TRANSNAV
http://www.transnav.eu
the International Journal on Marine Navigation and Safety of Sea Transportation
Volume 17 Number 2 June 2023
DOI: 10.12716/1001.17.02.09

Operationalising Automation Transparency for Maritime Collision Avoidance

K. van de Merwe[1,2], S. Mallam[2,3], Ø. Engelhardtsen[1] & S. Nazir[2]
[1] *DNV, Høvik, Norway*
[2] *University of South-Eastern Norway, Borre, Norway*
[3] *Memorial University of Newfoundland, St. John's, Canada*

ABSTRACT: Automation transparency is a means to provide understandability and predictability of autonomous systems by disclosing what the system is currently doing, why it is doing it, and what it will do next. To support human supervision of autonomous collision avoidance systems, insight into the system's internal reasoning is an important prerequisite. However, there is limited knowledge regarding transparency in this domain and its relationship to human supervisory performance. Therefore, this paper aims to investigate how an information processing model and a cognitive task analysis could be used to drive the development of transparency concepts. Also, realistic traffic situations, reflecting the variation in collision type and context that can occur in real-life, were developed to empirically evaluate these concepts. Together, these activities provide the groundwork for exploring the relation between transparency and human performance variables in the autonomous maritime context.

1 INTRODUCTION

1.1 *Human supervision in autonomous collision avoidance*

The last decade has shown an increasing interest in research and development efforts towards use of autonomy in the maritime industry. The purpose of increased automation is diverse, but improvements in cost, efficiency and safety for sharp-end personnel are major drivers [1]–[3]. Yara Birkeland, and the ASKO barges are examples of the ambition of the industry when it comes to the application of highly automated functions to support and/or substitute onboard personnel [4], [5]. In this development, remote-control centres are foreseen to play a role from where operators can perform oversight of autonomous ships and can make critical decisions with regards to the operations of the ship [6].

The purpose of remote-control centres is to provide shore-side support for autonomous ships, to be compliant with current regulations on minimum safe manning, and to provide an equivalent level of safety (or better) compared to conventional ship operations [7], [8]. The idea is that from a remote-control position operators can supervise the ship's operations and monitor, assist, and take over from the autonomous systems when the circumstances require this. In this case, it is assumed that humans can perceive and understand the information concerning the ship under supervision such that adequate situation awareness can be attained and maintained.

A key challenge to be resolved in moving towards autonomous, and potentially unmanned, shipping is how unforeseen circumstances, such as collision and grounding situations, are handled without the presence of navigators onboard the ship [9]–[12]. At present, navigators determine collision risk and perform relevant avoidance manoeuvres supported by a range of systems, e.g., radar, AIS, and ECDIS. Also, collision and grounding avoidance requires knowledge, skills, and experience to be performed in accordance with the collision regulations. When this task is performed by an autonomous Artificial Intelligence-powered collision avoidance system, adequate and sufficient contextual information is essential to support human oversight (see Figure 1) [13].

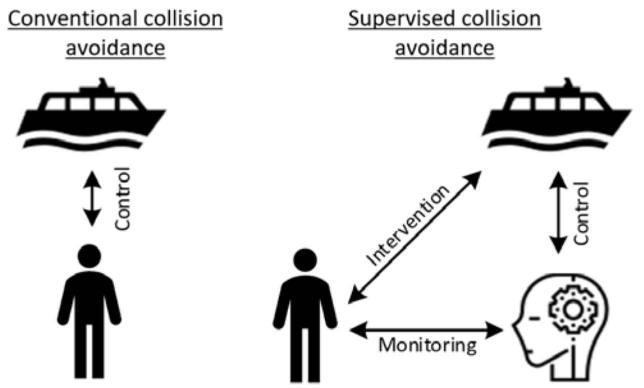

Figure 1. Conceptualization of control in conventional- and supervised collision avoidance.

An earlier study led by the first author identified the information required to supervise the performance of an autonomous collision avoidance system through a mapping and assessment of relevant cognitive tasks [12], [14]. This study concluded that adequately supervising an autonomous collision and grounding avoidance system requires insight into the system's information processing to understand its decisions and actions. Based on the knowledge that human supervision of automated functions has challenges in terms of human performance, keeping humans in the loop, or rather "on the loop", becomes an essential design requirement [15], [16]. Thus, providing sufficient information about the automated system's reasoning process has been proposed as one of the elements that could support humans in such a role. In other words, by disclosing the system's internal decision-making process to its supervisor, the system is made transparent with regards to its intent, performance, future plans, and reasoning process [17].

Automation transparency is concerned with making the inner reasoning of systems observable, such that its actions are understandable and predictable [15], [18], [19]. Therefore, transparency should make it clear to human supervisors what the system is currently doing, why it is doing it, and what it will do next [15]. Earlier reviews have indicated that transparency has a promising effect on human performance and situation awareness [20]–[22]. However, there is limited knowledge regarding transparency in the maritime domain, especially in relation to autonomous collision and grounding avoidance. To this end, further work is needed to investigate the role of transparency in supervised autonomous shipping and to explore its relationship with human performance in this context.

This paper discusses ongoing work towards performing an empirical evaluation to study differing levels and types of transparency concepts in a realistic traffic collision avoidance setting. An empirical evaluation is planned in which participants take the role of a supervisor of an autonomous collision avoidance system. An approach is used in which participants are tasked with evaluating traffic situations for their understandability, whilst being measured on human performance variables. The purpose of this evaluation is to better understand which levels and types of transparency information support human supervisors and how this knowledge can be applied to a dynamic collision avoidance context. This paper describes the groundwork for this evaluation by describing the systematic development process behind the traffic situations, as well as the levels and types of transparency concepts developed for this.

2 DEVELOPING TRAFFIC SITUATIONS

2.1 *Defining criteria to ensure variation*

To provide participants of the planned empirical evaluation with realistic conflicts, traffic situations were developed that reflected the variation in collision type and context that may occur in real-life. Also, to avoid familiarisation with the traffic situations, and thereby unintentionally influencing the results of the evaluation, multiple variants of traffic situations were developed based on a set of criteria (see Table 1).

Table 1. Criteria for establishing a varied set of traffic situations.

Criterion	Variation
Complexity avoidance manoeuvre own ship	Low - No limitations High - Limitations manoeuvre
Collision type	CR - Crossing HO - Head-on OT - Overtaking/overtaken NC - No collision
Avoidance actions own ship	Give-way Stand-on
Restrictions target	No restrictions Restricted in manoeuvrability
Traffic density	Few other ship and objects Many other ships and objects
Geography	Land Open water

Variability was ensured through differing levels of complexity, collision types, the avoidance actions of own ship, restrictions to target ships, traffic density, and geography. First, in high complex situations, own ship was restricted in its avoidance manoeuvring ability compared to low complex situations. That is, in low complexity situations, own ship was free to manoeuvre in any direction to avoid a collision, whereas in high complexity situations, there were obstacles prohibiting own ship to perform certain manoeuvres. Second, for collision type, traffic situations consisted of crossing-, head-on-, overtaking-/overtaken situations. Also, situations were developed in which no collision was present. Third, for avoidance actions, situations were developed for which own ship was the give-way vessel or the stand-on vessel. Fourth, for some situations, target ships were restricted in their manoeuvrability, e.g., because of ongoing bunkering.

Fifth, situations were developed with low- and high traffic densities. Finally, traffic situations were developed in which contextual factors were varied that were external to the traffic situations (i.e., land formations or open water).

To constrain the amount of variation and retain controllability in the traffic situations, some limitations were set in terms of number of ships posing a collision risk and the number of simultaneous collision situations. That is, own ship could only be in direct conflict with one other ship for one collision type (e.g., not in a crossing and head-on situation simultaneously), own ship could not be in both a give-way and stand-on situation simultaneously, and own ship was never restricted in its manoeuvrability. Also, although is it recognised that grounding avoidance is an essential part of collision avoidance, the traffic situations in this paper were limited to collision situations only. Finally, external factors that could affect the collision situation or own-ship's capabilities, such as weather or technical failures, were not included.

2.2 *Development process*

For each criterion in Table 1, two scenarios were created resulting in a set of 70 situations (see Table 2). The traffic situations were created in a desktop simulator from a popular equipment manufacturer by a navy-certified navigator with five years of navigational experience. Upon creating an initial set of traffic situations, a review was performed with independent navigators.

Table 2. The traffic situations created based on the set of criteria. Key: HO = Head-on, CR = Crossing, OT = Overtaking/overtaken, NC = No collision, L = Low, H = High, T = Total. *Note: in a head-on situation with one motorised target ship and no other exceptions, own ship cannot be stand-on.

	HO		CR		OT		T
Variant/Complexity	L	H	L	H	L	H	
Type (HO/CR/OT)	5	5	4	4	4	4	26
Type (NC)	2	2	2	2	2	2	12
Own ship stand-on*	0	0	2	2	2	2	8
Restrictions target	2	2	2	2	2	2	12
Geography (land)	2	2	2	2	2	2	12
Total	11	11	12	12	12	12	70

2.3 *Verification and validation workshop*

The final verification and validation were performed with two independent navigators holding active navigational licenses (D1/D2), with an average of 6.5 years of navigational experience (SD=2.1, min=5, max=8). The review was performed in the form of a 1,5-day workshop.

In the workshop traffic situations were shown on a display and participants were asked to state if own ship was in a collision situation, if yes, which type (HO/CR/OT), and the avoidance action required by own ship (give-way/ stand-on). In addition, three questions were asked, using a 7-point Likert scale, probing the situation's realism, complexity, and likelihood of occurrence. With these questions, a comparison between the situation's intended depiction and the navigator's perception was obtained. Discrepancies were discussed and suggestions for improving the design of the traffic situations were noted. A final set of traffic situations were produced, incorporating the inputs from the workshop (see Figures 2, 3, 4, and 5 for examples).

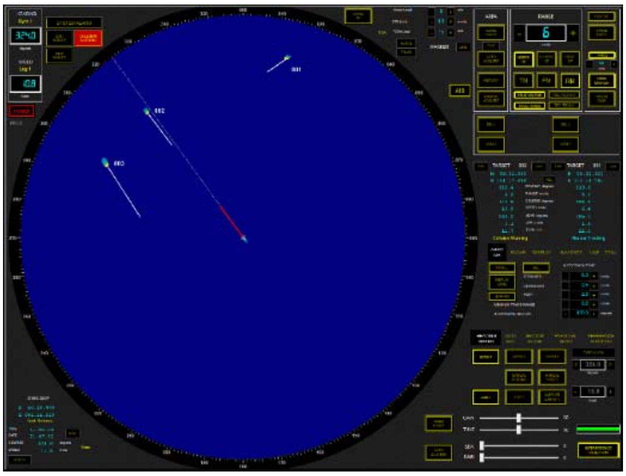

Figure 2. Own ship is in a head-on situation in open water where it is required to give-way. The situation is of low complexity as there are no restrictions to own ship's avoidance manoeuvrability.

Figure 3. Own ship is in an overtaking situation in open waters where it is required to give-way. The situation is of high complexity as there are restrictions to own ship's avoidance manoeuvrability (both port and starboard).

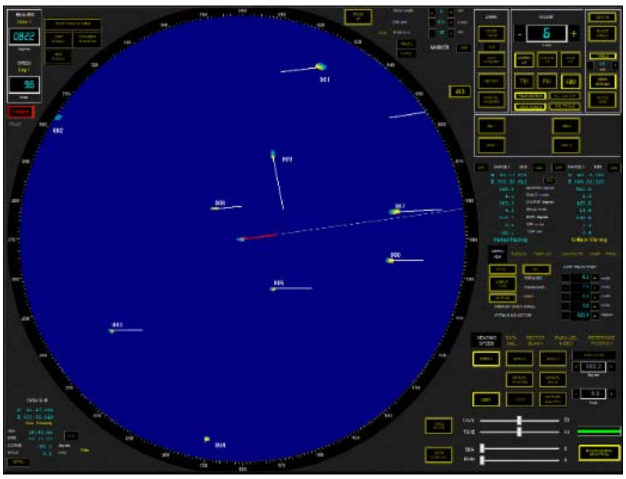

Figure 4. Own ship is in a crossing situation in open waters where it is required to stand-on. The situation is of high complexity as there are restrictions to the target ship's avoidance manoeuvrability (the ship crossing at port side).

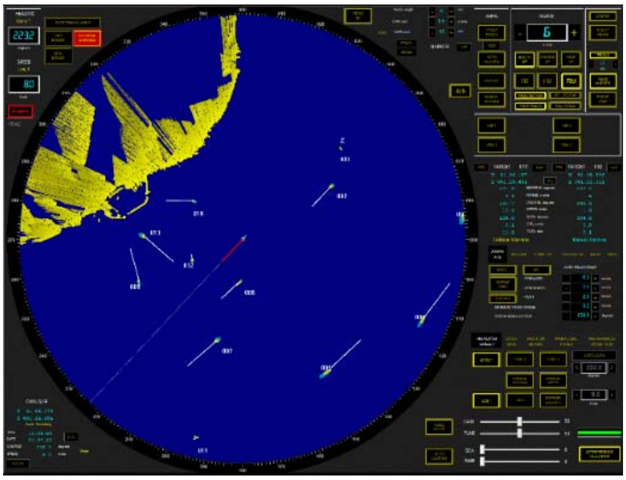

Figure 5. Own ship is in a crossing situation in open waters where it is required to give-way. The situation is of high complexity as there are restrictions to the target ship's avoidance manoeuvrability (a buoy).

3 DEVELOPING TRANSPARENCY FOR COLLISION AVOIDANCE

3.1 *Defining transparency layers*

An earlier study led by the first author performed a cognitive task analysis to identify the information required to perform supervision of a collision avoidance system [12], [14]. The analysis describes the information pertaining to the supervisory task and depicts which information should be disclosed to human supervisors to make the internal reasoning of the collision avoidance system observable. However, the analysis only describes what information should be made available and it does not dictate which type, or how much of the identified information should be disclosed. Simply depicting all information simultaneously will likely put too large a cognitive burden on the supervisor's information processing capabilities, resulting in high mental workload. At the same time, only limiting the information from the system to single information elements may not provide the full picture about the system's internal reasoning either. In addition, considering the dynamic nature of the collision avoidance task, the information needed to effectively supervise the system may vary given the circumstances and the task analysis does not define which information should be disclosed when. As such, providing transparency to supervisors means making choices as to which information is made available to allow supervisors to understand the system's behaviour.

The rationale for specifying what constitutes transparency information in a collision avoidance context, together with how this information can be categorised into distinct information types is discussed in a separate study [23]. In brief, a simple information processing model was used (see Figure 6), consisting of information acquisition, information analysis, decision selection, and action implementation stages, to identify and categorise the information into discrete steps [24]. As such, a layered approach to transparency was used allowing supervisors to observe the different facets of the system's input parameters, reasoning, decisions, and actions pertaining to the collision situation.

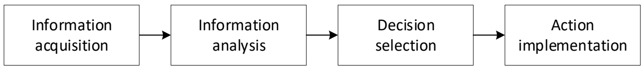

Figure 6. A simple model of human information processing adopted from [24].

This model provides, at minimum, a means to organize the information describing the system's information processing into several distinct parts. However, the model does not provide guidance as to which information takes priority over the other. A potential starting point is to try answer the question of what information supervisors would like to know at a minimum, before adding layers of transparency to allow for increased understandability. A plausible means for human supervisors to obtain an understanding of the collision avoidance system's performance is to be informed whether the system can avoid a potential collision at all. In other words, supervisors likely need to be informed about the system's decisions and actions first, before needing to "dig deeper" into the system's underlying reasoning. This indicates that the starting point for providing transparency to supervisors is thus the "decision selection" step of the information processing model depicted in Figure 6 and not the "information acquisition" step. (Note that in the "action implementation" step there is no information processing, only execution.) Further understanding of how and why the system has derived at its decision and planned actions can subsequently be obtained by "going backwards" through the model. That is, the "information analysis" stage of the model provides the relevant information pertaining to the analysis that underlie the system's decisions and actions. Finally, when the full picture is required for understanding the system's decisions and actions, the "information acquisition" stage of the model provides all the input data the system uses in its information processing.

3.2 *Development process*

A concept illustration is provided of a radar screen depicting a traffic situation in which own ship, in the centre of the radar screen, is involved in a head-on situation (see Figure 7). Own ship depicts its intended avoidance manoeuvre by drawing its planned track for the next three manoeuvring steps (each step corresponds to one vector length and equals six minutes). It also states "GW" indicating it intends to give-way. Additional information about current and next actions, including speed, are depicted on the left side of the figure. With this information, minimum transparency is provided to allow supervisors to understand that the system is about to initiate a 12-degree starboard turn and that it intends to give-way. The information provided in Figure 7 was proposed as the minimum information needed to obtain an understanding of the own ship's decisions and actions.

Figure 7. Traffic situation with transparency information overlaid (decision selection).

Figure 8 depicts that own ship considers two targets as especially relevant in this traffic situation. The target ship in red is depicted as the highest risk as this ship is the one considered to be on collision course with own ship (minimum predicted CPA exceeded). The target in orange is also highlighted as own ship has considered this target to be of importance during the avoidance manoeuvre. Further information regarding the targets that own ship considers is provided through the indicators next to the targets depicting the conflict situation (e.g., HO for head-on, and MV for motor vessel). In addition, further information regarding the system's reasoning is provided through a manoeuvrability indicator around own ship indicating where it can manoeuvre within one vector length. Finally, tables to the left of the radar screen depict additional target information and the variables own ship has considered in determining safe speed.

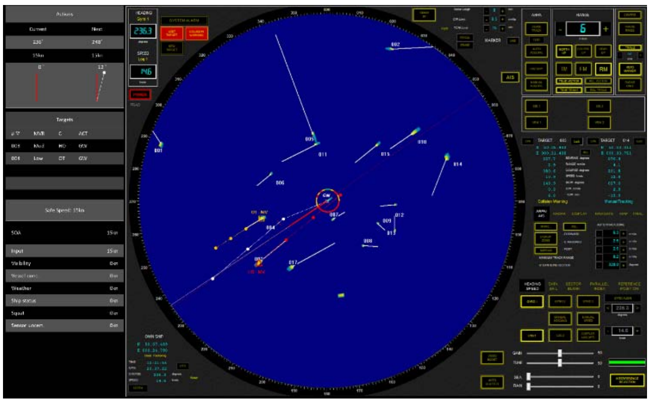

Figure 8. Traffic situation with transparency information overlaid (decision selection + information analysis).

Figure 9 provides a depiction of what a transparent collision avoidance system could look like when all transparency information described in the task analysis is provided. Here, all targets have received identifiers (green circles), and initial classifications (ship types and relevant conflict type indicators). In addition, information regarding the status of the system's sensors are provided in the tables to the left of the radar screen.

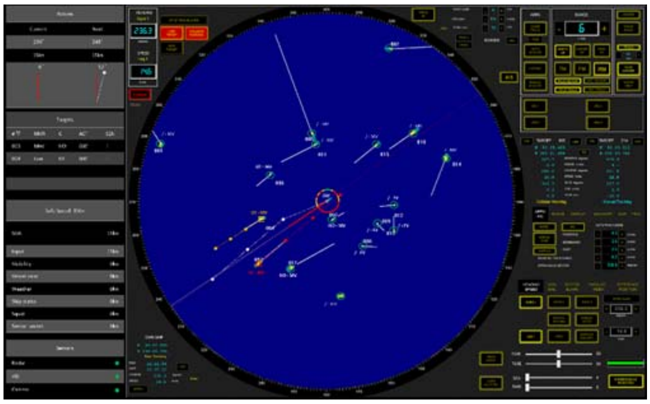

Figure 9. Traffic situation with transparency information overlaid (decision selection + information analysis + information acquisition).

3.3 *Verification and validation workshop*

The transparency concepts were developed through a series of iterations based on the information from the task analysis and the information processing model. Final verification and validation of the interfaces was performed in a second workshop with two independent navigators holding active navigational licenses with an average of 12 years of navigational experience (SD=9.9, min=5, max=19).

The purpose of this second workshop was to evaluate a selected set of traffic situations that included the transparency layers as described above. A representative subset of five traffic situations were included for review in this workshop, including head-on, crossing with own ship as stand-on, overtaken by a ship restricted in its manoeuvrability, crossing with speed-only as the avoidance manoeuvre and overtaking a slower ship when approaching a harbour. A talk aloud protocol was used where participants were asked to describe their interpretation of the traffic situation with primary focus on the information the system provided through the Human Machine Interface (HMI). In other words, the focus in the workshop was on how they perceived the collision avoidance system would solve the conflict situation, and not how they would solve it. The independent navigator's interpretations were noted, including all comments related to recommendations for improvement, corrections, and additions which were included in the final transparency iteration.

4 SUMMARY AND FURTHER WORK

When a collision situation occurs that requires human intervention, the collision avoidance system needs to facilitate human supervisors in gaining SA such that successful decisions can be made. This paper described the systematic development of a realistic and validated foundation for evaluating the relationship between automation transparency and human supervisory performance in an autonomous collision avoidance context. First, a set of traffic situations were developed based on navigational experience aimed at capturing the variability encountered in real-life situations. Second, a set of transparency concepts were developed based on a cognitive task analysis and a model for human information processing. Together, these preparations provide the groundwork for the planned empirical work to explore this relationship.

As the maritime industry moves towards increased use of automation, including deploying systems that can perform (part of) the collision and grounding avoidance functions, there is an urgent need to understand how humans will interact with these systems. Automation transparency has been proposed as a critical element that can support human supervisors in obtaining situation awareness of the system's behaviours and actions [16]. Conversely, without transparency, i.e., systems that have low degrees of observability and predictability, humans will be highly challenged in understanding what the system is doing, why it is doing it, and what it will do next. As such, given the critical nature of the supervisory task for autonomous maritime collision and grounding avoidance systems, it is pertinent that further understanding is needed with regards to the application of the transparency in this domain.

This paper aimed to address this need by investigating how an information processing model could be used to drive the development of transparency layers. Given the dynamic nature of collision and grounding avoidance the amount and type of information needed to understand the system may depend on the type of situation, the degree of human oversight, the complexity of the situation, or the time available to intervene. The transparency concepts discussed in this paper have attempted to address this. In addition, an empirical evaluation is underway in which the relationship between automation transparency and human performance variables are evaluated in a collision avoidance context. This way, the relation between transparency and human performance variables can be explored, and its practical benefits can be assessed.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to the navigators for their participation in the workshops. Also, we would like to express our sincere gratitude to Koen Houweling for his contribution in developing the traffic situations and the transparency illustrations.

REFERENCES

[1] L. Kretschmann, H. C. Burmeister, and C. Jahn, "Analyzing the economic benefit of unmanned autonomous ships: An exploratory cost-comparison between an autonomous and a conventional bulk carrier," Research in Transportation Business and Management, vol. 25, pp. 76–86, 2017, doi: 10.1016/j.rtbm.2017.06.002.

[2] I. Kurt and M. Aymelek, "Operational and economic advantages of autonomous ships and their perceived impacts on port operations," Marit Econ Logist, vol. 24, no. 2, pp. 302–326, Jun. 2022, doi: 10.1057/s41278-022-00213-1.

[3] K. Wróbel, J. Montewka, and P. Kujala, "Towards the assessment of potential impact of unmanned vessels on maritime transportation safety," Reliab Eng Syst Saf, vol. 165, pp. 155–169, 2017, doi: 10.1016/j.ress.2017.03.029.

[4] Kongsberg, "Kongsberg maritime and Massterly to equip and operate two zero-emission autonomous vessels for ASKO," Sep. 01, 2020. https://www.kongsberg.com/maritime/about-us/news-and-media/news-archive/2020/zero-emission-autonomous-vessels/ (accessed Nov. 18, 2020).

[5] Yara International, "Yara Birkeland," 2021. https://www.yara.com/news-and-media/media-library/press-kits/yara-birkeland-press-kit/ (accessed Jan. 02, 2023).

[6] DNV, "DNVGL-CG-0264: Autonomous and remotely operated ships." 2018. [Online]. Available: http://rules.dnvgl.com/docs/pdf/dnvgl/cg/2018-09/dnvgl-cg-0264.pdf

[7] IMO, Resolution A.1047(27) Principles of safe manning. 2011.

[8] IMO, "Guidelines for the approval of alternatives and equivalents as provided for in various IMO instruments," MSC.1/Circ.1455, Jun. 2013.

[9] K. Aylward, R. Weber, M. Lundh, S. N. MacKinnon, and J. Dahlman, "Navigators' views of a collision avoidance decision support system for maritime navigation," J. Navigation, pp. 1–14, Sep. 2022, doi: 10.1017/S0373463322000510.

[10] E. Hannaford, P. Maes, and E. Van Hassel, "Autonomous ships and the collision avoidance regulations: a licensed deck officer survey," WMU J Marit Affairs, vol. 21, no. 2, pp. 233–266, Jun. 2022, doi: 10.1007/s13437-022-00269-z.

[11] M. A. Ramos, I. B. Utne, and A. Mosleh, "Collision avoidance on maritime autonomous surface ships: Operators' tasks and human failure events," Saf Sci, vol. 116, pp. 33–44, 2019, doi: 10.1016/j.ssci.2019.02.038.

[12] K. van de Merwe, S. C. Mallam, Ø. Engelhardtsen, and S. Nazir, "Exploring navigator roles and tasks in transitioning towards supervisory control of autonomous collision avoidance systems," J. Phys.: Conf. Ser., vol. 2311, no. 1, p. 012017, Jul. 2022, doi: https://doi.org/10.1088/1742-6596/2311/1/012017.

[13] S. N. Mackinnon, Y. Man, M. Lundh, and T. Porathe, "Command and control of unmanned vessels: Keeping shore based operators in-the-loop," 18th International Conference on Ships and Shipping Research, NAV 2015, pp. 612–619, 2015.

[14] K. van de Merwe, S. Mallam, Ø. Engelhardtsen, and S. Nazir, "Supporting human supervision in autonomous collision avoidance through system transparency: a structured and systematic approach," under review.

[15] M. R. Endsley, "From Here to Autonomy: Lessons Learned from Human-Automation Research," Hum Factors, vol. 59, no. 1, pp. 5–27, 2017, doi: 10.1177/0018720816681350.

[16] M. R. Endsley, "Supporting Human-AI Teams:Transparency, explainability, and situation awareness," Computers in Human Behavior, vol. 140, p. 107574, Mar. 2023, doi: 10.1016/j.chb.2022.107574.

[17] J. Y. C. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia, and M. J. Barnes, "Situation Awareness-Based Agent Transparency," U.S. Army Research Laboratory, Aberdeen Proving Ground, ARL-TR-6905, Apr. 2014. doi: 10.21236/ADA600351.

[18] J. Y. C. Chen, S. G. Lakhmani, K. Stowers, A. R. Selkowitz, J. L. Wright, and M. J. Barnes, "Situation awareness-based agent transparency and human-autonomy teaming effectiveness," Theor Issues Ergon Sci, vol. 19, no. 3, pp. 259–282, May 2018, doi: 10.1080/1463922X.2017.1315750.

[19] M. R. Endsley, B. Bolté, and D. G. Jones, Designing for situation awareness: an approach to user-centered design. London ; New York: Taylor & Francis, 2003.

[20] A. Bhaskara, M. Skinner, and S. Loft, "Agent Transparency: A Review of Current Theory and

Evidence," IEEE Trans Hum Mach Syst, vol. 50, no. 3, pp. 215–224, Jun. 2020, doi: 10.1109/THMS.2020.2965529.
[21] F. Rajabiyazdi and G. A. Jamieson, "A Review of Transparency (seeing-into) Models," in 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Oct. 2020, pp. 302–308. doi: 10.1109/SMC42975.2020.9282970.
[22] K. van de Merwe, S. Mallam, and S. Nazir, "Agent Transparency, Situation Awareness, Mental Workload, and Operator Performance: A Systematic Literature Review," Hum Factors, p. 00187208221077804, Mar. 2022, doi: 10.1177/00187208221077804.
[23] K. van de Merwe, S. Mallam, Ø. Engelhardtsen, and S. Nazir, "Supporting human supervisory performance through information disclosure: establishing transparency requirements for maritime collision avoidance," in Proceedings of the Human Factors Society Annual Meeting, Orlando, FL, submitted.
[24] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, "A model for types and levels of human interaction with automation," IEEE Trans Syst Man Cybern, vol. 30, no. 3, pp. 286–297, May 2000, doi: 10.1109/3468.844354.