WHEN TRUST MATTERS

Teaming with automation in future maritime systems

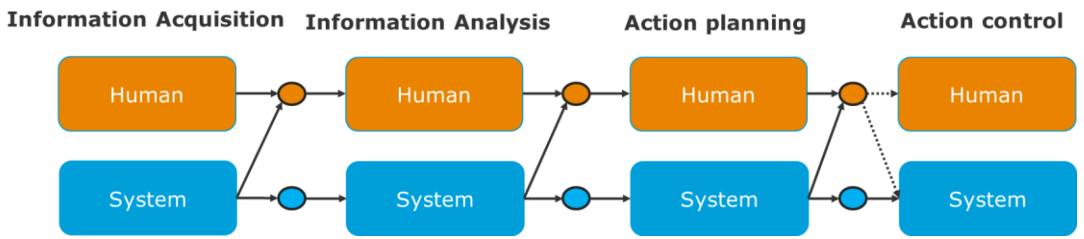
HFC forum 27.04.2023

Koen van de Merwe, Principal Researcher & PhD candidate

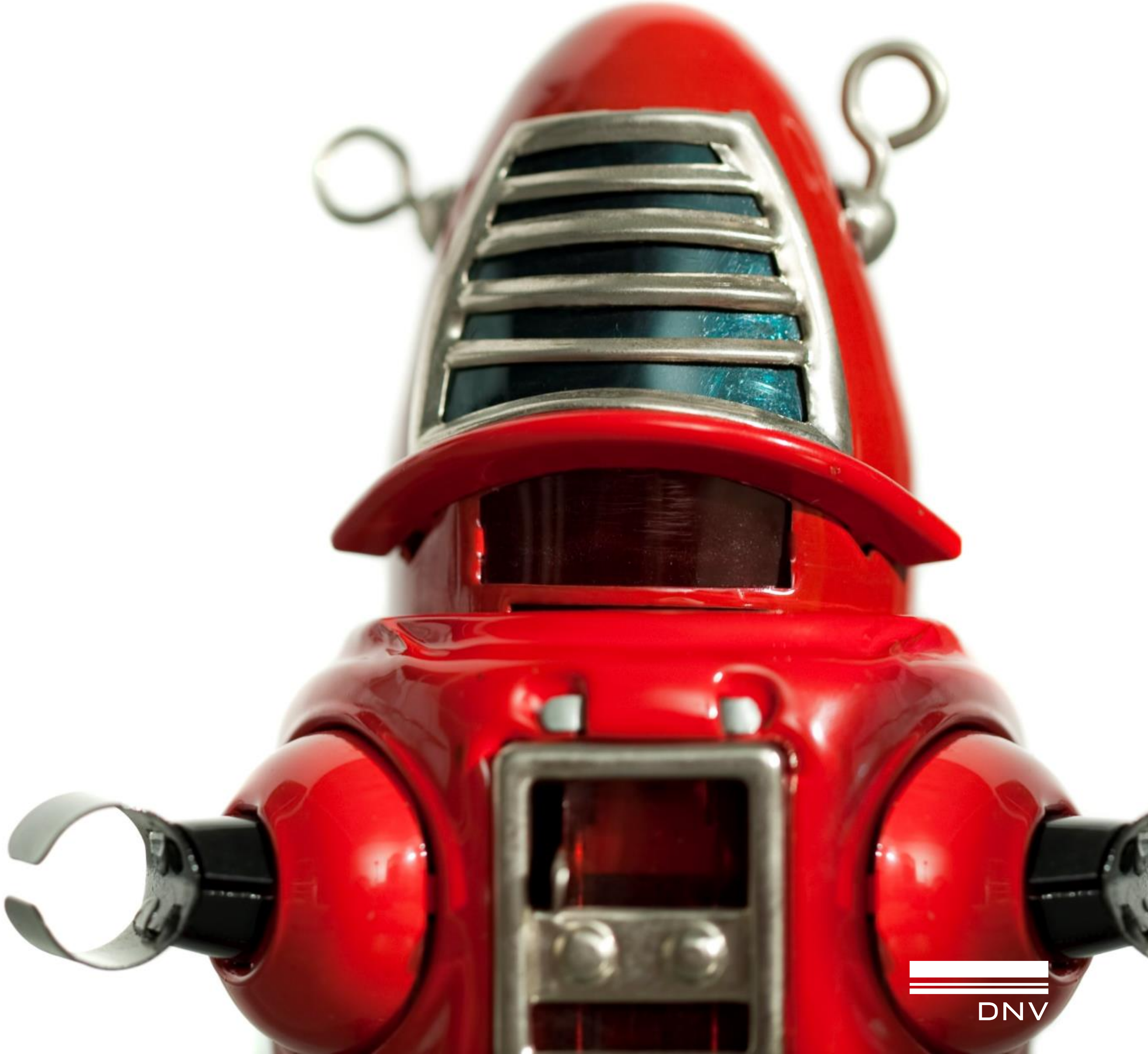
27 April 2023

Teaming with automation

"…the dynamic, interdependent coupling between one or more human operators and one or more automated systems requiring collaboration and coordination to achieve successful task completion." (Cuevas et al, 2007)



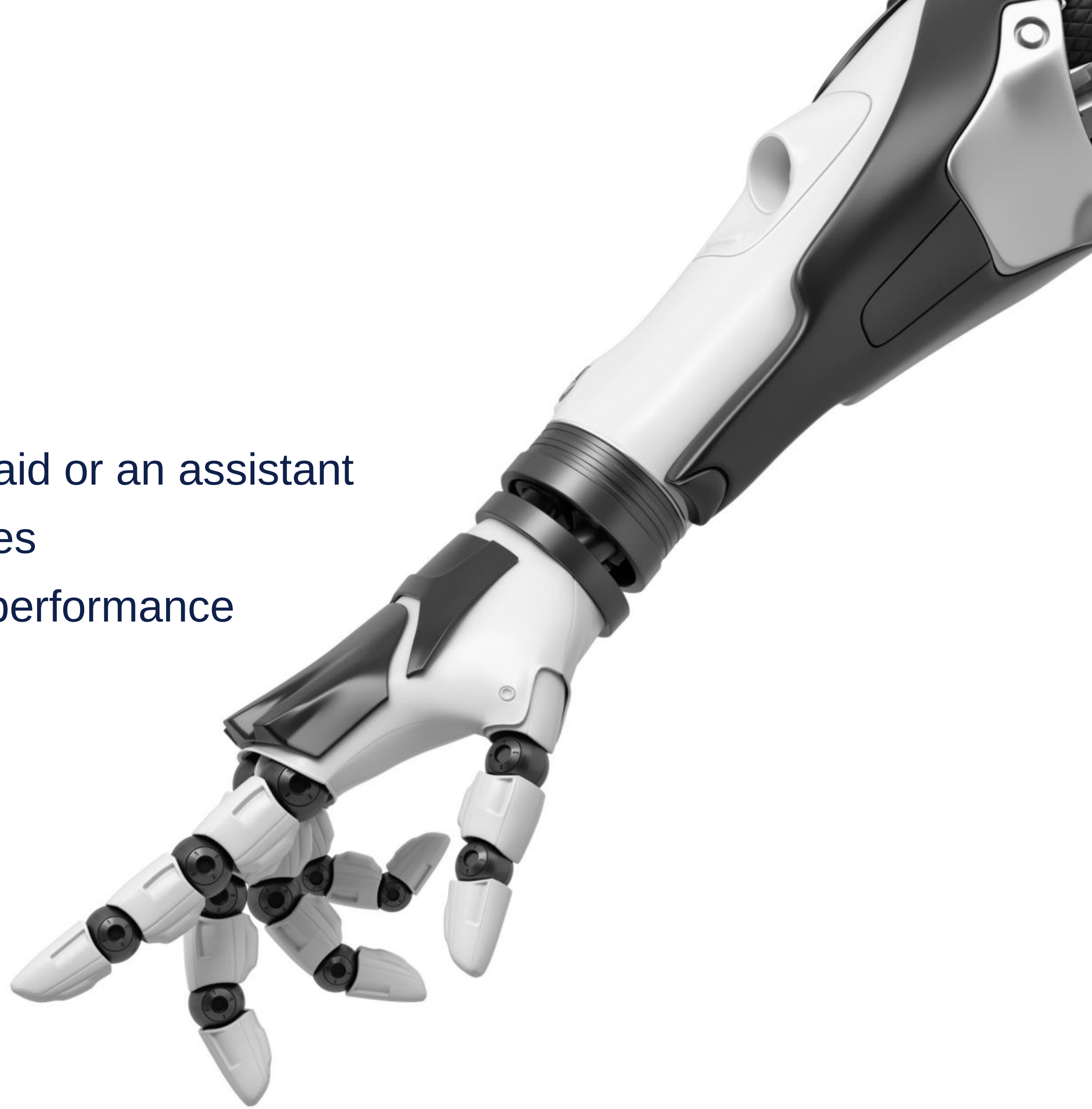
(DNV, 2018)



DNV

Some recent technological developments

Teaming with automation



- Concepts of operation (Endsley, 2017)
  - Human as a supervisor over automation that acts as an aid or an assistant
  - Humans and autonomy acting as collaborating teammates
  - Automation that oversees and acts as a limit on human performance

- Assumes
  - High levels of autonomy
  - New interfaces for shared situation awareness

Understanding your teammate's performance

Understanding of the agent's performance in its context

SA	Task Agent Team

Understanding of the agent's performance in general

Mental model	Logic, components Functioning Reliability Capabilities Limitations

(Endsley, 2023)

Understanding your teammate’s performance

Concept of operation:

- Supports supervision, intervention, and backup
- Teammates
- Limit

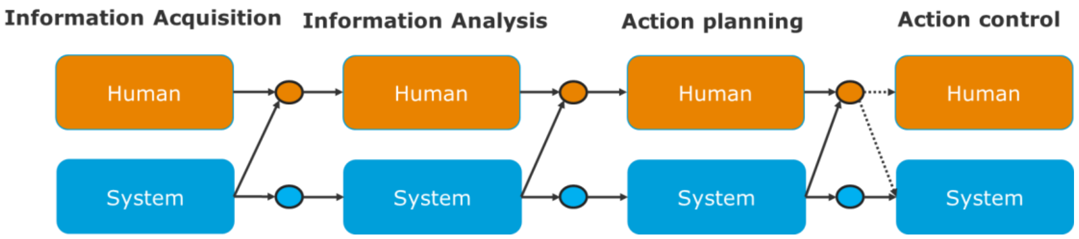
Transparency

SA

Task

Agent

Team



(DNV, 2018)

Mental model

Logic, components

Functioning

Reliability

Capabilities

Limitations

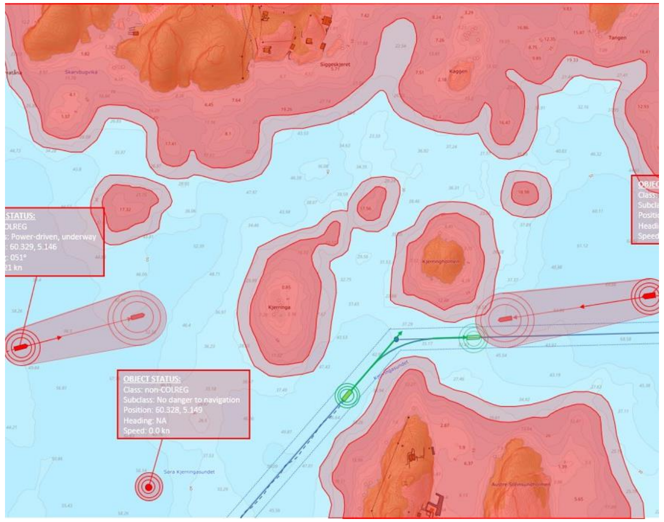
(Endsley, 2023)

Transparency: observability and predictability (MITRE 2018)

Making apparent what the system is currently doing but also why it is doing it and what it will do next

(Endsley 2017)

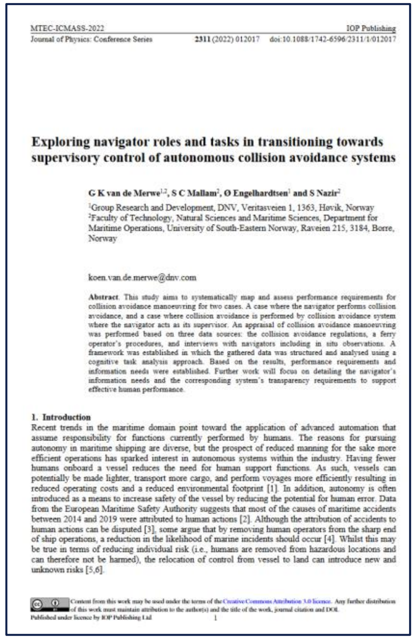
Transparency as a *system property*



Goal is to enable the operator to maintain proper SA of the system in its tasking environment without becoming overloaded

(Mercado et al, 2016)

Maritime collision avoidance


MTEC-ICMASS-2022 IOP Publishing
Journal of Physics: Conference Series **2311** (2022) 012017 doi:10.1088/1742-6596/2311/1/012017

Exploring navigator roles and tasks in transitioning towards supervisory control of autonomous collision avoidance systems

G K van de Merwe[1,2], S C Mallam[2], Ø Engelhardtsen[1] and S Nazir[2]
[1]Group Research and Development, DNV, Veritasveien 1, 1363, Høvik, Norway
[2]Faculty of Technology, Natural Sciences and Maritime Sciences, Department for Maritime Operations, University of South-Eastern Norway, Raveien 215, 3184, Borre, Norway

koen.van.de.merwe@dnv.com

Abstract. This study aims to systematically map and assess performance requirements for collision avoidance manoeuvring for two cases. A case where the navigator performs collision avoidance, and a case where collision avoidance is performed by collision avoidance system where the navigator acts as its supervisor. An appraisal of collision avoidance manoeuvring was performed based on three data sources: the collision avoidance regulations, a ferry operator's procedures, and interviews with navigators including in situ observations. A framework was established in which the gathered data was structured and analysed using a cognitive task analysis approach. Based on the results, performance requirements and information needs were established. Further work will focus on detailing the navigator's information needs and the corresponding system's transparency requirements to support effective human performance.

1. Introduction
Recent trends in the maritime domain point toward the application of advanced automation that assume responsibility for functions currently performed by humans. The reasons for pursuing autonomy in maritime shipping are diverse, but the prospect of reduced manning for the sake more efficient operations has sparked interest in autonomous systems within the industry. Having fewer humans onboard a vessel reduces the need for human support functions. As such, vessels can potentially be made lighter, transport more cargo, and perform voyages more efficiently resulting in reduced operating costs and a reduced environmental footprint [1]. In addition, autonomy is often introduced as a means to increase safety of the vessel by reducing the potential for human error. Data from the European Maritime Safety Authority suggests that most of the causes of maritime accidents between 2014 and 2019 were attributed to human actions [2]. Although the attribution of accidents to human actions can be disputed [3], some argue that by removing human operators from the sharp end of ship operations, a reduction in the likelihood of marine incidents should occur [4]. Whilst this may be true in terms of reducing individual risk (i.e., humans are removed from hazardous locations and can therefore not be harmed), the relocation of control from vessel to land can introduce new and unknown risks [5,6].

 1

Supporting human supervision in autonomous collision avoidance through system transparency: a structured and systematic approach

Koen van de Merwe[1,2*], Steven Mallam[2], Øystein Engelhardtsen[1], Salman Nazir[2]

[1] Group Research and Development, DNV, Veritasveien 1, 1363, Høvik, Norway
[2] Faculty of Technology, Natural Sciences and Maritime Sciences, Department for Maritime Operations, University of South-Eastern Norway, Raveien 215, 3184, Borre, Norway
*Corresponding author: Email: koen.van.de.merwe@dnv.com

Abstract

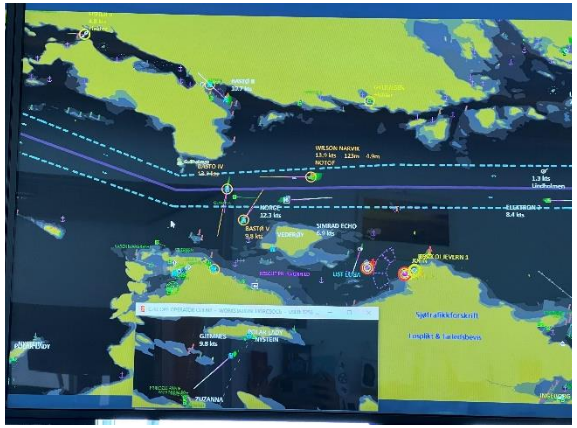
Given the foreseen supervisory role of the human in autonomous shipping, this study aims to explore the role of the human in supervised autonomous collision avoidance. A systematic analysis of goals, decisions, cognitive tasks, and information needs is performed by applying a Goal-Directed Task Analysis on conventional and human supervised collision avoidance cases. Data was obtained from in situ observations of collision avoidance manoeuvres, in situ interviews with navigators, an appraisal of the collision regulations, and company procedures. By using a structured and systematic approach, this study identified the information requirements for making autonomous collision avoidance systems transparent to its user. The results highlight the need for continuous, sufficient, and relevant information from the collision avoidance system to support human supervisory control in a dynamic context.

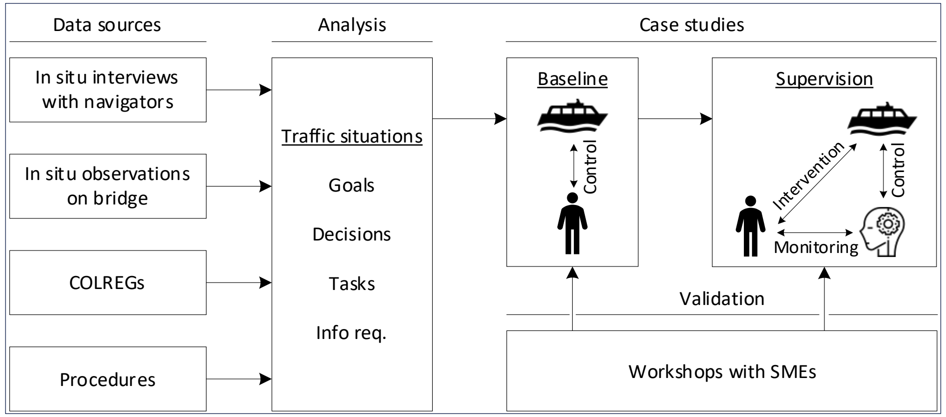
1 Introduction

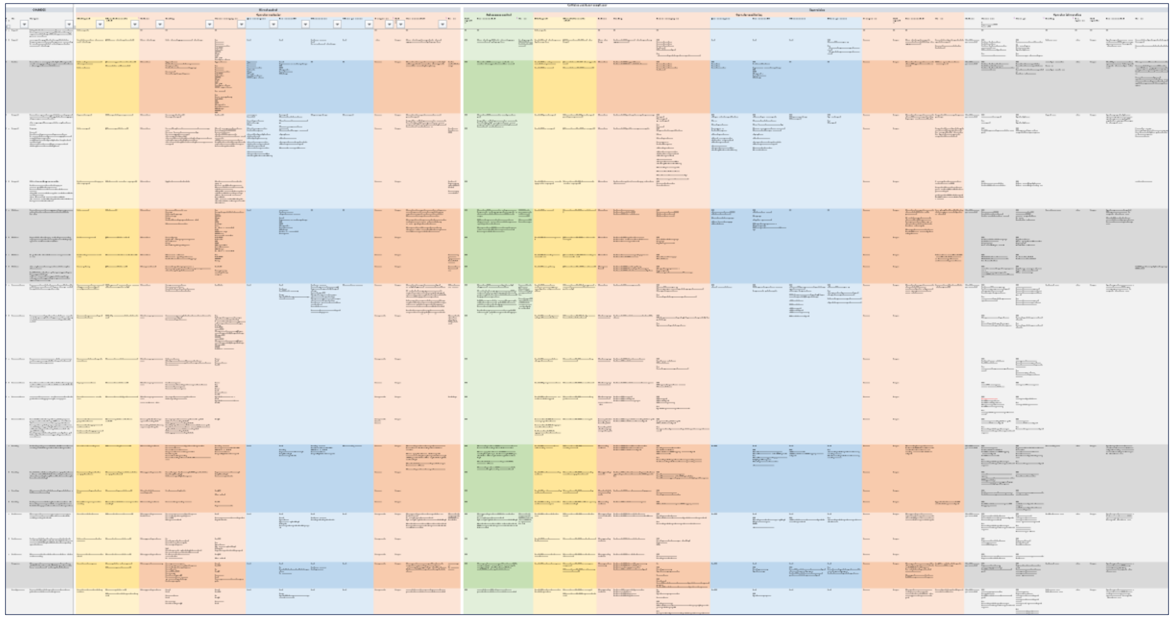
1.1 Towards autonomous shipping

In recent years, the maritime industry has shown increased interest in developing autonomous solutions with the aim to achieve more efficient, punctual, and safer operations (Kretschmann *et al.* 2017; Wróbel *et al.* 2017). Among others, the MUNIN research project (Maritime Unmanned Navigation through Intelligence in Networks) explored safety and autonomy in a dry bulk carrier for deep-sea shipping (Burmeister et al. 2014) and DNV demonstrated its ReVolt concept to explore crewless short-sea shipping (DNV 2018). Furthermore, Rolls Royce demonstrated an autonomous ferry in Finland showing its capabilities for fusing sensor information, detecting obstacles, avoiding conflicts and berthing automatically (Rolls Royce 2018). In Japan the commercial ship Suzaku conducted a 790-kilometre trial using a container ship, testing its autonomous navigation capabilities (NYK 2022). In Norway, the Yara Birkeland container ship has been launched and aims to start sailing autonomously in 2024 (Yara International 2022). Finally, the ASKO barges will commence service in 2022 and the aim is to sail autonomously, with remote supervision, in 2024 (ASKO 2022).

Although the reasons for pursuing autonomous operations are diverse, the prospect of reduced manning has sparked the interest of the industry. Autonomous and unmanned ships may allow for new and more efficient ship designs enabling lighter structures, reduced voyage costs, and/or increased payload capacity (Kretschmann *et al.* 2017; Kurt & Aymelek 2022). In addition, the prospects of reduced crew (Kooij & Hekkenberg 2020), and safer operations by removing humans from the sharp end of the operations are also motivating factors (Wróbel et al. 2017). One key challenge to be resolved in moving towards autonomous and potentially







van de Merwe, K., Mallam, S., Engelhardtsen, Ø., & Nazir, S. (under review). *Supporting human supervision in autonomous collision avoidance through system transparency: A structured and systematic approach*.

van de Merwe, G. K., Mallam, S. C., Engelhardtsen, Ø., & Nazir, S. (2022). Exploring navigator roles and tasks in transitioning towards supervisory control of autonomous collision avoidance systems. *Journal of Physics: Conference Series*, *2311*(1), 012017. https://doi.org/10.1088/1742-6596/2311/1/012017

A model for human information processing

Framework to decide at

- Which parts of the system should be automated

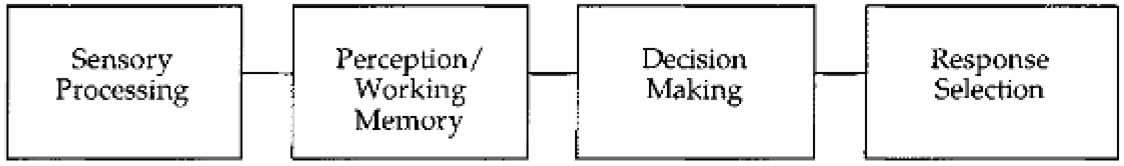
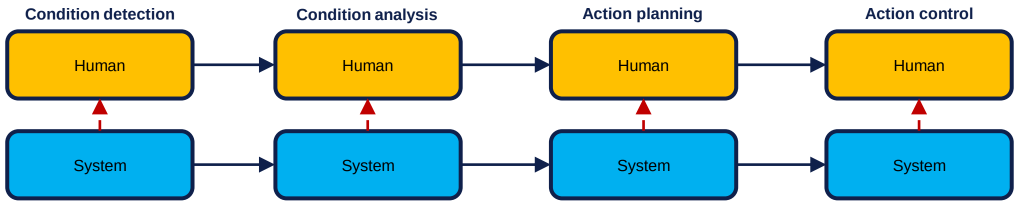


Fig. 1. Simple four-stage model of human information processing.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans.*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

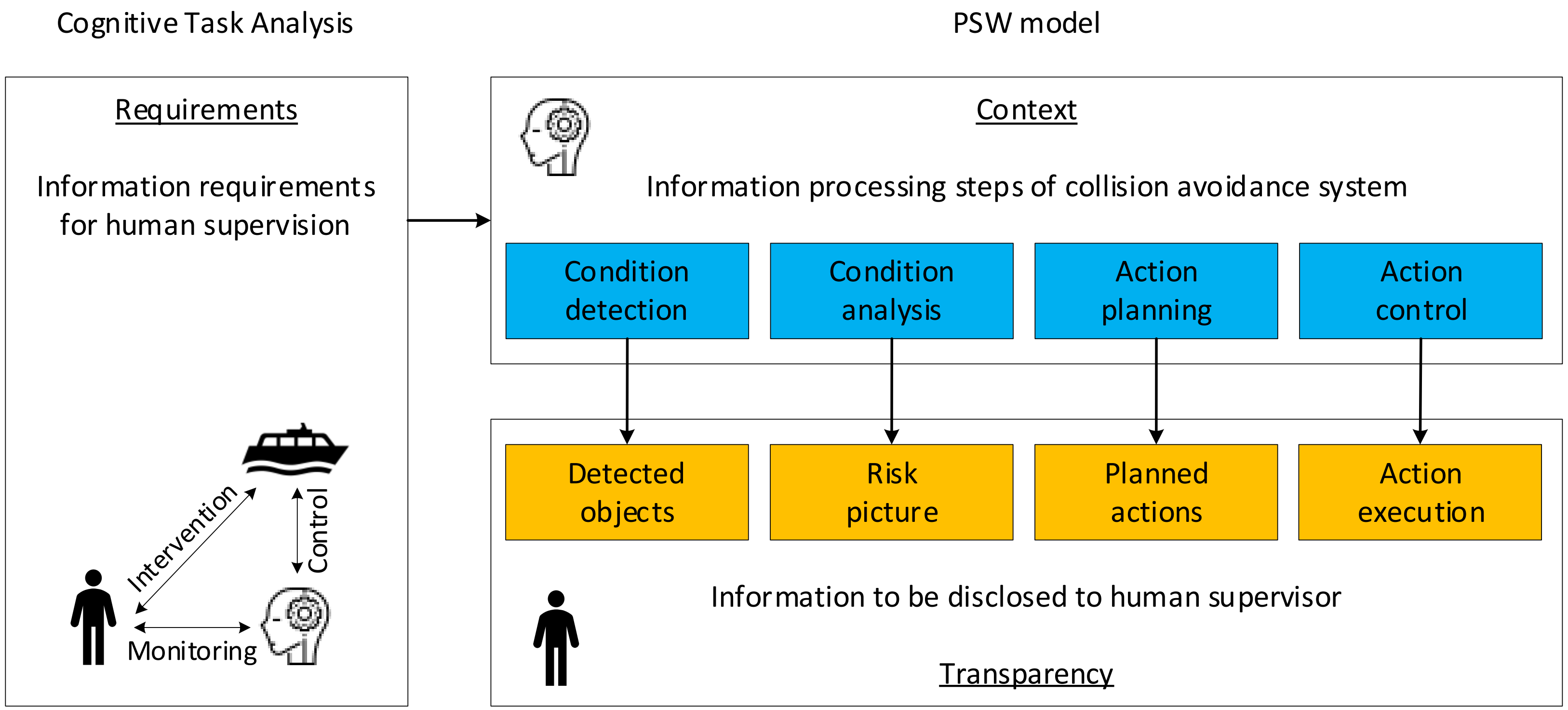
Or…

- Which information from the system should be disclosed to the human supervisor



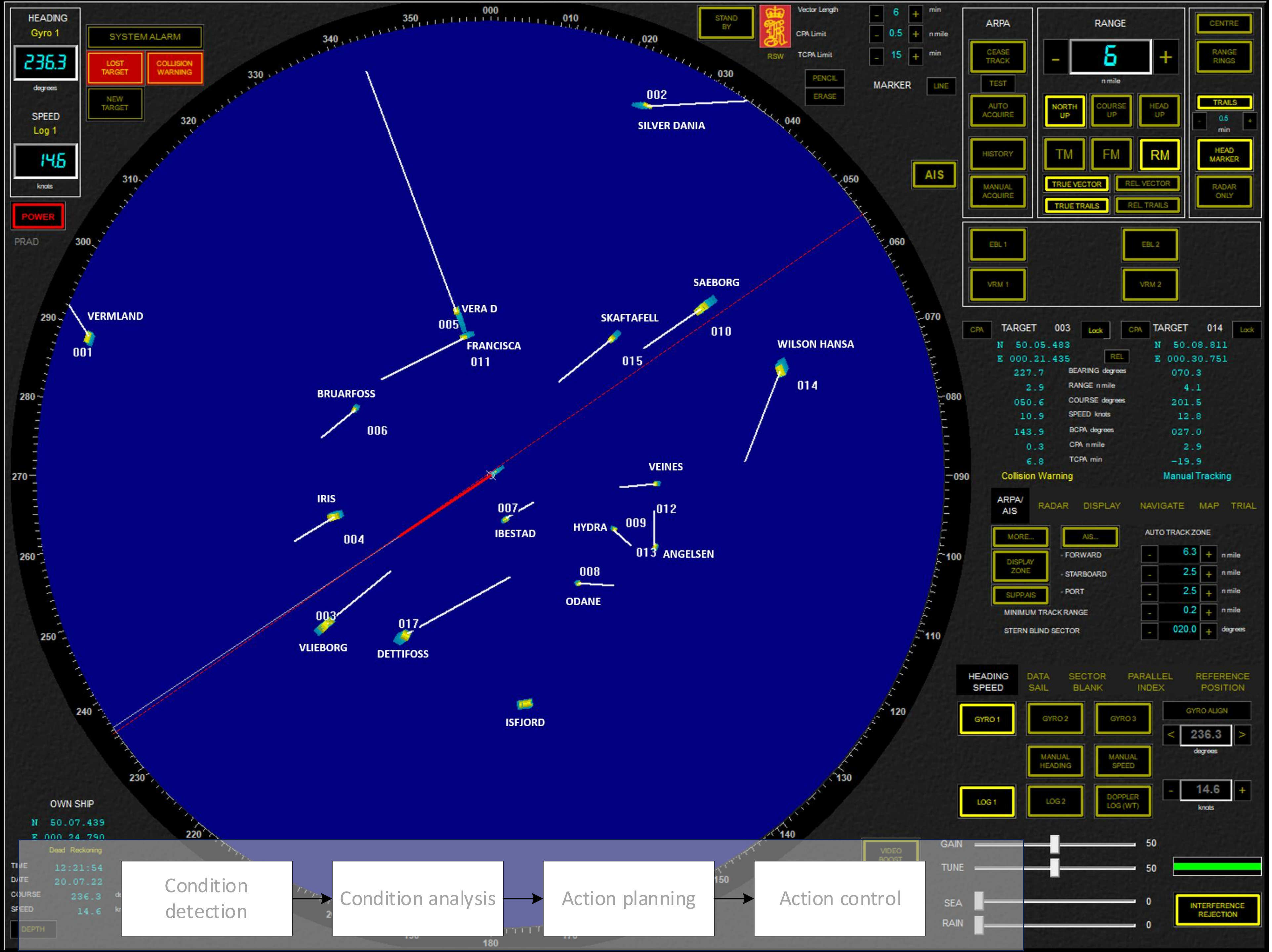
DNV. (2018). *DNVGL-CG-0264: Autonomous and remotely operated ships*. http://rules.dnvgl.com/docs/pdf/dnvgl/cg/2018-09/dnvgl-cg-0264.pdf

Analysis framework and processing model



(DNV, 2018)

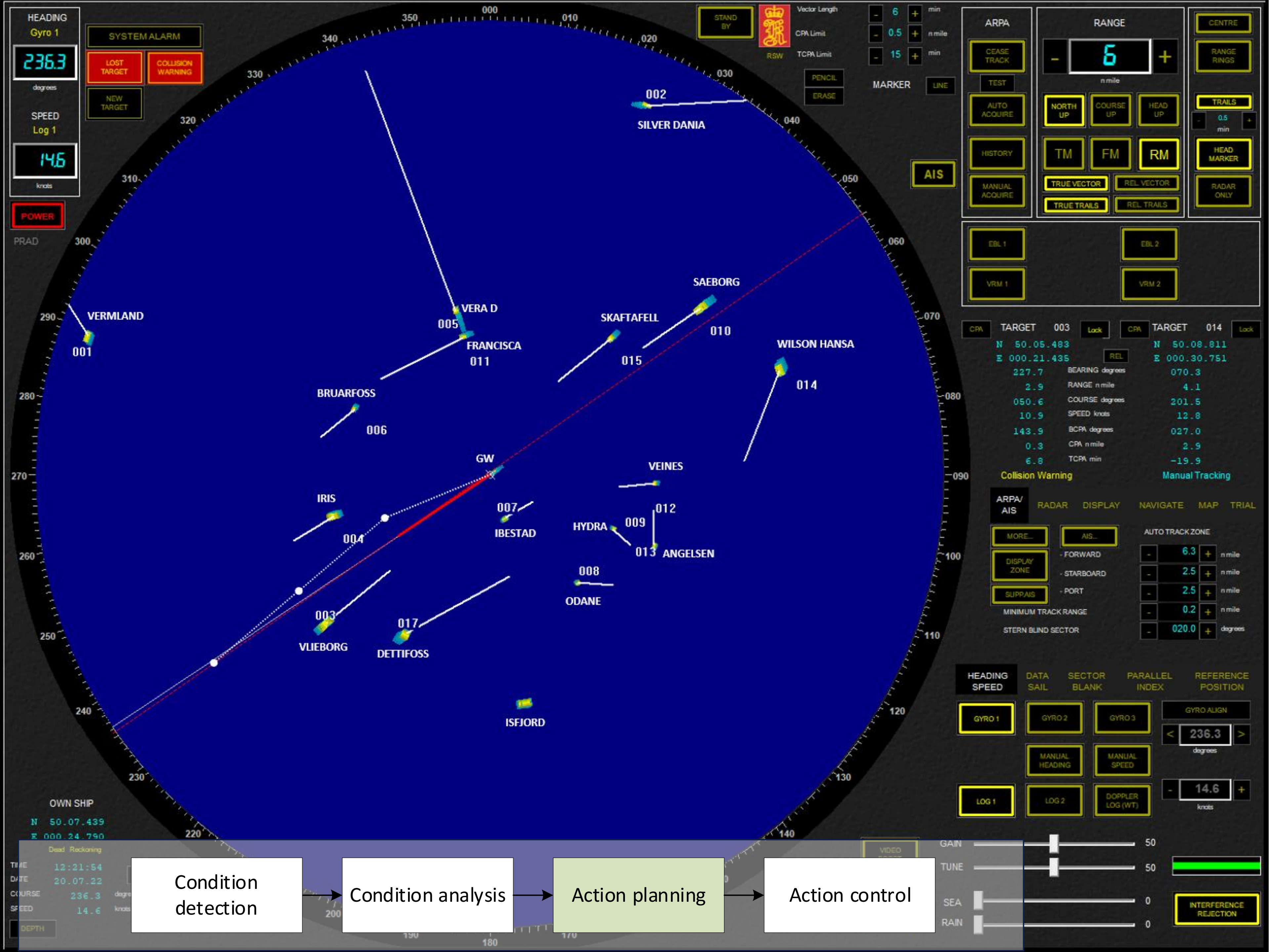
(Van de Merwe et al, under review)

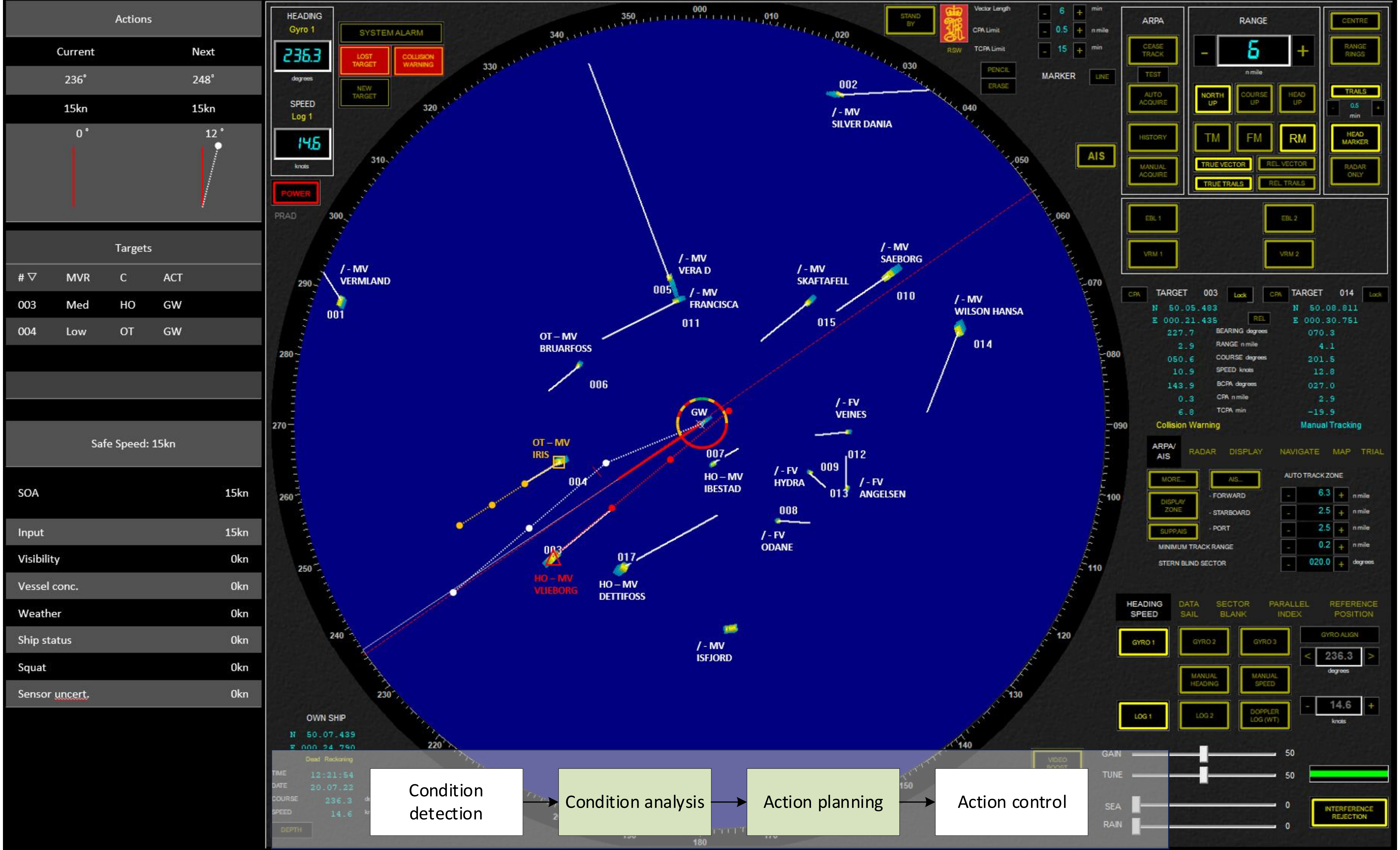

HEADING
Gyro 1
236.3
SPEED
Log 1
14.6
SYSTEM ALARM
LOST TARGET
COLLISION WARNING
NEW TARGET
POWER
ARPA
RANGE
6
NORTH UP
COURSE UP
HEAD UP
TM
FM
RM
TRUE VECTOR
REL VECTOR
TRUE TRAILS
REL TRAILS
AIS
SILVER DANIA
SAEBORG
SKAFTAFELL
WILSON HANSA
VERMLAND
VERA D
FRANCISCA
BRUARFOSS
VEINES
IRIS
IBESTAD
HYDRA
ANGELSEN
ODANE
VLIEBORG
DETTIFOSS
ISFJORD
Collision Warning
Manual Tracking
OWN SHIP
Condition detection
Condition analysis
Action planning
Action control

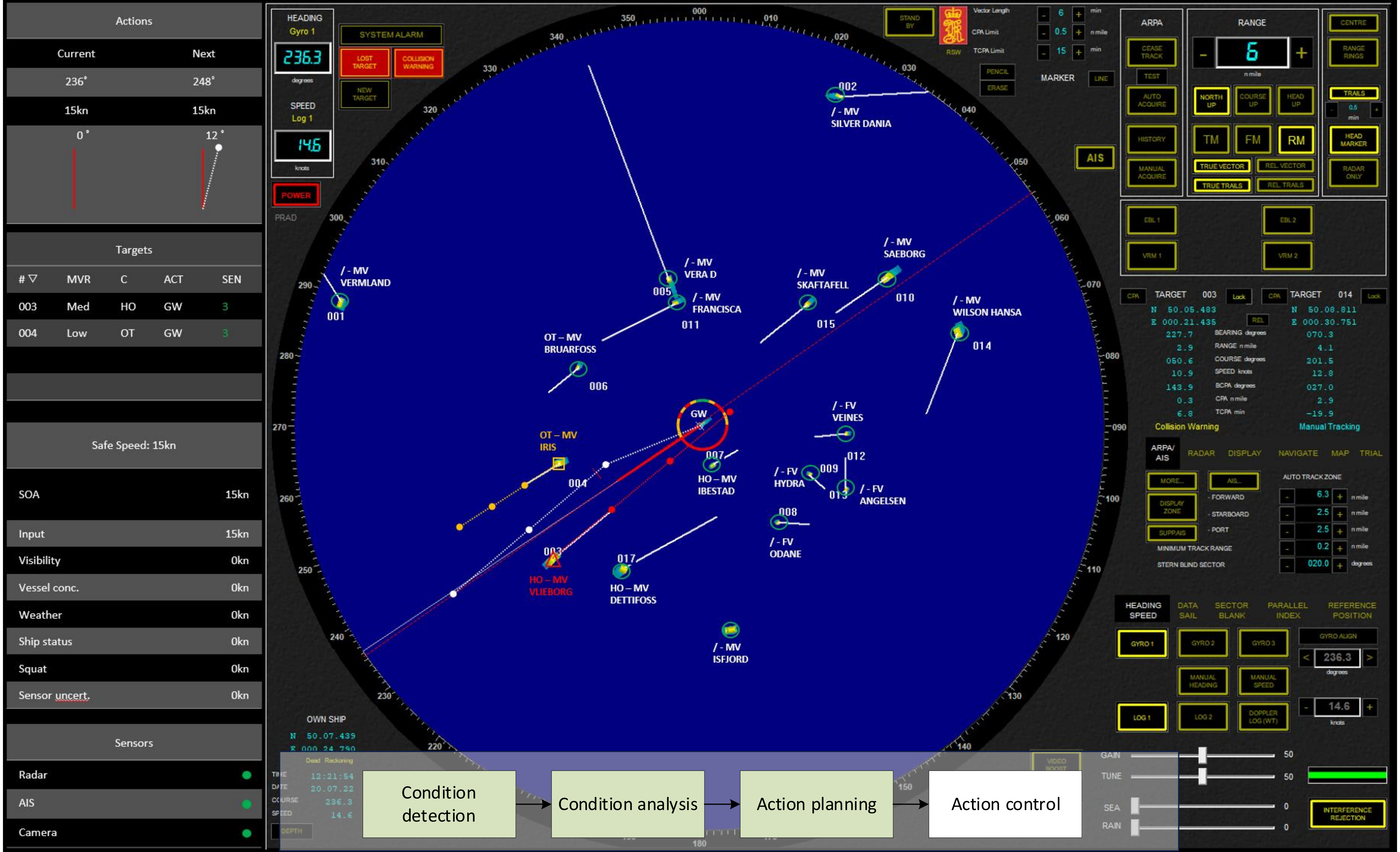
Actions

Current	Next
236˚	248˚
15kn	15kn
0˚	12˚

Safe Speed: 15kn




Actions
Current
Next
236°
248°
15kn
15kn
0°
12°
Targets
▽
MVR
C
ACT
003
Med
HO
GW
004
Low
OT
GW
Safe Speed: 15kn
SOA
15kn
Input
15kn
Visibility
0kn
Vessel conc.
0kn
Weather
0kn
Ship status
0kn
Squat
0kn
Sensor uncert.
0kn
HEADING
Gyro 1
236.3
SPEED
Log 1
14.6
SYSTEM ALARM
LOST TARGET
COLLISION WARNING
NEW TARGET
POWER
ARPA
RANGE
6
AIS
Collision Warning
Manual Tracking
OWN SHIP
GW
OT – MV IRIS
HO – MV VLIEBORG
HO – MV DETTIFOSS
HO – MV IBESTAD
OT – MV BRUARFOSS
Condition detection
Condition analysis
Action planning
Action control



Actions
Current
Next
236˚
248˚
15kn
15kn
0 ˚
12 ˚
Targets
▽
MVR
C
ACT
SEN
003 Med HO GW 3
004 Low OT GW 3
Safe Speed: 15kn
SOA 15kn
Input 15kn
Visibility 0kn
Vessel conc. 0kn
Weather 0kn
Ship status 0kn
Squat 0kn
Sensor uncert. 0kn
Sensors
Radar
AIS
Camera
HEADING
Gyro 1
236.3
SPEED
Log 1
14.6
SYSTEM ALARM
LOST TARGET
COLLISION WARNING
NEW TARGET
POWER
ARPA
RANGE
6
AIS
/ - MV VERMLAND
/ - MV VERA D
/ - MV FRANCISCA
/ - MV SILVER DANIA
/ - MV SKAFTAFELL
/ - MV SAEBORG
/ - MV WILSON HANSA
OT – MV BRUARFOSS
OT – MV IRIS
HO – MV IBESTAD
/ - FV VEINES
/ - FV HYDRA
/ - FV ANGELSEN
/ - FV ODANE
HO – MV VLIEBORG
HO – MV DETTIFOSS
/ - MV ISFJORD
GW
Collision Warning
Manual Tracking
OWN SHIP
Condition detection
Condition analysis
Action planning
Action control

Transparency: does it work?

Agent Transparency, Situation Awareness, Mental Workload, and Operator Performance: A Systematic Literature Review

Koen van de Merwe, DNV, Høvik, Norway, University of South-Eastern Norway, Borre, Norway, **Steven Mallam** and **Salman Nazir**, University of South-Eastern Norway, Borre, Norway

Objective: In this review, we investigate the relationship between agent transparency, Situation Awareness, mental workload, and operator performance for safety critical domains.

Background: The advancement of highly sophisticated automation across safety critical domains poses a challenge for effective human oversight. Automation transparency is a design principle that could support humans by making the automation's inner workings observable (i.e., "seeing-into"). However, experimental support for this has not been systematically documented to date.

Method: Based on the PRISMA method, a broad and systematic search of the literature was performed focusing on identifying empirical research investigating the effect of transparency on central Human Factors variables.

Results: Our final sample consisted of 17 experimental studies that investigated transparency in a controlled setting. The studies typically employed three human-automation interaction types: responding to agent-generated proposals, supervisory control of agents, and monitoring only. There is an overall trend in the data pointing towards a beneficial effect of transparency. However, the data reveals variations in Situation Awareness, mental workload, and operator performance for specific tasks, agent-types, and level of integration of transparency information in primary task displays.

Conclusion: Our data suggests a promising effect of automation transparency on Situation Awareness and operator performance, without the cost of added mental workload, for instances where humans respond to agent-generated proposals and where humans have a supervisory role.

Application: Strategies to improve human performance when interacting with intelligent agents should focus on allowing humans to see into its information processing stages, considering the integration of information in existing Human Machine Interface solutions.

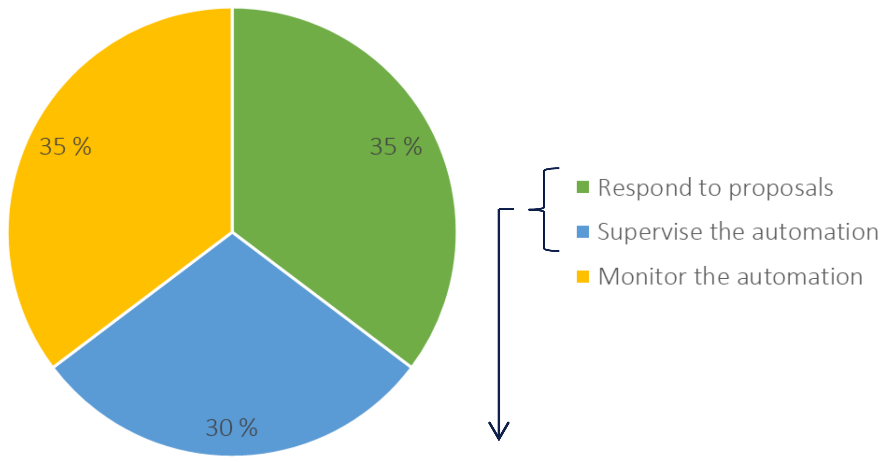
Keywords: PRISMA, human-automation interaction, automation transparency, information disclosure, seeing into

Address correspondence to Koen van de Merwe, Group Research and Development, DNV, Veritasveien 1, Høvik, Oslo 1363, Norway; e-mail: koen.van.de.merwe@dnv.com

HUMAN FACTORS
Vol. 0, No. 0, ■■ ■, pp. 1-29
DOI:10.1177/00187208221077804
Article reuse guidelines: sagepub.com/journals-permissions

INTRODUCTION

The human factors community has long had an interest in understanding the interactions between humans and automation, that is, the tasks executed by a machine agent of a function previously performed by a human (Parasuraman & Riley, 1997; Rasmussen, 1983). Central topics of research include understanding the benefits and concerns of replacing humans with automation (e.g., Bainbridge, 1983; Strauch, 2018), the need for appropriate design of automation (Norman, 1990), the effect of automation failures on human take-over responses (Endsley & Kiris, 1995), factors pertaining to automation use, disuse, and misuse (Parasuraman & Riley, 1997), human performance in taking over from automation (Eriksson & Stanton, 2017; Hergeth et al., 2017; Weaver & DeLucia, 2020), and the consequences of levels of automation on Situation Awareness (SA), mental workload, and operator performance (Endsley & Kaber, 1999; Jamieson & Skraaning, 2020; Onnasch et al., 2014). Combined, these studies culminate to the notion of an automation conundrum (Endsley, 2017), which is the problem that the more reliable and robust automation becomes, the less likely it is that a human supervisor will notice critical information and will be able to effectively intervene when required. This problem may be exacerbated with advanced automation or intelligent agents able to function independently, but still require human supervision. Considering the rapidly developing and ubiquitous presence of technology in our society, there is an urgent and continuous need of research into understanding and enhancing interactions between humans and automation such that collaboration and performance can be supported (Hancock et al., 2013; O'Neill et al., 2020; Warden et al., 2019).



Collision avoidance systems?

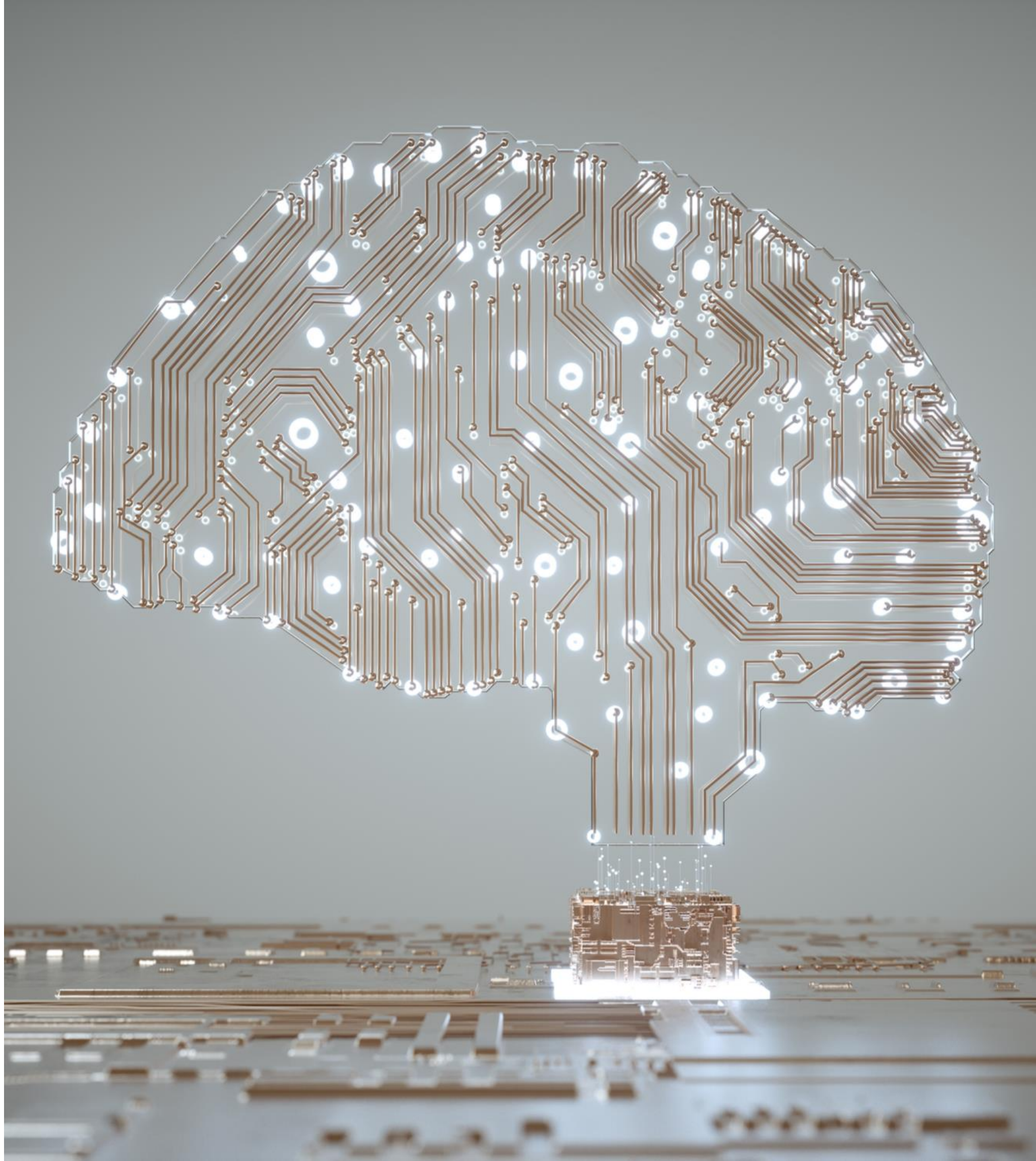
Apparent relation between task type and HF variables

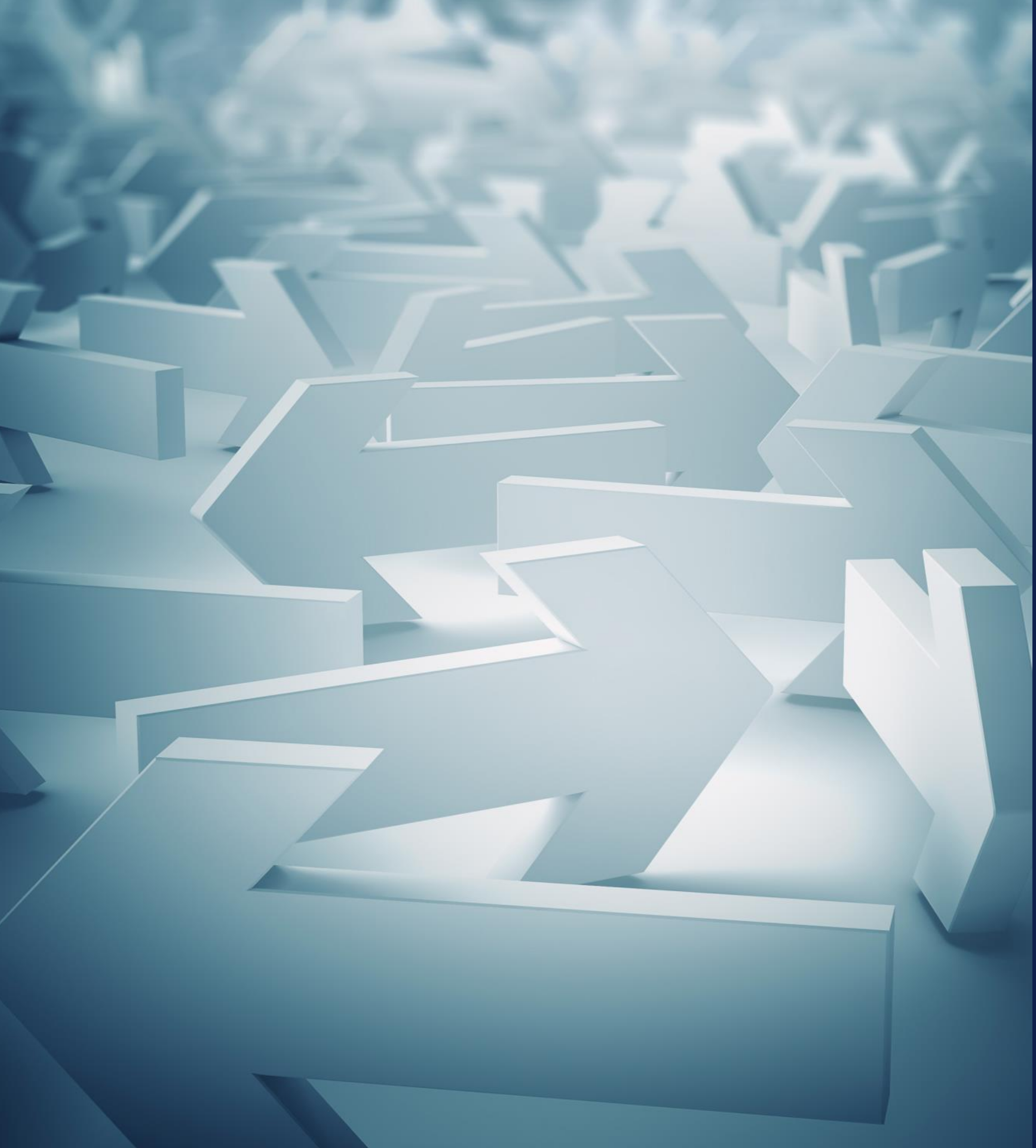
"…results…point towards a promising effect of automation transparency on operator performance, without the cost of added mental workload, for instances where humans respond to agent-generated proposals and where humans have a supervisory role." (Van de Merwe et al. 2022a)

Limits to transparency

- Risk of information to get lost in the noise
- Transparency information not integrated
- Transparency has limited effect if there is no means of control

(Endsley, 2023)





Challenges

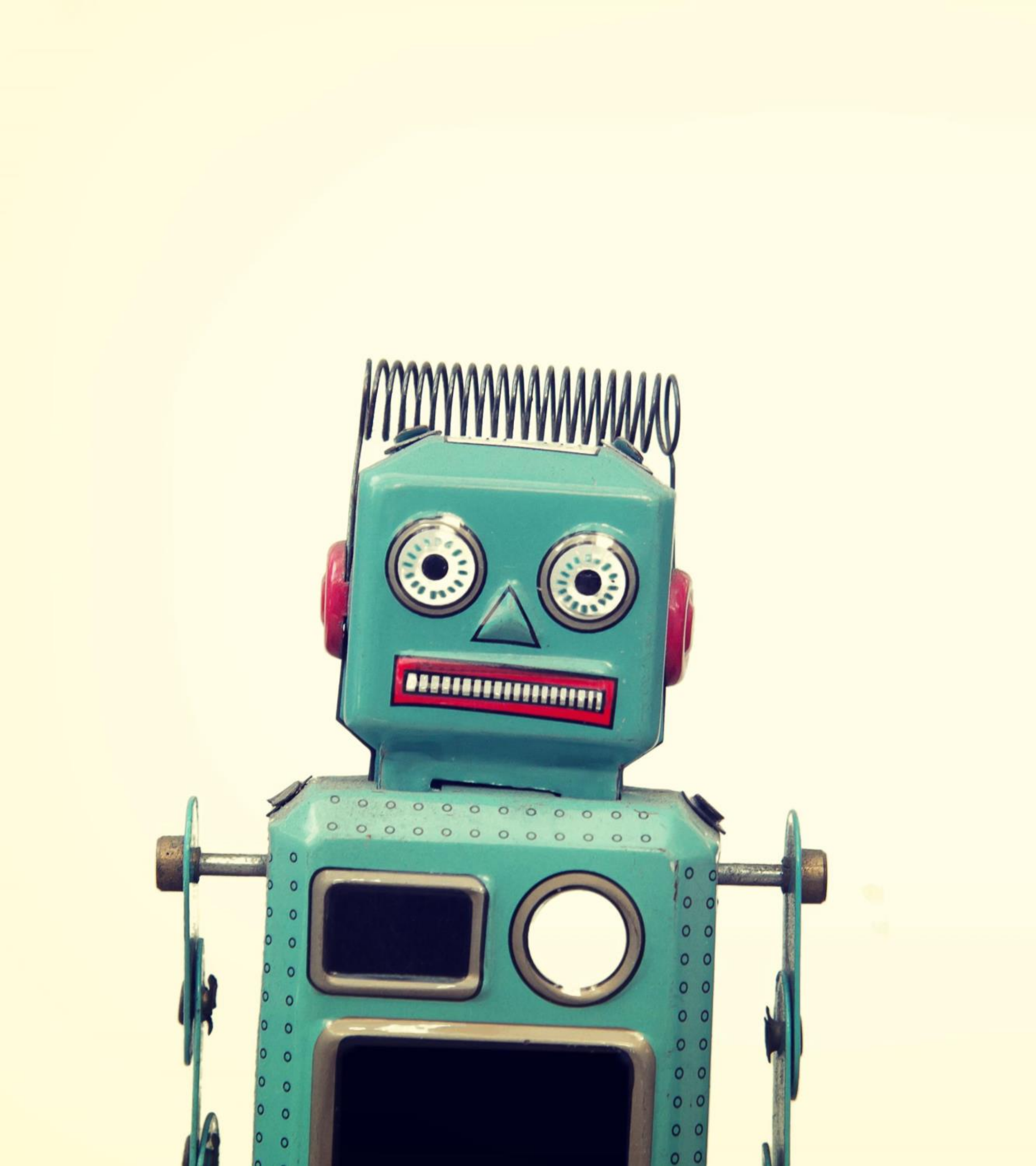
- When does a human supervisory task become an impossible task?
  - Detecting if system is outside its ODD?
  - Or should the system inform the supervisor?
  - What is the role of transparency herein?
- How to know who is the best performer?
  - Acceptance vs intervention
  - What is the role of transparency herein?
- “There is no I in team”
  - Agent transparency & human transparency?

Wrapping up

- Transparency as a means to support supervisory performance through agent
  - Observability
  - Predictability

- Can be applied to agents acting as
  - Decision-support tools
  - Teammates

- DNV class rules and notations
  - Process and product assurance
  - Continuous assurance (learning systems)
  - HF integration, e.g., transparency



Teaming with automation in future maritime systems

HFC forum spring 2023

Koen.van.de.merwe@dnv.com

www.dnv.com

Key references

- DNV. (2018). *DNVGL-CG-0264: Autonomous and remotely operated ships*. http://rules.dnvgl.com/docs/pdf/dnvgl/cg/2018-09/dnvgl-cg-0264.pdf
- Endsley, M. R. (2017). From Here to Autonomy: Lessons Learned from Human-Automation Research. *Human Factors*, *59*(1), 5–27. https://doi.org/10.1177/0018720816681350
- Endsley, M. R. (2023). Supporting Human-AI Teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior*, *140*, 107574. https://doi.org/10.1016/j.chb.2022.107574
- National Academies of Sciences, Engineering and Medicine. (2022). Human-AI Teaming: State of the Art and Research Needs. The National Academies Press. https://doi.org/10.17226/26355
- Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent Agent Transparency in Human–Agent Teaming for Multi-UxV Management. Human Factors, 58(3), 401–415. https://doi.org/10.1177/0018720815621206
- MITRE. (2018). *Human-Machine Teaming Systems Engineering Guide* (No. MP180941; p. 68). MITRE Corporation. https://www.mitre.org/publications/technical-papers/human-machine-teaming-systems-engineering-guide
- Oswald, F. L., Endsley, M. R., Chen, J., Chiou, E. K., Draper, M. H., McNeese, N. J., & Roth, E. M. (2022). The National Academies Board on Human-Systems Integration (BOHSI) Panel: Human-AI Teaming: Research Frontiers. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 66(1), 130–134. https://doi.org/10.1177/1071181322661007
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans.*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354
- van de Merwe, K., Mallam, S., & Nazir, S. (2022). Agent Transparency, Situation Awareness, Mental Workload, and Operator Performance: A Systematic Literature Review. *Human Factors*, 00187208221077804. https://doi.org/10.1177/00187208221077804
- van de Merwe, K., Mallam, S. C., Engelhardtsen, Ø., & Nazir, S. (2022). Exploring navigator roles and tasks in transitioning towards supervisory control of autonomous collision avoidance systems. *Journal of Physics: Conference Series*, *2311*(1), 012017. https://doi.org/10.1088/1742-6596/2311/1/012017
- van de Merwe, K., Mallam, S., Engelhardtsen, Ø., & Nazir, S. (Accepted for publication). Operationalising Automation Transparency for Maritime Collision Avoidance. TransNav: the International Journal on Marine Navigation and Safety of Sea Transportation, Gdynia, Poland.
- van de Merwe, K., Mallam, S., Engelhardtsen, Ø., & Nazir, S. (Under review). Supporting human supervisory performance through information disclosure: Establishing transparency requirements for maritime collision avoidance. *Proceedings of the Human Factors Society Annual Meeting*.
- van de Merwe, K., Mallam, S., Engelhardtsen, Ø., & Nazir, S. (Under review). *Supporting human supervision in autonomous collision avoidance through system transparency: A structured and systematic approach*. *Journal of Navigation*

DNV