follows a systematic procedure. However, the level of detail (i.e., to which degree simulations, checklist and specific guidewords are applied) depends on the scope and design phase.

The five articles are linked through the overall research question of the thesis: "What risk assessments are useful in the design phase of MASSs?" and build-up to the main contribution: a framework for a Human-centred Risk Assessment (HCRA) of SCCs for MASS operation. The articles address and discuss different aspects of the method. Hence there is little point in engaging in further theoretical discussion at this point. The framework presented in Articles 4 and 5 is based on the Scenario Analysis from the CRIOP framework and can be considered an HCRA. However, the articles only briefly describe an initial framework for Human-centred risk informed design and the HCRA method. There is a need for improved and more comprehensive method guidance, including adjusting the method to the different phases of the design process. It is, however, necessary to clarify what a HCRA can contribute to and why further research is needed.

The HCRA method was first presented in Article 4 and further detailed in Article 5. Figure 11 below provides a flowchart (stepwise approach) of the method.

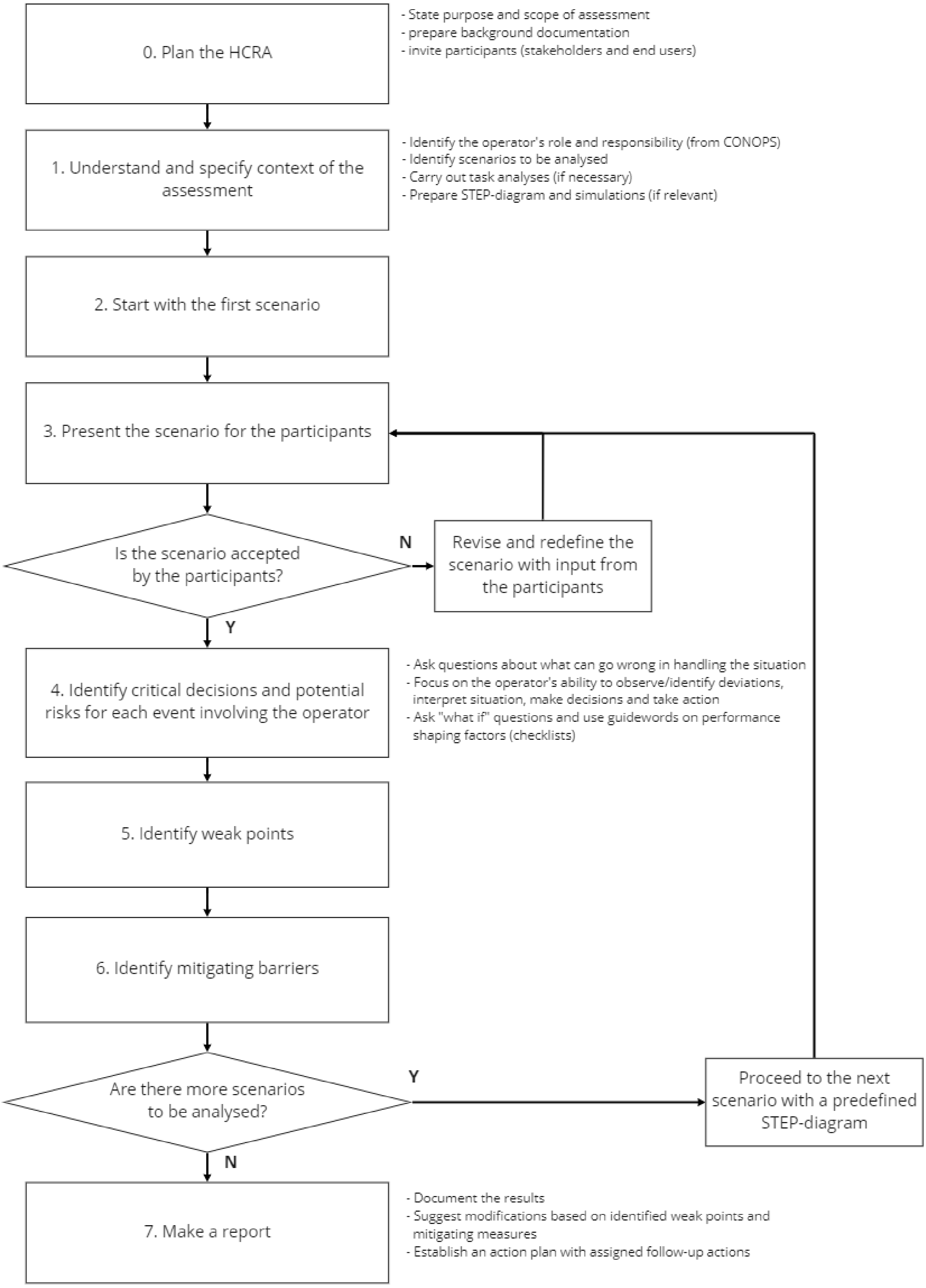

0. Plan the HCRA
- State purpose and scope of assessment
- prepare background documentation
- invite participants (stakeholders and end users)
1. Understand and specify context of the assessment
- Identify the operator's role and responsibility (from CONOPS)
- Identify scenarios to be analysed
- Carry out task analyses (if necessary)
- Prepare STEP-diagram and simulations (if relevant)
2. Start with the first scenario
3. Present the scenario for the participants
Is the scenario accepted by the participants?
N
Revise and redefine the scenario with input from the participants
Y
4. Identify critical decisions and potential risks for each event involving the operator
- Ask questions about what can go wrong in handling the situation
- Focus on the operator's ability to observe/identify deviations, interpret situation, make decisions and take action
- Ask "what if" questions and use guidewords on performance shaping factors (checklists)
5. Identify weak points
6. Identify mitigating barriers
Are there more scenarios to be analysed?
Y
Proceed to the next scenario with a predefined STEP-diagram
N
7. Make a report
- Document the results
- Suggest modifications based on identified weak points and mitigating measures
- Establish an action plan with assigned follow-up actions

*Figure 11 A flowchart of the HCRA method*

The risk analysis takes place in step 4 in the flowchart of the method. Depending on the design phase and maturity of the HMI, a checklist of questions related to performance shaping factors and specific guidewords can be systematically applied. In an early phase of the design, as in the case study of Article 5, the focus is on when and how the operator should intervene and how the initial HMI prototype will support the operator's situation awareness.

All risk assessments should start with a definition of the purpose of the analysis (Rausand, 2013). The purpose of the HCRA presented in this thesis is to contribute to the process of designing a safe and efficient HMI that supports the human operator in critical situations. In Article 5, the HCRA was placed in a more extensive iterative process to obtain a Human-centred risk informed designed solution. The process is visualised in Figure 12 below. The HCRA method (as presented in Figure 11) comprises steps 0, 1, and 2 in the figure below.

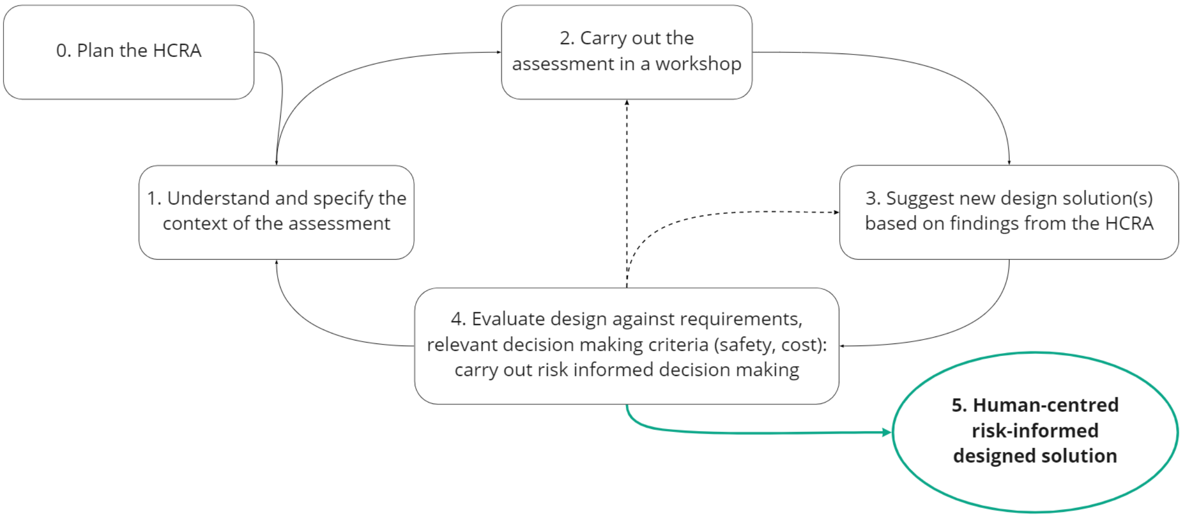

0. Plan the HCRA
2. Carry out the assessment in a workshop
1. Understand and specify the context of the assessment
3. Suggest new design solution(s) based on findings from the HCRA
4. Evaluate design against requirements, relevant decision making criteria (safety, cost): carry out risk informed decision making
5. Human-centred risk-informed designed solution

*Figure 12 The Human-centred risk informed design process, based on the HCD process (ISO9241-210 2019), adapted and adjusted from Article 5.*

A CONOPS and a prototype of the HMI, or at least a description of the task and functions envisioned for a SCC must be in place to carry out an HCRA. The method is iterative and should be applied at different phases of the design and development of MASS. The method can be considered a sort of "testing in design" method where the interactions between complex systems and the actual operator(s) are presented in a simulated environment. In this way, the designer can acquire reliable information about the nature and scope of the MASS concept and avoid the risks of irresponsible introducing a poorly designed HMI in the SCC. The risk analyst can retrieve additional information on safety performance for a risk informed decision making process and an overview of the risk knowledge among participants in the analysis. In short:

- The method is formative, focusing on improving the quality of the design by involving engineers and software developers responsible for different parts of a MASS system, HF experts, management, and end-users, in the process.
- The method takes a holistic approach to the sociotechnical system, as many exsisting risk assessments provides a fragmented risk picture when only parts of an overall system are analysed. The method helps view risk as a product of system complexity, where dependencies among technical, operational, human, and organisational elements of MASS are analysed in the context of realistic scenarios.
- The method focuses on Human factors. The method takes the perspective of the end-users and considers their contribution to and reduction of potential risks.

On an overall level, the HCRA can provide and facilitate:

- Identifying hazards and safety issues not covered by existing (technical) risk analyses.
- A pro-active design perspective where developers and designers can contribute to a design that avoids unreasonable risks and maintains human accountability by defining and redefining the responsibility of human vs automation (inspired by the principles of MHC (from van den Broek et al. (2020)).
- Discussions to understand what the ideal distribution of tasks between the human operator and the automated system would be:
    - From a technical perspective.
    - From a social and cognitive point of view, to help understand the role of the human operator in system safety (e.g. How to make sure that the operator will be able to do their part when requested).
- An arena for integrating HF early in the design and for organisational learning:
    - Describe the knowledge and lack of knowledge of the autonomous system, its performance, and interactions with the operator in different scenarios.
    - Help understand the system's novelty (e.g., new technology or new application of known technologies) by involving stakeholders, including the end user.
    - Discuss how to balance operational complexity with technical simplification.
    - Help improve the models of work throughout the design by applying the method at different stages of the design process (i.e., after the conceptual/preliminary design, the detailed design, and the built design).
    - Minimizing the gap between WAI and WAD by involving people with experience from similar systems, including the end-users, and considering the resources needed to execute the operations in the SCC.
- Supporting the requirements of integrating the "human element" in risk assessments as required in IMO's guidelines for Formal Safety Assessment (2018b) and their interim guidelines for MASS trials (2019).

The HCRA can be a valuable tool in addressing risks related to the HAI. The method suggests that the human operator is at the centre of the risk assessment. In terms of a risk assessment, this means that the operator is not only considered a source of error (similar to the malfunction of a technical component) but the operator's unsafe behaviour or action (i.e., human error) is examined in its operational context. The operators' presence is also considered a source of creative reasoning and a risk reducing capability of the system (i.e., a safety barrier). This is in line with Systems Engineering practise, where systems thinking[12] is applied to analyse the emergent properties of a system and where "human error" is examined in its context, unsafe control actions, and control mechanisms that shape human behaviour.

It is important to stress that all risk assessments have limitations and should not be used mechanically (Elms, 2019). Some limitations of the HCRA method are presented in Article 5 and further listed below:

- As the term indicates, the risk assessment focuses on HAI-related risks. The analysis may not reveal single component failures or functional failures. However, critical situations arising from such failures or unsafe control actions can be elaborated on in the forward causality analysis carried out in a STEP diagram.

[12] a holistic approach focusing on how a system's constituent parts interrelate and how systems work over time and within the context of larger systems.

- The method provides a valuable tool for discussing human behaviour in HAI. However, it does not fully represent the internal process of human cognition. The HCRA considers the human operator at a high level of abstraction. Other methods and models may be more accurate in assessing human information processing and could supplement the HCRA.
- For the HCRA, as for all risk assessments, it would simply be impossible to test all potential behaviour of a highly automated system under a wide range of possible circumstances. Due to the dynamic nature of the real-life environments in which they will operate, unpredictable outcomes are, in principle, always possible.
- Applying the HCRA requires a thorough knowledge of the systems being studied. Therefore, the quality of the results of this method is dependent on the amount of expert input (their expertise and experience).
- MASS concepts are in some way revolutionary, involving aspects of new operating paradigms and emerging regulatory, liability and security concerns (Mallam et al., 2020). Hence imagining work that is not currently done is challenging. In addition, there is limited knowledge about what types of requirements the operators of a MASS will have in terms of skills and training.
- A functional or operational approach to risk assessments is recommended by many researchers (Zhou et al., 2020). The HCRA can be considered an operational approach but may not be as systematic as a functional approach that breaks down the operation into a number of functions and task and assess each of them in terms of a functional hazard analysis.
- A coarser method is suggested because there is limited knowledge about MASS design solutions. As pointed out by Kari and Steinert (2021), “the working environment in the SCC is completely different from the traditional onboard bridge” (p.17), and accident reports in the maritime domain are not concerned with remote control.

# 6 Conclusion and Further Work

Where do we stand, and where do we go from here? This chapter reflects on the thesis's main research findings and contributions by revisiting the research questions raised in the introductory chapter (Chapter 1). The overall research process is summarised. With the final perspective on the scientific and practical implications of this PhD thesis, some direction for further research is pointed out.

## 6.1 Revisiting the research questions

**Research question #1:** *What types of risk assessments are suggested for the design phase of autonomous, unmanned or remotely controlled ships today?*
***1. a*)** What are the main issues and challenges of the risk assessment methods when identifying and addressing the accidental risks of MASS?
**1. b)** How is the human element included in these risk assessment methods?

Article 1 presented a literature review where the scope was guided by the need for a state-of-the-art on research within the field of risk assessment and MASS. The review revealed a few initial frameworks for risk assessment of MASS. Each of the methods represents a different investigative angle to the same underlying issue: how can we identify and assess the risk of MASS and its operation? Of the five ways identified, two specifically mention the remote operator (the human element) and their effect on the risk picture.

The main issues of the applicability of the risk assessment methods are the limited available operational data, knowledge of the system architecture, the role and responsibility of the operator and the high complexity introduced by software. All these aspects make the traditional risk analysis methods challenging for practical use.

With the limited findings on applicable risk assessment methods in Article 1 and the challenges of applying traditional risk assessment methods to MASS concepts, the next step was to look at the differencing factors between MASSs and conventional manned ships. We know that if autonomous unmanned ships are to become a success, they must be at least as safe as manned shipping. Thus, we must ask: how safe is then manned shipping today? And can autonomous shipping technology improve accident statistics? This is the background for the second research question addressed in Article 2.

**Research question #2:** *In the design of MASS: What will the main accidental risks be, and how can they be mitigated?*

***2. a)*** *What are the differentiating factors between MASS and conventional manned ships? How will the autonomous technology applied in MASS affect the known accidental risks in the maritime domain and what potential new risks will be introduced?*

One of the differentiating factors identified in Article 2 was the constrained autonomy, i.e., relying on intervention from a SCC to handle situations the autonomous ship cannot handle. The potential contribution to new risks is related to the shore control operators' performance and situational awareness, which typically manifest themselves in "out of the loop" performance issues due to poorly designed HMI. Article 2 positions HAI as a central factor that must be addressed in future risk assessments and design of MASS.

The limited operational experience with MASS and the chosen explorative research approach led/directed me to look outside the maritime domain to other transportation domains and their experiences with highly automated systems. Hence question 2. b was drafted:

***2. b)*** *What are the experienced risks from the operation of autonomous, unmanned or remotely controlled* transportation systems today? Are these risks applicable to MASS?

Article 3 presents lessons learned on safety issues from other domains and suggests incorporating design principles from the concept of Meaningful Human Control. That is, humans (supported by computers and algorithms) should ultimately remain in control and responsible for relevant decisions. The findings indicate that the development of autonomous or remotely controlled transportation systems (especially for AVs and MASS) appears to primarily be a technology push rather than considering and providing socio-technical solutions, including redesigning work, capturing knowledge, and addressing human factors. The primary safety issues across the domains are technical reliability and maturity, the need for automation transparency (including awareness of the decision made by automation), the need for defining what conditions the system can operate under, and assigning responsibilities to human operators and the automation. A risk assessment in the design phase should hence analyse how these issues and especially the assigned responsibilities (between automation and the human element/operator) will be handled during operation and especially during safety-critical situations.

From Article 3 and the initial research, it became clear that the human operator will be a critical element, often representing the final and most important barrier against accident occurrence. However, few risk assessment methods address the SCCO in the design of MASS today (Veitch & Alsos, 2022), and the classical technical risk assessment methods are insufficient to address human-automation interactions (Goerlandt, 2020). Hence, the idea of finding or developing an integrative approach combining risk assessment in the design phase of MASS with the need for socio-technical design principles and Human Factors arose. The third research question was outlined.

***Research question #3: How can the human element be integrated into risk assessment in the design phase of MASS?***

***3. a)*** *Are there any risk assessment methods focusing on the HAI suggested for MASS?*

The sub-research question 3.a is partly covered by RQ #1. However, the review in Article 1 was carried out in 2018, and in the later years, we have seen an increase in publications on the topic. For instance, the *Human System Interaction in Autonomy* method was suggested by Ramos et al. in 2020. This method focuses on human cognition modelling and human error propagation, making it an intricate, time- and resource-consuming method, as criticised by Endsley et al (2015) and Shneiderman (2020). Another method including the human element is the STPA, which is suggested by many as a promising method to be applied to MASS concepts (Banda et al. 2019; Thieme et al. 2018; Utne et al. 2020; Wróbel et al. 2017, 2018; Zhou et al. 2020). However, the method requires a hierarchical safety control structure, both on technical and organisational design, making the analysis complex and involving many steps that are not easy to follow or understand. Hence, both methods require a high level of system knowledge and method expertise, making them of limited value in an early design phase when developing an HMI for a SCC. RQ3.a is further covered by Article 5.

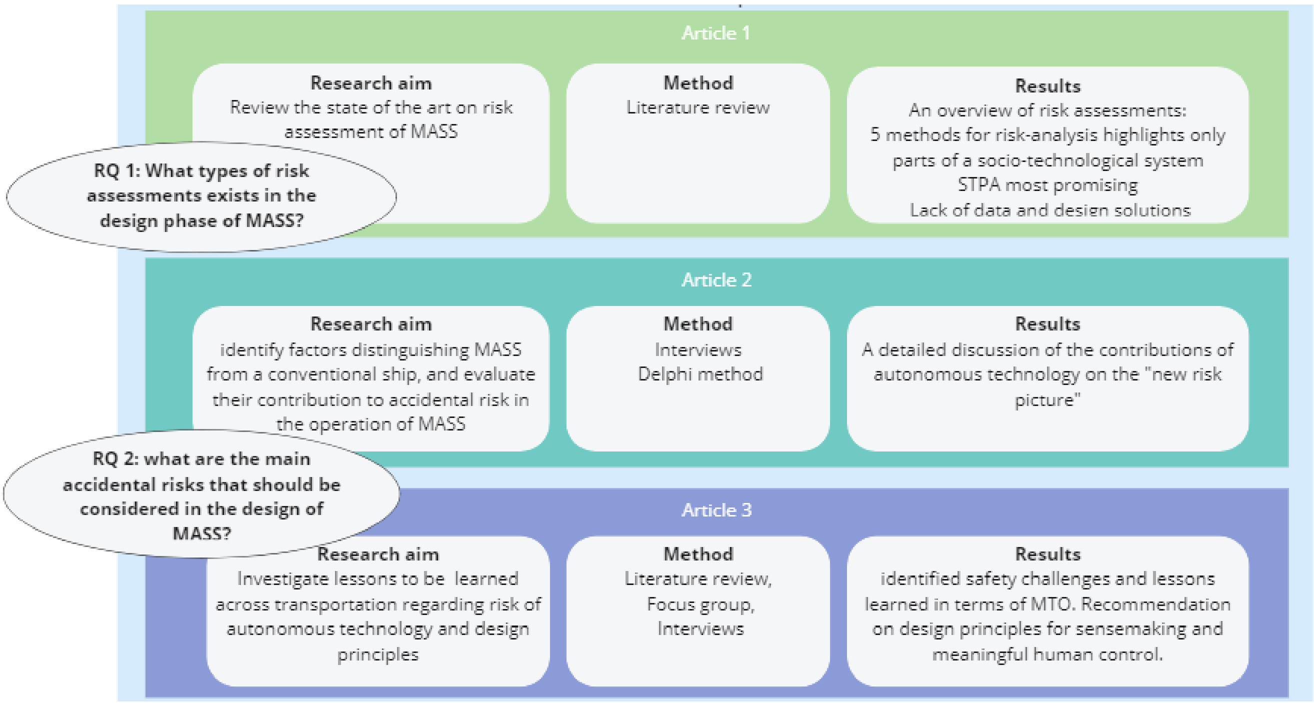
***3. b)*** *What could be a good risk assessment method for identifying and assessing HAI-related risk in the design phase of MASS?*

The sub-research question 3.b is first addressed by Article 4, which proposes using an adapted risk assessment method inspired by the Scenario Analysis in the CRIOP framework. CRIOP is recommended as a risk analysis method focusing on human aspects in DNVs Guidelines on Autonomous and remotely operated ships (p. 31, 2018). The method is an established, standardized scenario method primarily developed and used by the oil and gas industry. It is mainly used to verify and validate a remote control centre's ability to safely and efficiently handle all operational modes. The adapted Scenario Analysis is considered a Human-centred Risk Assessment (HCRA), and the proposed framework is considered the main contribution of the PhD research.

Article 5 follows the recommendations from Article 4 and attempts to validate the HCRA method in a case study on an actual prototype of an HMI in a SCC for an autonomous passenger ferry. Additional reviews of the method were carried out to further address RQ #3 and examine the Scenario Analysis considering its contributions to risk analysis and design research.

## 6.2 Overall research process

The overall research process of the thesis is illustrated in Figures 13, 14 and 15 below. The first part was a background study to establish the state of the art on risk assessment of MASS and the need to broadly define the area of interest for further research. This is illustrated in Figure 13 below.



Article 1

Research aim
Review the state of the art on risk assessment of MASS

Method
Literature review

Results
An overview of risk assessments:
5 methods for risk-analysis highlights only parts of a socio-technological system
STPA most promising
Lack of data and design solutions

RQ 1: What types of risk assessments exists in the design phase of MASS?

Article 2

Research aim
identify factors distinguishing MASS from a conventional ship, and evaluate their contribution to accidental risk in the operation of MASS

Method
Interviews
Delphi method

Results
A detailed discussion of the contributions of autonomous technology on the "new risk picture"

RQ 2: what are the main accidental risks that should be considered in the design of MASS?

Article 3

Research aim
Investigate lessons to be learned across transportation regarding risk of autonomous technology and design principles

Method
Literature review,
Focus group,
Interviews

Results
identified safety challenges and lessons learned in terms of MTO. Recommendation on design principles for sensemaking and meaningful human control.

*Figure 13 Summary of Article 1-3*

Based on the intermediate findings from the research conducted Article 1-3, the following main areas shown in Figure 14 were selected for further research/scope of work. In the intersection of the three main areas an integrative approach combining risk assessment in the design of a SCC including socio-technical design principles and Human Factors can be found.

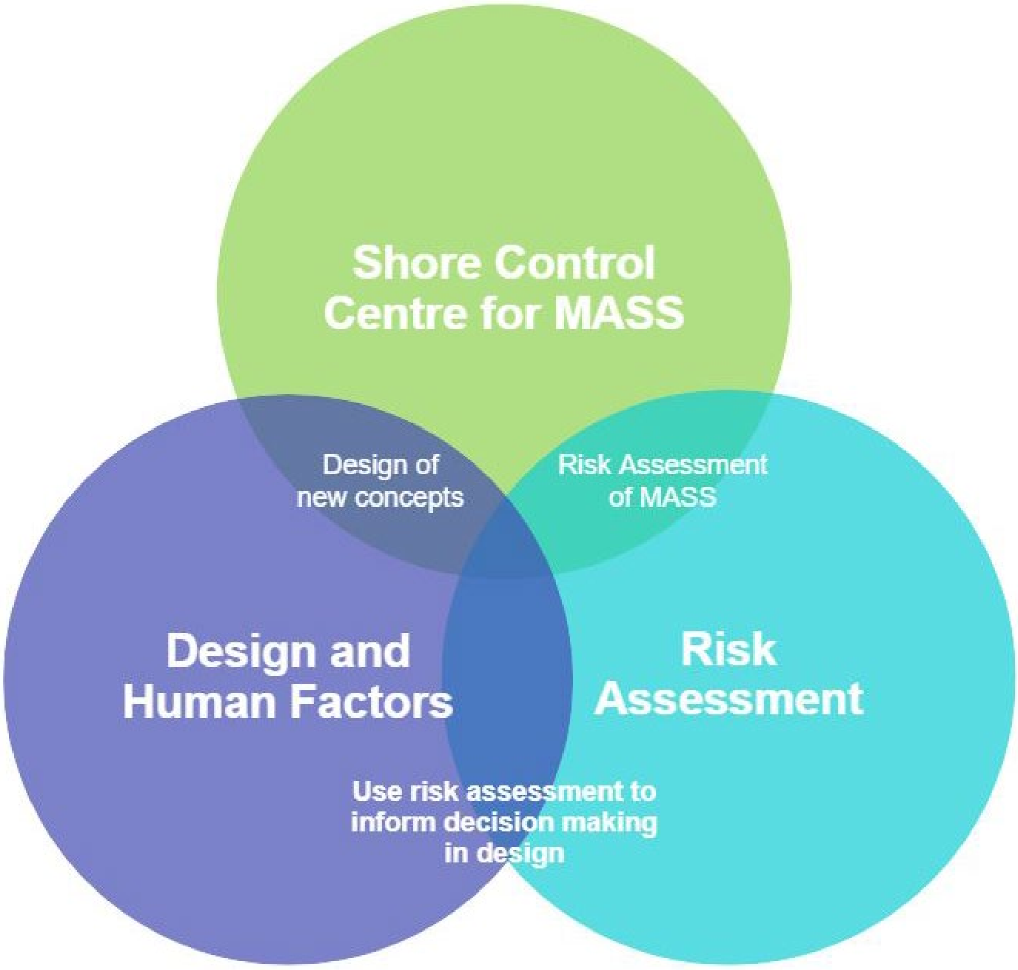
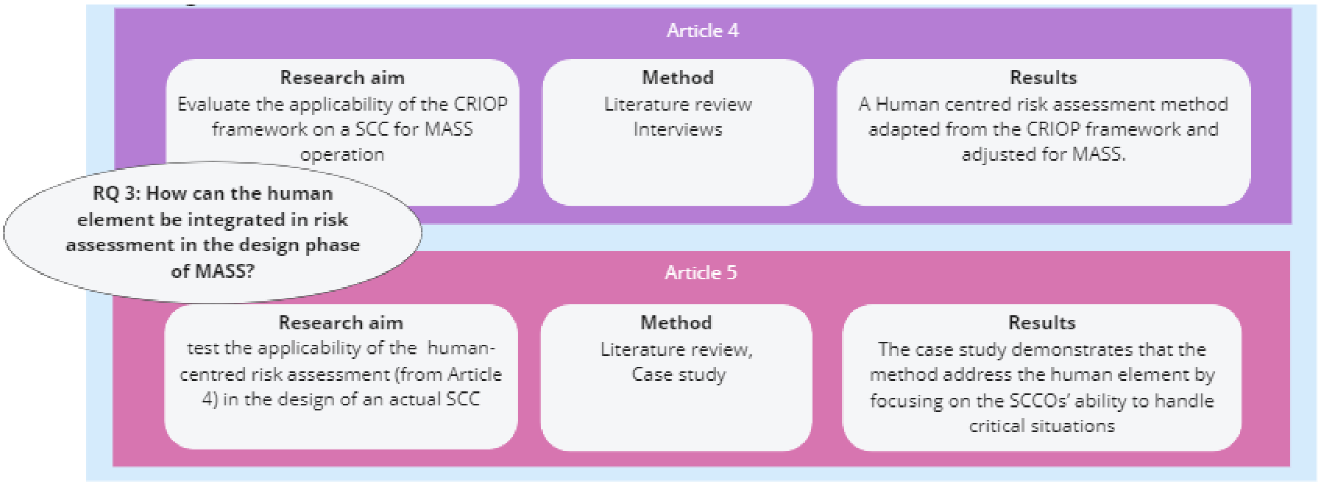

Shore Control Centre for MASS
Design of new concepts
Risk Assessment of MASS
Design and Human Factors
Risk Assessment
Use risk assessment to inform decision making in design

*Figure 14 The three main areas of research covered in the thesis*

The second part of the research process (Figure 15) involved testing out theories and concept development. The CRIOP framework was identified as a potential risk assessment method focusing on human aspects and hence selected for further investigation.



Article 4

**Research aim**
Evaluate the applicability of the CRIOP framework on a SCC for MASS operation

**Method**
Literature review
Interviews

**Results**
A Human centred risk assessment method adapted from the CRIOP framework and adjusted for MASS.

**RQ 3: How can the human element be integrated in risk assessment in the design phase of MASS?**

Article 5

**Research aim**
test the applicability of the human-centred risk assessment (from Article 4) in the design of an actual SCC

**Method**
Literature review,
Case study

**Results**
The case study demonstrates that the method address the human element by focusing on the SCCOs' ability to handle critical situations

*Figure 15 Summary of Article 4 and 5*

The thesis and the associated articles contribute to the field of risk assessment and to the design practice of MASS concept. The following sections summarise the scientific implication for each field.

## 6.3 Scientific Implications

The theoretical relevance of the PhD work is sought ensured by building on previous research within risk science, human automation interaction, human-centred design, and human factor theory, and by exploring what risk assessments focusing on the human element are useful in the design phase of MASS. The effect of the PhD research is in this section considered from a theoretical contribution to risk science and to the design practice of HMI in SCC.

### 6.3.1 Theoretical contribution to risk science

Contemporary risk science acknowledges that different approaches and methods can be used to measure, describe, or characterize risk (Aven & Renn, 2010). This thesis's main contribution, the Human-centred Risk Assessment of SCC for MASS operation, can be seen as a contribution to contemporary risk science. The human-centred scenario-based risk assessment is one way of assessing the risk by identifying hazards, evaluating how critical scenarios can be handled and henceforth identifying weak points in the design. The identified weak points can be design flaws in the HMI, but the method also addresses resilient aspects of the operation by highlighting weak points in the design and performance issues such as lack of training, missing procedures or other organisational aspects affecting the HAI.

The concept of integrative thinking was used as a method to explore the gap between engineering risk assessment and the need for human-centred risk informed design. As pointed out by Aven (2022), the idea presented here is that in order for risk assessment to be a solid and useful method for supporting risk informed decision making, a shift in the perspective from accurate risk estimation to knowledge and lack of knowledge characterisations is needed. Hence a qualitative risk assessment is suggested for the design of a SCC.

The terms *formative* and *summative analysis* are not commonly used within the field of risk science. The use of risk analysis can be broadly defined in the two ways, depending on the purpose of the analysis. QRA is a typical summative risk analysis, where the results in terms of probabilistic risk estimates are used to evaluate a designed solution against a risk acceptance criterion. In a design process where the goal is to “design out” potential risks and accidents, a formative use of risk assessment may provide a better way of recognising and roughly ranking the potential for improving safety issues (beyond system or component failure). The HCRA method may have its biggest advantage used as a formative assessment in the design phase.

### 6.3.2 Contribution to the design practice of HMI in SCC

Applying the HCRA in the design phase of SCC can contribute to a risk based design. In design research, like the HCD approach to interactive systems development, as described in ISO9241-210 (2019), the goal is an optimal HMI in terms of effectiveness, efficiency and satisfaction. Here, identifying hazards, estimating their occurrence, and designing safety controls for mitigating the risks are not necessarily considered design activities. However, In the literature on design methodologies for SCC for MASS, risk based design is identified as the most common approach among studies presenting practical design approaches to MASS systems (Veitch & Alsos, 2022). This is further supported by the recommendations from The Norwegian Maritime Authority (NMA, 2020) and classification societies, such as Bureau Veritas (2019), ClassNK (2020), DNV (2018) and Lloyd’s Register (2017).

## 6.4 Practical implications for the Maritime industry

Much effort is focused on the development of algorithms and control regimes for MASS (Thieme, 2018). However, these efforts should be accompanied by an assurance that the programmed algorithms/software follows principles of MHC and sociotechnical design principles. The strong focus on the technical aspects of MASS is running a risk of missing the critical human element, the operator in the SCC.

Today, risk assessments are mainly of summative use in the maritime industry. Typically to verify the capabilities and performance of the technology, demonstrate that a safety target is met, or performance standard is fulfilled for approval and verification. At the time being, MASS concepts are considered alternative designs and will have to be approved according to principles for "Alternatives and Equivalents" as outlined in IMO (2013). This is a risk based approach where the approval of the design is based on risk analysis identifying hazards, potential safeguards (barriers) and evaluation of the risk and risk control options by quantification at different phases of the design process. In the guideline, the design process and the risk assessment process are considered two separate and parallel tasks. However, new guidelines for the approval of MASS should acknowledge that QRAs have major limitations and that qualitative risk assessments can provide a better basis for risk based design.

On the regulation side, IMO refers to the FSA framework as the premier scientific and systematic way to assess risk. The guideline on FSA (IMO, 2018b) states that the human element is one of the most important contributory aspects to the causation and avoidance of accident, and that appropriate techniques for incorporating human factors should be used. As presented in Article 5, the HCRA covers many of the steps in the FSA by identifying hazards, events, and conditions that may lead to an accident or incident, how these may lead to different consequences, and suggesting measures to limit the impact by focusing on the capabilities of the operator.

The HCRA method is neither the most advanced nor the most systematic. It represents an opportunity for enabling risk informed decision making and is inspired by the body of human factors research that can be easily used by designers, engineers and system developers of all backgrounds. It aims at being practical, straightforward and safety focused and applicable at early phases of design. The method may be best used by interdisciplinary teams, where designers, managers, human factors experts and engineers of various backgrounds can collaborate.

## 6.5 Further work

The future is uncertain, and that is a precondition when we talk about risk concerning the future of sociotechnical systems. A risk assessment is performed to structure or knowledge about the system in order to perform decision making. It is, however, not only the future performance of the system that is uncertain. The results of risk assessments are uncertain. The risk assessment must be based on realistic assumptions about the system in its future context and operations (Johansen & Rausand, 2014). But also, the quality of the models and methods and the justification of the knowledge claims in the assessment can compromise the credibility of the assessment. In this context, more research is necessary to develop the HCRA.

A more detailed description of the adapted method and guidelines for applying the method should be developed. Practical guidelines for the application of the method at different design phases and for different MASS concepts (urban passenger ferries, cargo vessels, manned ships, etc.), where different types of SCC and considerations are involved. The method guidance should be based on a review of the performance shaping factors and the questions related to the Simple Model of Cognition (SMoC) by Hollnagel (1998), as presented in Table 5.4.1 – 5.4.4 in the CRIOP Report (2011). Additional factors and checklists related to HMI and fallback situations in other domains, such as aviation and automated driving, should be assessed for their applicability to and relevance for a SCC. Further research is also needed if a predefined list of key scenarios is to be made. Article 4 suggest some key scenarios (such as handover situations), but additional scenarios should be considered. For example, scenarios inspired by the Defined Situations of Hazard and Accidents (DHSA) applied in CRIOP studies of control centres for oil and gas installations.

On a higher level, research is needed to shed light on how people will work in a SCC, especially in safety-critical situations. There are still uncertainties related to how humans in a Maritime Autonomous Ship System will work together.

# 7 References

Adams, A., & Cox, A. L. (2008). *Questionnaires, in-depth interviews and focus groups*. Cambridge University Press.

AGCS, A. G. C. a. S. (2018). *Safety and shipping Review 2018 – an annual review of trends and developments in shipping losses and safety*.

Albert, B., & Tullis, T. (2013). *Measuring the user experience: collecting, analyzing, and presenting usability metrics, 2nd Edition*. Morgan Kaufmann, Waltham, MA, USA. https://doi.org/https://doi.org/10.1016/B978-0-12-415781-1.00017-0

APA, A. P. A. (2022a). *Online dictionary* https://dictionary.apa.org/focus-group

APA, A. P. A. (2022b). *Online dictionary* https://dictionary.apa.org/open-ended-interview

Archer, B. (1995). The nature of research. *Co-Design Journal*, *2*(11), 6-13.

Association, I. E. (2020). *What Is Ergonomics (HFE)?* Retrieved November 9 from https://iea.cc/what-is-ergonomics/

Aven, T. (2009). Perspectives on risk in a decision-making context–review and discussion. *Safety science*, *47*(6), 798-806.

Aven, T. (2012). The risk concept—historical and recent development trends. *Reliability Engineering & System Safety*, *99*, 33-44. https://doi.org/https://doi.org/10.1016/j.ress.2011.11.006

Aven, T. (2014). What is safety science? *Safety science*, *67*, 15-20. https://doi.org/https://doi.org/10.1016/j.ssci.2013.07.026

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, *253*(1), 1-13.

Aven, T. (2022). A risk science perspective on the discussion concerning Safety I, Safety II and Safety III. *Reliability Engineering & System Safety*, *217*, 108077. https://doi.org/https://doi.org/10.1016/j.ress.2021.108077

Aven, T., & Krohn, B. S. (2014). A new perspective on how to understand, assess and manage risk and the unforeseen. *Reliability Engineering & System Safety*, *121*, 1-10.

Aven, T., & Renn, O. (2009). The role of quantitative risk assessments for characterizing risk and uncertainty and delineating appropriate risk management options, with special emphasis on terrorism risk. *Risk Analysis: An International Journal*, *29*(4), 587-600.

Aven, T., & Renn, O. (2010). *Risk management and governance: Concepts, guidelines and applications* (Vol. 16). Springer Science & Business Media.

Aven, T., & Zio, E. (2011). Some considerations on the treatment of uncertainties in risk assessment for practical decision making. *Reliability Engineering & System Safety*, *96*(1), 64-74.

Aven, T., & Zio, E. (2014). Foundational issues in risk assessment and risk management. *Risk analysis*, *34*(7), 1164-1172.

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

Banda, O. A. V., Kannos, S., Goerlandt, F., van Gelder, P. H., Bergström, M., & Kujala, P. (2019). A systemic hazard analysis and management process for the concept design phase of an autonomous vessel. *Reliability Engineering & System Safety*, *191*, 106584.

Besnard, D., & Hollnagel, E. (2014). I want to believe: some myths about the management of industrial safety. *Cognition, Technology & Work*, *16*(1), 13-23.

Bjerga, T., & Aven, T. (2015). Adaptive risk management using new risk perspectives–an example from the oil and gas industry. *Reliability Engineering & System Safety*, *134*, 75-82.

Blanchard, B. S., & Fabrycky, W. (2013). Systems Engineering and Analysis: Pearson New International. In: Aufl. Harlow: Pearson Education Limited.

Blom, H. A., Everdij, M. H., & Bouarfa, S. (2016). Emergent Behaviour. *Complexity science in air traffic management*(Routledge), 83-104.

Boring, R. L., Hendrickson, S. M. L., Forester, J. A., Tran, T. Q., & Lois, E. (2010). Issues in benchmarking human reliability analysis methods: A literature review. *Reliability Engineering & System Safety*, *95*(6), 591-605. https://doi.org/https://doi.org/10.1016/j.ress.2010.02.002

Boylan, M., Coldwell, M., Maxwell, B., & Jordan, J. (2018). Rethinking models of professional learning as tools: a conceptual analysis to inform research and practice. *Professional Development in Education*, *44*(1), 120-139. https://doi.org/10.1080/19415257.2017.1306789

Breinholt, C., Ehrke, K.-C., Papanikolaou, A., Sames, P. C., Skjong, R., Strang, T., Vassalos, D., & Witolla, T. (2012). SAFEDOR–the implementation of risk-based ship design and approval. *Procedia-Social and Behavioral Sciences*, *48*, 753-764.

Bureau Veritas (2019). *Guidelines for autonomous shipping* https://marine-offshore.bureauveritas.com/ni641-guidelines-autonomous-shipping

Bureau Veritas, B. (2019). NI641 R01, Guidelines for Autonomous Shipping. In (Vol. October 2019).

Calhoun, S., & Stevens, S. (2003). Human factors in ship design. *Ship design and construction*, *2*, 1-27.

ClassNK. (2020). Guidelines for Automated/Autonomous Operation of ships – Design development, Installation and Operation of Automated Operation Systems/Remote Operation Systems. In.

Cooper, H. (2015). *Research synthesis and meta-analysis: A step-by-step approach* (Vol. 2). Sage publications.

Costa, N. A. (2016). *Human centred design for maritime safety: A user perspective on the benefits and success factors of user participation in the design of ships and ship systems*. Chalmers Tekniska Hogskola (Sweden).

Creswell, J. W. (2014). *A concise introduction to mixed methods research*. SAGE publications.

Cross, N. (2007). Designerly ways of knowing. Board of international research in design. *Basel: Birkhiuser*, *41*.

Danielsen, B. (2021). Making Sense of Sensemaking in High-Risk Organizations. In *Sensemaking in Safety Critical and Complex Situations* (pp. 53-70). CRC Press.

Denyer, D., & Tranfield, D. (2009). Producing a systematic review.

DNV, G. (2018). Autonomous and remotely operated ships. *Class Guideline DNVGL-CG-0264, https://rules. dnvgl. com/docs/pdf/DNVGL/CG/2018-09/DNVGL-CG-0264. pdf*.

DOD, U. D. o. D. (2012). MIL-STD-1472G. In *Design Criteria Standard: Human Engineering*.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science education*, *84*(3), 287-312.

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,

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of management review*, *14*(4), 532-550.

Elms, D. (2019). Limitations of risk approaches. *Civil Engineering and Environmental Systems*, *36*(1), 2-16.

EMSA. (2018). Annual overview of marine casualties and incidents. In: European Maritime Safety Agency Lisbon.

Endsley, M. (2019). Human Factors & Aviation Safety, Testimony to the United States House of Representatives Hearing on Boeing 737-Max8 Crashes. *Human Factors and Ergonomics Society, December*, *11*.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human factors*, *37*(2), 381-394.

European Commission. (2019). *Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Evans, J. H. (1959). Basic design concepts. *Journal of the American Society for Naval Engineers*, *71*(4), 671-678.

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 Engineering*, *202*, 107188. https://doi.org/https://doi.org/10.1016/j.oceaneng.2020.107188

Frankel, L., & Racine, M. (2010). The complex field of research: For design, through design, and about design.

Franzoni, C., & Stephan, P. (2021). *Uncertainty and Risk-Taking in Science: Meaning, Measurement and Management*.

French, S., Bedford, T., Pollard, S. J., & Soane, E. (2011). Human reliability analysis: A critique and review for managers. *Safety science*, *49*(6), 753-763.

French, S., & Niculae, C. (2005). Believe in the model: mishandle the emergency. *Journal of Homeland Security and Emergency Management*, *2*(1).

FAA, F. A. A. (2003). Human Factors Design Standard (HF-STD-001). In. Atlantic City International Airport, NJ: Federal Aviation Administration William J. Hughes Technical Center.

Giacomin, J. (2014). What is human centred design? *The Design Journal*, *17*(4), 606-623.

Gill, K. S. (1996). The Foundations of Human-centred Systems. In K. S. Gill (Ed.), *Human Machine Symbiosis: The Foundations of Human-centred Systems Design* (pp. 1-68). Springer London. https://doi.org/10.1007/978-1-4471-3247-9_1

Goerlandt, F. (2020). Maritime Autonomous Surface Ships from a risk governance perspective: Interpretation and implications. *Safety science*, *128*, 104758.

Goerlandt, F., & Montewka, J. (2015). Maritime transportation risk analysis: Review and analysis in light of some foundational issues. *Reliability Engineering & System Safety*, *138*, 115-134.

Grech, M., Horberry, T., & Koester, T. (2008). *Human factors in the maritime domain*. CRC press.

Grech, M. R., & Lutzhoft, M. (2016). Challenges and opportunities in user centric shipping: Developing a human centred design approach for navigation systems. Proceedings of the 28th Australian Conference on Computer-Human Interaction,

Greenwich, U. o. (2022). *Formative vs Summative*. Retrieved March 11 from https://www.gre.ac.uk/learning-teaching/assessment/assessment/design/formative-vs-summative

Hancock, B., Ockleford, E., & Windridge, K. (2001). *An introduction to qualitative research*. Trent focus group London.

Hansson, S. O. (2013). 4. Defining Pseudoscience and Science. In P. Massimo & B. Maarten (Eds.), *Philosophy of Pseudoscience: Reconsidering the Demarcation Problem* (pp. 61-78). University of Chicago Press. https://doi.org/doi:10.7208/9780226051826-005

Heikoop, D. D., Hagenzieker, M., Mecacci, G., Santoni De Sio, F., Calvert, S., Heikoop, D., & Calvert, S. (2018). Meaningful Human Control over Automated Driving Systems. 6th Humanist Conference,

Helander, M. (2005). *A guide to human factors and ergonomics*. CRC press.

Herkert, J., Borenstein, J., & Miller, K. (2020). The Boeing 737 MAX: Lessons for Engineering Ethics. *Science and Engineering Ethics*, *26*(6), 2957-2974. https://doi.org/10.1007/s11948-020-00252-y

Hirata, C., & Nadjm-Tehrani, S. (2019). Combining GSN and STPA for Safety Arguments. International Conference on Computer Safety, Reliability, and Security,

Hoem, Å., Johnsen, S., Fjørtoft, K., Rødseth, Ø., Jenssen, G., & Moen, T. (2021). Improving Safety by Learning from Automation in Transport Systems with a Focus on Sensemaking and Meaningful Human Control. In *Sensemaking in Safety Critical and Complex Situations* (pp. 191-207). CRC Press.

Hoem, Å. S., Fjørtoft, K. E., & Rødseth, Ø. J. (2019). Addressing the Accidental Risks of Maritime Transportation: Could Autonomous Shipping Technology Improve the Statistics? *TransNav, the International Journal on Marine Navigation and Safety of Sea Transportation*, *13*. https://doi.org/10.12716/1001.13.03.01

Hollnagel, E. (1998). *Cognitive reliability and error analysis method (CREAM)*. Elsevier.

Hollnagel, E. (2000). Looking for errors of omission and commission or The Hunting of the Snark revisited. *Reliability Engineering & System Safety*, *68*(2), 135-145.

Hollnagel, E. (2014). Is safety a subject for science? *Safety science*, *67*, 21-24.

Hollnagel, E. (2017). Can we ever imagine how work is done. *HindSight*, *25*, 10-13.

Hollnagel, E. (2018). *Safety-I and safety-II: the past and future of safety management*. CRC press.

Hollnagel, E., Wears, R. L., & Braithwaite, J. (2015). From Safety-I to Safety-II: a white paper. *The resilient health care net: published simultaneously by the University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia*.

Huang, Y., Chen, L., Negenborn, R. R., & Van Gelder, P. (2020). A ship collision avoidance system for human-machine cooperation during collision avoidance. *Ocean Engineering*, *217*, 107913.

Haavik, T. K. (2021). Debates and politics in safety science. *Reliability Engineering & System Safety*, *210*, 107547.

IACS, I. A. o. C. S. (2019). *2019 IACS Annual Review*. IACS. https://iacs.org.uk/news/2019-iacs-annual-review/

MSC.1/Circ.1455 Guidelines for the Approval of Alternatives and Equivalents as provided for in Various IMO Instruments, (2013).

IMO. (2018a). *IMO takes first steps to address autonomous ships*. Retrieved January 8 from https://www.imo.org/en/MediaCentre/PressBriefings/Pages/08-MSC-99-MASS-scoping.aspx

MSC/Circ.12 Revised Guidelines for Formal Safety Assessment (FSA) for the use in the IMO Rule-Making Process, (2018b).

Interim guidelines for MASS trials, (2019).

IMO. (2021). *Outcome of the Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS) (No. MSC.1/Circ.1638).*

Infinity, O. (2020). *A one-world view of remote operations at sea in a real-time, digital environment.* Retrieved December 18 from https://oceaninfinity.com/marine-robotics/

ISO9241-210. (2019). Ergonomics of Human-system Interaction: Part 210: Human-centred Design for Interactive Systems. In: International Organization for Standardization.

ISO11064. (2013). Ergonomic design of control centres In: International Organization for Standardization.

ISO/IEC31010. (2019). Risk management — Risk assessment techniques. In. Switzerland, Geneva: International Organization for Standardization.

ISO/IEC. (2014). GUIDE 51 Safety aspects - Guidelines for their inclusion in standards. In *Third edition*.

Janssen, C. P., Donker, S. F., Brumby, D. P., & Kun, A. L. (2019). History and future of human-automation interaction. *International Journal of Human-Computer Studies*, *131*, 99-107.

Johansen, I. L., & Rausand, M. (2014). Defining complexity for risk assessment of sociotechnical systems: A conceptual framework. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *228*(3), 272-290. https://doi.org/10.1177/1748006x13517378

Johansen, I. L., & Rausand, M. (2015). Ambiguity in risk assessment. *Safety science*, *80*, 243-251.

Johnsen, S. O., Bjørkli, C., Steiro, T., Fartum, H., Haukenes, H., Ramberg, J., & Skriver, J. (2011). *CRIOP: A scenario method for Crisis Intervention and Operability analysis.* http://www.criop.sintef.no/The CRIOP report/CRIOPReport.doc

Johnsen, S. O., & Porathe, T. (2021). *Sensemaking in Safety Critical and Complex Situations: Human Factors and Design*. CRC Press.

Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process safety progress*, *16*(3), 126-131.

Kaplan, S. (1997). The Words of Risk Analysis. *Risk analysis*, *17*(4), 407-417. https://doi.org/10.1111/j.1539-6924.1997.tb00881.x

Kaplan, S., & Garrick, B. J. (1981). On the quantitative definition of risk. *Risk analysis*, *1*(1), 11-27.

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.

Kim, T.-e., & Mallam, S. (2020). A Delphi-AHP study on STCW leadership competence in the age of autonomous maritime operations. *WMU Journal of Maritime Affairs*, *19*(2), 163-181.

Kirk, J., Miller, M. L., & Miller, M. L. (1986). *Reliability and validity in qualitative research*. Sage.

Kongsberg. (2017). *Autonomous ship project, key facts about Yara Birkeland*. Retrieved December 4 from https://www.kongsberg.com/no/maritime/support/themes/autonomous-shipproject-key-facts-about-yara-birkeland/

Kothari, C. R. (2004). *Research methodology: Methods and techniques*. New Age International.

Leedy, P. D., & Ormrod, J. (2001). Practical research: Planning and research. *Upper Saddle*.

Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of family medicine and primary care*, *4*(3), 324.

Leveson, N. (2020a). *Safety III: A systems approach to safety and resilience* [White Paper]. http://sunnyday.mit.edu/safety-3.pdf

Leveson, N. (2020b). A True Sociotechnical Approach to Safety in Complex Systems. In l. f. s.-H. F. i. A. I. O.-. Presentation at the Human Factors in Control (HFC) meeting "Learning from failures (Ed.).

Leveson, N. G. (1995). Safety as a system property. *Communications of the ACM*, *38*(11), 146.

Leveson, N. G. (2016). *Engineering a safer world: Systems thinking applied to safety*. The MIT Press.

Leveson, N. G., & Thomas, J. P. (2018). STPA handbook. *Cambridge, MA, USA*.

LR, L. s. R. (2016). *Risk Based Designs (RBD), ShipRight Design and Construction - Additional Design Procedures*

LR, L. s. R. (2017). ShipRight Design and Construction - Additional Design Procedures: Design Code for Unmanned Marine Systems. In (Vol. February 2017).

Lundberg, J., Rollenhagen, C., & Hollnagel, E. (2009). What-You-Look-For-Is-What-You-Find–The consequences of underlying accident models in eight accident investigation manuals. *Safety science*, *47*(10), 1297-1311.

Lutzhoft, M., Hynnekleiv, A., Earthy, J., & Petersen, E. (2019). Human-centred maritime autonomy-An ethnography of the future. Journal of Physics: Conference Series,

Lyons, J. B., Sycara, K., Lewis, M., & Capiola, A. (2021). Human-Autonomy Teaming: Definitions, Debates, and Directions. *Frontiers in psychology*, *12*, 589585-589585. https://doi.org/10.3389/fpsyg.2021.589585

Lützhöft, M. (2004). *“The technology is great when it works”: Maritime Technology and Human Integration on the Ship’s Bridge* Linköping University Electronic Press].

Lützhöft, M., Grech, M. R., & Porathe, T. (2011). Information environment, fatigue, and culture in the maritime domain. *Reviews of human factors and ergonomics*, *7*(1), 280-322.

MacKinnon, S., Man, Y., & Baldauf, M. (2015). D8. 8: Final Report: Shore Control Centre. *MUNIN (Maritime Unmanned Navigation through Intelligence in Networks) Consortium: Hamburg, Germany*.

Man, Y., Lundh, M., Porathe, T., & MacKinnon, S. (2015). From desk to field-Human factor issues in remote monitoring and controlling of autonomous unmanned vessels. *Procedia Manufacturing*, *3*, 2674-2681.

Maritime UK. (2019). *Maritime Autonomous Surface Ships (MASS) - UK Industry Conduct Princliples & Code of Practice*.  Retrieved from https://www.maritimeuk.org/media-centre/publications/maritime-autonomous-surface-ships-industry-conduct-principles-code-practice/

Martin, R. L. (2009). *The opposable mind: How successful leaders win through integrative thinking*. Harvard Business Press.

Maxwell, J. (1992). Understanding and validity in qualitative research. *Harvard educational review*, *62*(3), 279-301.

Maxwell, J. A. (2008). Designing a qualitative study. *The SAGE handbook of applied social research methods*, *2*, 214-253.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (Vol. 41). Sage publications.

Mosleh, A., & Chang, Y. H. (2004). Model-based human reliability analysis: prospects and requirements. *Reliability Engineering & System Safety*, *83*(2), 241-253. https://doi.org/https://doi.org/10.1016/j.ress.2003.09.014

MUNIN. (2016). *About MUNIN - Maritime Unmanned Navigation through Intelligence in Networks*. Retrieved September 11, 2017 from http://www.unmanned-ship.org/munin/about/

NMA, N. M. A. (2020). *Guidance in connection with the construction or installation of automated functionality aimed at performing unmanned or partially unmanned operations*. Retrieved from https://www.sdir.no/contentassets/2b487e1b63cb47d39735953ed492888d/rsv-12-2020.pdf

Norman, D. (2013). *The design of everyday things: Revised and expanded edition*. Basic books.

Norman, M. K., Hamm, M. E., Schenker, Y., Mayowski, C. A., Hierholzer, W., Rubio, D. M., & Reis, S. E. (2021). Assessing the application of human-centered design to translational research. *Journal of clinical and translational science*, *5*(1).

NRC, N. R. C. (2000). *Kvalitet i norsk forskning: En oversikt over begreper, metoder og virkemilder (In Norwegian)*. https://www.forskningsradet.no/om-forskningsradet/publikasjoner/?q=kvalitet%20i%20norsk%20forskning&year=2000

NTSB. (2018). *Collision Between a Sport Utility Vehicle Operating With Partial Driving Automation and a Crash Attenuator*. https://www.ntsb.gov/investigations/Pages/HWY18FH011.aspx

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & management*, *42*(1), 15-29.

Papanikolaou, A., & Soares, C. G. (2009). *Risk-based ship design: Methods, tools and applications*. Springer.

Parhizkar, T., Utne, I. B., & Vinnem, J.-E. (2022). Online Probabilistic Risk Assessment of Complex Marine Systems. *Springer Series in Reliability Engineering*.

Porathe, T. (2016). Human-centred design in the Maritime domain. *DS 85-1: Proceedings of NordDesign 2016, Volume 1, Trondheim, Norway, 10th-12th August 2016*, 175-184.

Porathe, T., Hoem, Å. S., Rødseth, Ø. J., Fjørtoft, K. E., & Johnsen, S. O. (2018). At least as safe as manned shipping? Autonomous shipping, safety and "human error". *Safety and Reliability–Safe Societies in a Changing World. Proceedings of ESREL 2018, June 17-21, 2018, Trondheim, Norway*.

Porathe, T., Prison, J., & Man, Y. (2014). Situation awareness in remote control centres for unmanned ships. Proceedings of Human Factors in Ship Design & Operation, 26-27 February 2014, London, UK,

Porathe, T., & Rødseth, Ø. J. (2019). Simplifying interactions between autonomous and conventional ships with e-Navigation. Journal of physics: conference series,

Ramos, M. A., Thieme, C. A., Utne, I. B., & Mosleh, A. (2019). Autonomous systems safety–state of the art and challenges. Proceedings of the First International Workshop on Autonomous Systems Safety,

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.

Ramos, M. A., Utne, I., & Mosleh, A. (2018). On factors affecting autonomous ships operators performance in a Shore Control Center. *Proceedings of the 14th Probabilistic Safety Assessment and Management, Los Angeles, CA, USA*, 16-21.

Ramos, M. A., Utne, I. B., & Mosleh, A. (2019). Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events. *Safety science*, *116*, 33-44.

Rausand, M. (2013). *Risk assessment: theory, methods, and applications* (Vol. 115). John Wiley & Sons.

Reason, J. (1997). *Managing the risks of organizational accidents (1st ed.).* Routledge.

Relling, T. (2020). A systems perspective on maritime autonomy: The Vessel Traffic Service's contribution to safe coexistence between autonomous and conventional vessels.

Relling, T., Lützhöft, M., Ostnes, R., & Hildre, H. P. (2018). A human perspective on maritime autonomy. International Conference on Augmented Cognition,

Reuters. (2016). *U.S. opens investigation in Tesla after fatal crash in Autopilot mode*. https://www.reuters.com/article/us-tesla-investigation-idUSKCN0ZG2ZC

Rumawas, V. (2016). Human factors in ship design and operation: Experiential learning.

Rumawas, V. (2021). Addressing Human Factors in Ship Design: Shall We? In *Sensemaking in Safety Critical and Complex Situations* (pp. 97-115). CRC Press.

Rødseth, Ø. (2021). Constrained Autonomy for a Better Human–Automation Interface. In *Sensemaking in Safety Critical and Complex Situations* (pp. 235-247). CRC Press.

Rødseth, Ø., Faivre, J., Hjørungnes, S., Andersen, P., Bolbot, V., Pauwelyn, A., & Wennersberg, L. (2020). *AUTOSHIP deliverable D3.1 - Autonomous ship design standard* (WP3 - Common challenges, methodologies, standards and tools for KETs Issue Final (v.13.07.2020)).

Rødseth, Ø., & Nordahl, H. (2017). Definition for autonomous merchant ships. Version 1.0, October 10. 2017. Norwegian Forum for Autonomous Ships. In.

Rødseth, Ø. J., Nordahl, H., & Hoem, Å. (2018). Characterization of autonomy in merchant ships. 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO),

SAE6906. (2019). Standard Practise for Human System Integration. In *Aerospace Standard*. SAE international

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. *Handbook of human factors and ergonomics*, *2*, 1926-1943.

Schneider, B. (2012). Design as practice, science and research. In *Design research now* (pp. 207-218). Birkhäuser.

Schuler, D., & Namioka, A. (1993). *Participatory design: Principles and practices*. CRC Press.

Selvik, J. T., & Signoret, J.-P. (2017). How to interpret safety critical failures in risk and reliability assessments. *Reliability Engineering & System Safety*, *161*, 61-68.

Sheridan, T., & Parasuraman, R. (2005). Human-automation interaction. Rev Human Factors Ergon 1 (1): 89–129. In.

Shneiderman, B. (2016). *The new ABCs of research: Achieving breakthrough collaborations*. Oxford University Press.

Skinner, R., Nelson, R. R., Chin, W. W., & Land, L. (2015). The Delphi method research strategy in studies of information systems.

Sloan, S. (2007). Risk Management vs. Safety Management: Can't we all just get along? ASSE Professional Development Conference,

Smith, S. W. (2003). Humans in the loop: Human-computer interaction and security. *IEEE Security & privacy*, *1*(3), 75-79.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of business research*, *104*, 333-339.

SRA, S. f. R. A. (2021). Core subjects of risk analysis. Discussion document. .

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2017). *Human factors methods: a practical guide for engineering and design*. CRC Press.

Steen, M. (2011). Tensions in human-centred design. *CoDesign*, *7*(1), 45-60.

Strauch, B. (2017). Ironies of automation: Still unresolved after all these years. *IEEE Transactions on Human-Machine Systems*, *48*(5), 419-433.

Thieme, C. A. (2018). Risk Analysis and Modelling of Autonomous Marine Systems.

Thieme, C. A., & Utne, I. B. (2017). A risk model for autonomous marine systems and operation focusing on human–autonomy collaboration. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *231*(4), 446-464.

Thieme, C. A., Utne, I. B., & Haugen, S. (2018). Assessing ship risk model applicability to Marine Autonomous Surface Ships. *Ocean Engineering*, *165*, 140-154.

Tian, W., & Caponecchia, C. (2020). Using the Functional Resonance Analysis Method (FRAM) in Aviation Safety: A Systematic Review. *Journal of Advanced Transportation*, *2020*, 8898903. https://doi.org/10.1155/2020/8898903

Tjora, A. (2012). *Kvalitative forskningsmetoder i praksis* (Vol. 2). Gyldendal akademisk Oslo.

Torraco, R. J. (2005). Writing integrative literature reviews: Guidelines and examples. *Human resource development review*, *4*(3), 356-367.

Tupper, E. C. (2013). *Introduction to naval architecture*. Butterworth-Heinemann.

Utne, I. B., Rokseth, B., Sørensen, A. J., & Vinnem, J. E. (2020). Towards supervisory risk control of autonomous ships. *Reliability Engineering & System Safety*, *196*, 106757.

Utne, I. B., Sørensen, A. J., & Schjølberg, I. (2017). Risk management of autonomous marine systems and operations. International Conference on Offshore Mechanics and Arctic Engineering,

van den Broek, J. H., Griffioen, J. J., & van der Drift, M. M. (2020). Meaningful Human Control in Autonomous Shipping: An Overview. IOP Conference Series: Materials Science and Engineering,

Veitch, E., & Alsos, O. A. (2021). Human-Centered Explainable Artificial Intelligence for Marine Autonomous Surface Vehicles. *Journal of Marine Science and Engineering*, *9*(11), 1227. https://www.mdpi.com/2077-1312/9/11/1227

Veitch, E., & Alsos, O. A. (2022). A systematic review of human-AI interaction in autonomous ship systems. *Safety science*, *152*, 105778. https://doi.org/https://doi.org/10.1016/j.ssci.2022.105778

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 RINA, Royal Institution of Naval Architects—International Conference on Human Factors,

Ventikos, N., Louzis, K., Sotiralis, P., Koimtzoglou, A., & Annetis, E. (2021). Integrating human factors in risk-based design: A critical review. *Ergoship 2021*.

Vicente, K. J. (2013). *The human factor: Revolutionizing the way people live with technology*. Routledge.

Ward, V., House, A., & Hamer, S. (2009). Developing a framework for transferring knowledge into action: a thematic analysis of the literature. *Journal of health services research & policy*, *14*(3), 156-164.

Wennersberg, L. A. L., Nordahl, H., Rødseth, Ø. J., Fjørtoft, K., & Holte, E. A. (2020). A framework for description of autonomous ship systems and operations. IOP Conference Series: Materials Science and Engineering,

Whittingham, R. B. (2004). *The blame machine: Why human error causes accidents*. Routledge.

Wikipedia. (2021). *Human-centered design*. Retrieved November 25 from https://en.wikipedia.org/wiki/Human-centered_design

Williams, C. (2007). Research methods. *Journal of Business & Economics Research (JBER)*, *5*(3).

Wróbel, K. (2021). Searching for the origins of the myth: 80% human error impact on maritime safety. *Reliability Engineering & System Safety*, *216*, 107942. https://doi.org/https://doi.org/10.1016/j.ress.2021.107942

Wróbel, K., Gil, M., & Montewka, J. (2020). Identifying research directions of a remotely-controlled merchant ship by revisiting her system-theoretic safety control structure. *Safety science*, *129*, 104797.

Wróbel, K., Montewka, J., & Kujala, P. (2017). Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. *Reliability Engineering & System Safety*, *165*, 155-169. https://doi.org/https://doi.org/10.1016/j.ress.2017.03.029

Wróbel, K., Montewka, J., & Kujala, P. (2018). System-theoretic approach to safety of remotely-controlled merchant vessel. *Ocean Engineering*, *152*, 334-345.

Yin, R. K. (2009). *Case study research: Design and methods* (Vol. 5). sage.

Yuh, J., Marani, G., & Blidberg, D. R. (2011). Applications of marine robotic vehicles. *Intelligent service robotics*, *4*(4), 221-231.

zeabuz. (2021). *The milliAmpere project*. Retrieved December 18 from https://www.zeabuz.com/projects

Zhou, X.-Y., Liu, Z.-J., Wang, F.-W., Wu, Z.-L., & Cui, R.-D. (2020). Towards applicability evaluation of hazard analysis methods for autonomous ships. *Ocean Engineering*, *214*, 107773.

# Part II

Article 1

# The present and future of risk assessment of MASS: A literature review

Hoem, Å. S. (2019). In Proceedings of the 29th European Safety and Reliability Conference (ESREL), Hannover, Germany (pp. 22-26).

# The present and future of risk assessment of MASS: A literature review

Åsa Snilstveit Hoem
*Department of Design, Norwegian University of Science and Technology, Norway. E-mail: aasa.hoem@ntnu.no*

Internationally, there is an increasing interest in autonomous and unmanned ships, so-called Maritime Autonomous Surface Ships (MASS). This represents a paradigm shift that is presently underway promising safer, greener and more efficient ship traffic. A hypothesis of increased safety is often brought forward as we know from various studies that "human error" is the most frequently reported cause of marine casualties. In the latest Allianz report, the cost of losses resulting from "human error" between 2011 and 2016 is equivalent to 1.6 billion USD. Important questions in this context are; if we replace the human with automation, can we then reduce the number of accidents? And how can we evaluate the potential for new types of accidents to appear? The paper "At least as safe as manned shipping" by Porathe et al. (2018) presents a new risk picture and highlights the need for risk assessment. This paper continues on the risk assessment part by presenting a literature review of carried out in March 2018. More specific this paper gives a summary of five risk assessment methods presented in eight papers, and discuss their strengths and limitations, before addressing the main issues for future risk assessments of MASS.

*Keywords*: MASS(s), Unmanned ship, Literature review, Risk Analysis, Risk Assessment.

## 1. Introduction

Shipping is currently on its way into its fourth technical revolution, called Shipping 4.0 or cyber-shipping (Rødseth et al. 2016). Failure to anticipate and design for the new challenges that are certain to arise following periods of technology change can lead to automation surprises (Cook and Woods 1996). MASSs are no exception. MASSs may be low manned or unmanned (Rødseth et al. 2017). In principle, MASSs are required to be, at least, as safe as conventional surface ships in similar service (Jalonen 2017, Earthy and Lützhöft 2018, Porathe et al. 2018) To demonstrate a certain level of risk and evaluate if the safety goal is fulfilled, risk assessment should be carried out (Rausand 2013). A risk-based design approach is recommended to be used for the development of MASS by Lloyd's Register (2016) and DNV-GL (2018). One important question is then what is risk-based design, and what risk analysis should be carried out (in the design phase)?

## 2. Research questions

This paper addresses the following questions:
1. What risk identification analysis and methods for MASS can be found in the literature today? (Primarily an assessment of models of risk identification).
2. What are the main limitations and challenges of these risk assessments?

It is important to be concise in what is meant by MASSs, risks and risk analysis, in this paper's context. The next section presents the background and definitions used, followed by the method.

## 3. What is MASS?

The International Maritime Organization (IMO) currently use the term MASS for any vessel that fall under provisions of IMO instruments and which exhibits a level of automation that is currently not recognized under existing instruments. In the following, the term "autonomous ship" is a merchant ship that has some ability to operate independently of a human operator. This covers the whole specter from automated sensor integration, via decision support to computer-controlled decision making. An "unmanned ship" is a ship without crew that needs a certain degree of autonomy, e.g. to handle situations where communication with a remote shore control center (SCC) is lost.

Within Autonomous Marine Systems (AMS), underwater vehicles, especially Unmanned Underwater Vehicles (UUVs), have existed for several decades and are characterized through their capability to survey the subsea environment on a larger scale than divers and submarines are able to (Yuh et al. 2011). A taxonomy for the different types of autonomous maritime vehicles is proposed by the Norwegian Forum for Autonomous Ships (NFAS).

Typically, an autonomous system is a set of automated tasks, added interactions with several systems and/or human interaction, with capabilities and factors deciding the degree/level of autonomy. Frameworks for degrees or levels of automation (LOA) have been discussed by several professionals, mostly within the area of motor vehicle automation (SAE 2016, Vagia et al. 2016). Within the maritime domain, IMO has started a Regulatory Scoping exercise on MASS. Rigors discussions regarding definitions and characterization of ship autonomy are outside of the scope for this work.

MASS is a relatively new concept, mostly dating back to the MUNIN project (started in 2012), hence the classifications and proposed taxonomy are still evolving. Three main concepts are currently differentiated for MASSs:

Proceedings of the 29th European Safety and Reliability Conference.
*Edited by* Michael Beer and Enrico Zio

ISBN: 981-973-0000-00-0 :: doi: 10.3850/981-973-0000-00-0 esrel2019-paper

a. Low manned vessels with a partly unattended bridge (Bertram 2016, Rødseth 2017).
b. A swarm of MASSs supervised by one manned ship, so-called master-slave (Bertram 2016).
c. MASSs supervised from SCCs (Rødseth and Tjora 2014, Rødseth 2017).

A MASS with low manning (a.) is an intermediate solution to unmanned autonomous ship during the transition period (Bertram 2016).

## 4. What are risk assessment and risk analysis?

Risk is a term used in many contexts and in many different fields with different meanings. In general, risk also covers positive consequences, while in the majority of industries the focus is on negative consequences such as the risk of accidents (accident risk or security risk). Risk is the effect of uncertainty on objectives (ISO, 2018). It can be further defined as a combination of the potential events, their consequences and their likelihood. Risk assessment consists of risk identification, risk analysis and risk evaluation (ISO, 2018).

Risk analysis is the process to comprehend the nature of risk and to determine the level of risk (ISO, 2009). A source of danger that may cause harm to an asset is called a hazard (Rausand 2011). Reviewing hazards may identify sources of potential harm to the system, which gives input to a risk analysis. Component failure accidents have received the most attention in engineering, but component interaction accidents are becoming more common as the complexity of our system designs increases (Leveson 2012). The traditional view on risk assessment is to define the risk as the product of consequence and probability (Rausand 2003). For MASSs, a traditional risk analysis will attempt to find the likelihood of events, such as collision, allision, grounding, or stranding, and the assessment of consequences such as damage to people, the environment or to other ships or infrastructures. However, it should be noted that recent definitions of risk analysis take a broader, qualitative perspective to emphasize that not all uncertainties can be probabilistically expressed (Aven 2009). A common operational definition of risk analysis is the process of answering the following three questions given by Kaplan and Garrick (1981): 1. What can happen/go wrong? 2. How likely is it? 3. If it does happen, what are the consequences? These questions translate into three tasks:

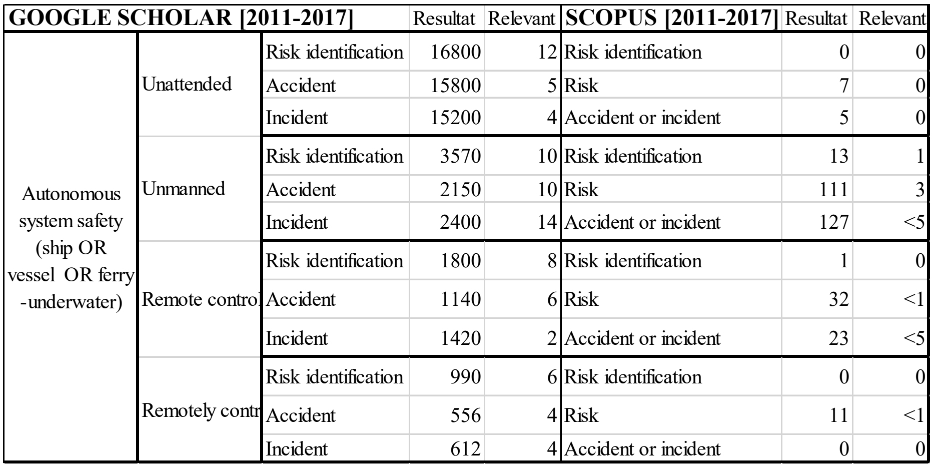
- Hazard identification (examples are HazId, Hazard and Operability Analysis (HAZOP), Failure Modes, Effects, and Criticality Analysis (FMECA), System Theoretic Process Analysis (STPA) and blended hazard identification methodology)
- Causal analysis (like fault tree analysis (FTA))
- Consequence analysis (a barrier or exposure analysis, like event tree analysis (ETA))

As input to risk analysis, we use historical data inputs from similar operations (experience and learning from accidents) and knowledge about the system structure and design. However, historical data is non-existing for MASSs and knowledge about the system structure is limited, as the development of the first MASSs are still in a conceptual phase. Hence, it is of interest to see how the literature is addressing this lack of information, operational data and experiences with MASS.

## 5. Method

An initial literature search was conducted in December 2017 to establish a picture of what type of definitions are the most common ones, in the sense of number of results. As mentioned, autonomous vessels can be unattended, unmanned, and/or remotely controlled. When considering maritime safety, it is of interest to look for publications on potential accidents and

Table 1. Preliminary literature search



| GOOGLE SCHOLAR [2011-2017] | | | Resultat | Relevant | SCOPUS [2011-2017] | Resultat | Relevant |
|---|---|---|---|---|---|---|---|
| Autonomous system safety (ship OR vessel OR ferry -underwater) | Unattended | Risk identification | 16800 | 12 | Risk identification | 0 | 0 |
| | | Accident | 15800 | 5 | Risk | 7 | 0 |
| | | Incident | 15200 | 4 | Accident or incident | 5 | 0 |
| | Unmanned | Risk identification | 3570 | 10 | Risk identification | 13 | 1 |
| | | Accident | 2150 | 10 | Risk | 111 | 3 |
| | | Incident | 2400 | 14 | Accident or incident | 127 | <5 |
| | Remote contro | Risk identification | 1800 | 8 | Risk identification | 1 | 0 |
| | | Accident | 1140 | 6 | Risk | 32 | <1 |
| | | Incident | 1420 | 2 | Accident or incident | 23 | <5 |
| | Remotely contr | Risk identification | 990 | 6 | Risk identification | 0 | 0 |
| | | Accident | 556 | 4 | Risk | 11 | <1 |
| | | Incident | 612 | 4 | Accident or incident | 0 | 0 |

incident involving autonomous vessels, in the literature. Hence, several Boolean searches were carried out with strings of the following keywords and results in Google Scholar and Scopus:

### 5. 1 *Evaluation criteria for relevance*

To identify suitable and relevant publications, the following criteria had to be fulfilled: 1) The article must be related to the maritime domain, 2) From the title or abstract of the paper, the words "risk(s)" or "accident" and some level of automation must be present. The number of results in Google Scholar was overwhelming but the ratio of relevant literature versus the total number of results was quite low. After reviewing the title and abstract of the first 40 articles, the subsequent papers were not within the scope of the search and the review was limited to the first 50 articles of each string. Scopus, on the other hand, gave a lower, more manageable number of results. Nevertheless, one important finding is that "unmanned" got the most hits in the database of Scopus.

### 5.2 *Selection of papers*

A second literature review was conducted in March 2018. This time the literature was obtained through Boolean searches in three interdisciplinary databases; Scopus, Google Scholar and Web of Science. Based on the findings in the first study, "Unmanned" was selected together with the keyword "risk identification".

## 6. Relevant literature

The second literature search resulted in 42 documents. Most of the reviewed papers are articles published in scientific journals and papers presented at international conferences. From these, 18 papers were of interest.

## 7. Findings

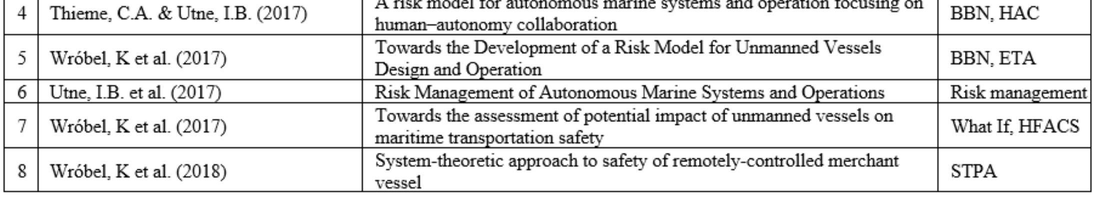
Considering the timeline of the publications of interest, the results clearly indicate that the topic *autonomous* and *unmanned shipping* has increased in popularity in terms of publishing during the last decade. The authors are mainly researchers at Nordic Universities, the Netherlands, Poland, and Japan. Many of the papers link to the MUNIN-project and the AAWA-project presented in Jalonen et al. (2017).

In the literature search result, only eight of the papers concerns topics related to risk models and/or risk identification directly. From the literature, it is possible to see strong progress from 2013 towards risk models that could be useful today. As the eight papers present different approaches to the methods for risk identification, risk analysis, and risk management, each method (and paper) are listed separately in Table 2 below. Risk models are used to assess the risk arising from ship traffic or during ship operation. Goerland et al. (2015) reviewed the use of risk definition of published maritime risk models and concluded that in many cases the models do not state the risk definition or risk measure. This is also the case for the reviewed paper here. As insufficient data are available for MASSs, quantification of models is difficult and the risk models in the paper are of a qualitative nature. The models do not present a high level of detail in the model description or structure, hence making it difficult to assess and compare them. Hence, the next sections present and discusses each model or method for risk analyses separately.

### 7.1 *The MUNIN project's risk assessment framework (HazId, paper 1, 2 and 3)*

The MUNIN project developed a technical concept for the operation of an unmanned merchant vessel and assesses its technical, economic and legal feasibility. To be more specific, the core concept was a dry bulk carrier operating completely unmanned for parts of an intercontinental voyage. The concept relies on a SCC to handle complex situations. Analysis of collision and foundering scenarios for the concept concluded that a decrease of risk of around ten times compared to manned shipping is possible, mainly due to the elimination of crews' fatigue issues. The final report (Burmeister et al. 2014) states that risks of engine and other system breakdowns are expected to be lower for unmanned ships if proper redundancy is

Table 2. Overview of the relevant reviewed literature



| No. | Author(s) (year) | Topic/Title | Risk model |
|---|---|---|---|
| 1 | Rødseth, Ø, & Tjora, A (2014) | A system architecture for an unmanned ship | HazId |
| 2 | Rødseth, Ø. & Burmeister, H.C. (2015) | Risk assessment for an unmanned merchant ship | |
| 3 | Rødseth, Ø. & Tjora, A (2015) | A risk based approach to the design of unmanned ship control systems | |
| 4 | Thieme, C.A. & Utne, I.B. (2017) | A risk model for autonomous marine systems and operation focusing on human–autonomy collaboration | BBN, HAC |
| 5 | Wróbel, K et al. (2017) | Towards the Development of a Risk Model for Unmanned Vessels Design and Operation | BBN, ETA |
| 6 | Utne, I.B. et al. (2017) | Risk Management of Autonomous Marine Systems and Operations | Risk management |
| 7 | Wróbel, K et al. (2017) | Towards the assessment of potential impact of unmanned vessels on maritime transportation safety | What If, HFACS |
| 8 | Wróbel, K et al. (2018) | System-theoretic approach to safety of remotely-controlled merchant vessel | STPA |

implemented and improved maintenance and monitoring schemes are followed.

In 2013, Rødseth et al. published an "Unmanned ships operational context relationship diagram," and in 2015, a risk assessment framework was published. They present a risk-based structured approach to the design by controlling the risk elements while providing solutions for problems and document evidence that the risk level will be acceptable. The method presented here adopts parts of the Formal Safety Assessment method from IMO (2014). The initial architect structure (seen in Figure 1 below) is used as a basis in a HazId exercise to systemize the search for dangerous situations or risks.

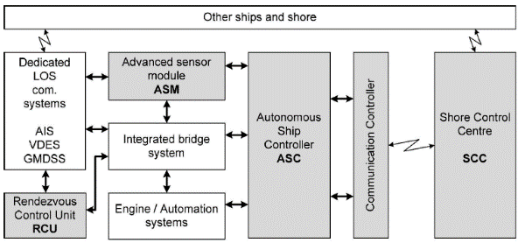


Fig.1: The MUNIN operational context relationship diagram derived from Rødseth et al. (2013).

In the risk assessment framework, different scenarios/accidents are considered, and hazards are identified together with mitigating actions, i.e. risk control options. They highlight the need to conduct a risk assessment before the system requirements are defined, in order to give input to the concept of operation (CONOPS) and verify the design. These risk control options aim at avoiding hazardous situations, but the interaction with the operator(s) is not given attention. Hence the method lack considerations of human autonomy collaboration. Given the paper is from 2015, the system architecture established here has formed the basis for other papers reviewed.

### 7.2 *A model describing the relationship between safety features of unmanned vessels (BBN, ETA, paper 5)*

With background in the MUNIN project and other sources where future anticipated design and performance are described (Burmeister et al., 2014; Rødseth and Burmeister, 2015), Wróbel, Krata, Montewka, and Hinz (2016) created a model describing the expected safety features in the paper "Towards the Development of a Risk Model for Unmanned Vessels Design and Operations". The risk model produced focuses on accidents' potential causes and failures within the system. The hazard analysis uses a Bayesian Belief Network (BBN) to describe the relationships between the safety issues from root cause to accident. The findings are structured into groups: Navigation, Engineering, Stability, and associated considerations, and Miscellaneous.

As the paper states, the model should be considered as a starting point to get an overview of relationships between safety features of unmanned vessels. There is no empirical data to support the likelihoods, so the validation is based on qualitative analysis. The model addresses several issues and potential accident types. However, addressing several accident types in one model may be a major challenge considering all different interactions and influencing factors. One major drawback is that the model does not include the communication connection to a SCC. In addition, the levels in the risk model are confusing; they are not levels of a technical system and should instead be considered as layers or steps or paths of an Event Tree Analysis (ETA). This assumption is made as the paper describe chains of consecutive events and conditions that may influence the consequences of potentially hazardous events. If the model were better structured, focusing on one accident type and explaining the levels and interactions, it would be useful as a basis for risk assessment of MASSs. Nevertheless, the paper addresses the challenge of uncertainties of the model due to unknown design and to the imperfection of brainstorming as a scientific method (Wróbel 2017, pp 7). From this publication in 2016, the model has been further developed.

### 7.3 *Review of marine accidents with "what if"-analysis and "HFACS" framework (paper 7)*

The same researchers (Wróbel et al., 2017) carried out a study of 100 marine accidents involving 119 vessels where the aim of the analysis was to assess whether the accident would have happened if the ship had been unmanned. It was also assessed whether its consequences would have been different. The assessment is based on a qualitative and subjective "what if"-analysis that ask: *1. If the ship were unmanned, how would that fact affect the likelihood of the particular accident?* and 2. *If the accident occurred anyway, would its consequences be more or less serious if there were no crew on-board?* The framework for Human Factors Analysis and Classification System for Marine Accidents (HFACS-MA) was set up to evaluate the causes of the accident. To answer the second question, the analysis of the accident's consequences was based on a simple check of whether the aftermath of maritime casualty affected people. The main challenge was the remoteness of the human operators, which has the benefit that the risk to the personnel is reduced. However, this remoteness implies that in case of an accident, like a fire, the human operator cannot recover the situation.

The "What-if" analysis and HFACS framework is not a method for risk assessment in the design phase, but the findings in the paper are of interest. It should be noted that the conditions for evaluating the safety of unmanned vessels are considering an unmanned vessel as a vessel where the bridge and crew is remote. The design and system architecture of autonomous systems might be completely different and new technology can cause accidents that we have not witnessed yet.. Another drawback of the study is the subjective evaluation of the effect of unmanned ships on the likelihood of the accidents and the many assumptions about which HFACS-MA causal category has the largest impact on an accident's occurrence. As a recommendation for further research the author emphasize the need to identify and list all anticipated hazards and their evaluated effects; only then can the level of safety associated with the unmanned ships operations be assessed (Wróbel et al., 2017, pp. 11).

### 7.4 *Systems-Theoretic Process Analysis, STPA (paper 8)*

In his latest paper, Wróbel argues that a framework building on the system-theoretic approach, STPA (Systems-Theoretic Process Analysis) is the best solution. This is also supported by Jalonen (2017). STPA is a hazard analysis technique based on STAMP (Systems-Theoretic Accident Model and Processes) first described by Leveson (2012). In order to perform a STPA, a safety control structure must be established. The safety control structure proposed in the paper is inspired by the many system architectures and models presented so far.

A list of hazards and correlated safety constraints related to different parts of the safety control structure is then presented. Furthermore, interaction mitigation of each control function is carried out in accordance with STPA principles. At the end of the paper (Wróbel et al., 2018) acknowledge the limitations of the system-theoretic method and present approaches on how to deal with uncertainties and so-called black swans. The modelling of the system is the most challenging part, causing a significant amount of uncertainty.

From the papers reviewed this is the most theoretically documented framework. However, from a safety perspective it could be beneficial to use a model that provides a quickly understood overview like the bow tie model that shows possible causal factors, consequences (outcome) and possible risk controls (barriers) linked to the hazardous events.

The analysis highlighted its preliminary status, addressing the uncertainty with respect to the design of MASSs. Technical issues have been identified as the factor contributing most to safety-related issues, followed by the interactions between SCC and the regulatory framework it needs to act under. Although interactions between operators are not covered. The authors of the paper try to add another dimension when including effectiveness and cost, which is not providing any useful information from a risk perspective. In an early design phase, knowledge of the system structure is limited. Hence, the mitigation of each interaction and their importance is valuable input to the evaluation and validation of design. Today, STPA are used for the assessment of dynamic positioning (DP) systems to identify hazards and for verification purposes (Rokseth et al., 2018).

### 7.5 *Bayesian Belief Network and Human Autonomy Collaboration (BBN, paper 5)*

As mentioned, Wróbel et al. (2016) suggest using Bayesian network for describing the relationships between the safety issues from root cause to accident. In a paper from 2017 Thieme and Utne investigate risk models focusing on human - autonomy collaboration. The main issue in the paper is that only a few risk models include human and organizational factors (HOFs). This aspect is illustrated in Figure 2 below.

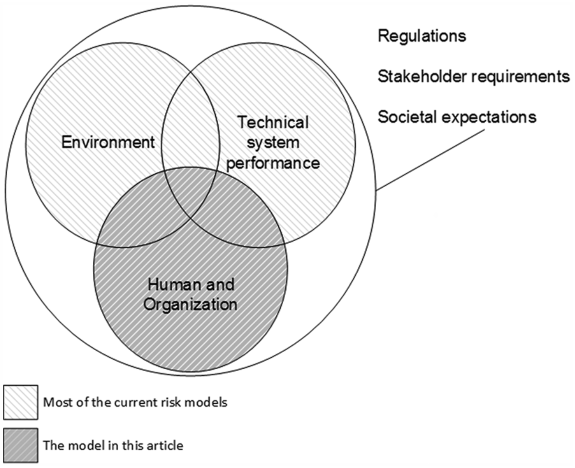


Fig. 2: The main aspects to include in an overall risk model. Derived from Thieme and Utne (2017).

The authors argue that risk models considering autonomous or remote operation should treat the human operators and the autonomous system as collaborators and not as individual or independent systems. The objective of the article is to present a BBN risk model focusing on human-Autonomy Collaboration (HAC) for AUV operation. Underwater vehicles are out of the scope of this paper. Nevertheless, as mentioned by the authors MASS may have similar requirements and demands as AUVs with respect to HAC, and the risk model could be adapted to other AMSs, as well (Thieme & Utne, 2017 pp. 1).

The paper provides a descriptive guideline for the steps involved in developing a BBN for risk modelling of HAC. This is a dynamic network of "nodes" which can be categorized as either Input-nodes, intermediate nodes and HAC nodes. The nodes have different states based on performance or status. The "arcs" connect the nodes (parent nodes to child nodes) and based on conditional probability tables (CPTs) for the parent node state, the child nodes' state are determined. This way the BBN can be quantified and the human–autonomy collaboration performance can be assessed in order to identify relationships between technical, human, and organizational factors and their influences on mission risk. However, it is a wide-ranging task and data on the human operator performance is not easy to evaluate, as in the case of workload perception variability from operator to operator. Trust and overreliance are other ambiguous terms, which are influenced by several factors, which are not possible to model in BBN (Thieme & Utne, 2017).

This is a systemic accident model that sees accidents as a result of concurrent interactions at the system level, rather than individual failures. It can be considered as an alternative option to the STPA but could also as a supplement. STPA can identify the nodes and interconnections between operators, technical systems, and HOF. However, the advanced method is detailed and intricate, and requires an understanding of BBN that is not easily acquired. It should include all dynamic interactions of components and subsystems which is, as mentioned, an extensive task.

### 7.6 *Risk monitoring and control (paper 6)*

Utne et al. (2017) suggest a concept for risk monitoring and control for an autonomous ship. There is a clear distinction between risk assessment during the design phase and the operation phase in their work. This paper mentions HazId and BBN but most of the paper discourses the definition of risk and what risk assessment of MASS should include. Figure 3 below shows the proposed structure of a risk management framework.

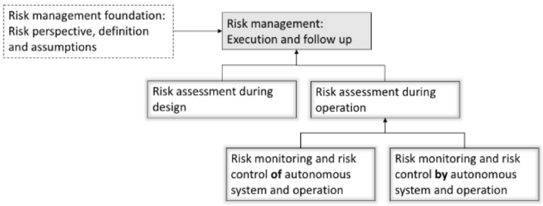

Risk management foundation: Risk perspective, definition and assumptions
Risk management: Execution and follow up
Risk assessment during design
Risk assessment during operation
Risk monitoring and risk control **of** autonomous system and operation
Risk monitoring and risk control **by** autonomous system and operation

Fig.3: Risk management of autonomous marine systems, derived from Utne et al. (2017).

The other papers mainly concern operational risk, which only covers one lifecycle of the system/parts of the risk management. In this paper, the most common definition of risk, is presented with the added measure "strength of knowledge" and uncertainty. When the strength of knowledge is low, the uncertainty is high. The complexity of autonomous systems and operations is also highly related to uncertainty, as the more complex a system and operation is, the more difficult it is to gain "perfect" knowledge of it. This is a good point, but there is no sufficient way of including this measure in a risk analysis, other than subjectively considering assumptions, data quality and information available. The paper suggests using the BBN model presented in previous section. The paper recommends identifying hazards and risk influencing factors (RIFs) in the design phase and include uncertainty as the main constituent part of risk (rather than probability alone), nevertheless the paper do not go into depth on how to include this.

## 8. Human factors

The literature search did not include the term "human factors" or "human error", however based on the findings in the literature listed in Table 2, this topic is given a section here. It is a consensus in the majority of the papers here that the contribution from human factors is important. Human factor issues and situation awareness are considered in five of the eight papers.

The explicit assumption is that with no humans on the bridge "human error" will go away (Porathe et al. 2018). The reason automation is safer is that they address human shortcomings like fatigue, limited attention span, information overload, normality bias etc. These issues are hypothesized to be reduced by increased ship autonomy by reducing the human involvement in direct control of ships, and by reducing the size of the crew on-board exposed to hazards of the hostile sea environment. However, it is important to remember that that our increasing dependence on information systems, and increasingly sharing of control of systems with automation, are creating a considerable potential for loss of information and control leading to new types of "human errors" (Leveson 2012).

There has been a cultural shift in the maritime industry toward increased levels of automation in tasks, particularly for navigation systems (Hetherington et al. 2006). This is partly because of reduced manning levels, as captains and crews are under increasing commercial pressure as supply chains are streamlined, and the availability of new technology. The paper "On Your Watch: Automation on the Bridge" by Lützhöft and Dekker (2002) discusses the qualitative consequences of automation on human work and safety. The paper propose that automation creates new human weaknesses and amplifies existing ones (Lützhöft and Dekker 2002 pp. 5). This is demonstrated by known accidents resulting from overreliance on machines. At the same time, automation can increase the cognitive demands on the reduced workforce.

In the discussion on "human error" it is important to remember that "human error" is not a cause but a result of other factors such as poor design, poor planning, poor procedures (Reason 2016). All human behavior is influenced by the context in which it occurs, and operators in high-tech systems are often at the mercy of the design of the automation they use. Hence, it might be more accurately to label an operator error as a flawed system or interface design instead. One example of this is a study of 27 collisions between attendant vessels and offshore facilities in the North Sea (Sandhåland et al. 2015). The study identified that errors due to reduced vigilance and misconceptions of the technical automation systems emerged as the primary antecedents of collisions.

Automation of human processes are expected to significantly reduce the number of incidents happening in shipping today. Nevertheless, the human element will not disappear. It will shift from ship to shore, where the remote operator exists and from where the software design and updating takes place. One must also assume that several potential accidents are adverted by the crew's actions and it is not clear if improved automation can match these numbers. Finally, one must also assume that some new types of incidents will occur because of the introduction of new technology and more automation.

## 9. More recent relevant literature

After the literature review was conducted several classification societies like DNV GL and Bureau Veritas (BV) have published guidelines on the topic of MASS and safety. DNV GL recommends the overall assurance process to be risk-based (DNV GL 2018), where minimum risk conditions (safe states) should be established based on structured risk analysis performed on several levels utilizing different methodologies; A preliminary hazard analysis (PHA) and a detailed risk analysis (FTA, ETA or FMEA), in addition to risk analysis method focusing on human aspects for operations from a SCC. BV also recommends assessing already available techniques for risk assessment (Veritas 2017).

## 10. Conclusions

It seems to be generally accepted that automation has the potential to decrease accidents that are due to human variability. However, automation has the potential of creating accidents, e.g. through transitions between automatic and manual control and the human having to rapidly assess the situation and make the right decisions. In the literature reviewed it seems that this challenge is not seemed to be included further than addressing situation awareness and human-machine (or autonomy) interaction.

Autonomy will create new types of accidents, partly due to accidents that was before averted by the human crew and partly due to introduction of new technology and corresponding new accident types. These types of accidents are challenging to include in the risk analysis as we lack statistical evidence for their probability. For further work it could be a good idea to make a database of the identified hazards and risks, and relate these to the dimensions of autonomy.

From the eight papers reviewed, it is difficult to conclude on one recommended practice for risk assessment of MASS. They all cover different topics, and some can be seen as overlapping and to some extent supplement each other. The papers highlight only parts of a socio-technical system, and a few scenarios. Some of the papers goes into depth in a case, while other papers highlight some perspectives and assumptions regarding the importance of safe operation and implementation. All risk analyses and models have different implications for how to analyze causes and consequences and target efforts. Comparing and discussing the results is hence challenging.

All eight papers acknowledge the lack of data on design solutions and system architectures and recognize that more work is necessary to develop approaches for risk analysis and assessment. Although, the STPA-method seems to be the most theoretically documented framework, it requires a high level of knowledge of the system architecture, and with a lack of empirical data subjective assumptions will be made to a greater extent. It should be stressed that all risk assessments and analysis have limitations. They also have different purposes and should be carried out both during the design and operation. Dynamic risk analysis will be important during operation, while risk assessment in the early design phase shall provide basis for constraints for the system, as pointed out by Utne et al. (2017). In the design phase it is beneficial to carry out a HazId/PHA and iterate it with the CONOPS until all relevant risks are managed.

As mentioned, no empirical studies have been performed to compare and evaluate the reviewed methods for risk analyses of MASSs. According to a study by Thieme et al. (2018), risk assessment and modelling of AUV and Autonomous Remotely operated vehicles are presented with operational data to some extent in the literature (Thieme et al. 2018 pp. 12-13). While this is not the case for MASS where less research has been conducted, both on the qualitative and quantitative side of risk analyses, as of today.

## 11. Recommendation for further research

From the review, the following main challenges, and hence request for further research within risk assessment in the design of MASSs, is listed:

- The need to cope with the lack of empirical (historical) data

- The need to include the human operator in the loop. In highly automated and autonomous systems, the influence of operators and other people interacting with the system is unneglectable (Bainbridge, 1983)
- The need for improved causal models to explicitly model organizational factors and software failures
- The need to consider dependencies between systems, including safety and security issues in complex control actions of MASS

## Acknowledgement

This work is supported by the SAREPTA (Safety, autonomy, remote control and operations of industrial transport systems) project, financed by the Norwegian Research Council with Grant No. 267860.

## References

Aven, T., & Renn, O. (2009). On risk defined as an event where the outcome is uncertain. Journal of risk research, 12(1), 1-11.

Bainbridge, L. (1983). Ironies of automation. Analysis, Design and Evaluation of Man–Machine Systems 1982, Elsevier: 129-135.

Burmeister, H. C., Bruhn, W., Rødseth, Ø. J., and Porathe, T. (2014). Autonomous unmanned merchant vessel and its contribution towards e-Navigation implementation: The MUNIN perspective. *International Journal of e-Navigation and Maritime Economy, 1, 1-13.*

Cook, R. and Woods, D. (1996). Adapting to new technology in the operating room. *Human factors, 38(4), 593-613.*

DNV-GL (2018). DNV GL Class Guideline 0264 Available:http://rules.dnvgl.com/docs/pdf/dnvgl/cg/2018-09/dnvgl-cg-0264.pdf [Accessed 31.10.2018]

Earthy, J. V. and Lützhöft, M. (2018). Autonomous ships, ICT and safety management. In Managing Maritime Safety (pp. 143-165).

Hetherington, C., et al. (2006). "Safety in shipping: The human element." *Journal of safety research 37(4).*

IMO, I. M. O. (2007). Guidelines for the application of Formal Safety Assessment (FSA) for use in the IMO rule-making process. MSC/Circ.1180-MEPC/Circ.

Jalonen, R. T., et al. (2017). Safety of Unmanned Ships: Safe Shipping with Autonomous and Remote Controlled Ship. Aalto University.

Kaplan, S. and Garrick, B. J. (1981). On the quantitative definition of risk. Risk analysis, 1(1), 11-27.

Leveson, N. (2012). Engineering a safer world: systems thinking applied to safety. Cambridge, MIT Press.

Lloyd's Register (2016, February), "Cyber-enabled ships. Deploying information and communications technology in shipping". First edition.

Lützhöft, M. H. and S. W. A. Dekker (2002). "On Your Watch: Automation on the Bridge." *Journal of Navigation 55(1).*

Porathe, T., et al. (2018). At least as safe as manned shipping? Autonomous shipping, safety and "human error". Safety and Reliability–Safe Societies in a Changing World, CRC Press: 417-425.

Rausand, M. (2013). Risk assessment: theory, methods, and applications, John Wiley & Sons.

Reason, J. (2016). Managing the risks of organizational accidents, Routledge.

Rokseth, B., Utne, I. B., & Vinnem, J. E. (2018). Deriving verification objectives and scenarios for maritime systems using the systems-theoretic process analysis. *Reliability Engineering & System Safety, 169*, 18-31.

Rødseth, Ø. and Tjora, Å. (2014). "A risk based approach to the design of unmanned ship control systems." Maritime-Port Technology and Development 2014.

Rødseth, Ø. J. and Burmeister, H. C. (2015). Risk assessment for an unmanned merchant ship. *International Journal on Marine Navigation and Safety Od Sea Transportation,* 9.

Rødseth, Ø. J., et al. (2016). "Big data in shipping-Challenges and opportunities."

Rødseth, Ø. J., et al. (2017). "Characterization of autonomy in merchant ships."

SAE (2016). "SAE J3016: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles."

Sandhåland, H., et al. (2015). "Situation awareness in bridge operations–A study of collisions between attendant vessels and offshore facilities in the North Sea." *Safety Science 79: 277-285.*

Thieme, C. A. and Utne, I.B (2017). "A risk model for autonomous marine systems and operation focusing on human–autonomy collaboration." *Journal of Risk and Reliability 231(4): 446-464.*

Thieme, C. A., Utne, I. B., and Haugen, S. (2018). Assessing ship risk model applicability to Marine Autonomous Surface Ships. *Ocean Engineering.*

Utne, I. B., et al. (2017). Risk Management of Autonomous Marine Systems and Operations. ASME 2017 *36th International Conference on Ocean, Offshore and Arctic Engineering.*

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,* 53, 190-202.

Veritas, B. (2017). Guidelines for autonomous shipping. Guidance Note NI, 641.

Woods, D., et al. (2017). Behind human error, CRC Press.

Wróbel, K., et al. (2017). "Towards the assessment of potential impact of unmanned vessels on maritime transportation safety." *Reliability Engineering & System Safety* 165: 155-169.

Wróbel, K., et al (2016). Towards the development of a risk model for unmanned vessels design and operations. *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, *10*.

Yuh, J., Marani, G. and Blidberg, D. R. (2011). Applications of marine robotic vehicles. *Intelligent service robotics, 4(4), 221.*

Proceedings of the 29th European Safety and Reliability Conference.
*Edited by* Michael Beer and Enrico Zio
Copyright ©2019 by ESREL2019 Organizers. *Published by* Research Publishing, Singapore
ISBN: 981-973-0000-00-0 :: doi: 10.3850/981-973-0000-00-0 esrel2019-paper

Article 2

# Addressing the accidental risks of maritime transportation: could autonomous shipping technology improve the statistics?

Hoem, Å. S., Fjørtoft, K., & Rødseth, Ø. J. (2019). In TransNav: International Journal on Marine Navigation and Safety of Sea Transportation, 13(3) (pp.487-494).

TRANSNAV
http://www.transnav.eu

the International Journal
on Marine Navigation
and Safety of Sea Transportation

Volume 13
Number 3
September 2019

DOI: 10.12716/1001.13.03.01

# Addressing the Accidental Risks of Maritime Transportation: Could Autonomous Shipping Technology Improve the Statistics?

Å.S. Hoem
*Norwegian University of Science and Technology, Trondheim, Norway*

K. Fjørtoft & Ø.J. Rødseth
*SINTEF Ocean, Trondheim, Norway*

ABSTRACT: A paradigm shift is presently underway in the shipping industry promising safer, greener and more efficient ship traffic. In this article, we will look at some of the accidents from conventional shipping and see if they could have been avoided with autonomous ship technology. A hypothesis of increased safety is often brought forward, and we know from various studies that the number of maritime accidents that involves what is called "human error" ranges from some 60-90 percent. If we replace the human with automation, can we then reduce the number of accidents? On the other hand, is there a possibility for new types of accidents to appear? What about the accidents that are today averted by the crew? This paper will present a method to assess these different aspects of the risk scenarios in light of the specific capabilities and constraints of autonomous ships.

## 1 INTRODUCTION

It is commonly believed that human errors are the main causation factor for maritime accidents and incidents. The term "human error" is a broad category covering a wide variety of unintentional unsafe behavior. From Allianz figures a range from 50 to 80% are often seen, with 75% being the figure used by Allianz (2018). With this background, it could be argued that an unmanned and fully autonomous ship should be much safer than a corresponding manned ship. However, there are several parameters which will determine the safety of an autonomous ship and this paper will attempt to present a more complete picture.

Section two will define the types of autonomous ships we believe is the most relevant in the near future, i.e. next 10 years. Section three will compare autonomous ships, as understood by the authors, with manned ships and list the main differences that can also be the basis for comparison of risk factors. Section four discusses types of accidents and causation factors and how this picture will be modified for autonomous ships. Sections five to seven discuss different classes of accidents and try to provide some quantitative expectations for how these classes will change when autonomy is introduced. Section eight will give a summary and conclusions.

## 2 WHAT IS AN AUTONOMOUS SHIP?

Autonomy literally means "self-governing" and comes in very different forms. Rødseth (2018) discusses this topic and provides a characterization scheme for autonomy in ships. Maritime Autonomous Surface Ship (MASS) is by IMO defined as a ship that, to a varying degree, can operate independently of human interaction. Autonomy is also closely connected to unmanned operation: Having a completely unmanned ship is desirable as it realizes significant gains by removing the hotel section and associated energy use, removing much safety equipment and reduces crew costs and by that also allows easier scaling down of ship sizes (Rødseth 2018b). Central in this is also the use of a shore control center (SCC) as discussed in Man et al (2015). In this context, autonomy is important to enable operators in the control center to monitor and control several ships and by that reduce costs of operations in the SCC.

It is theoretically possible to design a fully autonomous ship without any human oversight at all, but this is extremely unlikely in all but very special cases, due to the resulting extreme demands on the on-board technology. Being able to operate with "constrained autonomy" (Rødseth 2018) and having humans as back-up in cases where operational demands exceed the automation system's capabilities is a much more likely alternative. In addition, current public and private law and regulations associated with safety of ship operations as well as with the commercial issues related to shipping is also dependent on having a legally responsible person in charge of the ship. Changing laws and regulations will take a long time if it is at all possible (Rødseth 2017).

As the technology improves, the shipping community gets more experience with the operation of autonomous ships and when laws and regulations have been updated, it is very likely that fully autonomous ships will be launched, but this will take many years. Technology will be used for sensing, AI and IoT have been rapidly advanced and utilized in various fields. Automated operation systems of ships have been active with aims of further safe navigation by preventing human errors, improving working conditions of ship's etc. (Matsumoto 2018).

In line with the above discussion, in the following we will assume that an autonomous ship is a ship that is completely unmanned, but with a shore control center and limited (constrained) autonomy in the onboard control systems.

## 3 COMPARISON TO MANNED SHIPS

In the following paragraphs, we will attempt to identify the main factors that distinguish an autonomous ship from a conventional manned ship, based on the assumptions from the previous section: Fully unmanned cargo ship with constrained shipboard autonomy and a shore control center (SCC) to handle events that the automation cannot handle.

### 3.1 *Fully unmanned*

The most interesting autonomous ship projects are associated with fully unmanned operations as discussed in the previous section. While there will be provisions for having people onboard during maintenance and port operations, unmanned voyages have a number of important effects:

1 Higher demand on sensors, automation and shore control as operators in SCC lack some of the "personal touch", both on environment, ship and technical system's performance.
2 Much lower exposure to danger for the crew.
3 May be unable to inspect equipment or systems that report errors or problems.
4 Lower risk of fires in accommodation, galleys, laundry and waste systems, which is relatively high on manned ships.

### 3.2 *Constrained autonomy*

Autonomy will be limited for the onboard systems and the ship will be dependent on occasional support from the SCC. To avoid known problems with human-automation interfaces (HAI) in the shore control center, the ship automation will have "constrained autonomy" (Rødseth 2018). The assumption is that this also helps in testing and qualifying sensor and automation systems to specified performance level. This has a number of effects:

1 More limited, but also more deterministic action responses from sensors and automation.
2 Dependence on shore control operators' performance and situational awareness.
3 Dependence on communication link to shore.
4 Dependence on high quality implementation of fallback solutions and definition of minimum risk conditions for the ship.

### 3.3 *Shore control center*

The shore control center will be manned with supervision operators as well as specialist intervention teams that are activated in cases of special demands from a ship (Man et al. 2015). In addition to issues mentioned in the previous sections, this will have the following effects:

1 Dependent on good training and cooperation in the shore control center.
2 Intervention crew do not have to worry about personal risk and adverse conditions on board.

### 3.4 *Higher technical resilience*

Another important aspect is the reliability of technical systems onboard and increased redundancy in the same systems. As there is no crew available to provide a safety barrier in case of technical failures, it is necessary to add new technical barriers where necessary, e.g. by using increased redundancy. This requirement is already included in the guidelines published by DNV GL (2018).

Today's crew use much of their time on maintenance of the ship and its systems. This will not be possible on an unmanned ship and to avoid increased off-hire due to more and longer dry-dockings, it will be necessary to use systems with lower maintenance requirements. This can typically be diesel-electric energy and propulsion systems, no use of heavy fuel, improved coatings on the ship and in cargo holds etc. Effects are:

1 More technical barriers against technical faults.
2 Much improved technical systems with built in predictive maintenance functionality.
3 More dependent on maintenance at shore.

### 3.5 *Improved voyage planning*

Finally, unmanned ships will be used in liner type operations where they trade between a relatively limited number of ports where infrastructure and trained personnel are available to handle the unmanned ship safely and efficiently. In addition to infrastructure requirements, also the current legal systems rule out tramp type shipping where the unmanned ship calls on arbitrary ports: Until international regulations have been established, unmanned operation will need to be based on bilateral agreements between the involved flag, coastal and port states. This also means that operations of unmanned ship will be able to take advantage of better cooperation with coastal state authorities, better described fairways, possibly additional infrastructure in the fairways and improved planning of the voyage. The effects of this are:

1 Less chance of surprises during voyage.
2 More support from public functions on land.

## 4 ACCIDENT SCENARIOS

### 4.1 *Today's accident picture*

There are a number of different papers investigating accident statistics and causation factors in the available literature. They use different data sources and different methods and results vary quite widely. The publicly available databases of marine accidents have different database structure and approaches to analyze the accident causation and consequence mitigation. There are many reasons for this, among them large variations in accidents between geographic regions, types of ships, age of ships, flags and insurance, see e.g. Eleftheria et al. (2016), Equasis (2018) and Allianz (2018).

In this paper, we will use statistics from the European Maritime Safety Agency (EMSA 2018) and mainly figures from the period 2011 to 2017. This covers accident reports from EU and associated countries.

### 4.2 *Occupational fatalities*

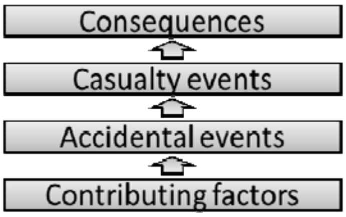
Working on a ship is in general considered more dangerous than similar jobs on land. In the UK, the fatality rate at sea is about 12 times higher than in the general population and in Poland it is eight times higher than that again (Allianz 2012).

From the EMSA statistics it can be seen a split between occupational fatalities, e.g. slipping, falling or being hit by objects, and fatalities caused by ship accident. In the period 2012 to 2017 such occupational fatalities amounted to 43% of a total of 683 fatalities in the period.

If a ship is operated without a crew, it is obvious that this will be a significant contribution to the safety of the voyage as seen from the now on-shore crew.

### 4.3 *Ship accidents*

EMSA uses a special classification system that is implemented in EMCIP (European Marine Casualty Information Platform). A much abbreviated version of the classification system is shown in Fig. 1.



Consequences

Casualty events

Accidental events

Contributing factors

Figure 1. EMCIP elements (EMSA 2018)

Most casualties should be seen as processes that involve a number of errors, failures and uncontrolled environmental impacts, and not just the more dramatic **Casualty Event** itself. This group of events will collectively be termed **Accidental Events** (Caridis 1999). The categories of accidental events used by EMSA are listed in Figure 2. **Contributing factors** is something that helps cause a result. The latter two are often called causal factors, which in general mean general actions, omissions, events or conditions, without which the marine casualty or marine incident would probably not have occurred or have been as serious (IMO 2008). Over the period 2011 to 2017, EMSA has analyzed 1645 accidental events with a distribution as shown in Figure 2 below.

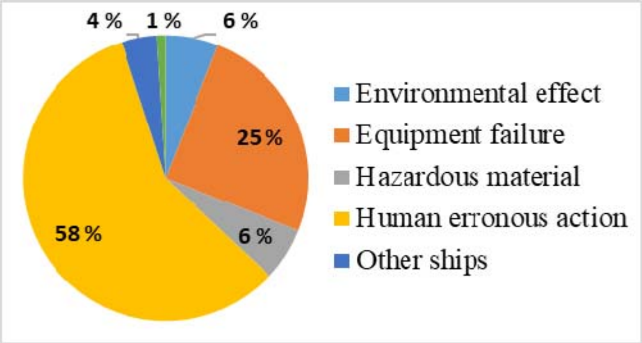


Figure 2. Accidental events from EMPIC (EMSA 2018)

This presents a lower percentage for human errors than what has been common in other literature (Allianz 2018, Baker 2009), but it is still a substantial contributing factor with 58%. It is also interesting to see that equipment failure represents 25% of the accidental events. We will come back to this in section 5.

### 4.4 *The human factor is still an issue*

Another statistics of interest is how respectively shipborne operations and shore management acts as a main contributing factor to the casualty events. This is rendered in Fig. 3, where around 2900 contributing factors have been analyzed.

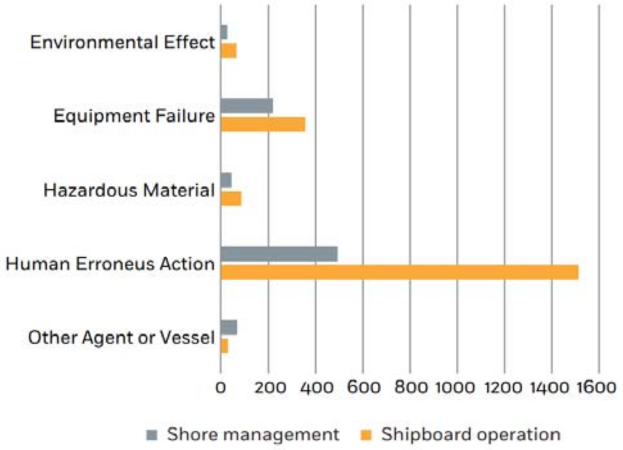


Figure 3. Relationship between ship and shore as contributing factor for marine casualties in general (EMSA 2018)

This may have an impact on expectations from a shore control center in the context of unmanned ships. However, shipboard operation is a main contributing factor to 70% of the casualty events.

This bring us to the human role in MASS operations. Humans still need to intervene with a MASS vessel, however the human element of the operations seem often to be forgotten when designing a MASS. The human, i.e. operator, still need to supervise and analyze the operations done by the autonomous systems, either from a SCC or when a MASS is manned. When looking at accident statistics of conventional shipping, we tend to look at the negative side of human intervening. In the design phase of MASS, the human machine interactions (HMI) should be addressed. A Concept of Operation (CONOPS) refer to the awareness of a situation. It gives the perception of an event with respect to time and condition, and the system behavior (actual and future). A CONOPS will address the human factors in the MASS operation aspect. Known relevant human factor challenges of remotely operated and automated systems that should be included (Karvonen 2018) are:

- Situation and automation awareness
- The understanding between automation and human role
- User experiences and usability of the solutions
- Trust in automation
- Graphical user interface and visualization

### 4.5 *Accidents in autonomous ships*

It is an expectation that more automation can remove some of the accidents today caused by human error: Automation address human shortcomings like fatigue, limited attention span, information overload, i.e. limits of the human working memory, normality bias etc. How much that automation can improve the accident statistic is still an open question. The full picture is also more complicated than this, as illustrated in Fig. 4. (Porathe et al. 2018).

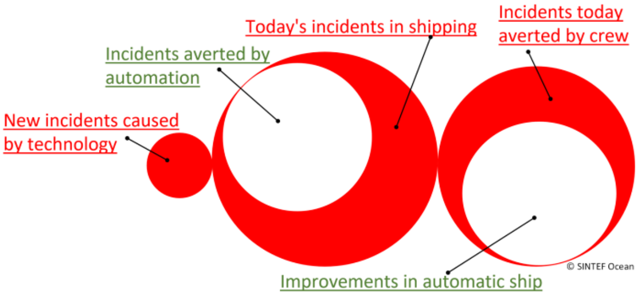

Today's incidents in shipping
Incidents today averted by crew
Incidents averted by automation
New incidents caused by technology
Improvements in automatic ship
© SINTEF Ocean

Figure 4. Three main groups of accidents and incidents

The middle circles represent today's incidents and accidents in shipping, which was discussed in section 4. The right circles represent the accidents that today's crew are able to avoid by being present onboard. The left circle represents new types of incidents that are caused by the advanced automation systems themselves. The dark circles are the damage potential and the white circles represent actions by the automation systems to avoid or minimize the effects of these incidents. This picture needs to include the effects of the SCC. Here, it is important to remember that our increasing dependence on information systems, and increasingly sharing of control of systems with automation, are creating a considerable potential for loss of information and control leading to new types of "human errors" (Leveson 2012). Which are contributing to the observed percentage of "human error" involved in the accident rates.

For the evaluation of the accident's causes, it is possible to apply Human Factors Analysis and Classification System for Marine Accidents (HFACS-MA). In a study by Wróbel et al. from 2017, 100 accidents reports were analyzed by applying this method and paying particular attention to the following two following aspects:

- If the ship were unmanned, how would that fact affect the likelihood of particular accident?
- If the accident occurred anyway, would its consequences be more or less serious if there were no crew on board?

According to HFACS-MA, the accident's causes are divided into 21 causal categories grouped in 5 levels:

- External Factors: Legislation gaps, administration oversights, and design flaws.
- Organizational Influences: Fallible decisions of upper-level management affecting supervisory practices as well as the conditions and actions of the operator (Scarborough 2005).
- Unsafe Supervision: Supervisory actions that influence the conditions of the operator and the type of environment in which they operate.
- Preconditions: latent unsafe conditions for unsafe acts that exist within a given work system (IMO 1999).
- Unsafe Acts: errors (slips and lapse), mistakes and violations performed by the operator.

The study concluded that the remoteness of the human operators and crew has the benefit of reducing the risk to the personnel significantly, and reducing the number of navigation-related accidents like collision or groundings (Wróbel et al. 2017, pp 10). However, the results also showed that the damage assessment and control is likely to be one of the biggest difficulties for the unmanned vessel.

One drawback of the study is that they evaluated an unmanned vessel as a vessel with the same design and technical systems in place, only with the bridge and crew being remote. The design and system architecture of autonomous systems will be completely different as discussed in section 3.4. Another drawback of the study is the subjective evaluation of the effect of unmanned ships on the likelihood of the accidents and the many assumptions about which HFACS-MA causal category has the largest impact on an accident's occurrence. As one of the recommendations for further research the author emphasize the need to identify and list all anticipated hazards and their evaluated effects; *only then can the level of safety associated with the unmanned ships operations be assessed* (Wróbel et al. 2017, pp. 11).

In this paper, we take a similar approach, but instead of analyzing accident investigation reports, we look at the larger picture and qualitatively evaluate the potential for the causal factors most common for the known accidents and incidents today.

### 4.6 *Experiences from accidents related to sensemaking and HMI*

The past decades we have seen a decline in marine accidents leading to loss of property, life and environmental damage. Particularly after 1980 the introduction of new technology has been accompanied by a steady and significant improvement in ship safety. These first steps towards greater use of automation in machinery spaces continued with advanced ships with smaller crews and increased operating efficiency through new technologies, particularly with regard to navigation system (Pomeroy 2017, Hetherington, Flin et al. 2006). However, more automation has also been related to the following issues: diminished ship sense, mishaps during changeovers and handoffs, latency and cognitive horizon, potential skill degradation, and resilience in abnormal situations.

One of the biggest challenges in highly automated systems is the disconnect, suggested as one of the ironies of automation (Bainbridge 1983), between the demand of the ships and its system and the skills and knowledge of the people operating it both at seas and ashore is causing new types of incidents and accident. In a review of 14 MAIB accident reports from 2005–2016, Kilskar and Johnsen (In Press) identified the following safety issues concerning automation at the bridge contributing to several of the accidents, hence contributing factors:

- Loss of situation awareness / poor sensemaking
- Insufficient training
- Alarm related issues
- Poor system design or display layout
- Poor (safety) management
- Poor or missing work load assessment
- Lacking or insufficient passage planning
- Missing, poor or unclear regulations or standards

Although these safety issues where identified in accident investigation reports, they concern HMI and will apply to operators in a SCC.

## 5 A QUALITATIVE COMPARISON OF AUTONOMOUS AND MANNED SHIPS

We have listed the main factors that distinguish an autonomous ship from a manned ship, as discussed in paragraph 3.1 – 3.5. With the identified causal and contributing factors, conditions, activities, systems, components, etc. that are critical with respect to accidental risk, presented in section 4 and 5, we attempt to classify the potential for higher or lower contributions to today's incidents in shipping.

Table 1 lists the characteristics of different technological solutions and shortly describes their strength and/or shortcomings. For each characteristic, a color indicates to which degree this is contributing to the three risk types listed in the three last columns: New accidents caused by new technology, Today's known accidents, and accidents adverted by crew) illustrated in Fig. 4. The factors contribution to the risk types is indicated by the following colors: increased risk (red - R), neutral impact (yellow - Y), or lesser impact/likelihood (green - G). Note that for the first type of accidental risks, new accidents caused by new technology, autonomous ships can obviously not be better than today. At best, it is neutral (Y). For a fully unmanned ship, one differentiating factor from manned ship is a higher demand and reliance on sensors, automation and shore control (row 1 in table 1 below). More advanced technology means a higher degree of system complexity causing new technological failures like unknown software failures for example. This contributes to a higher likelihood (risk) of new accidents caused by new technology, indicated by a red "R" under the column "New". For today's known incidents and accidents like collisions and allisions caused by human erroneous actions due to fatigue, new technology will be able to address such human shortcomings with collision detection and avoidance systems. Hence, a green "G" indicates the positive contribution on risk, as known accidents are avoided by new technology. However, accidents adverted by crew today should also be possible in autonomous operations by remote control and operation from the SCC. The technology in a fully unmanned ship and SCC shall be designed for remote operation, and the crew will still have impact, in order to avoid accidents and incidents. Hence, the contribution is neutral, indicated by a yellow "Y".

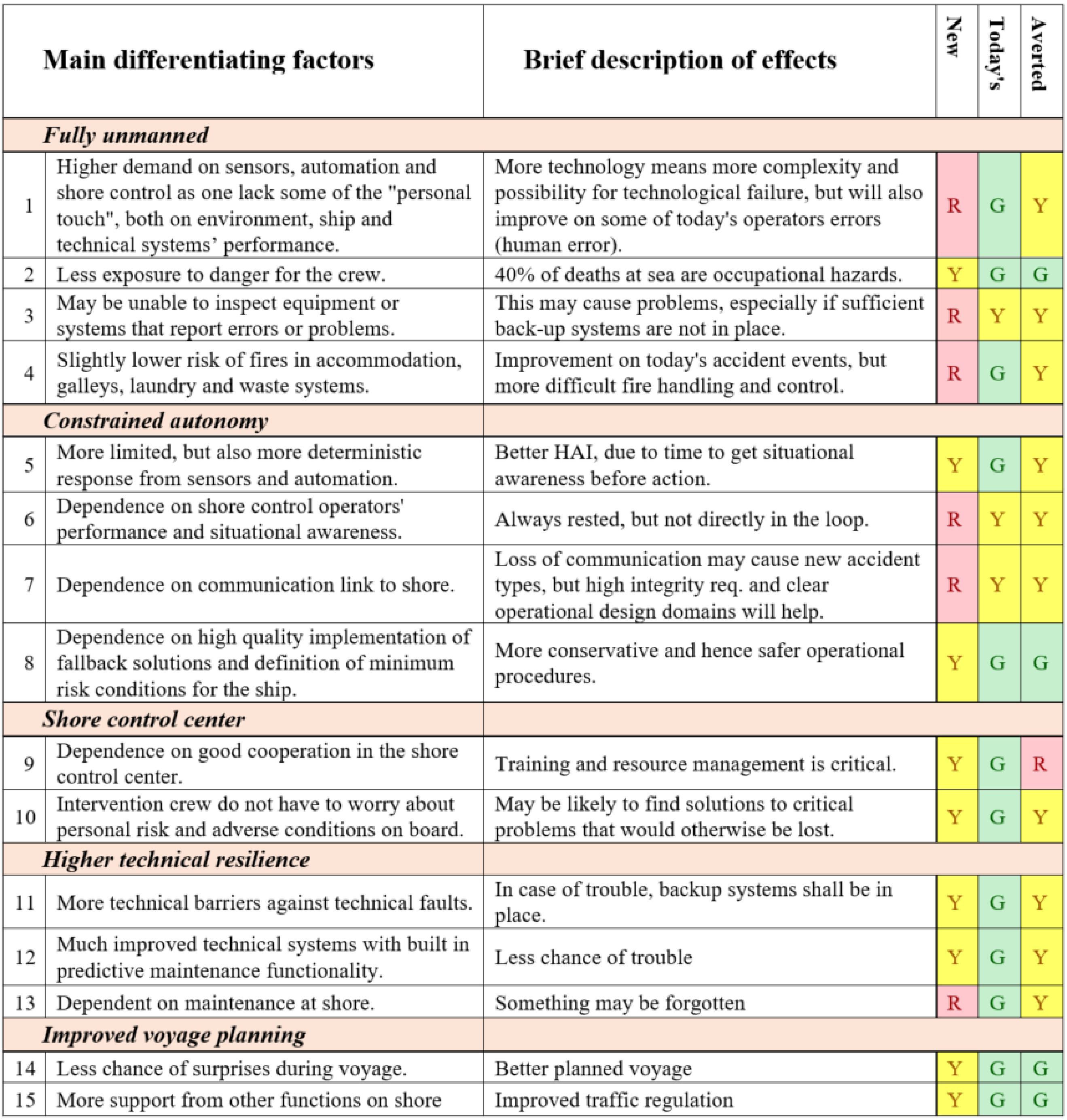
## 6 DETAILED DISCUSSION

First category, **fully unmanned,** points to a higher risk for software and technical failure. Due to for example:

- Sensor failure/degradation of hardware
- Insufficient redundancy
- Loss of propulsion or steering control
- Cyber security breaches
- Loss of communication with SCC

However, unmanned vessels will improve on some of today's operators' errors caused by human erroneous actions due to fatigue or other harsh working conditions.

Table 1. Qualitative comparison of autonomous and manned shipping



| Main differentiating factors | | Brief description of effects | New | Today's | Averted |
|---|---|---|---|---|---|
| ***Fully unmanned*** | | | | | |
| 1 | Higher demand on sensors, automation and shore control as one lack some of the "personal touch", both on environment, ship and technical systems' performance. | More technology means more complexity and possibility for technological failure, but will also improve on some of today's operators errors (human error). | R | G | Y |
| 2 | Less exposure to danger for the crew. | 40% of deaths at sea are occupational hazards. | Y | G | G |
| 3 | May be unable to inspect equipment or systems that report errors or problems. | This may cause problems, especially if sufficient back-up systems are not in place. | R | Y | Y |
| 4 | Slightly lower risk of fires in accommodation, galleys, laundry and waste systems. | Improvement on today's accident events, but more difficult fire handling and control. | R | G | Y |
| ***Constrained autonomy*** | | | | | |
| 5 | More limited, but also more deterministic response from sensors and automation. | Better HAI, due to time to get situational awareness before action. | Y | G | Y |
| 6 | Dependence on shore control operators' performance and situational awareness. | Always rested, but not directly in the loop. | R | Y | Y |
| 7 | Dependence on communication link to shore. | Loss of communication may cause new accident types, but high integrity req. and clear operational design domains will help. | R | Y | Y |
| 8 | Dependence on high quality implementation of fallback solutions and definition of minimum risk conditions for the ship. | More conservative and hence safer operational procedures. | Y | G | G |
| ***Shore control center*** | | | | | |
| 9 | Dependence on good cooperation in the shore control center. | Training and resource management is critical. | Y | G | R |
| 10 | Intervention crew do not have to worry about personal risk and adverse conditions on board. | May be likely to find solutions to critical problems that would otherwise be lost. | Y | G | Y |
| ***Higher technical resilience*** | | | | | |
| 11 | More technical barriers against technical faults. | In case of trouble, backup systems shall be in place. | Y | G | Y |
| 12 | Much improved technical systems with built in predictive maintenance functionality. | Less chance of trouble | Y | G | Y |
| 13 | Dependent on maintenance at shore. | Something may be forgotten | R | G | Y |
| ***Improved voyage planning*** | | | | | |
| 14 | Less chance of surprises during voyage. | Better planned voyage | Y | G | G |
| 15 | More support from other functions on shore | Improved traffic regulation | Y | G | G |

Important factors to address in the design and development of MASSs is robust sensor quality, redundancy on key technology, and good education for land-based operators, that builds the situational awareness based on technology. Next factor that has been pointed to is less exposure to danger for the crew. Statistics tells that about 40% of deaths at sea are occupational hazards. Another element is that it is expected that it will be slightly lower risk of fires in accommodation, galleys, laundry and waste system, because of no installation of such technology due to the fact that there is no need for it since there are no people on board. The expectations are fewer accidents, but when an accident happens, it might be more difficult to combat when people is not available and the only trust is technology, as addressed by Wróbel et al. (2017).

For a **constrained autonomy** vessel, we have pointed to better human-automation interfaces, due to time to get situational awareness before action. The design of SCC will learn from accidents where alarm related issues and poor HMI were major causal factors. It is likely that the humans are not directly in the loop (manually steering and navigating the vessel). To let the SCC take control there are dependencies to the infrastructure, such as the communication infrastructure, that will have enough coverage and bandwidth to bring data from the vessel to the SCC for awareness before decisions are taken. This also points to more conservative and safer operational procedures, to both operational practices and a higher safety degree.

**Shore control center** is another category that has been pointed to. The same applies for a SCC as on a vessel's bridge today, a good crew is those who collaborate and use each other's expertise in operations and problem solving. It is even more important at a SCC since the possibility to inspect the vessel is not the same. We assume here an increased risk of accidents that is today adverted by crew, as we know there will be controllability issues with a remote crew, and a high dependence on the SCC team's skills and knowledge. At the same time, the human risk factor is lower since the intervention crew do not have to worry about personal risk and adverse conditions on board. Training and resource management are important.

The category **Higher technical resilience** brings us back to the technology. It is important to build technical barriers towards technical failures with built-in predictive maintenance functionality.

Technical resilience is essential for MASS. The danger is that new unpredictable situations, that have not been thought of, can occur due to a high number of technical systems. Component interaction accidents are becoming more common as the complexity of system designs increases (Leveson 2012).

**Improved voyage planning** is a safety-critical function for autonomous vessels. Good planning means to prepare the voyage, the loads, the maintenance and all reporting during a voyage. This is a significant requirement compared with conventional vessels, were good planning is crucial for success, but often overlooked (NTSB 2015, DMAIB 2013, Bell 2006).

## 7 CONCLUSION

This paper provides a more realistic description of what an autonomous ship will be in the foreseeable future, i.e. unmanned, having monitoring and control personnel on shore, exhibiting constrained autonomy and having better operational planning and technical equipment than a manned ship.

While the overall risk picture for autonomous ships may look unpromising (Fig. 4), the differences in implementation have significant impacts on the individual risk types. The qualitative assessment done in Table 1 indicates that there is indeed a significant possibility to improve overall safety for autonomous ships compared to manned, although there are also areas that require special attention.

This paper only provides a cursory and qualitative analysis of the risk issues, but it is hoped that it can contribute to a more systematic process for risk assessment, also more accurately incorporating the positive technical contributions from autonomous ship designs.

## REFERENCES

Allianz Global Corporate and Specialty (2018), Safety and shipping Review 2018 – an annual review of trends and developments in shipping losses and safety, Munich, Germany, June 2018.

Allianz Global Corporate and Specialty (2012), Safety and Shipping 1912-2012: From Titanic to Costa Concordia, Munich, Germany, March 2012.

Baker, C., McCafferty, D. (2009). ABS Review and Analysis of Accident Databases.

Bainbridge, L. (1983). Ironies of automation. In Analysis, Design and Evaluation of Man–Machine Systems 1982 (pp. 129-135).

Bell, J., & Healey, N. (2006). The causes of major hazard incidents and how to improve risk control and health and safety management: A review of the existing literature. Health and Safety Laboratory.

DNV GL (2018), Class Guideline - Autonomous and remotely operated ships, DNVGL-CG-0264, September 2018.

DMAIB (2013). The Danish Maritime Accident Investigation Board: VEGA SAGITTARIUS Grounding on 16 August 2012. Marine accident report 2012003009. Issued on 27 March 2013. Valdby: DMAIB.

Eleftheria, E., Apostolos, P., & Markos, V. (2016). Statistical analysis of ship accidents and review of safety level. Safety science, 85, 282-292.

EMSA (2018), Annual Overview of Marine Casualties and Incidents 2018. EMSA, Lisbon, Portugal, 2018.

Equasis (2018), The World Merchant Fleet in 2017 – Statistics from Equasis, www.equasis.org, retrieved January 2019.

Caridis, P. (1999). CASMET. Casualty analysis methodology for maritime operations. National Technical University of Athens.

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: The human element. Journal of safety research, 37(4), 401-411.

IMO MSC/Circ.102/MEPC/Circ.392. 2002. Guidelines for Formal Safety Assessment (FSA) for use in the IMO Rule-Making Process. As amended. London: IMO 2002.

IMO MSC.255(84). 2008. Code of the international standards and recommended practices for a safety investigation into a marine casualty or marine incident (casualty investigation code). Adopted May 16, 2008. London: IMO 2008.

IMO Resolution A.884(21). Amendments to the code for the investigation of marine casualties and incidents (A.849(20)). London: IMO 1999.

Karvonen, I. 2018. Human Factors Issues in Maritime Autonomous Surface Ship Systems Development. The 1st International Conference on Maritime Autonomous Surface Ship.

Leveson, N. 2012. Engineering a safer world: applying systems thinking to safety.

Man, Y., Lundh, M., Porathe, T., & MacKinnon, S. (2015). From Desk to Field–Human Factor Issues in Remote Monitoring and Controlling of Autonomous Unmanned Vessels. Procedia Manufacturing, 3, 2674-2681.

Matsumoto T. et. al 2018. Guidelines for concept design of automated operation/autonomous operation of ships. International conference on maritime autonomous surface ship.

NTSB, National Transport Safety Board. 2015. Grounding of Mobile Offshore Drilling Unit Kulluk, near Ocean Bay, Sitkalidak Island, Alaska December 31, 2012. NTSB/MAB-15/10.

Pomeroy, R. V., & Earthy, J. V. (2017). Merchant shipping's reliance on learning from incidents–A habit that needs to change for a challenging future. Safety science, 99, 45-57.

Porathe T., Hoem Å., Rødseth Ø.J., Fjørtoft K., Johnsen S.O. (2018), At least as safe as manned shipping? Autonomous shipping, safety and "human error". XXXX

Rødseth Ø.J (2018). Defining Ship Autonomy by Characteristic Factors, Proceedings of ICMASS 2019, Busan, Korea, ISSN 2387-4287.

Rødseth Ø.J (2018b). Assessing Business Cases for Autonomous and Unmanned Ships. In: Technology and

Science for the Ships of the Future. Proceedings of NAV 2018: 19th International Conference on Ship & Maritime Research. IOS Press 2018 ISBN 978-1-61499-870-9

Rødseth Ø. J. (2017). From concept to reality: Unmanned merchant ship research in Norway. Proceedings of Underwater Technology (UT), IEEE, Busan, Korea, ISBN 978-1-5090-5266-0.

Rødseth Ø.J. & Nordahl H. (eds.). 2017. Definition for autonomous merchant ships. Version 1.0, October 10. 2017. Norwegian Forum for Autonomous Ships. http://nfas.autonomous-ship.org/resources-en.html. [Accessed 2018-12-12].

Scarborough, A., Bailey, L., & Pounds, J. (2005). Examining ATC operational errors using the human factors analysis and classification system (No. DOT-FAA-AM-05-25). Federal Aviation Administration Oklahoma City OK CIVIL AEROMEDICAL INST.

Wróbel, K., Montewka, J., & Kujala, P. (2017). Towards the assessment of potential impact of unmanned vessels on maritime transportation safety. Reliability Engineering & System Safety, 165, 155-169.

Article 3

# Improving Safety by Learning from Automation in Transport Systems with a Focus on Sensemaking and Meaningful Human Control

Hoem, Å. S., Johnsen, S. O., Fjørtoft, K., Rødseth, Ø. J., Jenssen, G., & Moen, T. (2021). In Sensemaking in Safety Critical and Complex Situations (pp. 191-207). CRC Press.

# 12 Improving Safety by Learning from Automation in Transport Systems with a Focus on Sensemaking and Meaningful Human Control

*Å. S. Hoem*
Norwegian University of Science and Technology

*S. O. Johnsen*
SINTEF

*K. Fjørtoft and Ø. J. Rødseth*
SINTEF Ocean

*G. Jenssen and T. Moen*
SINTEF

## CONTENTS

## INTRODUCTION

There is an increase in the use of automation and autonomous solutions within transportation. According to *The Oxford Dictionaries*, autonomy is the right or condition of self-government, and the freedom from external control or influence. Many researchers (Relling et al., 2018) have discussed that the term is used differently in colloquial language than in the technical definition and that it is interpreted in different ways across industries. In this chapter, we emphasise that autonomy does not necessarily mean absence of human interaction. Often there is a strong need to design how humans can make sense of automation failures and enact meaningful human control.

Automated systems operate by clear repeatable rules based on unambiguous sensed data. An autonomous system can be a set of automated tasks, with interactions with several sub-systems and/or humans, with a specific degree/level of autonomy. Autonomous systems obtain data about the unstructured world around them, process the data to generate information and generate alternatives and make decisions in the face of uncertainty. Systems are not necessarily either fully automated or fully autonomous but often fall somewhere in between (Cummings, 2019). For example, transportation can have different modes during a sea voyage. Outside the harbour, in heavy traffic, it can be closely operated either by the remote control centre (RCC) or a captain/driver, while in open waters with low traffic it can be controlled by the computers or the autonomous system. Within the road traffic segment, the Society of Automotive Engineers (SAE) has defined a taxonomy on the levels of automation describing the expectations between automated systems and the human operator (SAE, 2018). This is summarised in Table 12.1 below.

The levels apply to the driving automation feature(s) that are engaged in any given instance of operation of an equipped vehicle. As such, a vehicle may be equipped with a driving automation system that is capable of delivering multiple driving automation features that perform at different levels. The level of driving automation exhibited in any given instance is determined by the feature(s) that are engaged (SAE, 2018). Hence, autonomy is different across application areas; it varies over time and is affected by the context.

To get a better overview and understanding, we start by looking at experiences gained from ongoing research and/or industry projects in the four transportation

**TABLE 12.1**
**Levels of Automation – Simplified Description from SAE J3016 (2018).**

| LoA | Humans in control | Automation in control | Examples of automated features |
|---|---|---|---|
| 0: No driving automation | All operations | No automated task. Warns; protect | Blind-spot monitoring and lane-departure warning |
| 1: Driver assistance | All operations | Single automated systems: assists | Adaptive cruise control (ACC) |
| 2: Limited assist; auto throttle | Drives in-the-loop | Guides | Automated lane centring combined with ACC |
| 3: Assist, tactical; supervised | On-the loop human monitors all time | Manage movement within defined limits | "Traffic jam chauffeur" |
| 4: Automated assist strategic | Out-of-loop asked by system | Operates, but may give back control | Self-driving mode with geofencing |
| 5: Autonomous | Completely out-of-loop | Operates with graceful degradation | None are yet available to the general public |

domains: road, sea, rail and air. Through these case studies, we aim to explore safety, security, sensemaking and the human control of autonomous transport systems. We have adopted the term "meaningful human control" from discussion and debates from another area (lethal autonomous weapon systems; Cummings, 2019). The term addresses the concerns of a "responsibility gap" for harms caused by these systems, i.e. humans, not computers and their algorithms should ultimately remain in control of, and thus morally responsible for, relevant decisions about military operations. The same concern must be the result of autonomous systems in transportation, i.e. humans (supported by computers and algorithms) should ultimately remain in control and responsible for relevant decisions. The responsibility may be on the designer and producer of the autonomous systems, as Volvo and Mercedes Benz have stated for their autonomous cars (Chinen, 2019, p. 109).

## BACKGROUND: SAFETY OF AUTONOMOUS SYSTEMS

Safety is commonly defined as freedom from unacceptable risk (Hollnagel et al., 2008). For autonomous transportation to become a success, It must prove to be at least as safe and reliable as today's transport systems. By some, it is claimed that increased safety will be achieved by reducing the likelihood of human error when introducing more autonomy (Ramos et al., 2018). However, autonomy may create new types of accidents that before were averted by the human in control, as demonstrated by the Tesla fatal accident with Joshua Brown, NTSB (2017). Besides, the introduction of new technology will create new accident types, as explained by Porathe et al. (2018), Teoh and Kidd (2017), and Endsley (2019). The main safety challenges for autonomous systems are unexpected incidents not foreseen by automation, cybersecurity threats, technological changes (with increased complexity and couplings), poor sensemaking, lower possibility for meaningful human control (Human not in the loop) and limited learning from accidents.

The term "Human in the loop" means that the human is a part of the control loop, i.e. that the human receives information and can influence other parts of the chain of events (Horowitz and Scharre, 2015). A key issue is the ability of the actors to make sense of the situation. In our study, we define sensemaking in a pragmatic context as a continuous process of interpreting cues to establish situational awareness in a social context, as described in Kilskar et al. (2020).

When trying to scope risks of autonomous systems, we must include regulation, risk governance, organisational framework, interfaces between humans and the autonomous system, and the available infrastructure (software components and cyber-physical systems) to build a sense of the situation for humans and the automated system (Johnsen et al., 2019).

Autonomous systems are socio-technological systems. Hence, a holistic approach is necessary, rather than a reductionist approach looking at the system as isolated processes and components. We lack statistical evidence for the probability of accidents with autonomous transportation systems. However, several actors have started pilots with different levels of autonomy within different transport modes. There is a need to collect and systemise experiences from these. The following sections present a review of experiences from different transport modes. The main objective has been to gather experiences and status on different transport domains and to learn between the modes, by asking the following research questions:

1. What are the major safety and security challenges of autonomous industrial transport systems?
2. What can the various transport modes learn from each other regarding safety and security related to sensemaking and meaningful human control?
3. What are the suggested key measures related to organisational, technical and human issues?

## FINDINGS

### Autonomy at Sea

Several countries have developed test areas for testing Maritime Autonomous Surface Ships (MASS). The International Maritime Organisation (IMO) currently uses the term MASS for any vessel that falls under provisions of IMO instruments and which exhibits a level of automation that is currently not recognised under existing instruments. There are already several small-size unmanned and autonomous maritime crafts which have been engaged in surface navigation, scientific activities, underwater operations and specific military activities.

In Norway, three national testing areas have been established, with supporting infrastructure, with the aim to test out MASS in the same area as conventional ships. Norwegian Forum for Autonomous Ships (NFAS, 2020) is a network established for sharing experiences and research within the subject of autonomous ships, with the International Network for Autonomous Ships (INAS, 2020) as an extension of NFAS outside Norway. The research centre for Autonomous Marine Operations and Systems (AMOS, 2020) at NTNU was established in 2013 as a multidisciplinary centre for autonomous marine operations and control systems.

More extensive research projects, such as AAWA (2020), MUNIN (2020), Autosea (2020), Autoship (2020) and IMAT (2020), focus on specific concepts where unmanned, autonomous or smart ships are explored and tested. The world's first fully electric and autonomous container ship, Yara Birkeland (2020), is under construction. The ship is now planned to be in operation by 2022, earlier planned to start in 2020, and centres are scheduled to handle all aspects of remote and autonomous operation to ensure safety.

A newly established company, Zeabuz (2020), will test prototypes of an autonomous electric ferry system for urban waterways. Limited information is given about the concept other than it will be self-driving and electric. The remote and autonomous operational aspect of an RCC is not mentioned, but a remote support center is planned to operate in the initial phase.

Most of the projects above are in the initial stages with limited operational experience. Most safety concerns are related to the reliability of sensors and technical equipment and their ability to handle different situations.

## Experiences Related to Safety Challenges

In operation, MASS have only been tested in small scale without an interface for human supervision or control. We have examples of safety issues during early testing of autonomous technology (software and hardware) local in Norway in Trondheimsfjorden, with the small-scale version of the passenger ferry *AutoFerry*. One example is loss of control due to a technical failure, a so-called fallout, of the dynamic positioning system which made *AutoFerry* run into the harbour. However, there is no systematic data collection of failures or unforeseen events, and this is not a requirement from the Norwegian Maritime Authority (NMA) at present. Though, a Preliminary Hazard Analysis (PHA) has been carried out for the operation of the *AutoFerry* (Thieme et al., 2019), the main hazards were software failure; failure of internal and external communication systems; traffic in the channel (especially kayaks, difficult to discover); passenger handling and monitoring; and weather conditions. The practical challenges encountered in the ferry project were also listed. These challenges are related to available risk analysis methods and data, determining and establishing an equivalent safety level, and some of the prescriptive regulations currently in use by NMA. At present (start 2021) the *AutoFerry* project lacks an established plan on who should operate the ferry and how to intervene especially during emergencies. The human operator is said to be in the loop and able to intervene from an RCC. However, none of the projects have developed such a centre or made detailed plans for their operation so far. In the reviewed projects, the focus has been on technology development.

A literature review on risk identification methods for MASS (Hoem, 2019) identifies the uncertainty of the operational mode and context of the MASS operation (i.e. operational domain) to be a major challenge when identifying operational hazards and risks. There is a need to define what conditions the ship is designed to operate under. Rødseth (2018) proposed to use the "operational design domain" from SAE J3016 (2018) to define the context, i.e. the operational domain with its complexity.

This term is further described as an operational envelope (Fjørtoft and Rødseth, 2020). An operational envelope defines precisely what situation the MASS must be able to handle by assigning responsibilities to the human operators and the automation. It defines conditions of operations, describes the characteristics and requirements of the system and enables the design of Human–Autonomy Interface (HAI), based on specific task analysis, safety-critical tasks and challenges of sensemaking.

Several different guidelines are developed for autonomous shipping. IMO has published an Interim Guideline for MASS trials which aims to assist authorities and relevant stakeholders to perform autonomous tests. It includes risk management, how to comply with existing rules and regulations, safe manning, the human element and HMI, infrastructure, trial awareness, and communication and information sharing.

### Lessons Learned from Autonomy at Sea

Based on the preliminary testing and risk analysis, it is evident that MASS is a system of systems, depending on local sensor systems, automated port services, communication with RCC, other autonomous ships, conventional ships, Vessel Traffic Centres (VTS) and similar. These interactions are critical factors and should be addressed in design and operations. The degree of autonomy varies and is affected by the complexity of the operation. A MASS will operate in phases with transitions between human control and automation control. A well-defined operational envelope is key for addressing safety issues and carrying out a risk assessment. Potential hazards within each transition must be identified with fallback procedures in place, with focus on the sensemaking process and how humans should enter the control loop.

Challenges related to communicating the intent of a MASS in interactions between autonomous, unmanned ships and manned ships are addressed by Porathe (2019). The authors argue for "automation transparency" and methods allowing other seafarers to "look into the mind" of the autonomous ship, to see if they themselves are detected, and the present intentions of the MASS, i.e. sensemaking among all actors. This can be done by sharing information about the intention, what the automation knows about its surroundings, what other vessels are observed by its sensors and similar by a live chart screen accessible on-line through a web portal by other vessels, VTS, coastguard, etc. Such a common system could be the responsibility of the VTS and should be specified as a requirement for the operational design domain and the operational envelope.

In a guideline from the Bureau Veritas (2019), several hazards are listed as important: voyage, navigation, object detection, communication, ship integrity, machinery and related to systems, cargo and passenger management, remote control and security. Within each of them, a list of factors is mentioned. Using this, Hoem et al. (2019) identified a list of hazards comparing autonomous and manned ships. The scenarios were focussed on the following differentiating factors: fully unmanned, constrained autonomy, RCC, higher technical resilience and improved voyage planning. The paper gave a draft attempt to classify risk factors that can either be characterised as new types of incidents caused by technology, what is most characterised in regard to today's incidents in shipping and if the incidents are averted by crew today. As an example, the category fully unmanned points to a higher risk for technical failure but may improve some of today's operators' errors caused by poor design and lack of good human factor engineering practice. Important factors moving forward are robust sensor quality, redundancy on key technology and good education for land-based operators that support sensemaking and build situational awareness. It is likely that humans are not continuously monitoring one vessel at a time but will be needed to supervise and intervene when necessary. For a constrained autonomous vessel, the paper pointed to the need for better HAI due to the need of time to support sensemaking and get situational awareness before action.

## Autonomy in Air

Automation and autonomy in aviation have been implemented since World War II, where functions have been systematically automated and the manning has been systematically reduced. Incidents due to automation happen, but aviation safety (commercial passenger traffic) is extremely high.

In addition to increased automation in manned flights, the use of drones or unmanned aerial systems (UAS) has risen significantly in the last years. Examples of use are:

- Photography and video recording to support information and crisis management
- Inspection of (critical) components to improve safety, avoid human exposure, reduce costs or improve quality
- Detection and survey of environmental issues, such as gas emissions, ice detection in sea, overview and control of pollution
- Logistics – delivery of critical components or supplies (such as medicine, blood)

### Safety Challenges

Manned flights have a high level of safety, issues have often been a result of poor sensemaking and poor situational awareness of the crew. The reliability of the technical equipment is high. Automation accidents have happened lately where guidelines during design and certification have not been followed. This was the case in the Boeing 737 MAX fatal crashes (Cruz and de Oliveira Dias, 2020). After analysing the accidents, Endsley (2019) recommended ensuring compliance with human factors design standards and support for human factors assessment in aircraft testing and certification.

Safety challenges in UAS differ from the challenges in manned operations, due to the immaturity of technology. Looking at the use of large drones in the US, Waraich et al. (2013) documents that mishaps may happen more frequently (i.e. 50–100 mishaps occur every 100,000 flight hours vs human-operated aircraft where there is one mishap per 100,000 flight hours). The mishap rate is 100 times higher in UAS remotely piloted than in manned operations. The leading causes are poor attention to human factors science, such as poor design of human machine interfaces in ground control centres (Waraich et al., 2013; Hobbes et al., 2014).

In Petritoli et al. (2017), the mean time between failures (MTBF) estimated for UAS was around 1,000 hours, approximately 100 times higher than MTBF in manned flights. The dominant failures were in power systems, ground control system and navigation systems.

The risks of UAS operations are dependent on the operational domain, i.e. the type of operation (delivery, data collection, surveillance, inspection photography, etc.) and physical details of the drone such as weight, speed and height of operation. EASA (2016) has estimated the probability of fatality of different UAS weights and estimated probability of fatality as 1% with a UAS weight of 250 g, but 50% fatality with a weight of 600 g in case of a collision with a human when the drone drops.

Examples of undesired incidents from UAS are: collisions with personnel; interference with infrastructure (infrastructure such as airports is vulnerable and interference may lead to disruption of air traffic); actual damage to critical infastructure; damage to the drone; using the drone to spy or steal data (leading to loss of privacy, data theft and possible emotional consequences). Automated systems and UAS are vulnerable to attacks through the cyber-physical systems it consists of, such as sensors, actuators, communication links and ground control systems. As an example, an Iranian cyber warfare unit was able to land a US UAS based on a spoofing attack modifying the GPS data (Altawy et al., 2017).

There are several challenges of UAS operations in challenging climatic conditions such as low temperature, wind, winter with sleet and snow. Operational equipment may not be tested or hardened for these demanding conditions; thus, requirements, testing and certification are needed. Communication infrastructure is also demanding in the north, from 70° the quality of satellite communication is degraded. GPS spoofing may be a challenge and must be mitigated.

### Lessons Learned That May Be Transferred

Automation in aviation has succeeded in establishing a high level of safety, due to systematically automating simple tasks and reducing demands on the pilot: base development on the science of human factors, building infrastructure, to control and support flights, strong focus on learning from small incidents and accidents and support from control centres that have strict control of the operational domain/operational envelope. Thus, systematic development and stepwise refinement has had a huge success in terms of safety and trust, in addition to the strong focus on keeping the human in the loop supported by sensemaking. Even in this environment of high reliability, there is a strong need to ensure compliance with human factors design standards and support for human factors assessment in aircraft testing and certification to avoid fatalities by automation as seen in the Boeing 737 Max accidents.

The reliability of drones is lower than for manned planes, and there is a need to develop improved reliability of the new technology. Systematic risk assessment is needed to mitigate the areas with the most risks. The HMI between automation and the human operator is challenging. Design must use best human factors practices to support sensemaking and ensure that the operator can intervene and take control when needed.

## Autonomy in Rail

By automated metros (rail systems), we mean systems where there is no driver in the front cabin, nor accompanying staff, also called Unattended Train Operation

(UTO). UTO has been in operations since 1980. According to UITP (2013), there is 674 km of automated metros consisting of 48 lines in 32 cities. Examples of cities with UTOs are Barcelona, Copenhagen, Dubai, Kobe, Lille, Nuremberg, Paris, Singapore, Taipei, Tokyo, Toulouse and Vancouver. There is large infrastructure cost to ensure safe on and offloading of passengers and that the track is isolated from other traffic. Four distinct levels of automation are defined:

GoA1: Non-automated train operation, with a driver in the cabin.
GoA2: Automatic train operation system controls train movements, but a driver in the cabin observes and stops the train in case of a hazardous situation.
GoA3: No driver in the cabin but an operation staff on board.
GoA4: Unattended train operation, with no operation staff on board.

### Safety Challenges

Wang et al. (2016) list the following as arguments for UTO: increased reliability, lower operation costs, increased capacity, energy efficiency and an impressive safety record. We have at present not found normalised accident data for UTO (incidents based on person km), and no accidents have been reported. We have found reports in newspapers about minor incidents, without any fatalities reported. Based on data and experiences so far, it seems that the UTO has exceptionally high safety. However, more systematic analysis and normalisation of all international UTO transport incidents are needed.

Even though driverless trains have an impressive safety record, experience shows that they still face some challenges related to reliability and operability. One example of this is seen in Singapore. UTOs were introduced in Singapore's Mass Rapid Transits (MRT) system in 2003. Here, the operations were monitored remotely from an operations control centre. However, in 2018, most of these trains were manned again, for improving reliability. Some of the trains experienced technical issues and failures. In these cases, a driver on board a train will immediately be able to assess the problem, and, if necessary, push another disabled train out of the way. With a driverless system, a driver had to make his way to the unmanned train, which takes time. Nevertheless, the safety record of driverless trains is impressive, maybe due to the rail track as a system. Hence, further automation of railway systems is ongoing.

### Lesson Learned

As mentioned, it seems that the UTO has an exceptionally high level of safety. However, systematic analysis and normalisation of all international UTO transport incidents are needed. Thus, there is a need for systematic reporting and analysis of minor incidents/small accidents in order to support risk-based regulation and risk-based design of the technology.

A key issue related to safety is the focus on a restricted design domain and operational envelope. The environment/context of which the UTOs operates is typically underground, with few or no interaction with other traffic. Protection systems are in place at the embarkment area/platform preventing the most common incidents (people falling on tracks). There has been a focus on analysing personnel incidents when entering and leaving the UTOs and building safer infrastructure to minimise dangerous situations.

## Autonomy on Road

Cities worldwide are increasingly testing and implementing autonomy as the pace of autonomous vehicle innovation picks up. Norway has long-term experiences of autonomous transport systems such as Automated Guided Vehicles (AGVs) at St. Olav Hospital and autonomous shuttle buses used from January 2018 on public roads.

**Projects with autonomous vehicles (AVs):** Local governments must approve self-driving pilots. In the US, in California, all companies must deliver annual self-reports on incidents with highly automated vehicles. (This is one of the reasons why Uber and many other companies moved the testing of self-driving taxis to Arizona that has adopted a more liberal attitude.) This framework condition, i.e. legislation in California, has enabled the industry to document the level of safety and identify challenges.

Related to the present development trends, there are two clear trends that are different in nature:

1. a race to develop fully AVs, i.e. self-driving cars, aiming to replace today's private cars.
2. an effort to develop fully AVs to provide mobility-as-a-service (MAAS) or robotaxis.

The aim of the private self-driving car segment is to operate more safely than human drivers are able to in real-world conditions and at high speed. Here, the self-driving cars must be able to handle all types of obstacles and interactions with other road users in all kinds of weather and traffic conditions.

The MAAS segment focusses on small shuttle buses (or robotaxis) with geofencing to establish a safe route. Many of these are unable to go around an obstacle. They stop until the obstacle has moved or been removed. They operate at low speeds between 12 and 30 km/h.

There are many projects with self-driving vehicles on public roads operating around the world. According to Philantropies (2017), at least 53 cities are currently involved in testing AVs. Legal frameworks for the regulation of pilot testing are established in Singapore, the Netherlands, Norway and the UK (KMPG, 2018). Euro NCAP has designed a set of test procedures for testing automated vehicles on SAE level 2. The US Department of Transportation has developed a framework (NHTSA, 2018) for testing automated driving systems focussing on failure behaviour, failure mitigation strategies and fail-safe mechanisms.

**AGVs at St. Olav Hospital** have been in operation since 2006. Today, 21 AGVs operate at a speed of approximately 2 km/h (max speed is 5 km/h) and communicate with each other, open doors and reserve elevators. The automation is quite simple as they follow a predefined path, and when there are conflicts or problems with collisions/doors/elevators, a signal is given to the operational centre, always manned by an operator who can intervene or go to the place. Manned operators in the centre are necessary to ensure continuous operations. Even in this strict operational envelope, humans are critical components in the loop. Sensemaking has been in focus, examples are that the AGVs are "speaking" to hindrances/people – saying "please move" or "this elevator is reserved".

**Pilots with autonomous shuttle buses:** From 2017, testing of AVs was allowed in Norway. In the SmartFeeder (2019) research project, initial data are gathered from five test sites with MAAS pilots. Each pilot tests self-driving shuttle buses carrying up to six passengers, operating at an average speed of 15 km/h, and with an operator to monitor and take over control if necessary (during the test phase). These pilots are "fixed route autonomy", where the autonomous system follows a predefined route and processes a limited amount of sensor data along the route. The motivation varies, i.e. solving a last mile problem (connecting workplaces with public transportation), testing out technology and user acceptance or property and business development. In total, the buses in the pilots have driven almost 22,000 km, with approximately 40,500 passengers in both summer and winter conditions. Initial data have been collected regarding disengagement of the system and involvement of the operator in the pilots in three categories: "obstacle emergency stop" (sensors detect something and automatically stop), "soft stop" (operator overtakes system and decelerates the vehicle) and "Manual switch" (for manually driving the vehicle). The collected data are currently being processed and cleaned for more detailed analysis, and interpretations cannot be drawn yet. However, the reliability and robustness are challenging, and demands a restricted operating envelope in addition to the need for "humans in the loop" when the unanticipated is happening.

## Safety Challenges

Tesla with its autopilot has enabled automated driving at high speeds. Several severe accidents with Tesla autopilot have led Tesla to limit their autopilot functionality. These partially automated vehicle systems at SAE level 2 (SAE, 2018) always operate exclusively based on an attentive driver being able to control the vehicle. For fully automated driving (SAE level 4–5), the driver is no longer available as a backup for the technical limits and failures. Replacing human action and responsibility with automation raises questions of technical, ethical and legal risks, as well as product safety.

As far as we know from media and public accident reports there have been four fatal accidents worldwide: three with semi-automated (SAE level 2) autopilot and one with a more fully automated vehicle on public roads (SAE level 3), the Uber accident in Arizona where a Volvo refitted with Uber self-driving technology killed a pedestrian (NTSB, 2018). In all cases, the autopilot was engaged but without driver interaction or intervention with vehicle controls, highlighting the need for sensemaking and "meaningful human control".

There are few safety records (data) on SAE level 4 so far. Data from 2009 to the end of 2015 collected by Google's cars list three police reportable accidents in California while driving at 2,208,199 km (Teoh and Kidd, 2017). This is 1/3 of reportable accidents per km of human-driven passenger vehicles in the same area. In 2017, 19 of 21 reported accidents with Google-Waymo cars (level 4) were rear-ended accidents at signalised intersections. This is caused by ordinary drivers' misinterpretation of automated vehicle behaviour (as an example expecting that drivers are not halting when meeting a yellow light at an intersection.). Google-Waymo has now patented a software program allowing their vehicles to drive through yellow light. A look at accidents and incidents reported to the California Department of Motor

Vehicles (DMV) in 2019 shows that other 65 companies currently testing level 4 technology still have frequent rear-end collisions at signalised junctions. They also have trouble (and reported accidents) entering a motorway from the ramp. AVs have not yet learned the "nudging" that ordinary drivers do to see if traffic on the motorway yield and let you in.

**Experience from the autonomous shuttle buses:** For the pilots, it was mandatory to report incidents and accidents. No persons were injured, and only minor technical issues and malfunctions were reported. The following issues were revealed:

- Snow, heavy rainfall and fog are challenging for the sensors.
- Vegetation and light poles along the route of the bus is challenging as they interfere and disturb the sensors at times.
- The buses run along the same "track" with narrow wheels, causing significant wear and tear on the road along this track.
- Cyclists passing near the bus makes the bus stop abruptly.

These issues are related to the predefined operational envelope surrounding the vehicle, leading to abrupt stops when violated. As pointed out by Jenssen et al. (2019), AVs lack a sense of self, and software and sensors are still not designed to account for the discrepancy in the same way human drivers are able to.

When applying for testing, a mandatory risk assessment was carried out. The main risks listed were related to passenger injury as a result of an abrupt stop where passengers inside the bus are unprepared and can be harmed by falling. Risk-reducing measures are lowering the speed, installing seat belts, limiting the number of passengers and adding road signs.

**AGVs at St Olav:** A total of 100–130 minor incidents per year have been reported. Yearly, each AGV experiences around 15 emergency stops (Johnsen et al. 2019), where components must be changed. Reported incidents are minor crashes as a consequence of faulty navigation due to objects placed in the route, summarised in Johnsen et al. (2019). From interviews with the operators of the AGVs, the following main issues are identified:

- The AGVs ability to adapt to the surrounding infrastructure
- Keep the track of the AGVs clear of objects
- Make objects visible to the AGV: the AGVs are not able to detect all obstacles due to the sensor range
- Establish a control room with proper HMI design
- Maintain the interface to cyber physical systems: software updates has led to problems (due to poor testing and multiple vendors.)

## Lessons Learned

Vehicle automation can enhance safety but also introduces new risks due to poor technical implementation and the need for rapid response from the human actor. This is especially the case with SAE automation levels 2 and 3.

The accident data collected so far with automation (AGVs and level 1–4 vehicles) indicate safety hazards of human factors and technical issues, i.e. obstacle detection (sensors), programming (rule-based and not artificial intelligence, AI), prolonged attention (humans in the loop), HMI (Autopilot-engagement rules) and misuse. The list may become longer as more safety data are gathered and more in-depth information on accident causality of automated vehicles is established, e.g. overreliance and expectation mismatch.

Based on the experiences, there is a need to establish regulations that ensure systematic incident reporting, develop systems based on learning from incidents and invest in infrastructure to support automation, i.e. help the automation by focussing on an operational envelope that uses more data from infrastructure. The transport systems are automated but not autonomous. Autonomous systems are immature at present and must be further developed.

## A SUMMARY OF MTO SAFETY ISSUES

Based on the performed reviews, the suggested key measures are listed below.

**Humans:** As seen from all experiences, the uncertain and complex environment for autonomous systems must ensure the need for human intervention. Autonomous transportation systems will to a varying degree need human control if failures occur or under certain operational conditions. With today's UTOs and AGVs, an operator is still needed when there is a disruption and sensors fail to detect and recognise an obstacle or determine the next actions. However, in testing and developing autonomous transportation systems with drones, AVs and vessels, we see examples of projects where the human operator is not considered from the beginning. The industries' motivation seems to be to try to automate as much as possible and assume that humans will and can monitor it. Hence, HAI and how to keep the humans in the loop is often considered a challenge to be solved late in the project after knowing the limitations of the technology and by considering the humans as the adapting back-up. Most of the projects lack early incorporation of human factors in analysis, design, testing and certification process. Thus, there are costly challenges that should have been addressed earlier by starting with technology, human limitations and possibilities, and organisational and infrastructure needs. A key issue is to define the design conditions the system should operate under by defining the operational envelope and critical scenarios (such as sensor failures). Then specify how critical scenarios can be mitigated by infrastructure support i.e. surrounding systems such as other autonomous systems nearby (cars) or control infrastructure. If human intervention is needed to handle the scenarios, sensemaking must be supported within the existing limitation of human abilities.

As aviation is the industry with the most experience with safe automated systems, the list from Endsley (2019) with design principles for improving people's ability to successfully oversee and interact with automated systems should be a very useful element, allowing for manual overrides and sufficient training to users on automation to ensure adequate understanding and appropriate levels of trust.

**Technology:** To date, developing autonomous or remotely controlled transportation systems (especially for AVs and MASS) appears to primarily be about a technology push rather than considering and providing sociotechnical solutions including redesign of work, capturing knowledge and addressing human factors as we and others have seen (Lutzhoft et al., 2019).

Technology in autonomous systems and their interpretation (such as through AI) are not reliable at present – thus, there is a need to address poor reliability trough improving man/technology/organisation aspects. The reliability of drones is lower than for manned planes, and we have seen how sensors and technical equipment are causing safety issues in several projects. The systems must improve for an industrial setting and for safety-critical operations, i.e. become highly reliable and resilient to bad data and have automatic self-checking behaviour and avoiding single-point failures by checking across multiple inputs. Thus, there is a need to get support from other AVs with sensors, need for developing infrastructure (such as roads and seaways with sensors), in addition to establishment of control centres for road traffic and maritime traffic that must be responsible for supporting sensemaking among the actors (i.e. automated and not automated systems). Technical barriers must be in place to a larger extent on autonomous systems to avoid and reduce the outcome of failures and component interaction accidents, which are more common as the complexity increases.

Automation transparency is important for both sharing the situation awareness and communicating the intentions towards others and for the operator in an RCC to understand the behaviour of the automation. In complex systems, a wide range of alarm issues related to diagnostics, management and assessments of multiple input data will be challenging. Hence, alarms must be unambiguous and displayed with a clear message. This requires good human factor engineering practice, such as an alarm philosophy and relevant standards.

**Organisation:** Experience from the projects and pilots demonstrate a need to see the technological solution in a larger sociotechnical context. Autonomous transportation systems are a system of systems. We have seen that legislation is are needed to gather data and establish the operational context. There is a need for substantial investments in infrastructure: organisational interfaces are lacking and organisational/structural issues from the operator/company/area/society are often considered the last thing to get in place. Looking at the operational context, we have seen a need to limit the operational design domain and use operational envelopes, or safety envelopes to define situations, responsibilities and system characteristics during all conditions (especially in safety-critical conditions with sensor/data failures). Regulations and guidelines have slowly been established to support autonomous transportation systems. However, few of them require systematic reporting of accidents and incidents. Experience from accidents with AVs has given valuable insight, and hence all domains should prioritise and require reporting and systematic data collection of failures, hazards and unforeseen events. Not requiring reporting and sharing of safety-critical systems is a risk in itself.

## SENSEMAKING TO SUPPORT MEANINGFUL HUMAN CONTROL

Focus on the design of operational envelopes to reduce complexity and analysing the needs for cues and information to support sensemaking and meaningful human control, when needed, is a key issue. Defining operational envelopes answers the question of which functions and roles automation/autonomy should have, versus humans, when designing a complex system. This is also an important question for certification of the autonomous transportation system.

Sensemaking and the principle of meaningful human control should be used to verify that the proper functions are allocated to the human or the automation. According to Santoni de Sio and van der Hoven (2019), two design requirements should be satisfied for an autonomous system to remain under meaningful human control:

1. A "tracing" condition, according to which the system should be designed in such a way as to grant the possibility to always trace back the outcome of its operations to at least one human along the chain of design and operation.
2. A "tracking" condition, according to which the system should be able to respond to both the relevant moral reasons of the humans designing and deploying the system and the relevant facts in the environment in which the system operates.

From a safety perspective, this can be placed in the bowtie model, where the design principle of tracking are barriers preventing a technical fault, threat or unexpected situation to lead to a dangerous situation, as a human alway has established the possibility to intervene and take over control. On the other side of the bowtie, once a hazard has emerged, the outcome can be reduced by designing after a tracing condition making it possible to trace back the operation to a human who is in the position to understand the capabilities of the system and the possible effects in the world of its use and, hence, knows how to limit the consequences of an undesired event.

## CONCLUSION

We have given a summary of ongoing projects and safety issues. The main issues across the domains are technical reliability and maturity, the need for automation transparency (including awareness for the decision made by automation), the need for defining what conditions the system can operate under and assigning responsibilities to human operators and the automation. Experiences from known accidents involving a high level of automation, as in the cases of Boeing 737 MAX, Uber and Tesla, have shown overreliance on automation and poor understanding of capabilities and limitations. We need to collect and systemise data on accidents and incidents of autonomous transportation systems and design with human factor practice to support sensemaking and meaningful human control.

Design principles from meaningful human control should be used to verify if the interaction between automation and the human is safe. This can be used as an input to operational envelopes and to assist in the design of a good HAI supporting sensemaking.

## ACKNOWLEDGEMENT

This chapter has been funded by the Norwegian Research Council – project 267860 SAREPTA.

## REFERENCES

AAWA (2020). https://www.rolls-royce.com/media/press-releases/2016/pr-12-04-2016-aawa-project-introduces-projects-first-commercial-operators.aspx

AMOS (2020). https://www.ntnu.edu/amos/research

Autosea (2020). https://www.ntnu.edu/autosea

Autoship (2020). https://www.kongsberg.com/maritime/about-us/news-and-media/news-archive/2020/autoship-programme/

Bureau Veritas (2019). NI 641 R01 *Guidelines for Smart Shipping*.

Chinen, M. (2019). Law and Autonomous Machines. *Elgar Law, Technology and Society* (p. 109). Edward Elgar Publishing.

Cruz, B. S., & de Oliveira Dias, M. (2020). Crashed Boeing 737-MAX: Fatalities or malpractice? *GSJ* 8 (1), 2615–2624.

Cummings, M. L. (2019). *Lethal Autonomous Weapons: Meaningful human control or meaningful human certification?* IEEE Technology and Society.

Endsley, M.R. (2019). *Human Factors & Aviation Safety* Testimony to the United States House of Representatives. Hearing on Boeing 737-Max8 Crashes, December 11, 2019.

Fjørtoft, K. E., & Rødseth, Ø. J. (2020). *Using the operational envelope to make autonomous ships safer* Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio.

Hoem, Å. S. (2019). The present and future of risk assessment of MASS: a literature review. *29th European Safety and Reliability Conference*. European Safety and Reliability Association.

Hoem, Å.S., Fjørtoft, K., & Rødseth, Ø. (2019): *TransNAV 2019: Addressing the Accidental Risks of Maritime Transportation: Could Autonomous Shipping Technology Improve the Statistics?*

Hollnagel, E., Nemeth, C. P., & Dekker, S. (Eds.). (2008). *Resilience engineering Perspectives: Remaining Sensitive to the Possibility of Failure* (Vol. 1). Ashgate Publishing, Ltd.

Horowitz, M., & Scharre, P. (2015). *An Introduction to Autonomy in Weapon Systems*. Center for a New American Security (CNAS) Working Paper (CNAS: Washington, DC), p. 8

IMAT (2020). https://www.sintef.no/projectweb/imat/

INAS (2020). http://www.autonomous-ship.org/index.html#H2

Johnsen, S. O., Hoem, Å., Jenssen, G., & Moen, T. (2019). Experiences of main risks and mitigation in autonomous transport systems. *Journal of Physics: Conference Series* 1357 (1) 012012.

Kilskar, S. S., Danielsen, B. E., & Johnsen, S. O. (2020). Sensemaking in critical situations and in relation to resilience—a review. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, 6(1).

KMPG (2018). Autonomous vehicles readiness index. *Klynveld Peat Marwick Goerdeler* (KPMG) International.

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 (1), 012032.

MUNIN (2020). http://www.unmanned-ship.org/munin/

NFAS (2020). http://nfas.autonomous-ship.org/index.html

NHTSA (2018). *A Framework for Automated Driving System Testable Cases and Scenarios*. DOT HS 812 623. https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13882-automateddrivingsystems_092618_v1a_tag.pdf

NTSB (2017). 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. Highway Accident Report NTSB/HAR-17/02. Washington, DC.

NTSB (2018). *National Transportation Safety Board 2018*. Preliminary Report: Highway HWY18MH010.

Porathe, T. (2019). Interaction between Manned and Autonomous Ships: Automation Transparency. *Proceedings of the 1st International Conference on Maritime Autonomous Surface Ships*.

Porathe, T., Hoem, Å., Rødseth, Ø. J., Fjørtoft, K., & Johnsen, S.O. (2018). At least as Safe as Manned Shipping? Autonomous Shipping, Safety and "Human Error". *Proceedings of ESREL 2018*, June 17–21, 2018, Trondheim, Norway.

Ramos, M. A., Utne, I. B., Vinnem, J. E., & Mosleh, A. (2018). Accounting for Human Failure in Autonomous Ship Operations. Safety and Reliability–Safe Societies in a Changing World. *Proceedings of ESREL 2018*, June 17–21, 2018, Trondheim, Norway.

Relling, T., Lützhöft, M., Ostnes, R., & Hildre, H. P. (2018). A Human Perspective on Maritime Autonomy. *International Conference on Augmented Cognition* (pp. 350–362). Springer, Cham.

Rødseth, Ø. J. (2018). Defining Ship Autonomy by Characteristic Factors, *Proceedings of ICMASS 2019*, Busan, Korea, ISSN 2387–4287.

SAE International (2018). Standard, SAE J3016_201806. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. Revised

Santoni de Sio, F., & Van den Hoven, J. (2018). *Meaningful human control over autonomous systems: a philosophical account*. *Frontiers in Robotics and AI* 5, 15.

SmartFeeder (2019). https://www.sintef.no/prosjekter/smart-feeder/ (in Norwegian)

Teoh, E. R., & Kidd, D. G. (2017). Rage against the machine? Google's self-driving cars versus human drivers. *Journal of Safety Research* 63, 57–60.

Thieme, C. A., Guo, C., Utne, I. B., & Haugen, S. (2019, October). Preliminary Hazard Analysis of a Small Harbour Passenger Ferry–Results, Challenges and Further Work. *Journal of Physics: Conference Series* 1357 (1), 012024).

UITP (2013). *Observatory of Automated Metros World Atlas Report*. International Association of Public Transport (UITP), Brussels

Wang, Y., Zhang, M., Ma, J., & Zhou, X. (2016). Survey on driverless train operation for urban rail transit systems. *Urban Rail Transit* 2, 106–113. https://doi.org/10.1007/s40864-016-0047-8

Yara Birkeland (2020). https://www.kongsberg.com/maritime/support/themes/autonomous-ship-project-key-facts-about-yara-birkeland/

Zeabuz (2020). https://zeabuz.com/

Article 4

# Adopting the CRIOP Framework as an Interdiciplinary Risk Analysis Method in the Design of Remote Control Centre for Maritime Autonomous Systems

Hoem, Å. S., Rødseth, Ø. J., & Johnsen, S. O. (2021). In conference proceedings for the International Conference on Applied Human Factors and Ergonomics (pp. 219-227). Springer

# Adopting the CRIOP framework as an Interdisciplinary Risk Analysis Method in the Design of Remote Control Centre for Maritime Autonomous Systems

Åsa S. Hoem[1], Ørnulf J. Rødseth[2], Stig Ole Johnsen[3]

[1] Norwegian University of Science and Technology (NTNU), Department of Design,
7491 Trondheim, Norway
[2] SINTEF Ocean, Department of Energy and Transport,
7052 Trondheim, Norway
[3] SINTEF Digital, Department of Safety,
7052 Trondheim, Norway

Humans are increasingly asked to interact with automation in complex and large-scale systems. The International Maritime Organization (IMO) has started working on regulations for Maritime Autonomous Surface Ships (MASS). For the foreseeable future, unmanned ships will most likely be under supervision from a Remote Control Centre (RCC), called constrained autonomy. We see a need to include the end-user and carry out a risk-based design analysis, considering the operational quality of the RCC. This paper proposes an approach based on the CRIOP method, short for Crisis Intervention and Operability analysis. Could this framework be adapted to the evaluation of RCC used for MASS operations? What critical scenarios should be used for evaluations of the design/HMI of an RCC? The paper recommends Operational Envelopes to describe the constraints of the system and concludes with recommendations regarding an interdisciplinary, collaborative, and anticipatory analysis of the HMI to enhance operator performance and reliability.

**Keywords:** HAI, HMI, Remote Control Centre, Maritime Autonomous Systems, Risk Analysis, ConOps, Use Case, CRIOP, Scenario Analysis,

## 1 Introduction

On the topic of MASS, the majority of papers published to date focuses on technical aspects of the ship operations and design, indicating that most scholars focus on the high-end components of the system, while organizational and human-oriented issues remain under-explored [1]. Without changes in the regulatory framework, safe interactions between conventional ships and MASS will be a significant challenge. In the foreseeable future, it is doubtful that MASS can operate without human supervision and intervention [2]. Thus, a technology-centred approach will miss the critical human element in MASS operations. Focus on controls, software, and sensors will inevitably be of limited use if little attention is afforded to the human operators' needs in the larger system [3]. This article presents a method to facilitate risk analyses to ensure a safe and resilient design of an RCC and the human-automation interface (HAI).

## 2 Background

MASS could better be an abbreviation for Maritime Autonomous Ship *System*, as they are complex socio-technical systems consisting of equipment, machines, tools, technology, and a work organization. The system includes functions on the ship as well as onshore – not the least the RCC. Designing such a system should follow principals of socio-technical design, like involving the future users of the new systems. Some of the leading methods for assessing safety in complex systems (e.g. STAMP, FRAM), take the necessary systemic perspective that explores the relationships between causal factors within the systems and addresses the complexity known to be important for improving safety in modern organizations [4]. However, for novel systems like MASS, the knowledge level on detailed designs is low, and the uncertainty still high.

Consequently, it is not easy to apply such systemic safety models to support the initial design phase as they rely on detailed and high-qualitative data. Besides, the methods share a challenge of being time-consuming, resource-intensive and needing extensive expert knowledge to facilitate the analysis. In this early phase, we need a more straightforward cross-disciplinary method, including the end-user, to carry out a risk-based design analysis.

## 3 Risk-Based Design

According to current best practice, MASS will have to be approved according to principals for "Alternatives and Equivalents" [5], which is fundamentally a risk-based approach. In national guidelines, this is partly translated to a strong focus on the ship's intended operation that needs to be described in detail [6]. This description is part of the Concept of Operations (CONOPS) that most class societies and the Norwegian Maritime Authorities requires. Risk-based design (also known as Design for Safety) is a formalized methodology, introduced in the maritime industry as a design paradigm to help bestow safety as a design objective and not a constraint. In short, it means carrying out risk analysis and consider potential risk in the different phases of design and hence treat safety as a life cycle issue. The goal is to use the information obtained from the analysis to engineer or design out accidents before they occur. A risk-based approach is recommended by Lloyd's Register [7] and DNV [8]. Structured risk-analyses should be performed on several abstraction-levels, typically utilizing several different risk-analyzing methodologies [8]. One method is the CRIOP method, which can describe and model risk qualitatively and use best practices to ensure that human factors issues are integrated into the design.

## 4 CRIOP – Crisis Intervention and Operability Analysis

CRIOP is an established, standardized scenario method for Crisis Intervention and Operability analysis. The methodology was developed primarily for the oil and gas industry, back in 1990 [9]. The initial scope was a scenario-and-general-checklist method for evaluating offshore control centres (CC) focusing on the human aspects in terms of conditions for successful crisis handling. Since then, the methodology has developed through collaborations between regulatory authorities, operators, research institutions, contractors and consultants, to include/consider HMIs, best practices standards and Human Factors. Integrated operations and e-Operations are now included as remote support, or remote operations are more common, due to organizational and technical changes. Today, CRIOP is used to verify and validate an RCC's ability to handle all operational modes safely and efficiently, i.e. normal operations, maintenance, disturbance/deviations, safety-critical situations.

The key elements of CRIOP are checklists covering relevant areas in the design of a control centre, Scenario Analysis of critical scenarios and a learning arena where the operators, designers and managers can meet and evaluate the optimal control centre [9].

The CRIOP process consists of four major work tasks:

1. **Prepare and organize** by defining, gather necessary documentation, establish an analysis group, identifying relevant questions and scenarios and set a schedule.
2. **General Analysis (GA)** with checklists to verify that the CC satisfies the stated requirements based on best industry practice (a standard design review).
3. **Scenario Analysis** of critical scenarios. An experienced team of end-users should perform the analysis to validate that the control centre satisfies the actual needs.
4. **Implementation and follow up:** At the end of task 2. and 3. the findings and recommendations are documented, and an action plan is established.

The method can be applied at different phases of the lifecycle, as shown in Fig. 1 below.

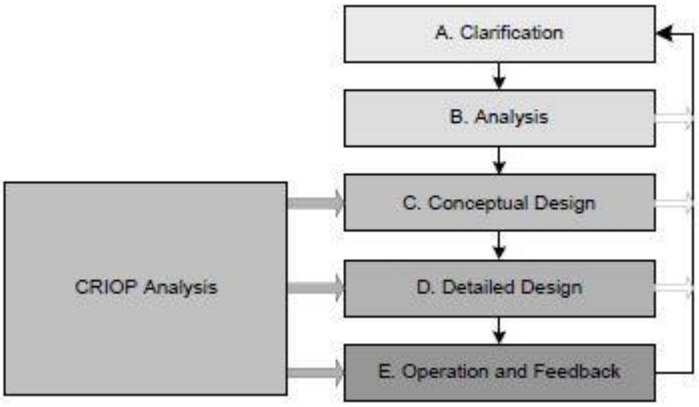


**Fig. 1.** Integration of CRIOP analysis in ISO 11064 design process (adapted from [9]).

This paper focuses on the methodology's applicability in the early phase, the Conceptual Design phase. Here, concepts, automation level, HMI/Alarms (displays, controls, and communication interfaces), and necessary layouts should be developed.

Results from preliminary task analysis, function allocation and job design analysis should be available before starting a CRIOP. However, RCC for MASS does not yet exist. Hence, such analyses are difficult to conduct due to the lack of established domains or users. Based on methods presented in [10], a pilot domain must be created. With a layout of a pilot domain for an RCC with operational envelopes in place, the CRIOP process can start.

We ask if the CRIOP framework could be adapted to the evaluation of RCC used for MASS operations. The general checklists must be updated, but the core ergonomics and risk-influencing factors (alarm philosophy, physical work environment, training) will be similar for an RCC for MASS and an offshore installation. Nevertheless, the risk analysis of a MASS and an offshore installation is quite different. We ask what key scenarios should be used for evaluations of the design/HMI of an RCC. Hence, we focus on the applicability of the work task 3 in the framework, the Scenario Analysis.

## 5 Operational envelope and use cases

The AUTOSHIP project has published an architectural concept [11], where the MASS' intended operations are broken down into smaller sets of generalized tasks, i.e. use cases. Each use case will be defined by operational constraints, e.g. geographic complexity, traffic complexity, worst-case weather, visibility conditions, etc. Together, these use cases define the MASS' Operational Envelope. This concept was first proposed in [13], calling it the operational design domain (ODD). The name was later changed to Operational Envelope to distinguish it from the ODD often used in the context of autonomous cars.

Each use case in the operational envelope describes and define both the automation's and the human's responsibilities, and the conditions that determine when responsibilities changes. [14] introduce two other important concepts, the maximum response time $T_{MR}$, and the response deadline $T_{DL}$. $T_{MR}$ is the maximum time interval a human operator need from an alert is raised to he/she is at the control position and has gained sufficient situational awareness to take safe action. $T_{DL}$ is defined as the minimum interval until a situation arises that the automation cannot handle. [12] introduced the idea of Constrained autonomy, which is now formally defined as a property of a sub-space of the operational envelope where the automation system at all times can calculate $T_{DL}$. By issuing an alert to the operator when $T_{DL} \leq T_{MR}$, one can assure that the operator will intervene in time when the automation can no longer handle a situation. The operational envelope also includes descriptions of what happens when the envelope is exceeded. The MASS must then fall back to a state that poses the least risk to life, environment, and property, so-called "Minimum Risk Condition" (MRC).

## 6 Remote Control Centre

As MASS are novel systems, one of the main challenges is that we have no experience from the operation or design of an RCC for MASS yet. We must base our experience from other domains such as aviation, automated road transport, or centralization of ship control done on the bridge. However, some basic principles are known:

i. Most of the time, ship operations are relatively easy to automate, e.g. transit in fair weather and non-complex traffic situations. These operations should be automated, and it is not necessary or desirable to have an operator in or on the control loop. It will be too boring for a human.

ii. More complex situations will typically develop slowly and can be identified early by the automation system, e.g. worsening weather or increasing traffic ($T_{DL}$ is known and relatively long - on the order of half an hour).

iii. Even in a more complex situation, it should be possible to automate operations, e.g. sailing in more congested waters. Automation should, in most cases be able to handle encounters between one other ship and the MASS. However, the situation becomes more ambiguous with two or more other ships ($T_{DL}$ is known but is shorter – on the order of minutes). The safe state could be to halt ships or reduce speed to mace the situation controllable – thus, controllability is a crucial issue.

iv. A primary driving factor for MASS is to operate many smaller ships rather than one large. Having smaller vessels increases the frequency of service, which is necessary to, e.g. transfer cargo from road to sea [12]. With crew onboard, this will not be economically feasible. There will be more than one ship to monitor from the RCC.

Based on these principals, the RCC operators will typically be in charge of several ships and not closely monitor only one ship. They will be alerted to situations that the automation cannot handle and will need to take the right action. Different types of ships and shipping operations may require other RCC configurations.

## 7 Review of the CRIOP Scenario Analysis

The Scenario Analysis is designed to verify that the CRO (Control Room Operator) can perform the task while considering cognitive abilities, human-system interaction and other performance shaping factors. The analysis is human-centred, focusing on the CRO's interaction with the system, including communication with other personnel. Emphasis is on how the systems support the operator's situation awareness and decision making in different situations.

The Scenario Analysis assesses the RCC's actions in response to possible scenarios. Based on the scenarios, a dynamic assessment is made of interaction between essential factors in the control room, e.g. presentation of information and time available. The methodology suggests using Sequentially Timed Events Plotting (STEP) diagrams for a graphic presentation of the scenario events. For each event, questions related to the SMoC (Simple Model of Cognition) should be asked. A checklist of performance shaping factors should also be used to ask additional questions to elaborate on answers received.

The Scenario Analysis follows four main activities:

1. Selection of a realistic scenario
2. Description of the scenario employing a STEP diagram
3. Identification of critical decisions
4. Analysis of the decisions and possible evaluation of barriers

## 7.1 Selection of realistic hypothetical scenarios

CRIOP recommend adapting scenarios based on incidents that have occurred and hypothetical incidents constructed by the analysis group. For MASS, when the operations are described in the operational envelope, the use cases will directly define scenarios. The challenge is to select the most critical ones and investigate if the use cases do not cover other critical scenarios in the operational envelope. One source for critical scenarios can come from hazard identification methods (e.g. HazId, HazOps, FMECA). It should consider both hazards like malfunctions of the system and hazards outside the control structure. A preliminary hazard analysis (PHA) is typically established in the general analysis of a concept design. Here, participants from different fields of expertise come together in brainstorming sessions to identify hazards and rank their impact. In the AutoFerry project, such analysis used a simple checklist-based approach and identified the most critical hazardous events to be related to the control system, communication between software and hardware components, the interaction between the ferry and recreational users of the channel and hacking and cyber-sabotage[15]. Wrobel made an assessment based on 100 ship accidents and suggested three prominent cases to be explored, i.e. groundings, collisions and fires [16].

MUNIN was the first project to develop a technical concept of a MASS back in 2015. Since then, several published papers discuss potential risks of MASS operations ([17],[18],[19]) contributing to a database of hazards and critical scenarios.

In reviews of risk analysis methods for MASS, the STAMP method [20] with STPA is recommended as it defines safety as a control problem, making it desirable for complex systems. The analysis identifies unsafe control actions and unsafe transition control actions that will lead to a hazard in a particular context and worst-case environment. These unsafe actions could also provide valuable input for scenarios.

### 7.1.1 Criteria for selecting scenarios

The CRIOP analysis should consider a few relevant scenarios, identified as key scenarios. In [9], the criteria for selecting these scenarios are listed. Adapted for MASS, the overall criteria should be operator involvement, hazard potential, complexity (to make sure the operators stress with peak workload) and acceptance (scenario accepted as possible by all participants).

An essential feature of MASS is the dynamic levels of autonomy that may change during a voyage depending on certain conditions. Hence the following types of human-automation interaction cases must be considered for Scenario Analysis:

1. Handover from automation to the operator. For both long and short $T_{DL}$.
2. Operator handling parts of the operational envelope that automation cannot handle.
3. Operator actions in the case of a fallback situation to MRC.

## 7.2 The STEP-model

STEP is relatively simple to understand and provides a clear picture of the course of the events to illustrate what can happen in a scenario. The graphic presentation is helpful for common ground to discuss possible hazardous events. A timeline on the horizontal axis keeps the events in order, and the connected "actors" are listed in a column. The relationship between events, what caused each of them is shown by drawing arrows to illustrate the causal links.

### 7.3 Identifying critical decisions

The analysis can start when the scenarios are documented. For each event involving an operator, questions are asked to identify how the systems support the operator's situation awareness and his/her ability to make decisions and execute actions. The CRIOP Handbook provides checklists with questions related to the scenarios and performance shaping factors depending on if the event relates to the operator receiving information (human-system interface) or making decisions (training, procedures and time available). The checklist helps identify potential error sources in the information systems, the operator's ability to achieve an adequate level of situation awareness, and whether sufficient information is available to allow the CRO to make decisions when required. Identified problems are called "weak points". Using the identified weak points, the Scenario Analysis's final step is to identify measures that should be taken to improve the identified weak points. Prior experiences suggest that CRIOP helps identify significant challenges between human operators and automation, as the best practice guidelines are used. Often mentioned issues are the ability to grasp the situation "at a glance", and simplifying automation steps such that the operator understands the action taken by the automation.

## 8 Summary

This paper presents an approach based on the CRIOP method. The framework can be adapted to the evaluation of RCC used for MASS operations. Experiences from implementing automation in other domains have found a strong need to base the development of best practices from Human Factors when there is a need for human control. CRIOP could be a risk analysis tool as we ask what can go wrong, why and how, and discuss different hazards and risks. Even though CRIOP is not based on probabilistic quantification, the participants' opinion on the scenarios is vital, contributing to a qualitative evaluation of risks. Critical scenarios for evaluations of the design/HMI should involve handover situations and fallback situations where the human operator is expected to intervene.

## 9 Need for further research

The next step is to test the feasibility of using an adapted version of CRIOP for hazard identification and assessment of a conceptual design of a real RCC. A case study with participants to validate the method focusing on the RCC and the HAI in a situation where the human is alerted to take control, is the HAI sufficiently well designed to satisfy $T_{DL}$? Furthermore, in the situations where the human operator has the responsibility for overall operations, will he/she be able to do this job at a satisfactory safety level?

**Acknowledgements**
This paper was published with support from the SAREPTA (Safety, autonomy, remote control and operations of industrial transport systems) project financed by the Norwegian Research Council with Grant No. 267860.

## References

1. Wróbel, K., Gil, M., & Montewka, J. (2020). Identifying research directions of a remotely-controlled merchant ship by revisiting her system-theoretic safety control structure. Safety Science, 129, 104797.
2. Porathe, T. & Rødseth, Ø. J. (2019). Simplifying interactions between autonomous and conventional ships with e-Navigation. Journal of Physics: Conference Series. vol. 1357:012041 (1)
3. Veitch, E., Hynnekleiv, A., & Lützhöft, M. (2020). The Operator's Stake in Shore Control Centre Design: A Stakeholders Analysis for Autonomous Ships. London, UK, DOI: 10.3940/hf.20
4. Relling, Tore (2020). A systems perspective on maritime autonomy: The Vessel Traffic Service's contribution to safe coexistence between autonomous and conventional vessels. *Doctoral Thesis at NTNU.*
5. IMO (2013). Guidelines for the Approval of Alternatives and Equivalents as provided for in Various IMO Instruments, MSC.1/Circ.1455, 24 June 2013.
6. NMA (2020), Norwegian Maritime Authorities circular RSV 12-2020, Guidelines for automation on fully or partly unmanned vessels (In Norwegian).
7. Lloyds register (2017) Code for Unmanned Marine Systems. ShipRight: Design and Construction. Additional Design Procedures
8. DNV GL (2018). Class Guideline-Autonomous and remotely operated ships. DNVGL-0264.
9. SINTEF report (2011). CRIOP Handbook from www.criop.sintef.no
10. Rutledal, Dag (2021). Designing for the unknown: Using Structured Analysis and Design Technique (SADT) to create a pilot domain for a shore control centre for autonomous ships. Submitted to the AHFE 2021 International Conference (in review).
11. Rødseth, Ø.J. (2019). Defining ship autonomy by characteristic factors. Proceedings of the 1st International Conference on Maritime Autonomous Surface Ships. SINTEF Academic Press.
12. Rødseth Ø.J., Faivre J., Hjørungnes S.R., Andersen P., Bolbot V., Pauwelyn A.S., Wennersberg L.A.L. (2020). "AUTOSHIP deliverable D3.1: Autonomous ship design standards".
13. Rødseth, Ø.J. & Nordahl, H. (2017), Definitions for autonomous merchant ships. In Norwegian Forum for Unmanned Ships, Version (Vol. 1, pp. 2017-10).
14. Rødseth, Ø.J., Psaraftis, H.N., Krause, S., Raakjær, J. and Coelho, N.F., (2020b). AEGIS: Advanced, efficient and green intermodal systems. In IOP: Materials Science and Engineering.
15. Thieme, C. A., Guo, C., Utne, I. B., & Haugen, S. (2019). Preliminary hazard analysis of a small harbor passenger ferry–results, challenges and further work. In Journal of Physics: Conference Series (Vol. 1357, No. 1, p. 012024). IOP Publishing.
16. Wróbel, K., Montewka, J., & Kujala (2017). Towards the assessment of potential impact of unmanned vessels on maritime transportation safety, *Reliability Engineering & System Safety*, *165*, 155-169.
17. Bureau Veritas (2019). NI641 guidelines for autonomous shipping.
18. Chaal, M., Banda, O. V., Basnet, S., Hirdaris, S., & Kujala, P. (2020). An initial hierarchical systems structure for systemic hazard analysis of autonomous ships. In Proceedings of the International Seminar on Safety and Security of Autonomous Vessels (ISSAV). Sciendo.
19. 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 Engineering, 202, 107188.
20. Leveson, N. G. (2016). Engineering a safer world: Systems thinking applied to safety (p. 560). The MIT Press.

Article 5

# Human-centred risk assessment for a land-based control interface for an autonomous vessel

Hoem, Å.S., Veitch, E., & Vasstein, K. (2022). In WMU Journal of Maritime Affairs 21, (pp. 179–211).

WMU Journal of Maritime Affairs (2022) 21:179–211
https://doi.org/10.1007/s13437-022-00278-y

ARTICLE

# Human-centred risk assessment for a land-based control interface for an autonomous vessel

Åsa S. Hoem[1] · Erik Veitch[1] · Kjetil Vasstein[2]

Received: 10 February 2022 / Accepted: 6 May 2022 / Published online: 15 June 2022

## Abstract

Autonomous ferries are providing new opportunities for urban transport mobility. With this change comes a new risk picture, which is characterised to a large extent by the safe transition from autonomous mode to manual model in critical situations. The paper presents a case study of applying an adapted risk assessment method based on the Scenario Analysis in the Crisis Intervention and Operability study (CRIOP) framework. The paper focuses on the applicability of the Scenario Analysis to address the human-automation interaction. This is done by presenting a case study applying the method on a prototype of a Human–Machine Interface (HMI) in the land-based control centre for an autonomous ferry. Hence, the paper presents findings on two levels: a method study and a case study. A concept of operation (CONOPS) and a preliminary hazard analysis lay the foundation for the scenario development, the analysis, and the discussion in a case study workshop. The case study involved a Scenario Analysis of a handover situation where the autonomous system asked for assistance from the operator in a land-based control centre. The results include a list of identified safety issues such as missing procedures, an alarm philosophy and an emergency preparedness plan, and a need for explainable AI. Findings from the study show that the Scenario Analysis method can be a valuable tool to address the human element in risk assessment by focusing on the operators' ability to handle critical situations.

**Keywords** Risk assessment · Scenario Analysis · Human factors · Autonomous ships · MASS · Shore control centre · Shore control centre operator

## 1 Introduction

Maritime Autonomous Surface Ships (MASS) are said to have a considerable impact on the shipping industry's sustainability, promising greener and safer solutions (e.g. Fan et al. (2020); Porathe et al. (2018)). However, because it will change the way

✉ Åsa S. Hoem
aasa.hoem@ntnu.no

Extended author information available on the last page of the article

work is done, the chance is that it will introduce new risks. Technological developments within software and hardware have led to rapidly increased automation in many systems and applications. IMO (2019) defines MASS as a ship which, to a varying degree, can operate independently of human interaction. IMO distinguishes four degrees of autonomy: (1) crewed ship with automated processes and decision support; (2) remotely controlled ship with seafarers on board; (3) remotely controlled ship without seafarers on board; and (4) fully autonomous ship. The MASS concept entails not only the ships in themselves but the complex socio-technical systems consisting of equipment, machines, tools, technology, and work organisation. Human operators have different roles and interactions with ship systems and functions in each of the listed degrees.

The degree/level of autonomy will vary in a dynamic way between full human-operated control and full machine control. This dynamic autonomy brings an additional layer of complexity to the systems and operations, especially regarding the interactions and handover between human operators and autonomous technology. For the foreseeable future, a human operator must in some way be “in the loop”, supervising the operation and on stand-by to take over control from a land-based control interface referred to as a shore control centre (SCC). Still, most of the research on MASS focuses on technical components of the system, running a risk of missing the critical human element in MASS operations.

In a study on the influence of human factors on the safety of a remotely controlled vessel, Wróbel et al. (2021) identified the shore control centre operators’ (SCCO) condition and their ability to correct known problems, to potentially have the most significant influence on the occurrence of accidents. The study also indicates that the SCCO’s action represents the final and most important barrier against accident occurrence. Designing such a system should follow principles for meaningful human control (Hoem et al. 2021; van den Broek et al. 2020) and socio-technical design, like involving the future users of the new systems, in the interim guidelines for MASS trials (IMO, 2019), the International Maritime Organization (IMO) stipulate that “for the safe, secure and environmentally sound conduct of MASS trials, the human element should be appropriately addressed.” In IMO’s guidelines for Formal Safety Assessment (FSA), it is stated that “the human element is one of the most important contributory aspects to the causation and avoidance of accidents…. Appropriate techniques for incorporating human factors should be used” (IMO 2018a). FSA is commonly seen as the premier scientific and systematic risk assessment approach. Per the latest revised guidelines (IMO 2018a), the FSA consists of five steps: (1) identification of hazards; (2) risk analysis; (3) risk control options; (4) cost–benefit assessment; and (5) recommendations for decision-making.

Risk definition and perspectives in the maritime domain are strongly tied to probabilistic methods (Goerlandt and Montewka 2015). This classical approach to risk analysis involves a process governed by data collection, processing, and calculating quantitative risk metrics using engineering and inferential statistics. Risk analyses are well established in situations with considerable data and clearly defined boundaries for their use. However, for MASS, we do not currently have sufficient empirical data. In addition, the complex and software-intensive technology of MASS, composed of not only hardware components but also logic control devices and a high number of sensors (Zhou et al. 2020), makes an accurate quantitative risk estimation extremely difficult to achieve. Furthermore, if achieved, the uncertainty related to these numbers will be high. Literature on risk assessment of MASS acknowledges the lack of data on design solutions and system architectures (Hoem 2019), making it challenging to apply probabilistic risk assessments. There are, however, arguments for seeing beyond expected values and probabilities in defining and describing risk. Over the last 20 years, there has been a shift from narrow perspectives based on probabilities to assessing a broader risk picture reflecting different views, assumptions, and ways of thinking that highlight events, consequences, and uncertainties (Aven 2009; 2012).

Risk assessment should be carried out both during the design and operation of MASS (Utne et al. 2017). During operation, risk monitoring and control are carried out both by the SCCO and by the technical system. MASS will have online risk control functionalities implemented in its control system, as described by Utne et al. (2017). In addition, the system shall visualise the risk monitoring through the HMI to the SCCO to support decision-making when human intervention is needed (for example, by real-time indicators on the systems' status, weather conditions, and presenting detected objects within the collision zone). With the integration of the SCC, interaction-associated hazards may lead to accidents if not well recognised and controlled (Yang et al. 2020).

Risk assessments in the design process are tools for decision-making. They can broadly be used in two ways: formative analyses (focused on the process, e.g., to improve the quality of a design) or summative (focusing on the results of the assessment, e.g., to evaluate if a safety target is met) (French et al. 2011; French & Niculae 2005). Some of the main reasons for carrying out a risk assessment are listed in Table 1 below. The activities represent some of the different phases of a product development cycle.

### 1.1 Risk-based ship design

According to IMO and current best practices and regulations, MASS will be approved according to principals for alternatives and equivalents (IMO 2013). This is fundamentally a *risk-based* approach rather than a *rule-based* approach where operational or functional requirements must comply with the statutory rules and

**Table 1** Formative and summative use of risk assessments

| Activities | Formative analysis | Summative analysis |
|---|---|---|
| Design | Proactively used to "design out" potential system failures and issues | Used to verify the capabilities and performance of the technology |
| Regulation and approval | Helps to choose between possible solutions | Demonstrates compliance and that a safety target is met |
| Licensing and verification | Helps understand modifications of the current design | Demonstrates fulfilment of a performance standard |

regulations. The regulatory framework for risk-based ship design (RBSD) was introduced with the primary objective to provide evidence on the safety level of a specific design of ships (Papanikolaou and Soares 2009), i.e. a summative approach to risk assessment. Meeting a particular level of safety (predefined risk acceptance criteria) implies that safety must be quantified using a formalised quantitative risk analysis procedure.

RBSD framework has mainly been applied for technical design (Ventikos et al. 2021). Applications including human element considerations are relatively fewer. This is most likely because guidelines on RBSD, such as Lloyds Register's procedures on risk-based design (2016) do not provide any guidance on including human and organisational aspects of risk. However, as IMO (2018b; 2019) states, the human element should be assessed as a part of an FSA and risk analysis in the design of MASS.

In the RBSD methodology, the human element is considered a factor that influences the causation probability. Quantifying the human element's contribution to risk is typically done by applying a human reliability analysis (HRA) method. HRA focuses on "human errors" as a cause. However, the "New view" of "human error" (Dekker 2014) focuses on "human error" as a symptom of problems rather than a source/cause of them. Typical problems are poor design or organisational issues. The classical view of "human error" is criticised by many as too narrow (e.g. Boring et al. (2010); Dekker (2014); Hollnagel (2000); Leveson (2016)). Quoting Leveson (2011), although the human element rests in the centre of socio-technical systems and guarantees their sustainability and viability, humans are seen as components with defined specifications. Hence, "human error" becomes a local problem rather than a symptom of flawed designs. Leveson (2011) presented a System Theoretic Process Analysis (STPA) method where "human error" is examined in its context, unsafe control actions, and control mechanisms that shape human behaviour. Many researchers see the STPA as a promising risk assessment to be applied to MASS concepts (Banda et al. 2019; Thieme et al. 2018; Utne et al. 2020; Wróbel et al. 2017, 2018; Zhou et al. 2020). However, this top-down method requires a hierarchical safety control structure, both on technical and organisational design (see Leveson and Stephanopoulos (2013)), making the analysis complex and involving many steps that are not easy to follow or understand (Hirata & Nadjm-Tehrani 2019). The control structure is dependent on what is included in the system and where the system boundary goes. For engineering, the most useful way to define the system boundary for analysis purposes is to include the parts of the system over which the system designers have some control.

### 1.2 Risk-informed decision-making in the design of MASS

In RBSD, only the ship is considered. For MASS, as mentioned, the vessel will be part of a more extensive system involving different components and actors, as shown in Fig. 1 below.

In the context of this paper, "risk-based design" simply means carrying out risk analyses that are not necessarily quantitative in the design process. The term

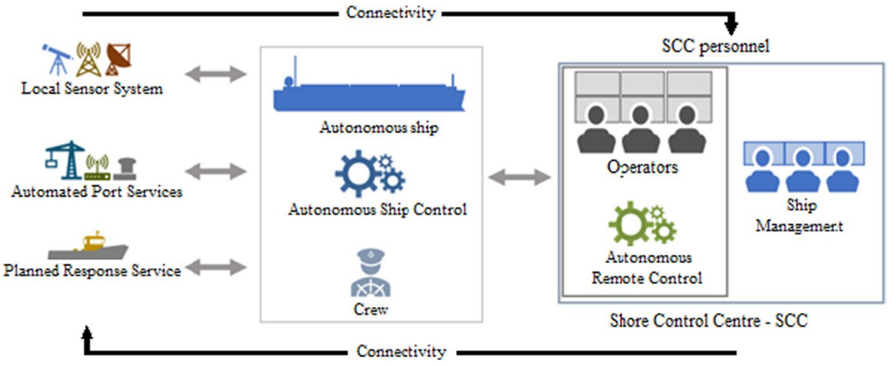


**Fig. 1** Examples of components and roles in an autonomous ship system, adapted and adjusted to the content of this paper from Wennersberg et al. (2020)

*risk-informed decision-making in design* may be a better phrase for explaining our approach. Papanikolaou and Soares (2009) have already described a similar risk-based design approach based on probabilistic functional requirements. Risk-based decision-making is criticised for focusing too much on probabilistic risk estimates and paying too little attention to design principles (Rausand 2013). Risk analysis is not the same as decision-making but is merely one tool in the process.

The Norwegian Maritime Authority NMA (2020) and classification societies, such as Bureau Veritas (2019), ClassNK (2020), DNV (2018), and Lloyd's Register (2017), have published guidelines on risk assessments of MASS. They all recommend applying a risk-based approach. DNV (2018) states that the design methodology should specifically address all functions of the auto-remote infrastructure needed to achieve an equivalent level of safety. The guideline mentions, explicitly, the CRIOP study as a risk analysis method focusing on human aspects. Hoem et al. (2021) presented an adapted version of the framework as an interdisciplinary risk assessment method in designing a SCC for the operation of MASS.

The CRIOP framework is an established, standardised scenario method primarily developed for the oil and gas industry in 1990 (Johnsen et al. 2011). Since then, the methodology has developed through collaborations between regulatory authorities, operators, research institutions, contractors, and consultants to include and consider HMI, best practices standards, and human factors, including principles from the ISO9241-210 (2019) and ISO11064 (2013) standards and the barrier management perspective (Johnsen et al. 2020). Today, CRIOP is used to verify and validate a control centre's ability to handle all operational modes safely and efficiently, i.e. normal operations, maintenance, disturbance/deviations, and safety–critical situations. The key elements of CRIOP are checklists covering relevant areas in the design of a control centre, scenario analysis of critical scenarios, and a learning arena where the operators, designers, and managers can meet and evaluate the optimal CC (Johnsen et al. 2011). The CRIOP process consists of four major work tasks:

1. **Prepare and organise** by defining, gathering the necessary documentation, establishing an analysis group, identifying relevant questions and scenarios, and setting a schedule.
2. **General analysis (GA)** with checklists to verify that the CC satisfies the stated requirements based on best industry practice (a standard design review).
3. **Scenario analysis** of critical scenarios. A cross-disciplinary team, including the end-users, perform the analysis to validate that the CC satisfies the actual needs.
4. **Implementation and follow-up:** At the end of tasks 2 and 3, the findings and recommendations are documented, and an action plan is established.

The scenario analysis's third work task is designed to verify that the control room operator (CRO) can perform the required task while considering cognitive abilities, human-system interaction, and other performance shaping factors. The analysis is human-centred, focusing on the CRO's interaction with the system, including communication with other personnel. Emphasis is on how the systems support the operator's situation awareness and decision-making in different situations.

The analysis considers a few relevant scenarios, identified as key scenarios. The methodology suggests using Sequentially Timed Events Plotting (STEP) diagram for a graphic presentation of each scenario and its events. Considering each event involving the operator, questions like "what can go wrong?" and "what if?" are asked to identify potential hazards and safety issues. Additional questions related to the simple model of cognition (by Hollnagel (1996)) can be asked to determine how the systems support the operator's situation awareness and his/her ability to make decisions and execute actions.

Furthermore, checklists on performance shaping factors are also used for additional questions to elaborate on the answers received. The questions and checklists help identify so-called weak points. Weak points comprise an identification of possible conditions, design issues, or safety problems in the achievement of operator tasks (involving identification, interpretation, planning and action on a situation). After identifying weak points, an evaluation of possible barriers and mitigating measures is initiated, and the results documented.

A CRIOP study can be applied at different phases of the design process. In the preliminary design, when the detail level is low, the method can assist in evaluating the assigned responsibilities between the autonomous system and the human operator. At this early design stage, it can also assist in identifying risks and ensuring end-user/operator involvement. At the final design stage, a CRIOP study can function as a tool for verification and validation by assuring the quality of documented task analyses, workload analyses, work environment/ergonomics, quality of alarms, and HMI (Johnsen et al. 2020).

### 1.3 Problem description and main ideas

The current RBSD framework is not adjusted for the design of MASS (including its integration with a SCC). For the risk-based design of MASS, the summative classical approach will be challenging for practical use as the background knowledge—the basis for the probability models and assignments—is weak, i.e. uncertain. The term uncertainty is used to capture the idea that a person or group does not know the true value of a quantity or the future consequences of an activity due to imperfect or incomplete knowledge (Aven 2019). In the face of uncertainties, the risk assessment of MASS may be better addressed by constructing scenarios that are validated according to logical consistency, psychological empathy with the main players involved, congruence with past trends, and narrative plausibility (see Aven and Renn (2009)). The main players involved in the operation of MASS are, as mentioned, the critical human element, i.e. the SCCO. However, few risk assessment methods address the SCCO in the design of MASS today (Veitch and Alsos 2022), and the classical technical risk assessment methods are insufficient to address human-automation interactions (Goerlandt 2020). Two recently developed methods, the STPA (Leveson, 2016) and *Human System Interaction in Autonomy* (Ramos et al. 2020), require a high level of system knowledge and method expertise. In addition, they can be quite time- and resource-consuming, making them of limited value in an early design phase when developing an HMI for a SCC.

Veitch and Alsos (2021) present a human-centred SCC design approach and bring in the concept of resilience by addressing the safety–critical interactions between the SCCO and the HMI. The authors acknowledge the need for building on this idea and use a systematic risk assessment method like the scenario analysis used in the CRIOP framework. This paper presents a case study where the adapted scenario analysis is carried out on an actual first prototype of an HMI for a SCC. The overall research question is as follows: can the scenario analysis support risk-informed decision-making in the design of a SCC?

This paper presents a method for carrying out a human-centred risk assessment and a use case where the method is applied in a workshop. The following section presents the scenario analysis methodology and the findings in a literature study of the method. Section 3 describes how we carried out a case study (the format of the case study-workshop, the HMI simulator, and the preparations) and the results, followed by a discussion of the applicability of the scenario analysis as a risk assessment tool in Section 1.3 and a conclusion in Section 2.

## 2 A review of the CRIOP scenario analysis

Hoem et al. (2021) presented how the scenario analysis in the CRIOP Framework (Johnsen et al. 2011) can be adopted, aiming to identify hazards and risks, assess them by identifying weak points in the design, evaluate existing barriers, and develop measures (mitigation actions) to improve the design. This section reviews the CRIOP scenario analysis in light of its contributions to risk analysis and design research.

We examined the research question by further asking in-depth questions like how the scenario analysis supports the idea of having the "human in the loop" during both design and operation and how the scenario analysis contributes to the FSA framework. Furthermore, what are the benefits and limitations compared to other risk analysis tools?

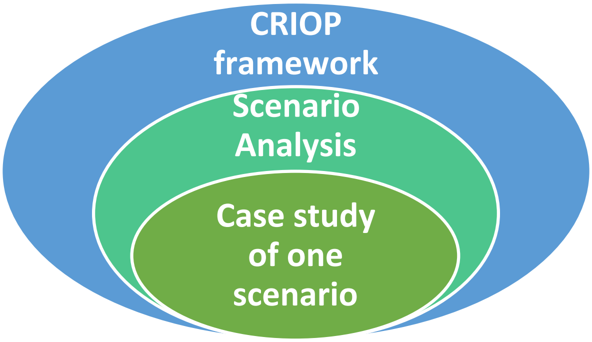


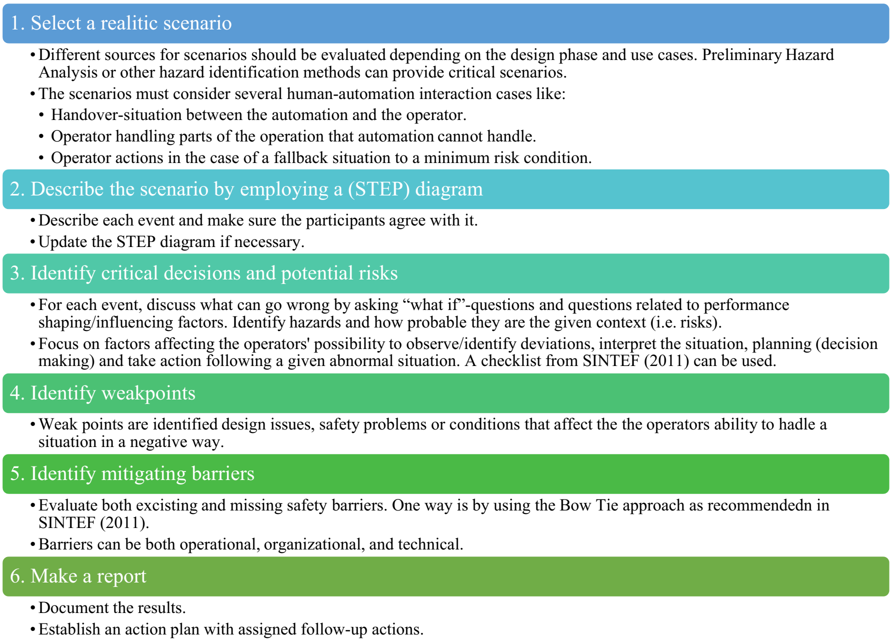
**Fig. 2** The scope of the paper is a case study of a scenario analysis  adapted from the CRIOP methodology

A case study (further discussed in Sect. 3) applying the scenario analysis method was carried out in the setting of a CRIOP-workshop, as shown in Fig. 2. We wanted to evaluate the applicability of the adapted scenario analysis and explore the contextual conditions of the method.

The adapted scenario analysis method can be summarised in the following steps in Table 2.

Steps 1 and 2 should be carried out prior to a workshop by experts from different fields of expertise. The scenario analysis group should consist of designers, end-users, engineers, software developers, human factors experts, and management.

**Table 2** The main activities of the adapted scenario analysis inspired by the CRIOP Framework



1. Select a realitic scenario

- Different sources for scenarios should be evaluated depending on the design phase and use cases. Preliminary Hazard Analysis or other hazard identification methods can provide critical scenarios.
- The scenarios must consider several human-automation interaction cases like:
  - Handover-situation between the automation and the operator.
  - Operator handling parts of the operation that automation cannot handle.
  - Operator actions in the case of a fallback situation to a minimum risk condition.

2. Describe the scenario by employing a (STEP) diagram

- Describe each event and make sure the participants agree with it.
- Update the STEP diagram if necessary.

3. Identify critical decisions and potential risks

- For each event, discuss what can go wrong by asking “what if”-questions and questions related to performance shaping/influencing factors. Identify hazards and how probable they are the given context (i.e. risks).
- Focus on factors affecting the operators' possibility to observe/identify deviations, interpret the situation, planning (decision making) and take action following a given abnormal situation. A checklist from SINTEF (2011) can be used.

4. Identify weakpoints

- Weak points are identified design issues, safety problems or conditions that affect the the operators ability to hadle a situation in a negative way.

5. Identify mitigating barriers

- Evaluate both excisting and missing safety barriers. One way is by using the Bow Tie approach as recommendedn in SINTEF (2011).
- Barriers can be both operational, organizational, and technical.

6. Make a report

- Document the results.
- Establish an action plan with assigned follow-up actions.

Depending on the scenario, it could also involve a broader range of stakeholders. By involving people with experience from similar systems and including the end-users, the analysis aims to minimise the gap between work as imagined (WAI) and work as done (WAD), considering the resources needed to execute the operations. WAI refers to the various assumptions, explicit or implicit, that people have about how work should be done. WAD refers to (descriptions of) how work is actually done, either in a specific case or routinely (Hollnagel 2017).

The purely technology-centred approach can sometimes lead to structural and functional rigidity in the design and operation processes. The consequences are that people must adapt to the system, and the HMI is something that is put on top of the system in the end after it has been built, which is the opposite of human system integration or human-centred design (HCD). This is also an issue of WAI vs WAD. In a scenario analysis, we are at the blunt end considering work as imagined (WAI). We use our expectations based on our experience of actual similar work (at the so-called sharp-end). It is practically impossible to predict or describe how work is done by others since it occurs at a different time and place (Hollnagel 2017). However, we can imagine how work is to be done and why in abstract terms. By including the actual end-user and presenting the scenario for them at their workstation, we can discuss if the imagined work is as close to the "work as done" as possible and be aware of the actual difference. As mentioned by Lützhöft (2004), people participate in integrating new technology into complex fields of practice—often in ways that are surprising to the designer, and involving the prospective users is necessary for providing knowledge of the current practice.

Furthermore, we avoid defining the design needs based solely on abstraction by carrying out scenario analyses at different stages of the design process (i.e. after the conceptual/preliminary designed HMI, the detailed design, and the built HMI). In this way, the method presents an opportunity to improve our models of work throughout the design.

The scenario analysis should be used as a formative method that recognises and roughly rank the potential for improving safety issues (i.e. the weak points) related to the HMI. The method can help improve the design of the HMI itself, the structure of the organisation, and the processes by which it is operated. Researchers have pointed to the need for bringing in human factors expertise early in the design process, e.g. Blackett (2021) and Johnsen and Porathe (2021), to avoid poorly designed solutions that are challenging and costly to change. The design process should be iterative and involve human factors and the end-user from the beginning, to support sensemaking and meaningful human control.

## 2.1 An iterative human-centred risk assessment

The CRIOP exercise, and hence the scenario analysis, is a participatory (multidisciplinary) iterative process. The methods support the HCD process activities for interactive systems (ISO9241-210 2019). By applying an HCD process, flexible and robust design solutions might be achieved where the operator situation awareness recovery, task switching support, and workload balancing are considered. The scenario analysis may work as both an analytic (in analysing the human–machine interactions) and an evaluative tool (evaluating the design against requirements). However, in the case of designing an HMI for a SCC, we do not have any predefined acceptance criteria available to measure a prototype against yet. The adapted scenario analysis can be considered a dual view approach to risk analysis, as both objective facts and subjective statements are considered. The result of a scenario analysis is a list of weak points and suggestions for improvements (i.e. mitigating barriers) and not a complete characterisation of a risk or a risk picture. The overall goal of a scenario analysis is to improve the design by enabling human-centred risk-informed decision-making. The adapted scenario analysis as a human-centred risk assessment process is visualised in Fig. 3 below.

## 2.2 Identifying hazards and safety issues not covered by existing risk analyses

Looking at what can fail or go wrong is a bottom-up approach whereby safety is treated as a failure prevention problem. Applying this methodology to complex systems has, in recent years, spurred vocal criticism (see, for example, Leveson (2020)). A hazard analysis is traditionally seen as a failure and malfunction of components. This is not necessarily the same as asking "what can go wrong?" and "what if?" questions. These questions imply that something surprisingly can happen due to a combination of performance variabilities.

The scenario analysis allows the participants to take the SCCO's role and experience how incidents and accidents can be handled based on available information presented to the SCCO. Hence, unlike many risk analysis methods, the scenario analysis focuses on the operational experiences of the HMI. Hazards and potential safety issues that are not necessarily revealed by traditional risk analysis can be identified by focusing on the

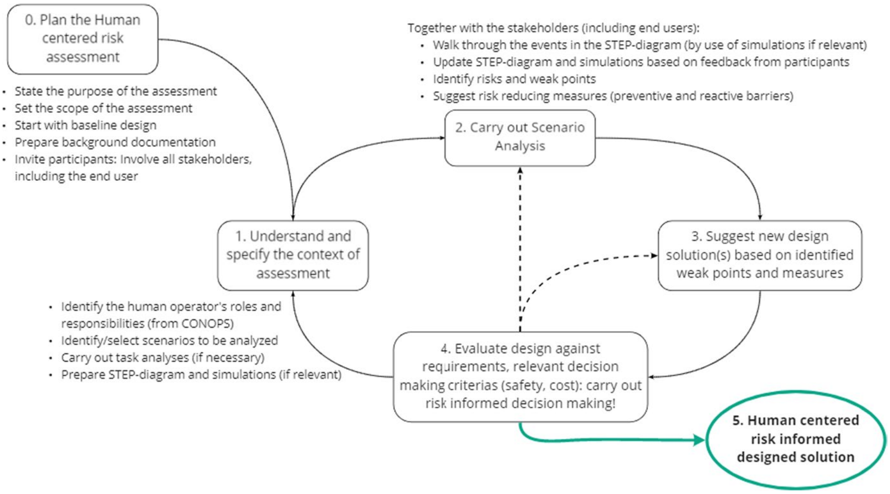

0. Plan the Human centered risk assessment
• State the purpose of the assessment
• Set the scope of the assessment
• Start with baseline design
• Prepare background documentation
• Invite participants: Involve all stakeholders, including the end user
Together with the stakeholders (including end users):
• Walk through the events in the STEP-diagram (by use of simulations if relevant)
• Update STEP-diagram and simulations based on feedback from participants
• Identify risks and weak points
• Suggest risk reducing measures (preventive and reactive barriers)
2. Carry out Scenario Analysis
1. Understand and specify the context of assessment
3. Suggest new design solution(s) based on identified weak points and measures
• Identify the human operator's roles and responsibilities (from CONOPS)
• Identify/select scenarios to be analyzed
• Carry out task analyses (if necessary)
• Prepare STEP-diagram and simulations (if relevant)
4. Evaluate design against requirements, relevant decision making criterias (safety, cost): carry out risk informed decision making!
5. Human centered risk informed designed solution

**Fig. 3** The iterative human-centred risk assessment approach based on the HCD process (ISO9241-210 2019)

SCCO's responsibilities, tasks, and capabilities. In this way, issues related to a poorly designed solution, a lack of explainable AI (see Veitch and Alsos (2021)), or deficient procedures and responsibility can be identified as weak points and handled early.

Like most traditional hazard analysis techniques, the STEP diagram provides a chain of events, addressing factors that affect how the accidental events are presented to the operator and propagate. The aim is not to track accidents back to a root cause or identify component failures but rather how different actors (including the end-user) experience a scenario sequence and which control and interaction issues may arise. In a setting where humans can brainstorm on possible interaction issues, the direct linear causality of the deterministic cause-effect relationship does not affect the risk analysis like a typical on paper risk model would. We can think more abstract than we can write down in a 2D model. The identified hazards, mitigating measures, and weak points do not need to be directly connected to an event in the STEP diagram.

The STEP diagram demonstrates how operational scenario sequences might be unambiguously specified by getting the workshop participants' second opinion. Furthermore, presenting the course of the scenario in a STEP diagram brings up an agreement between the designers, engineers, end-users, and the software developers on how the "behaviour" of the technical system and the SCCO's action should enfold, and hence (if necessary) redefine the system architecture at an early stage. The STEP diagram can further be translated into an event sequence diagram and give input to task analyses used in more advanced and comprehensive safety analyses, like the *Human System Interaction in Autonomy*-method proposed by Ramos et al. (2020).

Hazards and risks are in the scenario analysis considered in "two turns" by selecting scenarios based on a preliminary hazard analysis and further in the STEP diagram by asking what can go wrong, focusing on the SCCO's capabilities. This allows us to dive deeper into the challenging parts of the HMI and question how the HMI can support the SCCO to have quick detection and early response to a critical situation.

The method supports the underlying idea of resilience as the ability to sustain or restore its basic functionality following a stressor (Hollnagel et al. 2006). Increasing the resilience can be seen as a strategy for managing risk. In this case, we design for a safe and resilient HMI by identifying hazards and weak points and subsequently risk-reducing measures focusing on the SCCO's capabilities at an early stage. As recommended by Aven (2016), by applying a scenario-based risk analysis focusing on the capabilities of a SCCO, we are relating risk to performance and hence incorporate resilience dimensions.

### 2.3 Compared to the system theoretic process analysis

There are many risk assessment methods available to the designer. Their applicability depends on the purpose of the risk assessment (whether, for example, it is used to decide if an activity should be permitted, if a system is safe enough, if system improvements are necessary, or simply in choosing between competing options). For MASS, the effectiveness of the risk assessment varies with respect to different autonomous system properties (Bolbot et al. 2020). Zhou et al. (2020) have investigated the applicability of 29 hazard analysis methods for autonomous ship systems. The scenario analysis was not evaluated, as it is not targeted at ships or applied within the maritime domain. However, the adapted scenario analysis fulfils many of the evaluation criteria (EC) listed in the review (see Table 3 in Zhou et al. (2020)). The method can be used to analyse system-level hazards (EC1); can be used from the early system development phase (EC2), can be used to analyse the hazards resulting from HMI (EC6); consider the communication between ships and SCC (EC7); consider the communication among shore-based operators, or crew onboard (EC8); and consider different operational modes resulting from the change of levels of autonomy (EC10). The STPA fulfils all criteria listed in Zhou et al. (2020), and because several authors have recommended it as a promising method for risk assessments of MASS, it was selected for further comparison.

Banda et al. (2019) applied the method to carry out a hazard analysis in the concept design of autonomous passenger ferries. The study presented a systematic hazard analysis based on the STPA framework. However, a SCC was not part of the analysis. Still, several identified safety control actions were suggested involving a SCC to communicate with the passengers or remote monitoring and fault detection of the technical systems. All the identified hazards are related to the technical system (i.e. component failures), the environment (heavy weather, strong current), or the passengers on board (falling/jumping overboard, medical conditions, etc.). In the selected cases (two urban passenger ferries), it was not addressed who is responsible for the safety of the passengers when the vessels are in operation. The reason why the SCC was left out of the scope seems to be that the suggested design process adopts the foundations of the ship design spiral, a 60-year-old design concept without any human factor considerations, hence, neglecting an essential source of risks and operational issues.

STPA does not define any framework for an operational scenario-based analysis, although STEP diagrams could be used in such an analysis. In the STEP diagram of a critical scenario, the events that involve crucial decisions by the SCCO can be seen as safety control actions applied in the STPA. Unsafe control actions (another term used in the STPA) are defined by asking what can go wrong here. In STPA, an unsafe control action is an action leading to an identified hazard, and typically when:

a) A control action for safety is not provided or followed.
b) A safety control is provided too early or too late
c) A safety control is stopped too soon or applied too long
d) A safety control is degraded over time
e) An unsafe control action is provided

These aspects could beneficially be integrated into the scenario analysis and help define the scope and target of the analysis. As Zhou et al. (2020) suggest, possible combinations of the STPA and other risk assessment methods should be considered in future research. Explaining why and how these unsafe actions can occur is essential when identifying risks and weak points in the designed prototype. The result of an STPA is the list of accidents and hazards, the safety control structure, unsafe control actions, and causal factors. A scenario analysis can help identify the context that makes these results and the STPA claim that "the system is free from unacceptable risks leading to an accident" justifiable.

### 2.4 The method's contributions to the FSA framework

The scenario analysis method is in line with the FSA methodology. It can be seen as one framework to support the requirements of incorporating the "human element" in risk assessment, associating them directly with the occurrence of possible accidents, underlying causes, or influences (ref. Section 3.4 in IMO (2018a, b)). However, the scenario analysis does not introduce a risk matrix, or similar, to discuss the probability and consequences. The aim is to identify what could go wrong (identify hazards, events, and conditions that may lead to an accident or incident), how these may lead to different consequences, and suggest measures to avoid/limit the impact by focusing on the capabilities of the SCCO, hence, contributing to the majority of the steps in the FSA presented in Section 1.1.

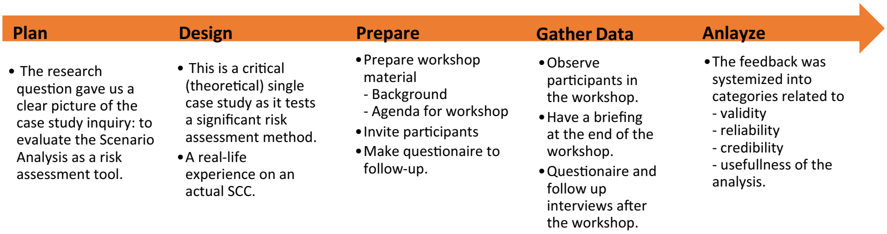
The adapted scenario analysis is in line with the request for risk-based design (ref. guidelines listed in Sect. 1.1), where several different risk-analysing methodologies are utilised. The method can be considered a risk analysis associated with the remote supervision and control of a MASS from a SCC, explicitly focusing on the SCC and its supporting systems (ref. DNV GL, 2018)). If the scenario analysis's tabletop exercise is carried out in a systematic manner, the assessment can provide valuable documentation and verification of a risk-based design process. It can also be seen as a tool used to represent and describe the knowledge and lack of knowledge of the autonomous system, its performance, and interactions with the SCCO.

It is important to remember that the primary goal of a CRIOP study is not the identified hazards but the identified weak points and the measures to improve them (with a correlating action plan). The identified hazards and risks can provide a basis for arguments on the need for design modification and contribute to risk-informed decision-making. Hence, the method provides decision support and could help the design team choose between alternatives, adjust the SCCO's activities, and implement risk-reducing measures, for example, in the case of a clear alarm philosophy or specific improvements for the HMI. In the design process of the SCC, the early scenario description and analysis exercise can also provide valuable discussions on how to balance operational complexity with technical simplifications.

## 3 Case study of the CRIOP scenario analysis method

A SCC for the operation of an autonomous urban passenger ferry, the milliAmpere2,[1] was the subject of the analysis carried out in a workshop with participants. The case study aimed to test the applicability of the scenario analysis framework by evaluating the validity, credibility, and reliability of the approach based on the exploration of a critical scenario in a simulated SCC together with experts from different disciplines. The goal of the scenario analysis was to improve the prototype HMI design by carrying out the risk assessment and identifying weak points (with suggested mitigating measures to improve them).

[1] MilliAmpere2 is a full-scale prototype of the world's first autonomous passenger ferry, milliAmpere, designed to become a living lab in Trondheim city, with capabilities and supporting infrastructure enabling trial passenger operation. https://zeabuz.com/miliampere/



**Plan**

- The research question gave us a clear picture of the case study inquiry: to evaluate the Scenario Analysis as a risk assessment tool.

**Design**

- This is a critical (theoretical) single case study as it tests a significant risk assessment method.
- A real-life experience on an actual SCC.

**Prepare**

- Prepare workshop material
  - Background
  - Agenda for workshop
- Invite participants
- Make questionaire to follow-up.

**Gather Data**

- Observe participants in the workshop.
- Have a briefing at the end of the workshop.
- Questionaire and follow up interviews after the workshop.

**Anlayze**

- The feedback was systemized into categories related to
  - validity
  - reliability
  - credibility
  - usefullness of the analysis.

**Fig. 4** Our case study research approach with activities and sub-activities

We used a qualitative case study methodology to do a process and outcome evaluation (Yin 2009). We have an initial descriptive theory about the case tentative to the study and a hypothesis about the expected characteristics of the case. We take an interpretive perspective to the case study by presenting our view as researchers on the scenario to be analysed (thus interpreting elements of the study). However, the approach is also relativistic as we aim to include the participant's multiple perspectives on the method: How do they interpret the risk analysis method? Do they find the scenario analysis helpful in identifying weak points? Moreover, do they believe the method is a good tool for risk-based design?

The case study research process is shown in Fig. 4. Our descriptive framework for organising the case study should not be confused with the scenario analysis method. Notwithstanding this important distinction, there are some natural overlaps in activities such as inviting participants and collecting feedback.

## 3.1 The SCC for remote operation of milliAmpere2

The shore control lab (SCL) (Fig. 5) is a test platform for research in highly automated ships. The lab is equipped with testing equipment to support research and development of the human control side of autonomous ships. One of the central

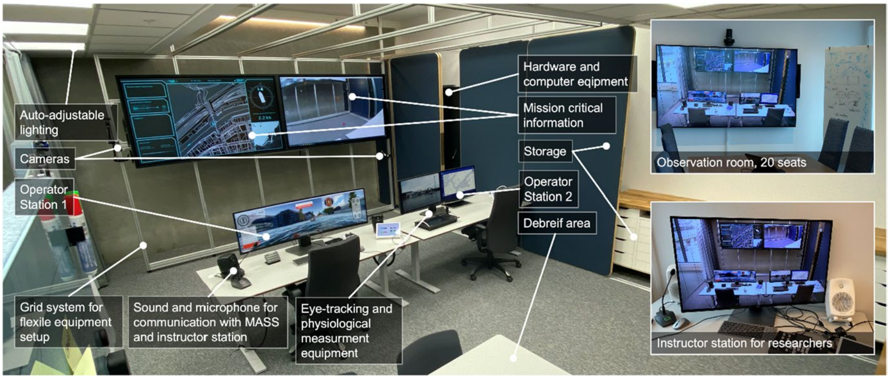

Auto-adjustable lighting
Cameras
Operator Station 1
Hardware and computer eqipment
Mission critical information
Storage
Operator Station 2
Debreif area
Observation room, 20 seats
Instructor station for researchers
Grid system for flexile equipment setup
Sound and microphone for communication with MASS and instructor station
Eye-tracking and physiological measurment equipment

**Fig. 5** The Shore Control Lab (SCL) at the Norwegian University of Science and Technology (NTNU)

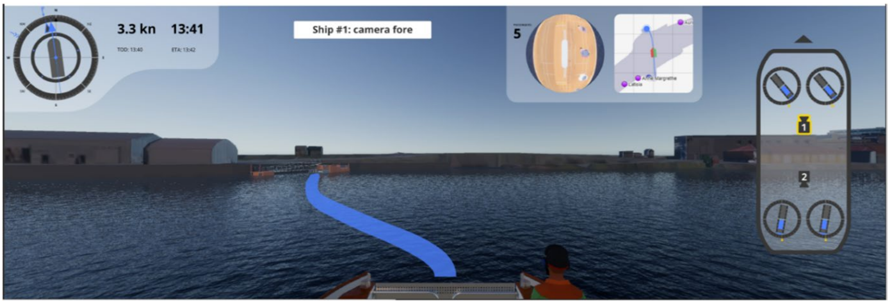


**Fig. 6** A screenshot of the simulator

research tools is a custom-built simulator based on the Gemini open-source platform (Vasstein et al. 2020). The simulator, built in Unity, allows for flexible testing in immersive environments with high-fidelity graphics and realistic physics engines.

The simulator presented in the workshop is illustrated in Fig. 6. The graphical user interface (GUI) displayed a simulated camera view from onboard the milliAmpere2 ferry in the approximate location where the physical camera is mounted. The GUI overlays show essential information like speed, heading, and the number of passengers. The central HMI consisted of a GUI and a control pad for handling actions (stop ferry and keep ferry in position by dynamic positioning (DP), drop anchor, manual control switch, communication with passengers, harbour authorities and emergency response, etc.). In addition, other peripherals like a joystick for manual control, keyboard, mouse, and speakers for alarms were available.

### 3.2 Preparations before the workshop

A concept of operation (CONOPS) for milliAmpere2 was developed by zeabuz (2021). The SCC was not included in the scope of the CONOPS, as a safety operator (responsible for safeguarding the passenger and the operation of the ferry) would initially be present onboard the ferry. The safety operator onboard would be able to initiate a safe state (set the ferry on DP or drop anchor), use a VHF radio, contact the harbour and emergency services, and manually control the ferry. An incremental approach to moving the safety operator to a SCC is suggested in the CONOPS. Three elements informed the tasks and responsibilities of the SCCO: the design of the ferry (including the autonomous and automation system), the tasks envisioned for the safety operator, and the experience from other domains (i.e. a remote-control centre for offshore oil and gas installations). The design team at the SCL made a background document presenting the operational domain, the design of the ferry, the HMI at the SCC (with peripherals), the SCCO's tasks and responsibilities, and the emergency organisation and response procedures. The document was an internal document shared with all participants before the workshop to build a shared understanding of the scope.

### 3.2.1 Selected scenario

The design team arrived at a scenario where an unexpected object (in this case, a partly submerged log) was floating in the pathway of the ferry, causing the ferry to stop, stay in position (automatically activate DP), and send a notification to the SCCO to assess the situation (see Fig. 7 below). The SCCO can then take manual control by switching a button on the control pad. A notification message "Manual control engaged" will be visible in the lower corner of the GUI until the switch is turned back to "Resume autonomy".

Before the workshop, a preliminary hazard analysis was carried out by the technology company zeabuz and facilitated by the classification society DNV. A list of critical situations was identified, one of them being the handover when the automated systems "ask" the safety operator to take over control. In a study by Dybvik et al. (2020), designing the HMI was identified as the most challenging part of a SCC design. In particular, this involves the handover from automation to human control. Knowing how to resolve this situation is a design issue and key to designing the interface. The simulator tool presented in the previous section made it possible to explain such a handover situation. The version shown here was specifically designed to confront users with a handover situation involving a shift of control from the automated system to the remote operator that occurred unbeknownst to the user-tester. The best GUI design, one that supplies the relevant information to the operator in the most appropriate way, will, when achieved, play a central role in enabling the coordination of operator actions to handle out-of-the-ordinary events.

In the case of manual control, passengers on board will be notified via audio announcements over speakers. When manual control is engaged, the ferry will stay in position until the operator uses the joystick to manoeuvre the ferry or pushes a button to resume normal operation. The operator can change the camera view between fore and aft on the ferry and land-based cameras with zoom function. Passengers can also provide information about the situation on the ferry and its surroundings by using a two-way communication link through an HMI display onboard milliAmpere2.

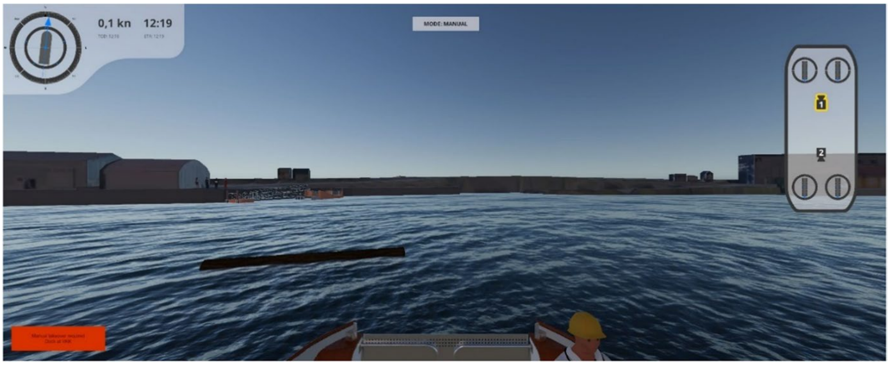


Fig. 7 A screenshot of the GUI showing a partly submerged log in the pathway of the ferry

### 3.2.2 Participants

An essential part of a CRIOP exercise is using experiences from similar control centres in operation and including the end-users as participants in the analysis. However, there is no SCC in operation for MASS yet, and we do not have an end-user (SCCO) in place at this point. It is still not agreed on what skills and qualifications are required for the SC operator. Findings in the HUMANE project (Lützhöft et al. 2019) point to the need for seafarer experience and the operator having certified navigational skills and seamanship. In our case, the operational domain of the autonomous ferry is limited to an urban canal, and it will not operate in harsh weather conditions. Still, the SCCO would need knowledge of the COLREGs rules and have a feeling for how a small ferry moves. Therefore, mariners with seagoing experience were invited to the workshop. In addition, we looked to other domains with remote control experience and invited participants from companies working with automated guided vehicles and autonomous shuttle buses. Furthermore, participants with system knowledge, including engineers and designers from the milliAmpere2 and Autoferry[2] project teams, were also invited to the workshop.

The invited participants were selected through convenience sampling and the SCL network at NTNU. In total, 12 participants attended the workshop. The characteristics of the participants are listed in Table 3. The circles indicate the expertise area. Black circles indicate participants with at least five years' experience; white circles indicate participants with less than five years' experience. A letter is added to the participant number to indicate the gender of the participants: F is female, and M is male.

**Table 3** The characteristics of the participants in the online workshop

| Participant no | **Disciplines and experiences** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Safety engineer | Naval architect | Interaction design | Marine cybernetics | Engineering cybernetics | Mariner w/ seagoing experience | Control room experience | Member of the SCL |
| 1 M | ○ | | ○ | | | ● | | x |
| 2 M | | | ● | | | ● | ○ | x |
| 3 M | ● | | ○ | | | | ● | |
| 4 M | | | ● | | | | | x |
| 5 F | | | ○ | | | | ● | x |
| 6 M | | | | ● | | ○ | | |
| 7 M | | ● | | ● | | | | |
| 8 M | | | | ● | | | | |
| 9 M | | ● | | | | | | x |
| 10 M | | | | | ● | | | |
| 11 F | | | | | ● | | ● | |
| 12 F | ● | | ○ | | | | ○ | x |

[2] A cross-disciplinary research project at NTNU on autonomous all-electric passenger ferries for urban water transport https://www.ntnu.edu/autoferry

### 3.3 The format of the case study

The study was carried out during the COVID-19 pandemic, and the need for social distancing made the workshop subject to a digital solution. This had both positive and negative aspects, further discussed in Sect. 3.6. We used Microsoft Teams' digital platform (Microsoft 2022) and the online whiteboard software Miro (miro 2022). Miro worked as a digital whiteboard for visual collaboration during the workshop (using digital "Post-it"-notes, adding comments and questions), collecting and analysing data after the workshop. The schedule and walkthrough process were presented with the STEP diagram and screenshots of the GUI showing the simulated scenario.

In the preparations for the workshop, a modification to the scenario analysis method was made. Due to a time constraint of two hours, the participants' limited knowledge of CRIOP studies, and the fact that this was a first design iteration of the HMI, we chose not to follow a strict stepwise approach in the scenario analysis. The sequential events in the STEP diagram plus the following list of considerations were merged into one brainstorming process focusing on the following:

- Ask "what can go wrong?" (Identify hazards) and "what would the operator wish to do in each situation?" Use questions related to performance influencing factors and Hollnagel's Simple Model of Cognition, such as "How is the SCCO notified? What information is presented? What happens if the information is not presented? How can the information be misunderstood? Which erroneous decisions can be made?"
- Identify weak points in the designed HMI.
- Identify mitigation actions by discussing existing barriers and missing barriers.

## 4 Data collection

As presented in Fig. 4, the results were evaluated by a short debrief at the end of the workshop and by sending out a survey to each participant. The questions assess the analysis's validity, reliability, credibility, and usefulness. We define these terms accordingly in Table 4.

### 4.1 Results from the case study

The hazard identification part of the scenario analysis was convened in a brainstorming session after the participants became familiar with the concept and the scenario analysis process (including the STEP diagram). The facilitator wrote "Post-it"-notes based on input from the participants. The data collection of hazards, weak points, and mitigating measures added to the STEP diagram during the workshop can be found in Figs. 8 and 9 in Appendix 7.

**Table 4** Definition of terms and questions in the questionnaire

| Term | Definition | Question(s) in the survey |
|---|---|---|
| External validity of the workshop settings | Whether the method "actually did what it aimed to do" (Salmon et al. 2020) | 1. Was the scenario realistic?<br>2. Based on the provided information, did you manage to identify hazards and assess risks? |
| Credibility (or internal validity) | When the results of the study mirror the views of the participants in the study: whether the participants believe the results are valid (Mills and Gabrielle 2010) | 3. Do you believe the scenario analysis results (identified weak points and mitigating barriers) are valid? |
| Reliability | If the study results can be reproduced under a similar methodology (Joppe, 2000) | 4. Are the results of the analysis confirmed by other similar studies? |
| Usefulness | Whether the participants found the scenario analysis to meet its goal of improving the design of the HMI<br>Usefulness is one of the many dimensions that influence and contributes to a product's usability (Trudel, 2021) | 5. How did you succeed in understanding and predicting the safety issues/weak points in the HMI?<br>6. Do you find the method helpful in including the human element in a risk assessment of the prototype design?<br>7. Did you learn anything from the workshop? Can you tell us more about what? |

For each event, the participants suggested hazards and how probable this was in the given context (the combination of which represented the event's risk according to the classical definition) and discussed potential barriers to avoid or mitigate the hazards. The participants were encouraged to focus on the SCCO's capabilities, tasks, sensemaking, and possible error sources and malfunctions in the HMI. They were allowed to drift around other topics, triggering discussions not directly related to the tasks and events in the STEP diagram, but instructed to not spend too much time on disagreements in assessing the severity of consequences or probabilities. Instead, they were encouraged to identify additional mitigating measures and discuss their potential effects.

From the discussion in the workshop and the identified risks and mitigating measures (barriers), the following identified weak points (design issues and safety problems) and mitigating actions (suggestions for improvements) can be summarised:

- The existing CONOPS does not address the responsibilities of the SCCO. The role of the SCCO is a missing priority! A list of situations where immediate SC intervention is required must be established:
- When and how should the operator intervene? Descriptions of tasks and supporting working procedures are needed.
- What are the needed skills and training for the SCCO?
- No alarm philosophy is established. Notifications on a screen alone are not enough.
- Recovering from a "safe state:"

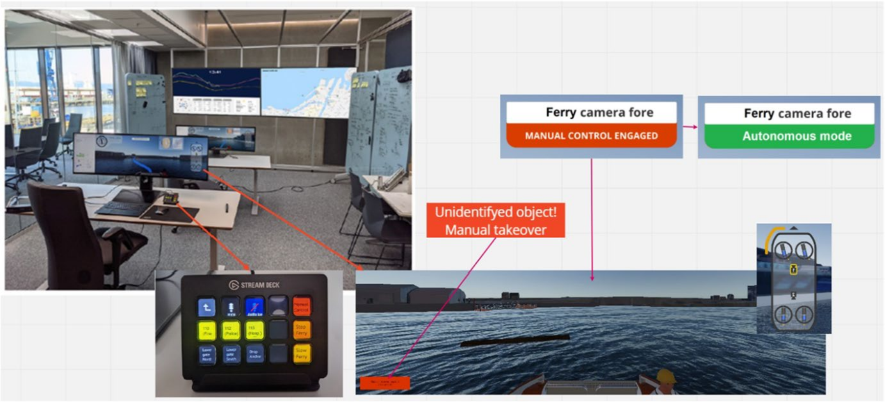


**Fig. 8** Screenshots of the simulated scenario in the GUI and the peripherals at the shore control centre

- After going to a safe state and dropping the anchor, what happens?
- What are the available resources to pick up passengers and resume operation? An emergency preparedness strategy is missing.
- Related to high-performance HMI: “ easy to discover”-notification messages should appear centred on the main screen and not down in the left corner.
- Develop the GUI to support explainable artificial intelligence:
- Bounding box around detected object to avoid misunderstanding which object is within the collision zone and detected by the ferry.

    o Implement layers showing the collision zone when necessary.

- A safety management system must be established: how can the SCCO report incidents (an unplanned, uncontrolled event that under different circumstances could have resulted in an accident), near-accidents (an event that could reasonably have been an accident but did not, typically due to the SCCO intervening), and accidents (an unintended sequence of events that lead to harm to people, environment, or other assets)?
- More data from actual experiments are needed, i.e. systematic recording of accidents and incidents in the testing phase.
- Maintenance issues: How will the SCC handle this? How is the status of the technical systems presented to the SCCO?

There were also some “loose” “Post-it”-notes considering general hazards, questions related to the overall structure, responsibility gaps, and the CONOPS (see Fig. 9 in Appendix 7.). These essential issues may not have been revealed by analysing individual hazardous events and their consequences.

### 4.2 The contextual conditions of the case study

The preconditions of the workshop gave some limitations to the applicability of the method:

- The autonomous passenger ferry milliAmpere2 was designed with a safety operator onboard. Hence, the CONOPS and preliminary hazard analysis were carried out with this precaution. In our case study, we provided a background document based on this CONOPS, but where the safety operator was transferred to the SCC and became a SCCO. The SCCO tasks and responsibilities were adjusted accordingly.
- An incremental approach was taken in this project. Ideally, the design of the SCC would be considered from the beginning of the project and not as an "add on" to be designed when the vessel and its technology are built and completed.
- Making retrofits to the ferry now will be expensive and challenging. Nevertheless, it is essential for good human system integration. Risks identified regarding communication, emergency preparedness, and other aspects may influence the final design, for example, the need for a two-way communication system and automatic fire detection and sprinkler system at the top side of the ferry. This is not in place on the milliAmpere2 today.

Ideally, we would have been physically present at the SCL experiencing the simulated scenario at the SC station. This would have made the scenario more tangible and closer to the actual operational environment. However, screenshots of the simulated scenario related to each event in the STEP diagram were presented to the participants. There were some benefits of having a digital workshop. Firstly, we were able to recruit participants located outside of the Trondheim area. We experienced that it was favourable to have a digital meeting, and we found it easier to get participants to spare two hours of their working hours when they could log on from their own office. During the workshop, the participants could (anonymously) type "Post-it"-notes and post them to events and tasks in the STEP diagram. Using the digital collaboration platform, Miro enabled us to more accessible collect data and document the process.

### 4.3 Evaluation by the participants

At the end of the workshop, a round of "criticism of the method" revealed some practical implications of the organisation of the tabletop exercise, like involving experts in the selection of scenarios and better structuring of the brainstorming process. In the questionnaire sent out to the participants after the workshop, open-ended questions related to the method's credibility, accuracy, reliability, and validity and results were asked (see Table 4). Based on the feedback from the participants and discussions among the authors, a summary of this, including potential future work to address identified threats and weaknesses, is listed in Table 5 in Appendix 8. The main feature mentioned by the participants was how the method provided a common platform for understanding the operations and how the SCCO could handle different situations. By visualising the scenario in a simulation of the HMI and structuring the discussion to events in a STEP diagram, the scenario became easy to comprehend. The method facilitated an open discussion and brainstorming around possible risks. Furthermore, the participants appreciated the possibility of exchanging experiences across disciplines and domains.

## 5 Discussion

Risk is about more than expected values. Expected value decision-making can be misleading, especially in the design phases of MASS, where risk and safety might be best understood and communicated in ways other than probabilistic risk analysis. One such way is by understanding and assessing risk in terms of knowledge and lack of knowledge and by identifying hazardous events and their tangible effects. We have presented a method combining HCD and risk assessment elements. The scenario analysis fulfils several criteria for a suitable hazard analysis method as defined by Zhou et al. (2020). Valuable features of the method include its ability to highlight possible issues of the SCC concept, as well as uncertainties, knowledge gaps, and missing priorities. This provides valuable input to a revised and more detailed CONOPS and system architecture. The analysis can also reveal interdependencies between subsystems not revealed by other risk assessments, helping the team agree on how to solve an issue and contribute toward the overall aim of improving the safety of the MASS system. This is also supported by the results in the case study, where we identified a wide range of weak points when analysing how the HMI would work in practice. Most of the identified weak points were crucial questions to the developers and the organisation that need to be answered before a new design iteration to the next development phase. The existing risk analysis and preliminary hazard analysis did not identify these weak points (design issues and safety problems).

An example of a risk assessment focusing solely on the technical aspects of MASS is presented by Banda et al. (2019). Here, the authors apply the STPA but do not integrate a SCC. The SCCO is left out of the scope and is only mentioned when identified as a barrier against accidents as if an addendum to designing the autonomous ferries. This undervalues the potential of incorporating an understanding of human needs and capabilities early in the design process. STPA is a systems-theoretic approach used, among other things, to analyse human-automation interaction. Applied to the SCC case, this includes identifying any unsafe control actions performed by the human operator. However, where the analyst chooses to set the system boundaries will strongly affect the outcome of the analysis, as the mentioned study implies. The adapted scenario analysis in this paper explicitly addresses the interactions between SCCO and MASS. Hence, the most considerable improvement of the scenario analysis, compared to established practices, is the strong involvement of the SCCO. Involving the SCCO addresses the requirements concerning the "human element" in the IMO's FSA and the interim guidelines for MASS trials.

A limitation is incurred by the want of SCCO taking part in the case study. At present, there are no certified SCCO, nor are no formal training standards available like there are for conventional seafarers. The CRIOP framework nonetheless explicitly states that the end-user must be included as a participant (Aas et al. 2009). This condition could only be partly met by selecting participants with relevant backgrounds, as judged by the workshop organisers. Inviting mariners with experience with highly automated bridges might be an option. However, this depends on the required qualifications (skills and education) and the responsibilities and tasks envisioned for the SCCO.

Our focus was on cognitive and not physical ergonomics related to workplace comfort and safety, typically included in an entire CRIOP exercise. The prototype was the first version of the initial design; hence, the complexity and fidelity of the analysis were consistent with the data and information we had available. The model (STEP-diagram) and analysis were of a high level and can be expected to mature in the following design phase. The systematic activities in a scenario analysis should be adjusted to the design phase in question.

Typically, a CRIOP study runs over several days, encompassing several critical scenarios (Johnsen et al. 2011). By contrast, our case study was more focused, and only one scenario was analysed. The method's validity is already proven. It is considered a "best practice" tool in the design process and for validating and verifying control centres in the oil and gas industry. The validity of testing the applicability in our case study was evaluated in terms of participants' feedback (summarised in Appendix 8.) and the method's ability to identify hazards, risks, and issues. All participants accepted the scenario as possible, and a long list of hazards and weak points were identified. The method's credibility is considered sufficient as the participants were recruited from different fields of expertise. None of the participants, except the facilitator, attended the preliminary hazard analysis. After the workshop, all participants reviewed the analysis report and confirmed that they believed the results were valid. In addition, several of the identified hazards and safety issues were mentioned in the preliminary hazard analysis carried out by zeabus. However, additional hazards were also identified. Threats to the validity, credibility, and reliability of the method are listed in Appendix 8. These are biases from the participants already involved in the HMI-design process, lack of having the actual end-user present, time constraints, and limited opportunities to modify the ferry's design, configurations, and technical solutions. In the case study, we applied the method on a prototype of the HMI during the early preliminary design phase of a SCC interface. This led us to apply a semi-structured approach where we combined some of the activities in the scenario analysis. The main focus was on the group discussion of safety issues, hazards, and possible mitigating measures.

In the case study, the scope is limited to one operator. In a future SCC, there could be several SCC stations with one operator at each station monitoring a fleet of unmanned ferries. An adapted scenario analysis should, in this context, also consider fleet operations and team collaboration. In our case study, the scenario analysis did not include events that involved such team cooperation. One of the reasons for this was that the CONOPS did not specify SCC organisational design.

One of the advantages of the method lies in its ability to generate discussions between stakeholders with different backgrounds on human factors issues, risks, and possible mitigating measures. The participants do not need extensive expert knowledge to facilitate the analysis, nor do they need to go through many complicated steps. The simplicity of the STEP diagram also makes it quicker for participants to familiarise themselves with the scenarios. We may risk simplifying the scenario analysis when trying to model complex systems. The analysis aims to be easy to understand and produces results, but its reliability and quality might be questionable for complex problems. In many ways, human-centred risk-informed decision-making must find the balance between making the risk analysis practicable and providing a sufficiently comprehensive scenario analysis for demonstrating safety.

All risk assessments have limitations and should not be used mechanically. As Brown and Elms (2015) stress, our perception of risk is constructed and affected by a range of typical biases and fallacies. For the scenario analysis, as for most risk assessments, the results are highly dependent on the expertise and experience of the participants. Another bias is the limitation of using input from a brainstorming session that gives the participants time to think and reflect on what they should do in a scenario. By doing so, what they say they will do may not be the same as what they would actually do. The facilitator should also be aware that individual operators present in the same context at the same time (i.e. in a situation or event) may ascribe different meanings to it. Hence, there is no objectively correct interpretation of what may go wrong. Different interpretations and perceptions of risks should be appreciated (see Goodman and Kuniavsky (2012)).

The scenario analysis has not relied on quantitative measures but instead reached its conclusions based on the qualitative information and contributions from the participants. In the analysis, we walked through critical events related to the SCCO and identified hazards and safety issues, aiming to improve risk understanding, which will provide valuable decision support. By focusing on the capabilities of a SCCO, we are relating risk to performance. This aligns with recently recommended practices where risk thinking is combined with principles and methods of robustness and resilience (Aven 2016).

## 6 Conclusions

Risk assessments can improve the understanding of the system, safety controls, and hazards of the activities under investigation. The traditional risk analysis methods applied in the maritime industry today may not be sufficient to address the complexity and emergent risk of MASS. Different risk analysis methods should be applied for different purposes at different phases of the design process. A risk analysis method focusing on human aspects should be used for risk-based design of MASS (DNV 2018). Examples of such methods are the CRIOP study, which can provide flexibility, and explainability and highlight safety issues by detailed identification of weak points when applied to an HMI design process. We have presented a qualitative case study of an interdisciplinary human-centred risk assessment method. The method is inspired by the scenario analysis in the CRIOP Framework. In this paper, we asked whether an adapted version of the scenario analysis could offer a helpful tool for supporting human-centred risk-informed decision-making in the design of a SCC.

The case study shows that the method could be applied for risk-informed decision-making in the design phase of SCC for MASS operation. In the case study workshop, a scenario of a handover situation where the simulated autonomous system asks for assistance from the SCCO was presented. The case study was carried out on a digital platform. Twelve people, including the design and engineering team of four, attended the workshop.

Findings from the study show that the scenario analysis method can be a valuable tool to address the human element in risk assessment by focusing on the operators' ability to handle the situation. Unlike traditional hazard analysis tools, the method is especially useful in identifying HAI-associated hazards. The method is cross-disciplinary and can be an arena for learning and sharing experiences. The simplicity of the method encourages an open discussion and involvement of several actors. Good design practice utilises human factor knowledge that emerges from the users sharing their experiences. The results reveal that the scenario analysis method could minimise the gap between WAI and WAD.

The experience of using the scenario analysis method for the evaluation and validation of control centres in the Norwegian oil and gas industry has been positive (Aas et al. 2009). In our study, the scenario analysis gave the workshop a necessary and efficient structure to analyse and discuss risks and mitigating measures. Hence, the analysis supports risk-based design for the human control element in autonomous ferries, allowing for human in the loop-capabilities. The design of an HMI supporting a safe and dynamic transition between autonomous and manual mode is a critical prerequisite for their implementation in urban waterways.

### 6.1 Further work

Based on the feedback on the issues of credibility, validity, and reliability of the scenario analysis (listed in Appendix 8.), there is a need for improved method guidance. Guidelines for the application at different phases of the design process should be developed, which is also a recommendation in a report on "Human-centred design and HMI in the development and implementation of autonomous systems in drilling and well" by (Johnsen et al. 2020). The performance shaping factors, checklists, and guidewords for the scenario analysis should be updated and specified for MASS operation.

With the increased opportunities and benefits of using simulations, a scenario analysis could benefit by having the scenarios presented in a simulation. The simulation would provide the participants with a more realistic understanding of the situation, different actions could be tested, and the scenario could be "paused" for elaboration on specific issues. The method should be applied to several cases and scenarios to increase its validity.

## Appendix 1

Figure 9 below is a modified screenshot of a part of the STEP diagram in the digital platform Miro. In the STEP diagram, the involved actors (denominates a person or object that affects the event (Johnsen et al. 2011)) are listed vertically, while the events are textboxes placed according to the order in which they occur. Arrows illustrate their relationship (causal links).

"Post-it"-notes were added during the workshop. Yellow "Post-it"-notes indicate hazards and safety issues, while green "Post-it"-notes indicate mitigating measures and possible solutions. The "Post-it"-notes are adjusted for better readability.

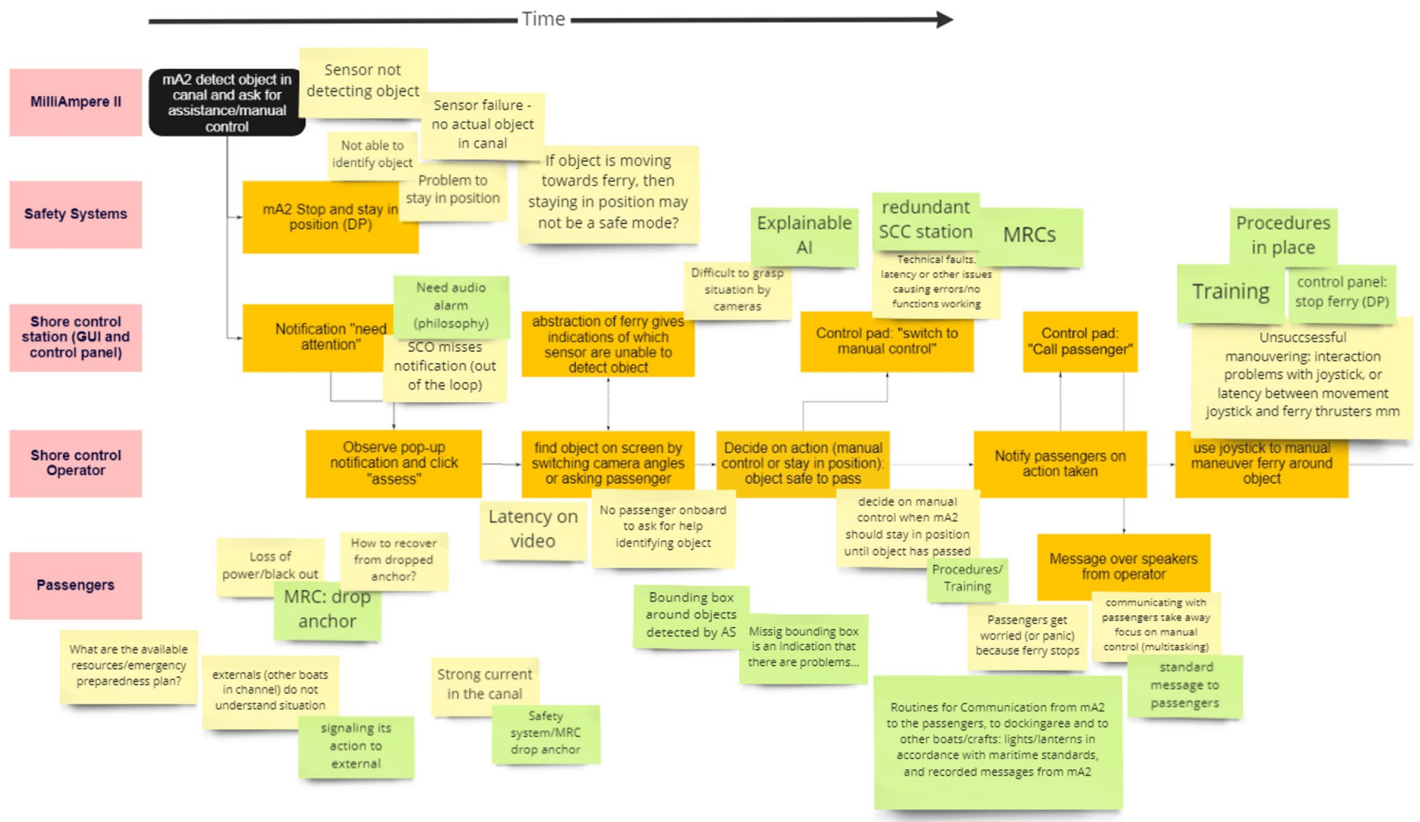

Time
MilliAmpere II
Safety Systems
Shore control station (GUI and control panel)
Shore control Operator
Passengers
mA2 detect object in canal and ask for assistance/manual control
Sensor not detecting object
Sensor failure - no actual object in canal
Not able to identify object
Problem to stay in position
mA2 Stop and stay in position (DP)
If object is moving towards ferry, then staying in position may not be a safe mode?
Notification "need attention"
Need audio alarm (philosophy)
SCO misses notification (out of the loop)
Observe pop-up notification and click "assess"
abstraction of ferry gives indications of which sensor are unable to detect object
find object on screen by switching camera angles or asking passenger
Latency on video
No passenger onboard to ask for help identifying object
Explainable AI
Difficult to grasp situation by cameras
Bounding box around objects detected by AS
Missig bounding box is an indication that there are problems...
redundant SCC station
MRCs
Technical faults, latency or other issues causing errors/no functions working
Control pad: "switch to manual control"
Decide on action (manual control or stay in position): object safe to pass
decide on manual control when mA2 should stay in position until object has passed
Procedures/ Training
Control pad: "Call passenger"
Notify passengers on action taken
Message over speakers from operator
Passengers get worried (or panic) because ferry stops
communicating with passengers take away focus on manual control (multitasking)
standard message to passengers
Routines for Communication from mA2 to the passengers, to dockingarea and to other boats/crafts: lights/lanterns in accordance with maritime standards, and recorded messages from mA2
Procedures in place
Training
control panel: stop ferry (DP)
Unsuccsessful manouvering: interaction problems with joystick, or latency between movement joystick and ferry thrusters mm
use joystick to manual maneuver ferry around object
Loss of power/black out
How to recover from dropped anchor?
MRC: drop anchor
What are the available resources/emergency preparedness plan?
externals (other boats in channel) do not understand situation
signaling its action to external
Strong current in the canal
Safety system/MRC drop anchor

**Fig. 9** A extract of the STEP diagram from Miro after the workshop

## Appendix 2

**Table 5** Summary of the scenario analysis methods' purpose and issues of credibility, validity, and reliability, including potential future work to address identified threats/weaknesses

| | General for the scenario analysis | Specific for the case study |
|---|---|---|
| Purpose and intent | The purpose depends on the context and design phase. The overall goal is to verify that the operator can perform the task at hand, considering cognitive abilities, human-system interactions, and other performance shaping factors | Risk analysis of the preliminary designed HMI to get input on a safe design of an SCC |
| External validity | The scenario analysis allows for tractability by employing a modelled scenario, making the actors, subsystems, and interactions visible to the analysts<br>The open questions allow for a greater level of discovery | All participants accepted the scenario as possible. Background information was provided. Participants unfamiliar with SCC and the operator's tasks were given time to ask clarifying questions. A long list of hazards and weak points were identified |
| Threats to validity | In a complex (intractable) system like MASS, we would never identify all scenarios of interest. There are limitations to the Scenario Analysis as a risk assessment<br>Not having the necessary knowledge of the systems to be assessed | There was a lack of detailed information on some technical solutions in the early design phase. Hence there is a risk for omissions, i.e., failing to identify crucial hazards and/or weak points |
| Future work to address validity | The validity (including limitations of the assessment) must be addressed and explained to the decision-makers as a part of the risk communication | |
| Credibility | The scenario analysis depends on the context. The scenarios to be analysed can be retrieved from a preliminary hazard identification. Identifying the risks by using different techniques increases the validity of the results | Participants were recruited from different fields of expertise. Engineers, interaction designers, human factors experts, and participants with seafarer experiences were part of the analysis group<br>The participants believed the results were valid. Participant check: The analysis report was taken back to the participants to be confirmed and to evaluate the method. In this way, the plausibility and truthfulness of the analysis were recognised and supported |
| Threats to credibility | Not having the "right" participants. The strength of the results depends on the expertise/knowledge of the participants | Biases: Some participants had already made decisions when designing the conceptual HMI<br>The time constraint limited the brainstorming process causing potential valuable input not to be considered |
| Future work to address credibility | Select the participants carefully<br>Give the participants sufficient time to write "Post-it" notes and post these anonymously<br>Furthermore, develop the digital platform for the analysis. The facilitator should create a good atmosphere and aim to be non-judgmental and not interrupt | |

**Table 5** (continued)

| | General for the scenario analysis | Specific for the case study |
|---|---|---|
| Reliability | The method is well-known in the oil and gas industry (Johnsen et al. 2011). It is seen as a "best practice" solution for integrating Human Factors into the design process<br>The purpose is not to attain the same results but to provide a methodological approach to identify hazards and weak points | The 12 participants came from various backgrounds: a cross-disciplinary group of hardware designers, human factors experts, programmers, engineers, and people with experience as seafarers<br>Several of the identified hazards and safety issues had been briefly mentioned in the Preliminary Hazard analysis carried out by zeabus |
| Threats to reliability | The results of an analysis are highly likely to vary depending on the participants and the design phase<br>If the STEP diagram is confusing and does not provide the participants with the proper explanation of the events, the quality of the analysis will be poor | The analyst team did not include the actual end-user, the SCCO<br>The time constraint and limited knowledge of the system architecture were mentioned as a challenge for the participants |
| Further work to address reliability | Make sure to invite the "right" participants. Have sufficient and updated CONOPS<br>Involvement of experts in selecting critical scenarios and making the STEP diagram | |
| Usefulness | The method can fill an "including the end-user" gap by providing a human (end-user) centred approach to the risk assessment<br>In studies applying the method in the oil and gas sector (Aas et al. 2009), users reported an increased understanding of the perspectives, needs, and requirements of the control room operators and potential hazards in the socio-technical system | Many participants reported that they could easily understand the scenario making it easy to discuss what-if questions, address hazards and evaluate the preliminary HMI design<br>Discussions between the programmers and the hardware designers on how the operator would handle surprises helped meet the goal of getting input on a safe(r) designed solution. Several participants reported that they had learned a new framework and considered it a supportive tool for risk assessment in the design phase |
| Threats to usefulness | Cost/benefit of the time/resources spent on the analysis<br>The usefulness of qualitative risk assessments to support defining a specific safety level is challenging<br>Whether the analysis generates enough information for the decision-makers depends on selected scenarios and the quality of the risk assessment | One participant questioned why the passenger ferry was in the final construction phase while the SCC was in the design phase. This limits the opportunity to modify the design and technical solutions<br>Several participants stressed the need for Scenario Analysis of several scenarios |
| Further work to address the usefulness | Need for better method guidance: The scenario analysis's performance shaping factors, checklists, and guidewords should be updated and specified for MASS operation (primarily focusing on navigational aspects) | |

**Acknowledgements** The authors thank all the NTNU Shore Control Lab and all participants in the workshops.

**Funding** Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital) The research is partly sponsored by the Norwegian Research Council under project numbers 267860 and 313921.

## Declarations

**Competing interests** The authors declare no competing interests.

## References

Aas AL, Johnsen S O, Skramstad T. (2009). CRIOP: a human factors verification and validation methodology that works in an industrial setting. In B. Buth, G. Rabe, & T. Seyfarth, Computer Safety, Reliability, and Security Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04468-7_20

Aven T (2009) Perspectives on risk in a decision-making context–review and discussion. Saf Sci 47(6):798–806. https://doi.org/10.1016/j.ssci.2008.10.008

Aven T (2012) The risk concept—historical and recent development trends. Reliab Eng Syst Saf 99:33–44. https://doi.org/10.1016/j.ress.2011.11.006

Aven T (2016) Risk assessment and risk management: review of recent advances on their foundation. Eur J Oper Res 253:1–13. https://doi.org/10.1016/j.ejor.2015.12.023

Aven T (2019) The cautionary principle in risk management: foundation and practical use. Reliab Eng Syst Saf 191:106585. https://doi.org/10.1016/j.ress.2019.106585

Aven T, Renn O (2009) The role of quantitative risk assessments for characterizing risk and uncertainty and delineating appropriate risk management options, with special emphasis on terrorism risk. Risk Anal Int J 29(4):587–600. https://doi.org/10.1111/j.1539-6924.2008.01175.x

Banda OAV, Kannos S, Goerlandt F, van Gelder PH, 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 191:106584. https://doi.org/10.1016/j.ress.2019.106584

Blackett, C. (2021). Human-centered design in an automated world. In D. Russo, T. Ahram, W. Karwowski, G. Di Bucchianico, R. Taiar, Intelligent Human Systems Integration 2021. https://doi.org/10.1007/978-3-030-68017-6_3

Bolbot V, Theotokatos G, Wennersberg LAL. et al. (2020). AUTOSHIP deliverable D2.4a: risk assessments, fail-safe procedures and acceptance criteria The Inland Waterway vessel analysis, December 2020.

Boring RL, Hendrickson SML, Forester JA, Tran TQ, Lois E (2010) Issues in benchmarking human reliability analysis methods: a literature review. Reliab Eng Syst Saf 95(6):591–605. https://doi.org/10.1016/j.ress.2010.02.002

Brown CB, Elms DG (2015) Engineering decisions: Information, knowledge and understanding. Struct Saf 52:66–77. https://doi.org/10.1016/j.strusafe.2014.09.001

ClassNK (2020) Guidelines for Automated/autonomous Operation on Ships (Ver. 1.0)

Dekker S (2014) The field guide to understanding "human error." Ashgate Publishing, Ltd. https://doi.org/10.1201/9781317031833

DNV GL. (2018). Autonomous and remotely operated ships. *Class Guideline DNVGL-CG-0264.*

Dybvik H, Veitch E, Steinert M. (2020). Exploring challenges with designing and developing shore control centers (SCC) for autonomous ships. In *Proceedings of the Design Society: DESIGN Conference* (Vol. 1, pp. 847–856). Cambridge University Press. https://doi.org/10.1017/dsd.2020.131

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. https://doi.org/10.1016/j.oceaneng.2020.107188

French S, Bedford T, Pollard SJ, Soane E (2011) Human reliability analysis: a critique and review for managers. Saf Sci 49(6):753–763. https://doi.org/10.1016/j.ssci.2011.02.008

French S, Niculae C. (2005). Believe in the model: mishandle the emergency. J Homeland Sec Emerg Manag 2https://doi.org/10.2202/1547-7355.1108

Goerlandt F (2020) Maritime Autonomous Surface Ships from a risk governance perspective: Interpretation and implications. Saf Sci 128:104758. https://doi.org/10.1016/j.ssci.2020.104758

Goerlandt F, Montewka J (2015) Maritime transportation risk analysis: Review and analysis in light of some foundational issues. Reliab Eng Syst Saf 138:115–134. https://doi.org/10.1016/j.ress.2015.01.025

Goodman E, Kuniavsky M (2012) Observing the user experience: a practitioner's guide to user research. Elsevier

Hirata C, Nadjm-Tehrani S. (2019).Combining GSN and STPA for safety arguments. Int Confe Comp Saf Reliab Sec https://doi.org/10.1007/978-3-030-26250-1_1

Hoem ÅS. (2019). The present and future of risk assessment of MASS: a literature review. Proceedings of the 29th European Safety and Reliability Conference (ESREL), Hannover, Germany,

Hoem Å, Johnsen S, Fjørtoft K, Rødseth Ø, Jenssen G, Moen T. (2021). Improving safety by learning from automation in transport systems with a focus on sensemaking and meaningful human control. In Sensemaking in Safety Critical and Complex Situations (191–207) CRC Press.

Hollnagel E (1996) Reliability analysis and operator modelling. Reliab Eng Syst Saf 52(3):327–337. https://doi.org/10.1016/0951-8320(95)00143-3

Hollnagel E (2000) Looking for errors of omission and commission or The Hunting of the Snark revisited. Reliab Eng Syst Saf 68(2):135–145. https://doi.org/10.1016/S0951-8320(00)00004-1

Hollnagel, E. (2017). Can we ever imagine how work is done. *HindSight*, *25*, p. 10–13. Retrieved from https://www.eurocontrol.int/sites/default/files/publication/files/hindsight25.pdf

Hollnagel E, Woods DD, Leveson N. (2006). Resilience engineering: concepts and precepts. Ashgate Publishing, Ltd.

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.

IMO. (2018a). Revised Guidelines for Formal Safety Assessment (FSA) for the use in the IMO Rule-Making Process (No. MSC-MEPC.2/Circ.12/Rev.2). IMO, London, UK.

IMO. (2018b). Regulatory scoping exercise for the use of maritime autonomous surface ships (MASS).

IMO. (2019). Interim guidelines for MASS trials (No. MSC.1/Circ.1604). IMO, London, UK.

ISO11064. (2013). Ergonomic design of control centres In International Organization for Standardization.

ISO9241–210. (2019). Ergonomics of human-system interaction: part 210: human-centred design for interactive systems. In: International Organization for Standardization.

Johnsen SO, Bjørkli C, Steiro T, Fartum H, Haukenes H, Ramberg J, Skriver J. (2011). CRIOP: a scenario method for crisis intervention and operability analysis

Johnsen SO, Holen S, Aalberg AL, Bjørkevoll KS, Evjemo TE, Johansen G, Myklebust T, Okstad E, Pavlov A, Porathe T. (2020). Automatisering og autonome systemer: Menneskesentrert design i boring og brønn (in Norwegian)

Johnsen SO, Porathe T (2021) Sensemaking in safety critical and complex situations: human factors and design. CRC Press

Joppe M. (2000). The research process, as quoted in understanding reliability and validity in qualitative research nahid golafshani. The Qualitative Report Volume, 8. https://doi.org/10.46743/2160-3715/2003.1870

Leveson NG (2011) Applying systems thinking to analyze and learn from events. Saf Sci 49(1):55–64. https://doi.org/10.1016/j.ssci.2009.12.021

Leveson NG (2016) Engineering a safer world: systems thinking applied to safety. The MIT Press

Leveson, N. G. (2020). Safety III: a systems approach to safety and resilience. *MIT Engineering systems lab*

Leveson NG, Stephanopoulos G. (2013). A system-theoretic, control-inspired view and approach to process safety

LR, LsR. (2016). *Risk based designs (RBD), shipright design and construction - additional design procedures*

Lützhöft, M. (2004). *"The technology is great when it works": Maritime Technology and Human Integration on the Ship's Bridge* Linköping University Electronic Press.

Lützhöft M, Hynnekleiv A, Earthy JV et al (2019) Human-centred maritime autonomy-An ethnography of the future. In Journal of Physics: Conference Series. 1357(1):012032. https://doi.org/10.1088/1742-6596/1357/1/012032

Microsoft. (2022). *Get started with Microsoft Teams*. https://support.microsoft.com/en-us/office/get-started-with-microsoft-teams-b98d533f-118e-4bae-bf44-3df2470c2b12

Mills AJD, Gabrielle W, Elden. (2010). Encyclopedia of case study research. https://doi.org/10.4135/9781412957397

miro. (2022). *The online whiteboard for real-time collaboration and asynchronous teamwork*. https://miro.com/online-whiteboard/

NMA. (2020). *Guidelines for the construction or installation of automated functionality, with the intention of being able to perform unmanned or partially unmanned operations*. Retrieved from https://www.sdir.no/contentassets/2b487e1b63cb47d39735953ed492888d/rsv-12-2020.pdf

Papanikolaou A, Soares CG (2009) Risk-based ship design: methods, tools and applications. Springer

Porathe T, Hoem ÅS, Rødseth ØJ, Fjørtoft KE., Johnsen SO. (2018). At least as safe as manned shipping? Autonomous shipping, safety and "human error". Safety and Reliability–Safe Societies in a Changing World. Proceedings of ESREL 2018, June 17–21, 2018, Trondheim, Norway.

Ramos MA, Thieme CA, Utne IB, Mosleh A (2020) Human-system concurrent task analysis for maritime autonomous surface ship operation and safety. Reliab Eng Syst Saf 195:106697. https://doi.org/10.1016/j.ress.2019.106697

Rausand M. (2013). Risk assessment: theory, methods, and applications (Vol. 115). John Wiley & Sons.

Register L. (2017). Cyber-enabled ships shipright procedure assignment for cyber descriptive notes for autonomous & remote access ships. Lloyd's Register, Guidance document Version 2.0.

Salmon PM, Read GJ, Walker GH, Stevens NJ, Hulme A, McLean S, Stanton NA. (2020). Methodological issues in systems human factors and ergonomics: perspectives on the research–practice gap, reliability and validity, and prediction. Human Factors and Ergonomics in Manufacturing & Service Industries https://doi.org/10.1002/hfm.20873

Thieme CA, Utne IB, Haugen S (2018) Assessing ship risk model applicability to marine autonomous surface ships. Ocean Eng 165:140–154. https://doi.org/10.1016/j.oceaneng.2018.07.040

Trudel CM. (2021). Useful, usable and used? In Recent Advances in Technologies for Inclusive Well-Being (pp. 43–63). Springer. https://doi.org/10.1007/978-3-030-59608-8_4

Utne IB, Rokseth B, Sørensen AJ, Vinnem JE (2020) Towards supervisory risk control of autonomous ships. Reliab Eng Syst Saf 196:106757. https://doi.org/10.1016/j.ress.2019.106757

Utne IB, Sørensen AJ, Schjølberg I. (2017). Risk management of autonomous marine systems and operations. International Conference on Offshore Mechanics and Arctic Engineeringhttps://doi.org/10.1115/OMAE2017-61645

van den Broek JH, Griffioen JJ, van der Drift, MM. (2020). Meaningful human control in autonomous shipping: an overview. IOP Conference Series: Materials Science and Engineeringhttps://doi.org/10.1088/1757-899X/929/1/012008

Vasstein K, Brekke E, Mester R, Eide E (2020) Autoferry Gemini: a real-time simulation platform for electromagnetic radiation sensors on autonomous ships. IOP Conference Series: Mater Sci Eng 929:012032. https://doi.org/10.1088/1757-899X/929/1/012032

Veitch E, Alsos OA (2021) Human-centered explainable artificial intelligence for marine autonomous surface vehicles. J Marine Sci Eng 9(11):1227. https://doi.org/10.3390/jmse9111227

Veitch E, Alsos OA (2022) A systematic review of human-AI interaction in autonomous ship systems. Saf Sci 152:105778. https://doi.org/10.1016/j.ssci.2022.105778

Ventikos N, Louzis K, Sotiralis P, Koimtzoglou A, Annetis E. (2021). Integrating human factors in risk-based design: a critical review. Ergoship 2021. ISBN 978–89–5708–330–7

Veritas, B. (2019). *Guidelines for autonomous shipping*

Wennersberg LAL, Nordahl H, Rødseth ØJ, Fjørtoft K, Holte EA. (2020). A framework for description of autonomous ship systems and operations. IOP Conference Series: Materials Science and Engineeringhttps://doi.org/10.1088/1757-899X/929/1/012004

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. https://doi.org/10.1016/j.ress.2017.03.029

Wróbel K, Montewka J, Kujala P (2018) System-theoretic approach to safety of remotely-controlled merchant vessel. Ocean Eng 152:334–345. https://doi.org/10.1016/j.oceaneng.2018.01.020

Yang X, Utne IB, Sandøy SS, Ramos MA, Rokseth B (2020) A systems-theoretic approach to hazard identification of marine systems with dynamic autonomy. Ocean Eng 217:107930. https://doi.org/10.1016/j.oceaneng.2020.107930

Yin RK (2009) Case study research: Design and methods, vol 5. SAGE

zeabuz. (2021). *ConOps for autonomous passenger ferry in Trondheim, rev. C*.

Zhou X-Y, Liu Z-J, Wang F-W, Wu Z-L, Cui R-D (2020) Towards applicability evaluation of hazard analysis methods for autonomous ships. Ocean Eng 214:107773. https://doi.org/10.1016/j.oceaneng.2020.107773

Wróbel K, Gil M, Krata P et al (2021) On the use of leading safety indicators in maritime and their feasibility for Maritime Autonomous Surface Ships. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability. https://doi.org/10.1177/1748006X211027689

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

**Åsa S. Hoem**[1] · **Erik Veitch**[1] · **Kjetil Vasstein**[2]

Erik Veitch
erik.a.veitch@ntnu.no

Kjetil Vasstein
kjetil.vasstein@ntnu.no

[1] Department of Design, The Norwegian University of Science and Technology (NTNU), Trondheim, Norway

[2] Department of Engineering Cybernetics, The Norwegian University of Science and Technology (NTNU), Trondheim, Norway