Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 12–19
https://doi.org/10.54941/ahfe1008236

"They Will Move": Confirmation Bias, Expectation Frames, and Premature Monitoring Drop Off in COLREG Encounters

Felix-Marcel Petermann[1,2], Jooyoung Park[2], Ole Andreas Alsos[2], and Erik Styhr Petersen[1,2]

[1]Western Norway University of Applied Sciences (HVL), Haugesund, Norway
[2]Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ABSTRACT

Collision avoidance depends not only on detecting cues but also on interpreting them against expectations about other vessels' intentions, compliance with rules, and local traffic norms. This paper examines confirmation bias in maritime watchkeeping: the tendency to discount cues that contradict an initial expectation and reduce monitoring of a developing hazard. We analyse Applied Cognitive Task Analysis interviews from five structured simulator scenarios with seven experienced mariners. In 439 coded excerpts, 14 expectation-driven breakdowns ('Over Trust Assumptions') were identified, nine co-occurring with situation-awareness failures, particularly misinterpreted cues and weak projection. Qualitative themes show how reliance on local knowledge, stereotyped vessel behaviour, and presumed International Regulations for Preventing Collisions at Sea (COLREG) compliance can lead to premature closure and delayed corrective action. We identify design directions supporting 'frame checking' through explicit assumption states, anomaly cues tied to expected COLREG behaviour, and continued monitoring in stand-on situations.

Keywords: Maritime watchkeeping, Confirmation bias, Navigation, Cognitive task analysis

INTRODUCTION

Watchkeepers routinely form rapid expectations about how an encounter will unfold: the ferry will pass close but safely; the give-way vessel will alter early; the Vessel Traffic Service (VTS)-coordinated ship will wait; and the overtaker will keep clear. These expectations are not in themselves errors. They are a hallmark of maritime expertise, supporting fast, recognition-primed judgement under time pressure (Klein, 1993, 2008). The problem arises when expectations become resistant to revision in the face of contradictory cues, suppressing attention that would otherwise prompt earlier action and collapsing sensemaking prematurely (Klein et al., 2006; Weick, 1995).

We frame this phenomenon as **confirmation bias in maritime watchkeeping**: the tendency to discount detectable cues that contradict an already-formed expectation about encounter trajectory (Nickerson, 1998; Wason, 1960).

Received August 28, 2026; Revised September 28, 2026; Accepted October 5, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

In dynamic environments, expectations shape which cues are sampled, how ambiguous cues are interpreted, and how much contradictory evidence is required before a prior belief is revised. In maritime watchkeeping, this process unfolds over relatively slow cue-update rates and through automation-mediated information from radar, Automatic Radar Plotting Aid (ARPA), Automatic Identification System (AIS), and Electronic Chart Display and Information System (ECDIS). Precise numerical outputs such as closest point of approach (CPA) and time to closest point of approach (TCPA) can further encourage an extrapolation to be treated as an encounter outcome rather than a prediction contingent on continued course and speed (Lee & See, 2004; Parasuraman & Manzey, 2010).

The International Regulations for Preventing Collisions at Sea (COLREG) provide an additional expectation framework. Rule knowledge shapes the watchkeeper's forward projection of who should give way and what action should follow (Cockcroft & Lameijer, 2012). This is particularly consequential under Rule 17: the stand-on vessel maintains course and speed until it becomes apparent that the give-way vessel is not taking appropriate action (International Maritime Organization (IMO), 1972). An Officer of the Watch (OOW) expecting the other vessel to alter may therefore be slow to conclude that this expectation has failed. Local knowledge and recurrent vessel categories can further strengthen such expectations.

Sensemaking provides a useful account of this transition. A frame is continually tested against incoming information until, in cases of **premature closure**, subsequent cues are assimilated into the existing interpretation rather than allowed to challenge it (Klein et al., 2006; Weick, 1995). In bridge watchkeeping, an early reading such as "he'll pass astern of us" can become the running frame. As compatible cues reinforce it, contradictory cues may be explained away and active cross-checking can gradually decline. We refer to this behavioural pattern as **premature monitoring drop-off**.

This paper examines expectation-driven breakdowns in Applied Cognitive Task Analysis (ACTA) interviews conducted around five structured simulator scenarios with seven experienced mariners. From 439 coded excerpts, 14 were identified as '*Over Trust Assumptions*'; nine co-occurred with *situation-awareness failures*. We examine the expectation scaffolds underlying these cases and derive design directions for **frame checking**: supporting watchkeepers in recognising when an encounter has stopped conforming to what their current expectation predicts.

METHOD

Approach and ACTA Protocol

We used Applied Cognitive Task Analysis (ACTA) (Militello & Hutton, 1998) to elicit the cognitive basis of collision-avoidance watchkeeping, including the cues experts attend to, patterns they recognise, and expectations they form. Each 90–120-minute session comprised the three ACTA phases: a task diagram of a typical bridge watch, a knowledge audit probing cognitively demanding sub-tasks, and a simulation interview grounded in five navigational

scenarios. A shared Miro board allowed participants to point, annotate, and correct in real time. Interviews were conducted through Microsoft Teams and audio- and video-recorded with consent.

Scenarios and Participants

Five scenarios set in the Haugesund and Karmøy Sound area were designed to span different COLREG encounters, visibility conditions, and waterway constraints (Table 1). Each was presented as a two-stage time-series stimulus comprising an out-the-window bridge view, radar, and ECDIS display. Environmental conditions varied between clear day (C0), fog with visibility below 0.5 NM (C1), and clear night (C2), counterbalanced so that every scenario was presented in at least two conditions across the cohort. Participants were probed about situation assessment, critical cues, intended actions, anticipated novice failure modes, and desired decision support.

Table 1: Simulation-interview scenarios.

Scenario	Rule	Core Cognitive Challenge
S1 - Confined meeting	R14	Cross-checking cues in a confined channel
S2 - Starboard narrow pass	R15	Give-way obligation vs. practical seamanship
S3 - Port not giving way	R17	Stand-on vigilance; threshold for own action
S4 - Constrained vessel	R18	Correct status identification
S5 - Overtaken from astern	R13	Rearward SA with an obstructed aft view

Seven experienced mariners participated: three Captains, two Chief Officers, and two 2nd Officers, with four to more than thirteen years of sea service. Participants were recruited through industry partners as an expert convenience sample. Several had familiarity with the Haugesund and Karmøy Sound area represented in the stimuli.

Coding and Analysis

Interviews were transcribed verbatim and segmented at the turn level, yielding 439 codeable excerpts. Coding combined deductive and inductive passes following reflexive thematic analysis (Braun & Clarke, 2006). The deductive codebook covered situation-awareness levels (perception, comprehension, projection) (Endsley, 1995) naturalistic-decision-making constructs (pattern recognition, mental simulation, if-then rules) (Klein, 1993, 2008), and information sources (visual, radar, ECDIS, AIS, VHF, local knowledge). An inductive pass added recurrent patterns, including ***'Over Trust Assumptions'***, defined as excerpts in which a participant described forming or acting on an expectation about another vessel's behaviour, such as intention, rule compliance, or conformity to local traffic patterns, which was not warranted by the cues available at the time.

The present analysis examines the 14 excerpts coded 'Over Trust Assumptions' and their co-occurrence with situation-awareness failure codes. Co-occurrence was counted when both codes occurred in the same

or directly adjacent excerpts within a narrative episode. Qualitative themes were developed through iterative re-reading against transcript context and the corresponding scenario stimulus. Twenty percent of excerpts were double-coded by a second researcher, with discrepancies resolved through discussion and codebook refinement. Ethical approval was obtained from SIKT (ref. 398976), and all participants provided written informed consent.

RESULTS

Distribution and Co-Occurrence

'*Over Trust Assumptions*' was identified in 14 of the 439 coded excerpts (3.2%), spanning scenario walkthroughs (n = 7), task-diagram discussions (n = 4), and knowledge-audit probes (n = 3). Six of seven participants contributed at least one excerpt, indicating that the pattern was not driven by a single participant.

Nine of the 14 excerpts co-occurred with at least one situation-awareness failure. The most frequent failure modes were Misinterpreted Cue (Levels 1–2) and Weak Projection (Level 3). Missing a cue entirely was not the dominant pattern. Rather, cues were available and attended to but interpreted through an expectation frame that over-wrote their implication for how the encounter would evolve.

Expectation Scaffolds

Qualitative analysis identified three recurrent scaffolds underlying '*Over Trust Assumptions*', together with a fourth pattern in which display state shaped the inferred narrative of the encounter. The patterns were not mutually exclusive.

Presumed VTS Coordination and Local-Area Handling

Participants described assumptions that a VTS-controlled or otherwise supervised area guaranteed coordination:

"… this is a VTS area. We would be informed of the vessel passing …" (P6)

"… they are still almost in the same position … most likely the VTS has already told me that this vessel will wait for me…" (P1)

Here, the expectation replaces rather than supplements cross-checking: in the absence of an explicit VTS instruction, target behaviour is interpreted as evidence that coordination has already taken place.

Vessel-Category Stereotypes

Recurrent traffic participants, particularly high-speed passenger ferries operating on known routes, produced baseline expectations derived from previous encounters:

> "… you recognize the same vessels… These appear big but manoeuvre quickly and pass closely, so avoidance often isn't needed." (P2)

Such expectations raise the threshold for what counts as anomalous behaviour: a familiar vessel must deviate further from its expected pattern before prompting additional monitoring or action.

Presumed COLREG Compliance

Stand-on situations were particularly vulnerable to expectations that the give-way vessel would alter course. Participants described continuing to expect this action as time-to-CPA shortened, potentially deferring their own intervention under Rule 17. The threshold at which non-compliance becomes "apparent" is therefore itself shaped by the expectation that the rule will be followed.

Display-State-Derived Expectations

Participants also described how degraded or incompletely configured displays can generate expectations that outlive their basis. For example, the absence of an ECDIS vector could lead an unfamiliar operator to infer a collision course. Although this represents an overly pessimistic rather than benign expectation, the mechanism is similar: a display state is treated as the outcome of the encounter rather than a representation whose completeness and validity must be checked.

From Expectation to Premature Monitoring Drop-Off

Across the excerpts, a consistent pathway connected expectation frames to delayed intervention. An expectation, grounded in VTS coordination, vessel category, or rule compliance is formed early and stabilised by neutral cues. As the encounter progresses, cross-checking can decline because the established frame no longer demands active verification. The transition from monitoring to intervention is correspondingly delayed, compressing the time available for measured action and increasing the likelihood of a late manoeuvre or close-quarters situation.

This pathway is a composite pattern across the coded excerpts and corresponding scenario contexts rather than a directly observed temporal sequence. We refer to its behavioural signature as **premature monitoring drop-off:** the tapering of active cross-checking that follows, rather than precedes, the formation of a confident expectation.

DISCUSSION

Confirmation Bias at the Boundary of Expertise

Our data support treating confirmation bias in watchkeeping as a boundary condition of expertise rather than a generic cognitive flaw. The expectation scaffolds identified here, VTS coordination, vessel categories, and presumed rule compliance, are the same scaffolds that, in typical cases, make experienced watchkeepers fast, efficient, and broadly accurate (Klein, 1993, 2008) . What differentiates the problematic excerpts from competent scenario reasoning is

not the presence of an expectation but its relation to subsequent cue sampling. In these cases, expectations substitute for cross-checking, explain away neutral cues, raise the threshold for what counts as anomalous, and compress the set of hypotheses under active test (Klein et al., 2006; Weick, 1995).

A design response aimed at removing expectations would therefore be both infeasible and counterproductive. A more tractable target is **frame checking**: making it easier to recognise when the current situation has stopped conforming to what an expectation predicts. Three design directions follow from the identified patterns.

First, the ***state of an assumption should be explicit***. In the VTS and vessel-category cases, whether an expected action is confirmed, for example through a VTS instruction or VHF agreement, or merely inferred from target behaviour is not represented. Distinguishing ***confirmed*** from ***inferred*** expectations could prevent an assumption from being treated as an established encounter state.

Second, ***anomaly cues should relate to expected behaviour, not only geometric thresholds***. CPA/TCPA alerts indicate when encounter geometry has already degraded. A rule-based anomaly cue could instead indicate when a nominated give-way vessel has not acted as expected or when observed behaviour increasingly diverges from the current prediction. The aim is to surface the point at which the expectation begins to fail rather than only the consequence of its failure (Woods, 1995).

Third, stand-on situations require ***projection support for the remaining intervention window***. Rather than presenting CPA/TCPA only as current numerical values, an interface could communicate their development and the remaining opportunity for a measured alteration before a close-quarters situation requires more substantial action. This would transform the expectation that the give-way vessel will act from an open-ended wait into a time-bounded expectation whose deterioration is visible.

These directions are also relevant to highly automated bridges and shore-control operations, where operators' expectations about other traffic are combined with expectations about the autonomy itself (Parasuraman & Manzey, 2010). Additional automation layers may therefore increase rather than remove the need to expose assumptions, behavioural divergence, and uncertainty. Frame checking should support recognition-based expertise rather than replace it.

LIMITATIONS AND FUTURE WORK

Three limitations qualify these findings. First, the study used a small exploratory sample (N = 7) recruited through industry partners and did not apply a standardised expertise scale; the findings therefore characterise patterns rather than their prevalence. Second, participants discussed two-stage static scenario packets rather than navigating a dynamic simulator. Premature monitoring drop-off is consequently reconstructed from verbal protocols and coded episodes rather than directly measured behaviour. Third, the present analysis focuses on excerpts coded ***'Over Trust Assumptions'***; related expectation-based patterns captured under adjacent codes were outside its scope.

Ongoing simulator studies use the same scenarios with expert and novice participants, combining eye tracking, dynamic bridge behaviour, and SAGAT freeze probes to examine monitoring drop-off directly. Interface prototypes implementing assumption state, rule-tied anomaly cues, and projection support will subsequently allow the proposed frame-checking concepts to be evaluated with experienced watchkeepers and shore-control operators.

CONCLUSION

Confirmation bias in maritime watchkeeping can be understood as a boundary condition of expertise: the same expectation frames that enable experienced mariners to act efficiently can make them slow to revise an established interpretation. In our ACTA data, expectation-driven breakdowns clustered around presumed VTS coordination, vessel-category stereotypes, presumed COLREG compliance, and display-state-derived narratives, and were associated primarily with interpretation and projection failures rather than simple perceptual misses. We describe the resulting risk pathway as ***premature monitoring drop-off***. Designing for ***frame checking***, making assumptions explicit, identifying divergence from expected behaviour, and supporting projection of the remaining intervention window provides a concrete direction for supporting continued monitoring without removing the recognition-based expertise on which maritime decision making depends.

REFERENCES

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706QP063OA

Cockcroft, A. N., & Lameijer, J. N. F. (2012). A Guide to the Collision Avoidance Rules. In *A Guide to the Collision Avoidance Rules, Seventh Edition*. Elsevier. https://doi.org/10.1016/C2010-0-68322-2

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

International Maritime Organization (IMO). (1972). COLREG – Convention on the International Regulations for Preventing Collisions at Sea. In *COLREG – Convention on the International Regulations for Preventing Collisions at Sea*. International Maritime Organization. https://doi.org/10.62454/KC904E

Klein, G. (1993). A Recognition Primed Decision (RPD) Model of Rapid Decision Making. In G. Klein, J. Orasanu, R. Calderwood, & C. Zsambok (Eds.), *Decision Making in Action*. https://www.researchgate.net/publication/235418838_A_Recognition_Primed_Decision_RPD_Model_of_Rapid_Decision_Making

Klein, G. (2008). Naturalistic Decision Making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 456–460. https://doi.org/10.1518/001872008X288385

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intelligent Systems*, *21*(4), 70–73. https://doi.org/10.1109/MIS.2006.75

Klein, Moon, & Hoffman. (2006). Making Sense of Sensemaking 2: A Macrocognitive Model. *IEEE Intelligent Systems*, *21*(5), 88–92. https://doi.org/10.1109/MIS.2006.100

Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Militello, L. G., & Hutton, R. J. B. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics*, *41*(11), 1618–1641. https://doi.org/10.1080/001401398186108

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Parasuraman, R., & Manzey, D. H. (2010). Complacency and Bias in Human Use of Automation: An Attentional Integration. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *52*(3), 381–410. https://doi.org/10.1177/0018720810376055

Wason, P. C. (1960). On the Failure to Eliminate Hypotheses in a Conceptual Task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129–140. https://doi.org/10.1080/17470216008416717

Weick, K. E. (1995). Sensemaking in Organizations (Foundations for Organizational Science). *Star*, 235. https://books.google.com/books/about/Sensemaking_in_Organizations.html?id=nz1RT-xskeoC

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, *38*(11), 2371–2393. https://doi.org/10.1080/00140139508925274