Supervising Autonomous Ships: How Display Modality and Encounter Geometry Shape Trust and Situation Awareness

Zhichen Lu[1,2,3], Zhegong Shangguan[4], Xiaoxuan Hei[1], Benoit Clement[1,3]
Matthew Stephenson[2,3], Adriana Tapus[1]

***Abstract*— Maritime Autonomous Surface Ships (MASS) are shifting navigation from onboard operation to shore-based remote supervision. In this paradigm, operators must evaluate autonomous collision-avoidance behavior through interface-mediated information, but it remains unclear whether onboard display configurations can be directly transferred to supervisory systems. We conducted a 2×3 within-subjects simulator experiment ($N = 26$) to investigate how crossing angle (20° vs. 70°) and display modality (ECDIS-only, bridge-view-only, and combined display) influence supervisory trust and intervention behavior. Objective collision-risk distributions were matched across conditions. Results showed that 70° encounters reduced trust and increased takeover intention. Display modality did not consistently affect trust ratings, but it altered behavioral safety thresholds. The combined display increased trust under 20° encounters but reduced trust under 70° encounters. These results indicate that supervisory trust is shaped by encounter geometry, and that multimodal MASS interfaces should prioritize geometric interpretability over information density alone.**

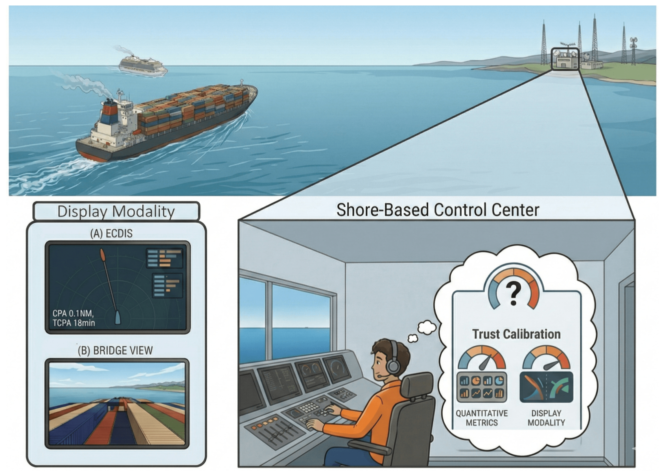


Fig. 1. Operational overview of the remote monitoring system. (Top) Ship-to-ship encounter scenarios; (Bottom-Left) Information display modalities: (a) Electronic Chart Display and Information System (ECDIS); (b) First-person bridge view; (Bottom-Right) Shore-Based Operator performing supervisory control tasks at a Remote Control Center (RCC).

I. INTRODUCTION

In recent years, as the technological maturity of Maritime Autonomous Surface Ships (MASS) has significantly improved, shore-based remote supervision has emerged as a viable alternative to traditional onboard staffing [1][2]. Under this paradigm, operators monitor autonomous ship behaviors via system interfaces and intervene when necessary [3]. Unlike traditional crews, however, shore-based operators may lack extensive practical sailing experience [4][5], meaning their understanding of encounter scenarios relies entirely on the information provided by the interface rather than accumulated navigational intuition [6]. Consequently, the ability of these interfaces to support stable scenario comprehension and accurate safety judgments becomes a critical determinant of remote supervision safety [7]. Current onboard systems primarily provide spatial representations centered on the Electronic Chart Display and Information System (ECDIS), which presents ship positions, tracks, and associated risk parameters [8][9]. In a traditional bridge environment, crews can supplement this with their direct visual observation to obtain a first-person perspective. Compared with onboard seafarers (Figure 1), shore-based operators inherently lack the immersive realism of direct environmental perception. Shore-based operators receive the same navigational information only through displays; they lack the direct sensory cues of ship motion available to onboard crews. In a remote supervision context, this onsite perspective is typically transformed into a video stream via onboard cameras, introducing a new information modality. Since these representations were not originally designed for remote supervision, it remains to be systematically verified whether they can adequately support inexperienced operators in maintaining stable situation awareness and calibrated trust [9][10].

Maritime collision avoidance is governed by the International Regulations for Preventing Collisions at Sea (COLREGs[11]), where collision risk is conventionally quantified using metrics derived from the Closest Point of Approach (CPA), namely Distance at CPA (DCPA) and Time to CPA (TCPA) [12][13]. However, how an encounter unfolds visually is governed by its geometry, in particular the initial heading difference between the two ships [14][15]. As a result, two encounters with comparable objective risk metrics can still develop very differently on screen when their crossing angles differ [16]. For inexperienced shore-based operators, there is currently no direct evidence on whether these geometric variations independently influence

[1]Authors are with Autonomous Systems and Robotics Lab/U2IS, ENSTA, Institut Polytechnique de Paris, Palaiseau, France {zhichen.lu, xiaoxuan.hei, benoit.clement, adriana.tapus}@ensta.fr
[2]Authors are with Flinders University, Tonsley, Australia matthew.stephenson@flinders.edu.au
[3]Authors are with the French Australian Laboratory for Humans-Machine Teaming (IRL Crossing), Adelaide, Australia
[4]The author is with Cognitive Robotics Lab, Department of Computer Science, University of Manchester, Manchester M13 9PL, U.K. zhegong.shangguan@manchester.ac.uk

their trust in the autonomous system and their subsequent intervention decisions [7][16]. This study distinguishes two levels of trust. Explicit trust is what operators report when they rate the system. Implicit trust is what their behavior reveals: whether and when they intervene, which we treat as behavioral safety calibration. Furthermore, while combining ECDIS and bridge views theoretically provides richer spatial information, integrating representations from different reference frames imposes an additional cognitive transformation load [17]. Whether this integration load significantly impacts operator trust also lacks empirical investigation [6][10].

To address these gaps, this study conducts a controlled simulation experiment to systematically investigate how crossing angle and display modality jointly affect the trust calibration and intervention behavior of shore-based operators. By employing a deterministic controller compliant with COLREGs and manipulating encounter geometry while maintaining comparable objective risk distributions, we simultaneously measure subjective trust ratings and behavioral intervention thresholds. This approach allows us to distinguish between explicit trust evaluation and implicit safety calibration, providing an empirical foundation for the design of future shore-based MASS interfaces.

Based on the theoretical background and related research, we propose the following hypotheses: (i) H1: The crossing angle significantly affects operators' trust ratings and intervention behavior; (ii) H2: Display modality influences both overall trust and behavioral safety thresholds; and (iii) H3: An interaction effect exists between display modality and crossing angle on trust, reflecting that the cognitive load required for the integration of reference frames varies with the geometric complexity of the encounter.

II. Related Work

A. Remote Supervision and the Emerging Operator Profile

Remote supervision fundamentally alters the human role in maritime navigation. Instead of direct control, operators are positioned as monitors of autonomous decision processes whose authority is triggered only under uncertainty or perceived risk. Regulatory roadmaps, including the IMO MASS Code framework, formally incorporate Shore Control Centers into supervisory architectures, and industry actors have obtained classification approval for relocating critical control functions from onboard crews to remote facilities [18]. In such settings, the operator's responsibility is contingent on the autonomy level of the ship [19]. Crucially, this supervisory role may not necessarily require extensive seafaring experience [5]. Emerging analyses suggest that future shore based operators may rely primarily on representations mediated by interfaces rather than embodied maritime practice [20]. However, most maritime human factors studies continue to recruit certified mariners as participants [21][6], leaving open the question of how supervisory users with less experience construct situational understanding and trust judgments under encounters with varying geometries.

B. Trust Calibration in Autonomy Guided by Motion

Trust calibration—defined as the alignment between perceived reliability and actual system state—has long been recognized as central to safe interaction between humans and automation [7]. Miscalibration manifests as overtrust or undertrust, both of which can lead to inappropriate reliance or premature intervention. Empirical work in automated driving and UAV supervision demonstrates that transparency manipulations influence not only trust magnitude but also its alignment with system performance [16][3].

In maritime contexts, structured presentation of avoidance intent and indicators related to CPA has been shown to improve comprehension of encounter dynamics [6][22], and transparency layers can enhance situation awareness in simulated environments [23], [24]. Yet existing studies predominantly manipulate explanation level or information granularity, rather than properties of the encounter itself. Whether geometric characteristics of relative motion systematically shape trust calibration remains largely unexplored.

C. Display Modality, Reference Frames, and Spatial Integration

Spatial representation plays a central role in supervisory judgment. ECDIS encodes ship position and collision parameters in an allocentric chart format, supporting route monitoring but abstracting dynamic relative motion [25][8]. Accident analyses indicate that breakdowns in situation awareness frequently occur at perceptual and comprehension levels in tasks mediated by ECDIS [25]. Conversely, visual modalities from cameras approximate egocentric perception but lack explicit quantitative structure.

Research on multimodal integration suggests that combining representations across reference frames imposes additional cognitive transformation demands [17][26]. In teleoperation contexts, egocentric dominance can induce cognitive tunneling, reducing sensitivity to global spatial relationships [10]. Despite the coexistence of chart formats and visual modalities in maritime environments, no empirical work has examined how encounter geometry interacts with representational format to shape supervisory trust and intervention behavior, particularly among operators without professional backgrounds.

III. METHODOLOGY

A. Participants

A total of 26 participants (11 female, 15 male; mean age = 22.08, SD = 1.27) took part in the study. All participants have basic maritime theoretical education and a certain level of experience in ship bridge simulators. Participants were current maritime students undergoing academic training. While they have foundational knowledge of navigation principles and maritime procedures, they had not yet accumulated any practical seafaring experience. Participation was strictly voluntary. All individuals read and signed an informed consent form prior to the experiment. They were explicitly informed of their right to withdraw from the study at any time without penalty. Each participant completed a single experimental session lasting approximately forty minutes. A ten-minute break was provided halfway through the session to help reduce fatigue.

B. System and Interface

The experiment ran on LOTUSim, an open-source real-time maritime simulator that provides physics-based surface-ship dynamics [27]. We built the shore-based operator interface on top of it, rendering the encounter through a Unity front end (Figure 2). A deterministic COLREGs-compliant kinematic controller governed ship motion, initializing encounters with predefined crossing angles (Figure 3), and executing rule-based avoidance behavior to ensure encounter geometries remained identical across participants and conditions. Navigation behavior was kept constant to isolate the effects of interface design on human perception.

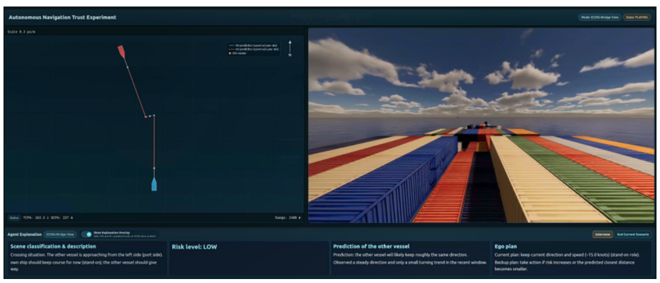


Fig. 2. Inside of experimental interface: Upper Left: chart-based ECDIS view showing own ship (blue) and target ship (red), predicted tracks, and the CPA marker. Upper Right: first-person bridge view. Lower: system explanation panel displaying scene classification (Crossing situation. The other vessel is approaching from the left side (port side). own ship should keep course for now (Stand-on); the other vessel should give way.), risk level (LOW), predicted target ship behavior (Prediction: the other vessel will likely keep roughly the same direction.), and the current plan (Current plan: keep current direction and speed (15.0 knots) (Stand-on role). Backup plan: take action if risk increases or the predicted closest distance becomes smaller.).

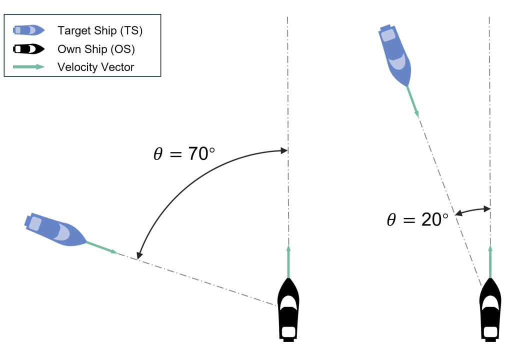


Fig. 3. Definition of crossing angle. The crossing angle θ is defined as the initial heading difference between Own Ship (OS) and Target Ship (TS). Two encounter geometries were evaluated: 70° (Left) and 20° (Right).

Objective collision risk was computed continuously from simulated geometry at each timestep using combined Distance at CPA (DCPA) and Time to CPA (TCPA) thresholds. Each scenario was assigned an objective risk level based on the values of TCPA and DCPA recorded at the moment the participant terminated observation. Three ordinal risk levels were defined and calibrated to the maneuvering envelope of the simulated ship and the geometric bounds of the scenario space. Threshold values were determined through iterative calibration against scenario geometry and verified by two experienced mariners (Second Officer and Third Officer (Officer in Charge of a Navigational Watch, OOW)): **High risk**: TCPA $< 40\,\text{s}$ *and* DCPA $< 200\,\text{m}$; **Medium risk**: TCPA $< 115\,\text{s}$ *and* DCPA $< 330\,\text{m}$, and **Low risk**: all remaining events.

Three display configurations were evaluated:

- **ECDIS-only**: a chart-based spatial representation showing ship motion and geometric information.
- **Bridge-only**: a first-person onboard navigational view.
- **Combined**: a split configuration presenting both chart-based and onboard views.

In ECDIS-enabled conditions, participants could toggle an explanation overlay that displayed predictive trajectories and CPA-related markers. The underlying chart background remained static, and the ego ship always initialized from a fixed region near the lower portion of the display to maintain a consistent spatial reference.

A structured situation analysis panel updated at ten-second intervals, deriving its content from current and historical geometric states of both ships. Every ten seconds, the panel read the current geometric state of both ships: relative bearing, range, DCPA, TCPA, and relative velocity. It matched these values against a fixed set of rules. Each rule wrote one of four output fields: scene type, risk level, predicted target-ship behavior, and the ego ship's planned action. Two experienced mariners checked every rule against COLREGs. We updated the panel every ten seconds so operators could read it without losing responsiveness.

In addition, the panel and the predictive overlay were available only in the two ECDIS-enabled conditions. The Bridge-only condition did not show them. This means the three conditions differed not only in viewpoint but also in how much explicit predictive information they exposed.

C. Experimental Design

The study adopted a within-subject design. Display modality (three levels), crossing angle (20° vs. 70°), Give-way ship maneuver timing (early vs. late), and own-ship role (Give-way vs. Stand-on) were fully crossed within participants (the crossing scenarios are illustrated in Figure 4). The angles were chosen to avoid boundary cases such as near head-on or near perpendicular encounters, while still producing clearly distinguishable encounter dynamics. Own-ship role and maneuver timing were included to increase the ecological validity of the encounter scenarios, and to prevent participants from forming simple expectations about ship behavior. These variables were balanced across trials but were not the primary focus of the study.

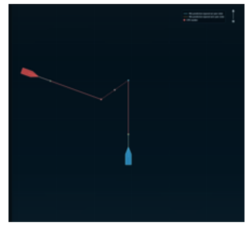
Responsibilities in encounters involving risk of collision are asymmetrically allocated in COLREGs. In a crossing situation (Rule 15), where two ships are on intersecting courses and a risk of collision exists: i). If the other ship is on a ship's starboard side, that ship is the Give-way ship and is required to take early and substantial action to keep well clear (Rule 16). ii). If the other ship is on a ship's port side, that ship is the Stand-on ship and shall, in principle, maintain her course and speed (Rule 17). Early and late evasion are operationalized as differences in the timing of maneuver initiation by the Give-way ship: i). Early evasion denotes a maneuver initiated after collision risk has been identified, while sufficient temporal and spatial margins remain, consistent with the "in ample time" requirement. ii). Late evasion denotes a maneuver initiated only after safety margins have substantially decreased, potentially leading to a close-quarters situation or increased interaction uncertainty. We focus on the timing of action by the Give-way ship, rather than the type or magnitude of the maneuver. Each participant completed twelve trials. Within these trials, six were experienced from the Give-way role and six from the Stand-on role. Crossing angle and timing were balanced across trials. The eight scenario templates defined in Table I were each instantiated under three display modality conditions, yielding a total of 24 unique scenario instances. Each participant completed 12 of these instances, drawn from one of two counterbalanced stimulus sets. No participant viewed the same encounter under the same display-modality and viewpoint combination, reducing familiarity effects. All encounters ended with successful avoidance. However, geometric collision risk could emerge and evolve during the scenario.

TABLE I
SCENARIO MATRIX DEFINITION

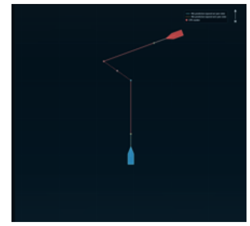
Scenario ID	Crossing Angle	OS Role	Maneuver
L20 early	20° Port	Stand-on	Early evasion
L20 late	20° Port	Stand-on	Late evasion
L70 early	70° Port	Stand-on	Early evasion
L70 late	70° Port	Stand-on	Late evasion
R20 early	20° Starboard	Give-way	Early evasion
R20 late	20° Starboard	Give-way	Late evasion
R70 early	70° Starboard	Give-way	Early evasion
R70 late	70° Starboard	Give-way	Late evasion

D. Procedure

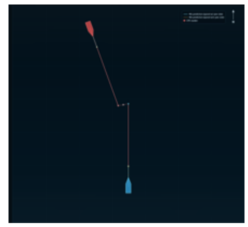
Participants first received a brief interface introduction and completed a practice trial. They were instructed to observe the encounter and determine whether collision risk existed at any time during the scenario. During each trial, participants could terminate the scenario early by pressing an "End Current Scenario" button once they believed risk had been resolved. They could also press an "Intervene" button at any time to indicate takeover intention. The system automatically recorded the geometric state at the moment of termination or intervention, including Range, TCPA, and DCPA. After each trial, participants completed a short questionnaire assessing trust, predictability, perceived mental effort, information sufficiency, and takeover intention.



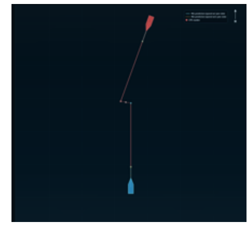
(a) 70° Port × Stand-on



(b) 70° Port × Give-way



(c) 20° Port × Stand-on



(d) 20° Port × Give-way

Fig. 4. Overall crossing situations (a)–(d).

E. Dependent Measures

To evaluate supervisory trust, calibration behavior, and workload, we administered structured questionnaires during the experiment and after its completion. The questionnaire items were designed with the assistance of two experienced mariners. The complete set of items is provided in Table II. Scenario measures assessed risk perception, confidence, trust in system analysis, predictability, perceived cognitive effort, information sufficiency, and intervention intent. Post experiment workload was measured using a six item NASA-TLX [28], which asks operators to rate six sources of workload, including mental demand, physical demand, temporal demand, performance, effort, and frustration, and combines them into a single workload score. A final evaluation survey assessed global trust, deployment comfort, perceived transparency, interface clarity, attention to status information, and information reliance. Trust was operationalized at both scenario and global levels. Calibration behavior was quantified through intervention timing and termination geometry relative to DCPA/TCPA thresholds. Cognitive demand was assessed through self-reported effort and NASA-TLX scores. Single item scenario measures were used to minimize interruption during repeated trials. Geometric variables provided objective indicators of implicit safety thresholds. Statistical analyses used linear mixed effects models with participant and scenario template as random effects.

IV. RESULTS

We analyzed each dependent variable with a 2 (crossing angle) × 3 (display modality) repeated-measures design. Before testing, Shapiro-Wilk checks confirmed approximate normality in every condition cell, so we used parametric models throughout. For the three-level modality factor we tested sphericity with Mauchly's test and applied the Greenhouse-Geisser correction where it was violated. We followed each significant omnibus effect with Bonferroni-corrected pairwise comparisons, which raise the significance

TABLE II

FULL QUESTIONNAIRE ITEMS USED IN THE STUDY.

ID	Construct	Item
Scenario Level Measures		
Q1	Risk perception	Is there a collision risk?
Q2	Confidence	How confident are you in your judgment?
Q3	Trust	I trust the system's analysis.
Q4	Predictability	The system's behavior was predictable.
Q5	Cognitive effort	Understanding the system's behavior required mental effort.
Q6	Info sufficiency	The information provided was sufficient for making my judgment.
Q7	Intervention intent	If possible, I would take over control during this encounter.
NASA-TLX (Post Experiment)		
Final System Evaluation		
F1	Global trust	I trust this autonomous navigation system.
F2	Deployment comfort	I feel comfortable relying on this system for operation.
F3	Info sufficiency	The information provided was sufficient for understanding the situation.
F4	Decision understanding	I understand how the system makes decisions in crossing situations.
F5	Display clarity	The display configuration was clear and intuitive.
F6	Status helpfulness	The system's status information helped me interpret the situation.
F7	Attention	I actively paid attention to the system's status information during the trials.

threshold to control false positives across the multiple contrasts. We ran the repeated-measures ANOVA in SPSS and refit the key effects as linear mixed-effects models in Python (statsmodels 0.14), adding a random intercept per participant to absorb individual baseline differences. The two methods agreed. We report effect sizes as η_p^2 and set $\alpha = .05$. Descriptives appear in Table III and estimates in Table IV.

A. Effects of Crossing Angle (Table IV A)

Crossing angle had a strong and consistent effect on all measures. Participants rated the autonomous system as less trustworthy and less predictable under 70° encounters than under 20° encounters, while reporting greater mental effort and lower information sufficiency (all $p < .001$). These subjective differences were mirrored behaviorally: participants observed 70° encounters for 36 s longer on average before terminating ($p < .001$, $\eta_p^2 = .271$) and were substantially more likely to express takeover intention.

B. Effects of Display Modality (Table IV B)

Display modality did not significantly affect trust, predictability, or takeover intention (all $p > .10$). The Combined display was associated with lower perceived mental effort relative to ECDIS-only, while Bridge-only yielded lower information sufficiency ratings. Despite the absence of subjective trust differences, both Bridge-only and Combined participants terminated scenarios at smaller DCPA values than ECDIS-only participants.

C. Display Modality × Crossing Angle Interaction (Table III & Table IV C)

Significant modality-by-angle interactions emerged across trust, mental effort, information sufficiency, and takeover intention (Figure **??**). The Combined display showed a geometry-dependent reversal in trust. At 20°, Combined participants reported the highest trust of any condition, yet at 70° they reported the lowest, yielding the steepest decline across conditions. Information sufficiency followed the same pattern, with the Combined × 70° interaction representing the largest effect in the study. The Bridge-only condition did not exhibit comparable declines in either measure ($p > .10$ for both interactions). For mental effort, the pattern was reversed: Bridge-only participants reported *lower* effort under 70° than under 20°, opposite to the increase observed in the other two conditions. Correspondingly, Bridge-only participants were less likely to intend takeover at 70° relative to the ECDIS-only baseline.

D. Information Use

Ship position and predicted trajectory were the most frequently cited cues (61.5% each), followed by overall impression (53.8%) and the system's stated plan (53.8%). Only one participant reported using TCPA. In the Combined condition, overlay use rose sharply with geometric complexity. Participants opened the ECDIS overlay 0.8 times per trial at 20° but 3.3 times per trial at 70°. The overlay lives in the chart frame while the encounter is watched in the bridge frame, so each reopening marks a switch between the two frames. This fourfold rise in switching under 70° is the concrete behavior behind the reference-frame integration cost we invoke for H3, and it tracks the higher mental effort and the sharper 70° trust drop seen for the Combined display.

E. Post-Experiment Workload (NASA-TLX)

We administered the NASA-TLX once, after the full session, so it indexes overall workload rather than workload in any single condition. Averaged across the 26 participants, overall workload stayed below the scale midpoint (Raw TLX = 3.17, SD = 0.79, on a 1–7 scale; midpoint = 4). Physical demand sat near the floor (M = 1.50, SD = 0.58), as expected for a seated monitoring task, while mental demand (M = 3.38, SD = 1.33) and effort (M = 3.27, SD = 1.61) were the main load components. The global score tracked the per-trial mental-effort ratings: participants who scored higher on the NASA-TLX also reported more mental effort across scenarios (r = .55, p = .005). This convergence indicates that the condition-level mental-effort effects reflect genuine workload differences rather than a single-item artifact.

F. Effects of Role and Timing

We tested own-ship role and maneuver timing, both crossed within participants. Role produced no main effect: trust, takeover intention, and termination DCPA did not differ between Give-way and Stand-on trials (all $p > .13$). The crossing-angle and display-modality effects therefore hold across both roles.

TABLE III

DESCRIPTIVE STATISTICS (MEAN ± SD) PER DISPLAY MODALITY AND CROSSING ANGLE. LIKERT ITEMS USE A 7-POINT SCALE.

	ECDIS only		Bridge only		Combined	
Measure	20°	70°	20°	70°	20°	70°
Trust	5.97 (1.09)	4.67 (1.91)	5.47 (1.40)	5.12 (1.89)	6.50 (0.83)	3.56 (1.15)
Mental Effort	3.28 (1.49)	4.44 (1.58)	3.72 (1.41)	3.36 (1.39)	2.34 (1.15)	4.72 (0.89)
Info Sufficiency	5.72 (1.22)	5.39 (1.54)	5.22 (1.46)	4.43 (1.87)	6.41 (0.84)	4.06 (1.35)
DCPA (m)	288.70 (95.35)	285.87 (124.28)	280.50 (110.78)	182.47 (7.28)	275.98 (83.21)	179.10 (8.59)

TABLE IV

SELECTED LINEAR MIXED-EFFECTS MODEL RESULTS (RANDOM INTERCEPT: PARTICIPANT). BOLD = $p<.05$.

Outcome	**Predictor**	$\hat{\beta}$	p	η_p^2
A. Crossing Angle (70° vs. 20°)				
Trust	-	−1.55	< .001	.225
Mental Effort	-	+1.10	< .001	.145
Info Sufficiency	-	−1.12	< .001	.137
Takeover Intention	-	+0.65	< .001	.397
B. Display Modality Main Effects				
Mental Effort	Combined vs. ECDIS	−0.50	.046	.027
Info Sufficiency	Bridge vs. ECDIS	−0.69	.008	.046
DCPA (m)	Bridge vs. ECDIS	−39.89	.027	.031
	Combined vs. ECDIS	−45.27	.012	.040
C. Display × Crossing Angle Interaction				
Trust	Bridge × 70°	+0.95	.107	.017
	Combined × 70°	−1.64	.003	.055
Mental Effort	Bridge × 70°	−1.39	.025	.034
Info Sufficiency	Combined × 70°	−1.97	< .001	.070
Takeover Intention	Bridge × 70°	−0.47	.016	.048

V. DISCUSSION

A. Encounter Geometry and Supervisory Trust

The results support $H1$: crossing angle significantly influenced supervisory trust and intervention behavior independent of objective risk magnitude. Despite comparable DCPA/TCPA distributions across conditions, encounters at 70° consistently produced lower trust ratings, higher intervention frequency, longer observation durations, and increased perceived mental effort relative to 20° encounters. This pattern indicates that operators did not primarily rely on formal risk metrics when evaluating system reliability. Instead, supervisory judgment appears to be driven by the perceptual unfolding of relative motion—the evolving pattern of lateral displacement, bearing change, and trajectory convergence that distinguishes large angle from small angle encounters even when quantitative risk indices are equivalent. This interpretation is consistent with research on human collision perception [29] demonstrating that observers often rely on the optical structure of motion, such as visual expansion and the rate of change of visual angle, to estimate time-to-collision rather than explicitly computing distance- or velocity-based quantities. From this perspective, different encounter geometries generate distinct perceptual motion patterns in the operator's visual field. Large-angle crossings produce more pronounced lateral displacement and complex bearing changes, potentially increasing perceptual ambiguity and cognitive processing demands even when formal risk indicators remain constant. Consequently, supervisory trust calibration may be influenced not only by objective safety metrics but also by the perceptual characteristics of the encounter geometry through which the automation's behavior is interpreted.

In maritime supervisory HRI, encounter geometry therefore constitutes an independent determinant of trust calibration, not merely a context in which risk metrics are applied.

B. Display Modality and Behavioral Safety Thresholds

The results partially support $H2$. Display modality did not produce a consistent main effect on overall trust magnitude: subjective trust ratings were broadly similar across ECDIS-only, Bridge-only, and Combined conditions. However, modality altered behavioral safety thresholds. Participants terminated encounters at larger DCPA values in the ECDIS-only condition than in the other two conditions, which produced nearly identical, smaller termination distances. The common factor of the latter two is the first-person bridge view, suggesting that the egocentric viewpoint, rather than chart availability, drove the shift: a diverging target ship visually clears the bow before the chart geometry confirms a safe margin, leading operators to end observation earlier. This is consistent with egocentric dominance in teleoperation, where first-person viewpoints reduce sensitivity to global spatial relationships. The termination geometry distribution supports this pattern, with many termination events occurring outside formally HIGH risk regions.

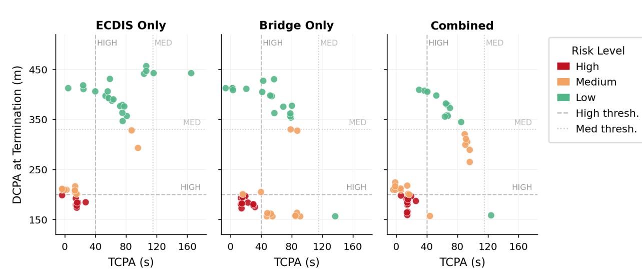


Fig. 5. Distribution of termination geometry across trials. Points represent DCPA and TCPA values at scenario termination, colored by objective risk level and shaped by display modality. Dashed lines indicate experimental risk thresholds.

This pattern is consistent with prior research showing that subjective trust and behavioral reliance are related but distinct constructs. Operators may report similar levels of trust in an automated system while adopting different reliance strategies in risk-sensitive decisions [30]. Similarly, models of calibrated trust suggest that reliance behavior is dynamically adjusted based on available system cues and environmental information, rather than solely reflecting explicit trust attitudes [31]. Recent studies of human–automation interaction also report that users can maintain stable trust ratings while adapting their operational strategies in response to interface characteristics [32]. Together, these findings suggest that display modality may influence how operators translate spatial interpretation into action decisions even when explicit trust evaluations remain unchanged. Consequently, evaluations based solely on trust questionnaires may overlook meaningful changes in supervisory behavior, highlighting the importance of incorporating behavioral measures when assessing supervisory interface design.

C. Geometry Dependent Effects of Display Modality

The results support $H3$: a significant interaction between display modality and crossing angle was observed for trust, mental effort, information sufficiency, and intervention intention. Under 20° encounters, the Combined condition yielded higher trust than single modality conditions. Under 70° encounters, however, trust declined more sharply in the Combined condition than in ECDIS-only. This asymmetry is consistent with $H3$'s framing in terms of reference frame integration load: when crossing angles produce more pronounced changes in relative bearing and trajectory orientation, reconciling a first-person bridge view with an allocentric chart view requires greater cognitive effort. Prior research on spatial cognition has shown that switching between egocentric and allocentric frames of reference imposes measurable cognitive costs and increases mental workload during spatial tasks [33]. Operators' reports on information use after the experiment further clarify this pattern. Ship position and predicted trajectory were cited as primary information sources by 60% of participants each, while TCPA was selected by only one participant. This limited reliance on temporal risk metrics, even when they were explicitly available through the overlay, indicates that supervisory judgment is grounded primarily in spatial geometric interpretation rather than formal metric evaluation.

Similar tendencies have been observed in transportation safety contexts, where operators tend to rely more strongly on visual spatial cues when responding to automation prompts [34]. In human automation systems, display design also plays a critical role in shaping trust and situation awareness, and additional information sources may increase cognitive demands when task environments become more complex [35]. The present findings therefore do not imply that multimodal display is generally detrimental; rather, its effectiveness is contingent on the geometric complexity of the encounter being observed.

D. Implications for Interface Design

The three findings converge on a shared design implication: supervisory interfaces for autonomous maritime systems should prioritize the direct representation of encounter geometry over the provision of additional abstract risk indicators.

First, interfaces foregrounding the bridge view should retain an explicit geometric margin cue rendered within the visual scene, since the first-person viewpoint alone led operators to terminate at smaller safety margins. Second, multimodal interfaces should address the coherence of the reference frame rather than simply juxtaposing multiple views. Designs that provide consistent spatial orientation, synchronized anchors, or unified coordinate projections can reduce the integration load associated with trust instability under geometrically complex encounters. Emerging immersive technologies such as VR/AR offer a further alternative: embedding predictive trajectories within a shared three dimensional scene could reduce cross frame reconciliation demands entirely. Third, the dissociation between declarative trust and procedural calibration identified under $H2$ suggests that interface evaluation frameworks should incorporate behavioral measures alongside subjective ratings.

E. Limitations

Several limitations should be acknowledged. First, participants were maritime students with theoretical training but without professional seafaring experience, and the sample size was relatively small due to the limited availability of maritime students, which may limit the generalizability of the findings to certified bridge officers or experienced shore-based operators. Second, the experiment examined two-ship crossing encounters under controlled geometric conditions; real-world maritime environments often involve multi-vessel traffic, environmental disturbances, and communication uncertainties that may alter supervisory calibration behavior. Third, although objective risk distributions were matched across conditions, the simulator-based setting cannot fully reproduce the perceptual and contextual complexity of live operations. Fourth, transparency was not equal across conditions. Only the ECDIS-enabled conditions carried the predictive overlay and the situation panel, so viewpoint and information richness are partly confounded.

VI. CONCLUSION

This study examined how the crossing angle and the display modality jointly shape the supervisory trust and intervention behavior in autonomous maritime collision avoidance, with particular relevance to emerging shore-based supervision contexts where operators may lack extensive navigational experience. Three hypotheses were tested using a controlled real time simulation with a deterministic COLREGs compliant controller. The crossing angle significantly affected trust ratings, perceived workload, and intervention behavior despite comparable objective risk distributions, indicating that supervisory judgment is primarily driven by the perceptual geometry of unfolding relative motion rather than abstract risk metrics. The display modality did not produce a consistent effect on overall trust magnitude but significantly altered behavioral safety thresholds, revealing a dissociation between declarative trust evaluation and procedural safety calibration that trust questionnaires alone cannot capture. An interaction between display modality and crossing angle further showed that multimodal representation benefits trust under small angle encounters but reduces trust stability under larger angle encounters, consistent with increasing reference frame integration demands. Together, these findings indicate that trust calibration in maritime supervisory HRI is shaped by encounter geometry and that interface effectiveness cannot be evaluated independently of the spatial complexity of the scenarios being supervised. The results provide an empirical foundation for designing shore-based supervisory interfaces that prioritize geometric interpretability over informational density.

REFERENCES

[1] C.-H. Chang, I. B. Wijeratne, C. Kontovas, and Z. Yang, "Colreg and mass: Analytical review to identify research trends and gaps in the development of autonomous collision avoidance," *Ocean Engineering*, vol. 302, p. 117652, 2024.

[2] L. Ma, X. Ma, H. Lan, Y. Liu, and W. Deng, "A data-driven method for modeling human factors in maritime accidents by integrating dematel and fcm based on hfacs: A case of ship collisions," *Ocean Engineering*, vol. 266, p. 112699, 2022.

[3] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors*, vol. 39, no. 2, pp. 230–253, 1997.

[4] M. Bachari-Lafteh and A. Harati-Mokhtari, "Operator's skills and knowledge requirement in autonomous ships control centre," *Journal of International Maritime Safety, Environmental Affairs, and Shipping*, vol. 5, no. 2, pp. 74–83, 2021.

[5] G. R. Emad and S. Ghosh, "Identifying essential skills and competencies towards building a training framework for future operators of autonomous ships: a qualitative study," *WMU Journal of Maritime Affairs*, vol. 22, no. 4, pp. 427–445, 2023.

[6] K. van de Merwe, S. Mallam, S. Nazir, and Øystein Engelhardtsen, "Supporting human supervision in autonomous collision avoidance through agent transparency," *Safety Science*, vol. 169, p. 106329, 2024.

[7] J. D. Lee and K. A. See, "Trust in automation: Designing for appropriate reliance," *Human Factors*, vol. 46, no. 1, pp. 50–80, Spring 2004.

[8] F. Motz, H. Widdel, S. MacKinnon, A. Patterson, and L. Alexander, "Experimental investigation for presentation of ais symbols on ecdis in a motion-based ship bridge simulator," in *Ergonomie und Mensch-Maschine-Systeme*, L. Schmidt, J. Grosche, and C. M. Schlick, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 147–159.

[9] K. van de Merwe, S. Mallam, and S. Nazir, "Agent transparency, situation awareness, mental workload, and operator performance: A systematic literature review," *Human Factors*, vol. 66, no. 1, pp. 180–208, 2024, first published online 2022.

[10] M. R. Endsley, "Supporting human-ai teams:transparency, explainability, and situation awareness," *Computers in Human Behavior*, vol. 140, p. 107574, 2023.

[11] "Convention on the international regulations for preventing collisions at sea, 1972 (colregs) - colregs operator guidance framework," 2022.

[12] J. Kearon, "Computer program for collision avoidance and track keeping," in *Proc. Conf. Math. Aspects Mar. Traffic*, 1977, pp. 229–242.

[13] J. M. Mou, C. van der Tak, and H. Ligteringen, "Study on collision avoidance in busy waterways by using ais data," *Ocean Engineering*, vol. 37, no. 5, pp. 483–490, 2010.

[14] T. A. Johansen, T. Perez, and A. Cristofaro, "Ship collision avoidance and colregs compliance using simulation-based control behavior selection with predictive hazard assessment," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3407–3422, 2016.

[15] E. M. Goodwin, "A statistical study of ship domains," *Journal of Navigation*, vol. 28, no. 3, p. 328–344, 1975.

[16] K. E. Schaefer, J. Y. C. Chen, J. L. Szalma, and P. A. Hancock, "A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems," *Human Factors*, vol. 58, no. 3, pp. 377–400, May 2016, epub 2016 Mar 22.

[17] C. D. Wickens, "The trade-off of design for routine and unexpected performance: Implications of situation awareness," *Situation awareness analysis and measurement*, pp. 211–225, 2000.

[18] DNV, "Autonomous and remotely operated ships," DNV, Høvik, Norway, Tech. Rep. DNV-CG-0264, 2024.

[19] S. I. Sezer, S. I. Ahn, E. Akyuz, R. E. Kurt, and P. Gardoni, "A hybrid human reliability analysis approach for a remotely-controlled maritime autonomous surface ship (mass-degree 3) operation," *Applied Ocean Research*, vol. 147, p. 103966, 2024.

[20] X. Li and K. F. Yuen, "A human-centred review on maritime autonomous surfaces ships: impacts, responses, and future directions," *Transport Reviews*, vol. 44, no. 4, pp. 791–810, 2024.

[21] M. R. Grech, T. Horberry, and A. Smith, "Human error in maritime operations: Analyses of accident reports using the leximancer tool," in *Proceedings of the human factors and ergonomics society annual meeting*, vol. 46, no. 19. Sage Publications Sage CA: Los Angeles, CA, 2002, pp. 1718–1721.

[22] K. van de Merwe, S. Mallam, S. Nazir, and Engelhardtsen, "The influence of agent transparency and complexity on situation awareness, mental workload, and task performance," *Journal of Cognitive Engineering and Decision Making*, vol. 18, no. 2, pp. 156–184, 2024.

[23] A. Brandsæter and A. Madsen, "A simulator-based approach for testing and assessing human supervised autonomous ship navigation," *Journal of Marine Science and Technology*, vol. 29, no. 2, pp. 432–445, 2024.

[24] Z. Shangguan, X. Han, Y. E. Mrhasli, N. Lyu, and A. Tapus, "Factors influencing emotional driving: examining the impact of arousal on the interplay between age, personality, and driving behaviors," *Frontiers in Psychology*, vol. 16, p. 1487493, 2025.

[25] M. Barić, L. Grbić, L. Periĉin, and R. Jelic, "The role of the ecdis on the development of situational awareness–a study on grounding accidents," *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 17, no. 1, pp. 219–225, 2023.

[26] L. Sanneman and J. A. Shah, "The situation awareness framework for explainable ai (safe-ai) and human factors considerations for xai systems," *International Journal of Human–Computer Interaction*, vol. 38, no. 18-20, pp. 1772–1788, 2022.

[27] C. Buche, J. Grosset, H. Lechêne, M. Dubromel, P. Havez-Bodivit, M. Neo, and J. Prodhon, "Lotusim: Multi-domain simulator for marine robotics," 2026. [Online]. Available: https://arxiv.org/abs/2607.03072

[28] S. G. Hart and L. E. Staveland, "Development of nasa-tlx (task load index): Results of empirical and theoretical research," in *Advances in psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[29] J.-J. Yan, B. Lorv, H. Li, and H.-J. Sun, "Visual processing of the impending collision of a looming object: Time to collision revisited," *Journal of Vision*, vol. 11, no. 12, p. 7, 10 2011.

[30] A. S. Finn, P. B. Kalra, C. Goetz, J. A. Leonard, M. A. Sheridan, and J. D. E. Gabrieli, "Developmental dissociation between the maturation of procedural memory and declarative memory," *Journal of Experimental Child Psychology*, vol. 142, pp. 212–220, 2 2016.

[31] C. Lebiere, L. M. Blaha, C. K. Fallon, and B. Jefferson, "Adaptive cognitive mechanisms to maintain calibrated trust and reliance in automation," *Frontiers in Robotics and AI*, vol. 8, p. 652776, 2021.

[32] Z. Xu, T. Jiang, D. Xiao, Y. Fang, and N. Zheng, "Analyzing riders' behavioral adaptation to driving patterns of advanced autonomous vehicles: A virtual reality simulation study," *International Journal of Human–Computer Interaction*, vol. 41, no. 12, pp. 7273–7291, 2025.

[33] R. Orti, T. Iachini, E. D'Agostino, F. Ruotolo, and G. Ruggiero, "Cognitive load in switching between egocentric and allocentric spatial frames of reference: a pupillometry study," *Scientific Reports*, vol. 15, no. 1, p. 26451, 2025.

[34] Y.-K. Ou, W.-X. Huang, and C.-W. Fang, "Effects of different takeover request interfaces on takeover behavior and performance during conditionally automated driving," *Accident Analysis Prevention*, vol. 162, p. 106425, 2021.

[35] C. Wickens, "Designing for situation awareness and trust in automation," *IFAC Proceedings Volumes*, vol. 28, no. 23, pp. 365–370, 1995, iFAC Conference on Integrated Systems Engineering, Baden-Baden, FRG, 27-29 September 1994.