BEYOND FAMILIARISATION: EXERCISING, STRATEGY TESTING AND CRITICAL MARITIME INFRASTRUCTURE

DOI: https://doi.org/10.37458/nstf.27.2.6
Review paper
Received: October 4, 2026
Accepted: November 14, 2026

Sergiu Mitrescu*

Abstract: The tabletop exercise has been central to the adaptation of NATO and EU member states to hybrid threats, enabling practitioners to familiarise themselves with unfamiliar threat configurations and develop responses that have coalesced into strategies, resilience baselines, and regulatory frameworks. Yet the instrument itself has not kept pace with the process it enabled: exercises remain oriented toward

* Sergiu Mitrescu is an associate expert at the New Strategy Center with a background in international relations and security studies. His research operates at the intersection of hybrid threats and maritime security, with a particular focus on critical maritime infrastructure and state adaptation. His practitioner work involves the design and delivery of tabletop exercises and scenario-based simulations. He currently leads the implementation of a NATO Science for Peace and Security project on hybrid threats to critical maritime infrastructure in the Black Sea region. He holds an MA in security studies from the University of Birmingham. Contact: sergiu.mitrescu@newstrategycenter.ro

scenario-based threat exposure and solution elicitation rather than testing whether the strategies that resulted hold up across conditions or what they produce downstream. This article proposes a reorientation of the tabletop exercise planning and analysis phases through the grammar of decision-making under deep uncertainty and robust decision-making. Drawing on morphological and heuristic instruments developed specifically for critical maritime infrastructure and the maritime domain, it operationalises this reorientation through the conventional three-phase TTX lifecycle — a plea for planners, not players. In doing so it proposes shifting the onus from testing the scenario to testing the strategy, and from surfacing responses to surfacing the conditions under which strategies fail and the consequences they generate. Critical maritime infrastructure serves as the paradigmatic case throughout.

Keywords: tabletop exercises, critical maritime infrastructure, decision-making under deep uncertainty, robust decision-making, gray zone, hybrid threats, strategy testing

Acknowledgement: This article was partly developed in the context of NATO Science for Peace and Security SAFE-INF project (Grant 8495) dedicated to the creation of an operating concept for the protection of critical maritime infrastructure against hybrid threats in the Black Sea through exercising.

Introduction

Hybrid threats occupy a central place in contemporary national security strategy, public discourse, and resource allocation. As a result, a substantial adaptation process is under way across NATO and EU member states, spanning capability development, legal frameworks, societal preparedness, education, and doctrine, within which exercising has been recognised as 'no longer optional but a strategic imperative' (Hybrid

CoE, 2026). Tabletop exercises (TTX) have been central to this effort, serving as the principal instrument through which practitioners familiarise themselves with unfamiliar threats, bridge institutional silos, and develop responses under simulated conditions. An expanding ecosystem of centres of excellence, research projects, and regulatory requirements, examined in detail in section 3, attests to their importance.

The measures that have emerged from this adaptive process, together with insights drawn from academic research, operational experience, and on-the-ground observation, have naturally coalesced into strategies: resilience baselines, legislative frameworks, national policies, and operational concepts. Yet a paradox has emerged within this otherwise productive process. The instrument that proved so effective at familiarising practitioners and eliciting responses is not being used to test the strategies that resulted from it. TTXs remain oriented toward scenario-based threat exposure and solution elicitation instead of branching out towards evaluating whether what was developed holds up across different conditions, or what it produces downstream.

This matters because gray zone competition generates unintended consequences structurally, not accidentally. In a tightly coupled competitive space characterised by manufactured ambiguity, multi-domain pressure, and adversarial adaptation, responses alter the conditions they act upon. A measure designed to protect critical maritime infrastructure may shift commercial incentives, trigger jurisdictional friction, or signal escalatory intent. If strategising in this environment is inherently littered with such consequences, then contending with them is not optional but a professional obligation — and the adaptive instruments through which practitioners prepare must be structured to surface them.



This article proposes a reorientation of the tabletop exercise through the grammar of decision-making under deep uncertainty (DMDU), and specifically robust decision-making (RDM). It is a plea for planners, not players: the nature of play remains unchanged, with practitioners receiving the same kind of vignette, and responding as is customary within such settings. What changes is what happens before and after they arrive, how exercises are structured, what they are designed to discover, and how their outputs are analysed. In doing so, the framework supports the decision-making process, the very capability that gray zone competition is designed to degrade. Critical maritime infrastructure serves as the paradigmatic case throughout, for reasons of both analytical salience and the availability of morphological and heuristic frameworks developed specifically for the domain (Hofmann et al., 2026; Albrecht and Harmsen, 2025).

The article proceeds as follows. Section 2 situates the gray zone as the competitive space within which adaptation takes place and positions critical maritime infrastructure as the domain where its dynamics are most visible. Section 3 introduces the tabletop exercise, examines what it achieves, surveys the expanding ecosystem around CMI, and identifies the limitations that call for a reorientation. Section 4 presents DMDU and RDM, establishes their qualitative foundations, and operationalises them through the three-phase TTX lifecycle using morphological and heuristic instruments developed specifically for the maritime domain. Section 5 discusses the pedagogical contribution, the diffusion potential, the generalisability of the approach, and its limitations.

The gray zone and critical maritime infrastructure

Three concepts are often used interchangeably, though they capture different things. Hybrid threats are the instruments. Hybrid warfare is the mode of their employment. The gray zone is the strategic space in which both operate — 'the space of competition short of war' (Mumford and Carlucci, 2022, p. 197). It is within this space that decision-makers and planners are compelled to campaign — contending with manufactured ambiguity, multi-domain pressure, and consequences that flow from their own responses as much as from the threat itself (Mitrescu, 2025b). The instruments deployed within it, regardless of actor, tool, or domain, converge on a single target: the decision-making capability of the adversary (Giannopoulos, Smith and Theocharidou, 2021) (Fig. 1).



Figure 1. The hybrid threat landscape (adapted from Giannopoulos, Smith and Theocharidou, 2021).

Critical infrastructure, and particularly critical maritime infrastructure, has emerged as a salient in gray zone competition. Infrastructure is more than the service it provides. Recent scholarship (Bueger, Liebetrau and Stockbrügger, 2023) identifies three dimensions: the material — the functions infrastructure fulfils and the structural interdependencies it creates; the agential — its capacity to constrain and enable socio-political behaviour, to embody ambitions, to hold multiple identities simultaneously; and the ideational — the place it occupies in national self-perception (Mitrescu, 2025a). It is not solely the piece of infrastructure that is critical but the convergence of these properties, stemming from an axial relationship between material and immaterial factors evaluated in context.

For CMI, the maritime domain carries its own legal, praxological, and geographical characteristics. UNCLOS governs a jurisdictional patchwork — EEZs are neither territorial seas nor high seas, sovereign rights over resources coexist with navigational freedoms, and the resulting ambiguity maps onto the strategic ambiguity of the gray zone itself. This legal patchwork has a direct alliance dimension: infrastructure in the EEZ falls outside the territorial scope of NATO's Article 5, even when the installations are national territory. Offshore installations sit far from shore-based response capabilities, in spaces where responsibilities are distributed across navies, coast guards, civilian regulators, and private operators. The maritime domain thus offers ample opportunities for (im)plausible deniability (Sukin & Hedgehock, 2025).

Bueger and Liebetrau (2023) map the threat spectrum to CMI along a continuum from natural disasters and accidents through blue crimes and terrorism to gray zone activities and acts of war (Fig. 2). The study of CMI protection has historically moved along this spectrum.

Initial concerns centred on resilience against the natural environment, seeking the address the environmental conditions within which offshore infrastructure operates. During the 1990s and 2000s, attention shifted to man-made threats: piracy and maritime terrorism. The gray zone end of the spectrum came to the forefront after 2014, as awareness of hybrid threats grew in the wake of Crimea, and especially after the Nord Stream sabotage in 2022, which gave birth to a prolific field of study at the intersection of CMI and hybrid threats.

The protection of CMI brings its own dilemmas, with measures intended to safeguard critical maritime infrastructure inevitably generating their own vulnerabilities. Examining CMI protection after Nord Stream, Neal (2026) identifies four persistent dilemmas in which protective responses paradoxically produce new risks: charting infrastructure locations for maritime safety exposes them to targeting; enhanced surveillance blurs civil-military boundaries and risks escalation; amplified risk perceptions deter the investment on which the infrastructure depends; and securitising the threat advertises which assets are most valued.

The characteristics of CMI and the characteristics of the maritime domain thus coalesce to form a paradigmatic case in which both the threat and the adaptive response generate consequences that resist linear assessment. It is this complexity — where planning must contend simultaneously with what the adversary does, what the response produces, and the interaction between the two — that has driven the development of a dedicated exercising ecosystem. Section 3 examines that ecosystem, what it has achieved, and where it reaches its limits.

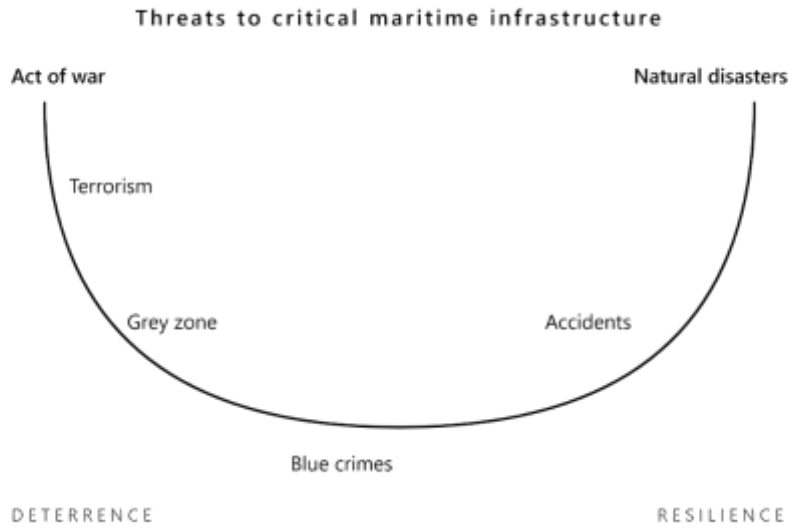


Figure 2. Threats to critical maritime infrastructure (reproduced from Bueger & Liebetrau, 2023)

Adaptation through exercising

Tabletop exercises have been a cornerstone of hybrid threat preparedness after 2014 with an emerging ecosystem around CMI after the Nord Stream sabotage in 2022. These exercises enabled practitioners to contend with unfamiliar threats and eliciting the measures that naturally make their way into strategies. Yet as the field has advanced, from initial familiarisation to legislation, resilience baselines, and national policies, the instrument itself has not advanced with it. TTXs remain oriented toward the same expert elicitation that justified their early use, even as the adaptive process they enabled has moved beyond it. This section introduces the tabletop exercise as a format, examines what it does well, surveys the expanding ecosystem around critical maritime infrastructure, and identifies the limitations that call for a reorientation.

An emerging ecosystem

A sizable exercise and research ecosystem has developed around the issue, with institutions approaching the problem of critical maritime infrastructure from their respective angles. Centres of excellence have expanded or reoriented their portfolios to include it — from energy security (ENSEC CoE, Vilnius) to hybrid threat preparedness (EU Hybrid CoE, Helsinki), maritime and asymmetric threat perspectives (MARSEC CoE, Istanbul; COE-DAT, Ankara), and engineering and physical security (DLR Institute for the Protection of Maritime Infrastructures), to name a few. Alongside these, NATO- and EU-funded research projects featuring tabletop exercises as part of their methodological design address specific dimensions of CMI. These institutional efforts are of course complemented by regional and national initiatives, in ever increasing numbers. For example, ten Baltic Sea littoral states and the European Union signed the Memorandum of Understanding on the Protection of Critical Undersea Infrastructure at Vihula in May 2025, while Romania, Bulgaria, and Türkiye extended the mandate of their Black Sea Mine Countermeasures Task Group to include critical undersea infrastructure protection at the 2026 NATO Summit in Ankara.

A considerable volume of exercising also takes place at the national level, driven by countries whose energy and communications systems depend on undersea infrastructure. There is, in short, an already well-defined TTX ecosystem operating at the NATO and EU level, one that will only expand further as the CER Directive's resilience requirements, whose full enforcement begins in 2027, oblige critical entities to conduct crisis simulations and stress tests as part of their compliance oblige critical entities to conduct crisis simulations and stress tests as part of their compliance obligations. Yet

even within this mature and expanding ecosystem, the prevailing exercise paradigm remains oriented toward familiarisation, creative problem-solving, and practitioner networking, strengths whose limits the following section examines.

Tabletop Exercises

Wargaming in its modern form originated in Prussia in the 1820s, when the von Reiswitz father and son developed a set of instructions for the representation of tactical manoeuvres under the guise of a *Kriegsspiel*. By the interwar period it had become standard practice. In a 1950 speech to the Naval War College, Admiral Nimitz observed: 'The war with Japan had been re-enacted in the game room here by so many people and in so many different ways that nothing that happened during the war was a surprise, absolutely nothing except the Kamikaze tactics towards the end of the war; we had not visualized those'(Development, Concepts and Doctrine Centre, 2017, p. 4).

The tabletop exercise is a direct descendant of this tradition. The Homeland Security Exercise and Evaluation Program (HSEEP) defines it as 'a discussion-based exercise in response to a scenario, intended to generate a dialogue of various issues to facilitate a conceptual understanding, identify strengths and areas for improvement, and/or achieve changes in perceptions about plans, policies, or procedures.' In practice, a scenario is presented at a simulated time, participants apply their knowledge to a list of problems presented by a facilitator, and resolution is reached through group discussion. The environment is deliberately no-fault — participants are encouraged to contribute without consequence, exploring decisions rather than performing them. The exercise cycle (Fig. X) situates this within a broader preparedness architecture

of design, conduct, evaluation, and improvement planning.

'Exercises play a vital role in preparedness. A well-designed exercise provides a low-risk environment to familiarize personnel with roles and responsibilities; foster meaningful interaction and communication across jurisdictions/organizations; assess and validate plans, policies, procedures, and capabilities; and identify gaps in resources needed to improve capabilities' (U.S. Department of Homeland Security, 2020, p. v).



Figure 3. Exercise Cycle (adapted from the Homeland Security Exercise and Evaluation Program, 2020)

Where they succeed

The core benefit of TTXs is that they are praxeologically embedded within current strategic and operational planning — they are, in other words, the instrument of choice. Their contextual value lies precisely in their

familiarity. In Jervis's (1976) words, they are an evoked set, concept or tool so embedded in professional practice that it is the default frame through which practitioners approach a problem, without deliberation over whether to use it. Bakken, Hærem and Lund-Kordahl (2025) observe that exercise participants can generate mental models that would not be attainable in a real-life learning environment. Hybrid threats, they note, often represent small deviations from the normal situation that are difficult to detect, with the most demanding challenge being the projection of future developments subject to both own and adversary actions. Central to this is 'the ability to discover, attribute, and understand whether one is in fact subject to an attack' (Bakken et al., 2025, p. 234).

Second, TTXs create an auspicious environment for creativity through their no-fault nature. They are definitionally simulative, unlike operational reality, where avoiding faults is the primary objective and failure is unacceptable, TTXs permit and even encourage exercising to failure. Andreassen, Elvegård, Villanger and Johnsen (2025) draw attention to this space but also to its limits: exercise evaluation is often unsystematic, top-down, and focused on performance rather than learning, in many cases limited to symbolic value. When established procedures prove inadequate during play, the observed default is 'more of the same' rather than reconsidering the assumptions behind them. Exercise are often plagued by an absence of criteria for what constitutes a good solution.

Third, TTXs are intrinsically pedagogical and ludic, a setting of hands-on learning infused with a degree of ludus, particularly conducive to creative thinking and a central element of adaptation, since adaptation means learning. Vassileva and Zwilling (2018) argue that hybrid warfare demands a cognitive rather than material

approach to education: developing not what to think but how to think. Participants should be 'trained for certainty and educated for uncertainty' (p. 223). Magnussen, Torgersen, Boe and Saeverot (2025) ponder the development of competencies for hybrid threats, that by definition, are not always yet directly known or visible. They argue that preparedness requires 'bricolage' — the capacity to solve unforeseen challenges with resources available at the time — and that hybrid threats demand relational models built around the general characteristics of the challenges to be solved rather than specific predetermined content.

In his article aptly titled ‘Life in the “Gray Zone”: observations for contemporary strategists’ Wirtz (2017, p. 112) concludes that 'organisations need to adapt so that planning, budgeting, and procurement timelines begin to correspond to unfolding events.' These contributions demonstrate that the pedagogical community has identified both what hybrid threat exercises should develop in participants and why organisations need to evolve. What they have not addressed is how the content of exercises, the scenarios, the threat configurations, the evaluation criteria, should be structured to enable systematic exploration of the conditions under which strategies fail.

Their limitations

The ecosystem identified above, together with academic research, operational observation, and wider developments in the study of hybrid threats, is advancing a substantial body of knowledge on the protection and resilience of critical maritime infrastructure. Strategies are maturing. The CER Directive, NATO resilience baselines, national hybrid threat policies, specialist guidance, these are further moves in what is essentially a chess opening, each

CC BY NC ND

building on the familiarisation and solution elicitation that the initial exercises enabled.

Yet the field has reached a point where the task is no longer familiarisation but rather a form of strategising, an endeavour that is inherently littered with unintended consequences (Freedman, 2013).

Going this far into an adaptive process — and with the benefit of medium-term hindsight in a field that barely existed before 2014 — deontologically compels planners to contend with the unintended consequences this adaptation produces. The question is how to do so with the resources already embedded in the adaptive process. The answer points back to the tabletop exercise. Its institutional trust, procedural familiarity, established resource base, and practitioner buy-in make it the obvious candidate — an evoked set that any new instrument would find hard to replicate. In this sense, there is a compelling case for reconsidering its function.

Once one does reconsider, certain intrinsic limitations come into view. First, the unit of analysis is the scenario, not the strategy. The prevailing paradigm is scenario-based risk ranking: the perceived most likely threat is presented, a response is generated, performance is evaluated. The exercise asks what ought to be done in this situation rather than whether what has already been decided holds up across different conditions, or what it produces downstream.

Second, the scenario is treated as the endpoint rather than a probe. Practitioners respond to it, are evaluated against it, and leave. The exercise never asks: what would have happened under adjacent conditions? The threat configuration is fixed for the duration of play, and so the strategy is tested against one reading of the environment rather than against the range of readings the environment might produce.

Third, exercises are episodic. Each is self-contained: designed, played, and evaluated on its own terms. There is no structured relationship between exercises whereby the findings of one inform the design of the next. The learning is additive at best, not iterative.

Fourth, evaluation is oriented toward performance rather than discovery. Andreassen et al. (2025) observe that exercise evaluation is often unsystematic, top-down, and focused on whether practitioners performed well rather than on surfacing the assumptions their performance rested on. What comes out of the exercise is what practitioners decided. What is systematically ignored is what they assumed, the dependencies their decisions required that were never surfaced and tested.

Decision Making Under Deep Uncertainty (DMDU)

> *'As we know, there are known knowns — these are things we know we know. We also know there are known unknowns — that is to say we know there are some things we do not know; but there are also unknown unknowns — the ones we don't know we don't know…. It is the latter category that tends to be the difficult one.'* (Rumsfeld, 2002)

In a broad sense, uncertainty may be defined as limited knowledge about future, past, or current events. This uncertainty involves subjectivity, since it relates to satisfaction with existing knowledge, which is coloured by the underlying values and perspectives of the assessor (Marchau et al., 2019). Decision-making under deep uncertainty (DMDU) concerns the subset of problems where this dissatisfaction is profound: parties to a decision cannot agree on the appropriate models to describe interactions among a system's variables, the probability distributions to represent uncertainty about

key parameters, or how to value alternative outcomes (Lempert, 2019).

Marchau et al. (2019) formalise this intuition across four levels (Fig. X). At Level 1, the future is clear enough to support point predictions. At Level 2, alternate futures can be assigned probabilities. At Level 3, a bounded set of plausible futures can be identified but not ranked probabilistically. At Level 4, the set of possible futures is itself unbounded and unknown — what is known is that one does not know. Beyond these lies total ignorance.

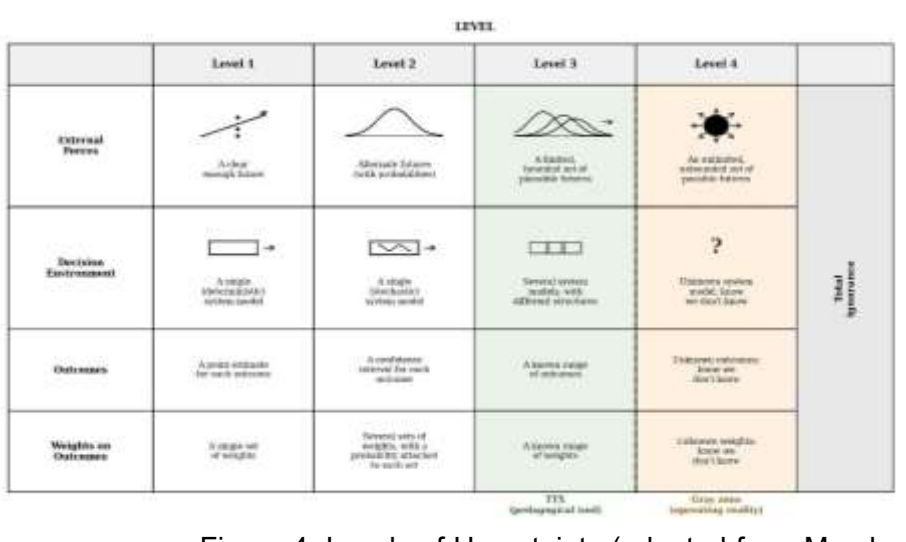


Figure 4. Levels of Uncertainty (adapted from Marchau et al., 2019)

The gray zone operates firmly at Level 4 because ambiguity is manufactured, not residual, because the external conditions facing decision-makers are designed to resist the very quantification that lower levels of uncertainty permit. How threats produce their effects is also contested with no single causal model commands consensus. The metrics by which outcomes should be evaluated are themselves subject to disagreement among stakeholders. And the set of possible futures is unbounded, because the adversary

actively reshapes it in response to every move. In other words, the gray zone is defined by a generalised deep uncertainty.

Decision-making under deep uncertainty can be understood through a general framework for decision support (Marchau et al., 2019). At its core is a decision environment, a system comprising physical and human elements and their mutual interactions, that decision-makers seek to influence through policies in order to achieve desired goals. The results of these interactions are the outcomes of interest: the criteria against which policies are evaluated. The valuation of outcomes reflects the weights, not necessarily quantitative, assigned to them by relevant stakeholders, including decision-makers themselves. Alongside policies, external forces act upon the decision environment. These are developments outside the system that decision-makers cannot control but that may significantly influence system behaviour: technological, societal, economic, and political developments, among others. Both policies and external forces may affect the relationships among the system's elements and hence the structure of the decision environment itself, as well as the outcomes of interest.

Applying this framework to conditions of deep uncertainty reveals several locations where uncertainties arise. First, there is scenario uncertainty: it is difficult to identify which external developments will prove relevant for long-term system performance, let alone their magnitude and direction. Second, even if external developments were known, there would remain structural uncertainty about how the decision environment responds to them. The causal mechanisms determining system performance may be insufficiently understood, or competing hypotheses of causation may coexist. Finally, the valuation of outcomes is itself

CC BY NC ND

contested. Stakeholders differ on the importance of future problems and the objectives to be achieved; over time, new stakeholders may emerge, existing ones may leave, and preferences may shift, all of which affect the evaluation of outcomes and the viability of policies.

'Integral to DMDU approaches is the use of analytic methods for decision support [...] a set of processes intended to create the conditions for the production and appropriate use of decision-relevant information' (National Research Council, 2009, cited in Marchau et al., 2019). What distinguishes DMDU from conventional decision support is that these methods do not seek to reduce uncertainty to a manageable prediction. They seek to identify choices that perform adequately across the range of futures that cannot be predicted.

Robust Decision Making (RDM)

The DMDU literature encompasses a wide family of approaches for structuring decisions under deep uncertainty. The present article focuses on Robust Decision Making (RDM) for two reasons. First, its central criterion, robustness, understood as performing adequately across many plausible futures rather than optimally in one predicted future, is particularly relevant for how strategies against gray zone pressure should be evaluated. Second, as will become apparent, RDM's origins in stress testing, assumption-based planning, and iterative scenario discovery share a deep structural kinship with the logic of simulations.

The departure from conventional decision analysis is fundamental. Where traditional approaches 'begin by seeking agreement regarding the likelihood of future states of the world and then use this agreement to provide a prescriptive ranking of policy alternatives' (Lempert, 2019, p. 27), RDM inverts these steps. It

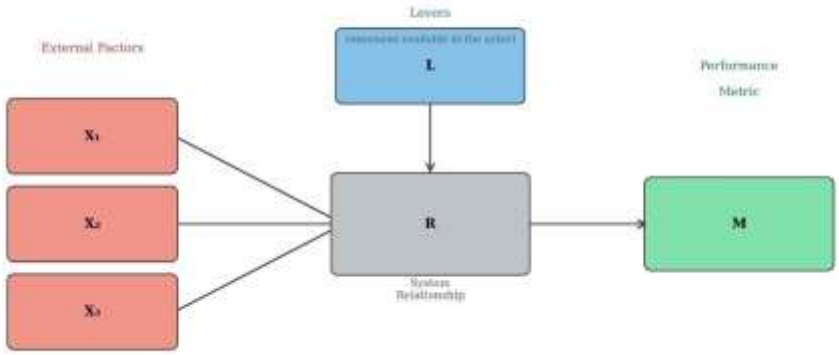
begins with one or more strategies under consideration, uses models and data to stress-test those strategies over a wide range of plausible futures, and then uses the results to identify vulnerabilities and, where possible, to improve the strategies. The criterion is robustness, performing adequately across many futures, rather than optimality in one predicted future.

As part of an 'agree-on-decisions' approach, RDM draws on the related concepts of stress testing and red teaming (Lempert, 2019). The latter, often associated with military planning), involves forming an independent group to identify means to defeat an organisation's plans. Both aim to reduce the deleterious effects of overconfidence in existing systems and plans by improving understanding of how and why they may fail (Lempert, 2007 in Lempert, 2019, p. 27). RDM also draws from scenario analysis the concept of organising information about the future into a small number of distinct cases that help people engage with, explore, and communicate deep uncertainty. Scenarios are often developed and used in deliberative processes with stakeholders. Deemphasising probabilistic ranking, focusing on a sense of possibility rather than probability, helps stakeholders expand the range of futures they consider, allowing them to contemplate choices from a wider range of vantages, including futures they find uncomfortable or dissonant (Lempert, 2019, p. 28).

It is convenient to group the dimensions to be explored in RDM analysis into four classes, the XLRM framework (Fig. 5), where X represents uncertain factors outside the control of the decision actor, L represents the levers for action — the measures available to the actor that may be combined into candidate strategies or courses of action, R allows for different specifications of the causal relationships through which different strategies (L) operating in alternative states of the world (X) yield

CC BY NC ND

measurable outcomes, and M represents the metrics to be used in evaluating the nature of those outcomes



(Lempert et al., 2003).

Figure 5. XLRM Framework for decision support under deep uncertainty (adapted from Lempert et al., 2006)

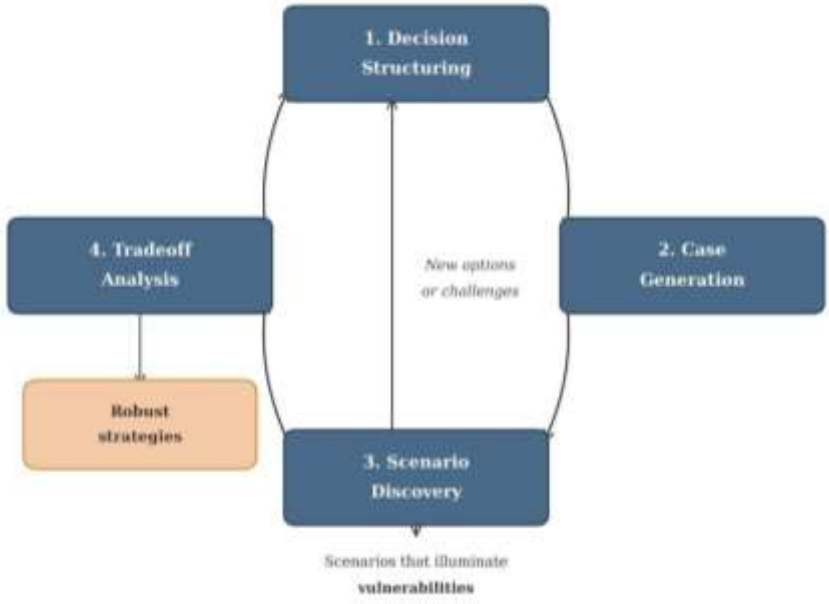


Figure 6. Ideal RDM analytical cycle (adapted from Popper, 2019)

CC BY NC ND

The RDM process is iterative (Fig. 6). Decision structuring defines the problem dimensions. Case generation produces the ensemble of outcomes. Scenario discovery identifies the vulnerability conditions. Tradeoff analysis evaluates whether and how strategies can be modified to address those vulnerabilities. The cycle then returns to decision structuring as new options or challenges emerge from the analysis.

'RDM analyses have the virtue of placing the decision-maker explicitly in the modelling framework and so the outputs of such studies are framed precisely in the terms that confront the decision-makers being supported' (Popper, 2019, p. 3). They do not provide greater reliability in predicting futures. Rather, they allow explicit definition of what assumptions one would need to believe to recommend one course of action over others.

The qualitative foundations of RDM

RDM as described above is intended for computational application — simulation models, ensemble runs, algorithmic scenario discovery. But the logic that underpins it — stress testing strategies against a structured uncertainty space, identifying the conditions under which they fail, and iterating — is separable from the computational apparatus through which it is conventionally delivered.

Jonathan Rosenhead's *Rational Analysis for a Problematic World* (1989) is widely cited as a foundational text for robust decision frameworks. Rosenhead introduced systematic ways to evaluate strategic choices based on keeping future options open under deep uncertainty, relying on qualitative structuring rather than complex computer models. The criterion was not prediction but flexibility: a good decision was one

CC BY NC ND

that preserved the decision-maker's capacity to adapt as uncertainty was resolved. This logic, robustness over optimality focused on structuring over modelling, predates the computational turn in DMDU by over a decade. RDM's orientation toward long-term consequences finds a natural fit in the gray zone when viewed through the campaign prism (Monaghan & McDonald, 2023): threats are not momentary instances of disequilibrium but elements of a sustained, systematic competitive effort whose effects accumulate and interact over time.

Marchau et al. (2019) themselves acknowledge this lineage. Their treatment of DMDU approaches recognises that expert-based methods constitute a legitimate mode of analysis alongside computational ones. The distinction is not between rigorous and informal but between different means of achieving the same end: the systematic exploration of how strategies perform across conditions that cannot be predicted. Where computational RDM achieves this through simulation ensembles, expert-based approaches achieve it through structured deliberation, eliciting assumptions, testing them against alternative framings, and identifying the conditions under which they fail.

Popper (2019) provides the most direct demonstration. Working on a U.S. Army security cooperation planning problem, he applies RDM to a domain that lacks formal models, sufficient scientific foundations, or even the formalisms for framing relationships in terms amenable to quantitative analysis. In the absence of formal models, the analysis tests candidate actions against different explicit statements of causal relationships to allow more systematic reasoning over choices and outcomes. It results in a different instantiation of the same logic that makes assumptions explicit and stress-tests strategies against alternative theory of causation

and where scenario discovery identifies the conditions under which strategies fail.

'There has been a persistent problem in applying the power of RDM in the absence of formal models' (Popper, 2019, p. 4). The paper's contribution is to show that this problem is soluble. The theories of causation that practitioners carry, implicitly, informally, but in considerable detail, can themselves serve as the system model when made explicit and testable. 'Robust decision making is designed to provide more purposeful use of models to support decisions taken in the face of deep uncertainty. But the method appears to be limited to those cases for which formal models either exist or may be constructed from existing knowledge' (Popper, 2019, p. 4). His analysis demonstrates that this apparent limitation is not inherent to the method but to the assumption that models must be computational.

'Importantly, RDM is conceived as a human-in-the-loop system, using advanced computing techniques not in pursuit of artificial intelligence but rather to create decision support tools for observation of phenomena and connections otherwise rarely visible to planning teams' (Popper, 2019, p. 2). Workman (2026) captures precisely the tension between understanding the complexity of the operating environment and acting on it — a tension that, as the following section will argue, is first and foremost a qualitative issue. Examining UK defence strategic planning, he identifies a persistent discrepancy: strategic documents exhibit a deep understanding of the complex, non-linear and deeply uncertain operating environment, yet the decision-making systems through which strategies are developed remain anchored in linear, predict-and-act approaches. This, as the following section argues, is first and foremost a qualitative problem — one that a well-

designed tabletop exercise, structured through DMDU grammar is well positioned to address.

Infusing exercising with RDM

The starting point is the questions planners ask. A conventional exercise design asks: what happens if a subsea cable is disrupted? How do we respond? A planner working within an RDM logic asks different questions: under what combinations of threat characteristics does our response to cable disruption stall? At what point does our protection framework generate jurisdictional friction that delays the very coordination it was designed to enable? Which assumptions about attribution speed, legal authority, or commercial operator behaviour need to hold for our strategy to work — and what happens when they don't? The shift is from testing the scenario to testing the strategy, and from surfacing responses to surfacing the conditions — including the consequences of the response itself — under which the strategy fails.

Exercising brings an intrinsic conceptual tension in that it is a Level 3, scenario-based approach contending with a Level 4 phenomenon. First and foremost, any exercise is a pedagogical compromise: it needs to hold a finite and identifiable future for the duration of play to provide players with a learning context. For the discussion at hand, given that exercises are envisaged here as pedagogical tools that also enable structured strategy testing, it is also a question of falsifiability — a strategy can only be tested against conditions that can be specified. Whether the slices, taken iteratively across exercises, accumulate meaningful coverage of the space depends on whether each slice is structurally distinct from those that preceded it.

Two recently developed analytical instruments provide the morphological and heuristic foundations for structuring the XLRM framework in the domain of critical maritime infrastructure. Hofmann, Stockbrügger and Sill Torres (2026) provide a morphological matrix (see table 1) that deconstructs hybrid threat scenarios across seven independent dimensions: strategic goal, operational goal, target, means of impact, degree of state control, temporal structure, and ambiguity as a cross-cutting dimension (see table below). Within the XLRM matrix, it constitutes part of the X, the uncertainty. Rather than writing a single plausible narrative, planners systematically select configurations across these dimensions, producing structurally distinct threat scenarios that test different aspects of the strategy. The morphological approach is modular: the same characteristic — say, high ambiguity combined with state-tolerated control — can manifest through many different means of impact, and conversely, the same means of impact can carry very different strategic intentions. It is this modularity that enables characteristic-based play rather than narrative-bound play.

The paper also develops a four-phase model of conflict escalation — preparation and grey zone activity, active hybrid confrontation, open military force, and high-intensity war — where each phase is defined as a characteristic configuration of the same dimensions, particularly ambiguity, degree of state control, and temporal structure.

Table 1. Morphological dimensions of hybrid threats to CMI (reproduced from Hofmann, Stockbrügger and Sill Torres, 2026).

Dimension	**Description**
Strategic goal	The overarching political objective: intelligence gathering, testing, signalling, destabilisation, deterrence, or escalation
Operational goal	The direct technical or physical objective: reconnaissance, manipulation, disruption, damage, sabotage, or destruction
Target	The affected infrastructure: energy, information and sensor systems, maritime transport, or military-relevant
Means of Impact	The instruments used: ranging from civilian vessels and small craft to unmanned systems, cyber operations, navigation disruption, and kinetic means
Degree of state control	The actor's relationship to the state: directly state-led, state-controlled, state-tolerated, or autonomous
Temporal structure	The chronological pattern: singular, episodic, or systematic
Ambiguity (cross-cutting)	The degree of attributable responsibility: from very high to low — shaping the interpretation of all other dimensions

Albrecht and Harmsen (2025) provide the second part of X — the information environment in which those threat characteristics are perceived, contested, and acted upon. Their heuristic builds on an integrative approach to security that moves beyond both purely objectivist readings of threats as measurable facts, and

purely constructivist ones where threats are seen as products of discourse and securitisation. Instead, it treats hybrid threat situations as dynamic processes in which the objective event and its subjective interpretation are inseparable: what happened and how it is perceived, by whom, and through which institutional lens shape each other reciprocally.

From this foundation, Albrecht and Harmsen derive six fields of analysis. How do threats emerge as side effects of maritime modernisation and technological dependence? Which aspects exceed the scope of traditional risk analysis? How is security co-produced across state and non-state actors — and how is insecurity co-produced alongside it? Whose expertise is considered legitimate, and what interpretive conflicts arise? What institutional routines have developed to manage uncertainty, and do they help or create blind spots? And which decision-making logics — linear planning or adaptive learning — shape the behaviour of the actors involved? Overlaid onto Hofmann et al.' morphological dimensions, these fields produce the information friction that practitioners actually experience in the gray zone: diffuse attribution, contested expertise, jurisdictional ambiguity, and the possibility that the response itself is altering the conditions it was designed to address.

The uncertainty space (X) is the totality of conditions under which the strategy plausibly has to operate. To construct the Level 3 slices required by the exercise, the planning team needs a way to decompose that space into dimensions that can be systematically recombined, and a way to layer onto each combination the epistemic conditions that make the gray zone distinct from conventional threat environments. The morphological matrix (Hofmann, Stockbrügger and Sill Torres, 2026) provides the first: seven dimensions of hybrid threat

activity whose combinations generate structurally distinct configurations rather than ad hoc scenarios. The six-field heuristic (Albrecht and Harmsen, 2025) provides the second: the contested expertise, jurisdictional ambiguity, and diffuse attribution under which practitioners must respond. Given they were both were developed specifically for critical maritime infrastructure and the maritime domain, they are adopted here to provide the analytical grammar through which the planning team probes the uncertainty space.

Table 2. Heuristic fields for analysing hybrid threats in the maritime domain (reproduced from Albrecht and Harmsen, 2025).

Field of Analysis	**Key questions**
(1) Risk development through modernization	How do hybrid threats emerge as side effects of maritime globalization and/or technology development?
(2) Uncertainty and not knowing	Which aspects are beyond the scope of traditional risk analysis?
(3) Co-production of security	What role does the interpretation of the geopolitical situation play? Which state and non-state actors are involved in security management operations? How is responsibility shared?
(4) Change in expertise	Whose knowledge is considered legitimate for assessment? What conflicts are there in the interpretation?
(5) Institutionalization of uncertainty	What routines and standards emerge to deal with uncertainty?

(6)Reflexive rationalities	Which decision-making logics shape the behavior of actors: linear- planning or adaptive learning?

The exogenous uncertainties (X) represent the full range of conditions under which the strategy might have to operate — conditions the planning team does not define in advance but discovers iteratively during the exercise planning phase. Each exercise probes a bounded slice of this space through the morphological configurations and information environment conditions described above; which conditions prove consequential for the strategy is what the process reveals. The performance metrics (M) are the objectives of the strategy — what it is designed to achieve. Within M sit the minimum acceptable thresholds: the point below which the strategy is no longer meeting its objectives. Unlike linear, risk-based planning, which asks how well a strategy performs, the approach proposed here asks under what conditions it fails. The levers (L) are the strategy, blueprint, or operational concept under test — held constant so that the exercise can reveal where it is not robust. The system relationships (R) emerge from the interaction of L and M within a given slice of X. They are constructed by asking which dependencies L, evaluated against M, requires — and because they are contingent on the configuration being tested, the same strategy evaluated against the same objectives may surface different dependencies under different conditions. A first instantiation of R emerges from the pre-exercise phase as a hypothesis derived from L and M: a way of unearthing and articulating the assumptions planners carry about how the system will behave. The post-exercise phase tests that hypothesis: $R(T_1)$ captures which relationships held under play and which did not. The gap between $R(T_0)$ and $R(T_1)$ is the exercise's finding.

TTXs are conventionally structured in three phases, pre-exercise planning, live play, and post-exercise analysis, each serving a distinct function. This tripartite structure maps directly onto the RDM analytic cycle. During the pre-exercise phase, the planning team assembles the XLRM specification. The threat space (X) is structured through Hofmann's morphological matrix, with Albrecht's heuristic layered onto each configuration to produce the information environment practitioners will navigate. The strategy (L) is held constant. The metrics (M) are derived from the testable assumptions embedded in L. The system relationships $R(T_0)$ are constructed as a hypothesis, which dependencies the strategy, evaluated against those metrics, relies on. The result is a complete XLRM configuration, embedded within a realistic vignette that practitioners experience as a scenario rather than a matrix.

During the live play, practitioners enter a jargon-free environment that does not reference DMDU or RDM extensively to avoid distraction. They receive the vignette — a narrative that carries the characteristic-based configuration designed in the pre-exercise phase without exposing its morphological structure. Practitioners experience a scenario, not a matrix. They respond using L; their responses are recorded. The live play produces the raw material for analysis: the strategy is being tested against a specific configuration of uncertain conditions, and the friction points that emerge are the qualitative artefacts from which patterns will later be drawn.

During the post-exercise phase, the planning team maps the friction points recorded during live play back onto the morphological grid. The analysis asks two distinct questions. First: where did the strategy fail because the threat configuration fell outside what it was designed to address — a gap in coverage? Second:

where did the strategy fail because the response itself produced consequences that undermined it — jurisdictional friction, altered commercial incentives, delayed coordination, escalatory signaling? The first contends with scope, while the second one with effects. It is this second question that surfaces the unintended consequences identified in section 2 as structural features of gray zone competition. Together, these two questions produce $R(T_1)$ — revealing which of the relationships assumed in $R(T_0)$ held under play and which did not. The qualitative approach is valuable on its own — but it also paves the way for computational integration. Because the pre-exercise phase encodes every vignette as an XLRM-compatible configuration, the qualitative outputs of the post-exercise phase are structured data. Where resources permit, computational tools can batch-process the full morphological space, expanding coverage beyond what any sequence of exercises can explore on its own.

Pilot

This section presents a hypothetical worked example. An internationally funded project brings together four countries, two NATO and EU coastal states (Alpha and Bravo), a landlocked EU-aspirant dependent on transit infrastructure (Charlie), and a coastal state aligned but not integrated into alliance structures (Delta), to organise tabletop exercises on the protection of critical maritime infrastructure against hybrid threats in a shared maritime space.

The exercises are structured as five steps across the three phases of the exercise cycle, following the RDM analytical cycle (Popper, 2019), decision structuring, case generation, scenario discovery, and tradeoff analysis. The morphological analysis (Hofmann, Stockbrügger and Sill Torres, 2026) and six-field heuristic (Albrecht and Harmsen, 2025) introduced

CC BY NC ND

above provide the threat characterisation and information environment. Three additional works supply the remaining elements: a RAND study on undersea deterrence (Bohnert, Kochhar, Justice and DeLuca, 2024) provides the strategy under test — a decision matrix mapping attribution confidence against a reveal/conceal posture to determine available response actions; Gill et al., 2026 provide operationally realistic threat profiles for the maritime domain, from which the exercise configuration is drawn; and the Comprehensive Resilience Ecosystem (Jungwirth et al., 2023) provides the domain-level structure from which system dependencies are read.

Phase I — Pre-exercise

STEP 1 The planning team arrives with two elements already defined: the strategy under test (L) and the objectives against which it will be evaluated (M). For this pilot, L is the RAND decision matrix introduced above (reproduced below as Figure 7). The exercise holds this framework constant so that play can reveal where it is not robust. The strategy's objectives (M) are critical infrastructure deterrence, calibration of response without unnecessary escalation, management of the reveal/conceal trade-off, maintenance of attribution capacity sufficient to justify the chosen response, and coordination across the four states on which the undersea infrastructure depends. For each, the planning team defines minimum acceptable thresholds — the point below which the strategy is no longer meeting its objectives. Four illustrative thresholds are specified: decision latency, attribution confidence relative to the matrix threshold, action coverage, and escalation pathway. The uncertainty space (X) represents the full range of conditions under which the strategy might have to operate. The planning team does not define these conditions in advance but discovers them through expert elicitation during the exercise planning phase,

uncovering the gray zone dimensions relevant to the strategy that shape the operating environment. Each exercise probes a bounded slice of this space; which conditions prove consequential for the strategy is what the process reveals. The system relationships (R) — the dependencies that must hold for L to meet M — are derived once a specific slice of X is selected; they are the work of the next step.

Figure 7. Decision Support Matrix for detering threats

Decision Support Matrix

	Actions and Reactions						
Reveal and Conceal Choice	Nothing	Reveal	Passive Defensive	Active Defensive	Respond in Kind or Active Status Quo	Horizontally Escalate	Vertically Escalate
Nothing	None	N/A	N/A	N/A	N/A	N/A	N/A
Defensive and planning	None	None	None	None	Categorical	Inferred	Disputable
Unnotified but covert	Categorical	Categorical	Categorical	Categorical	Inferred	Disputable	Clear
Unacknowledged but militarily discoverable	Categorical	Inferred	Inferred	Inferred	Disputable	Clear	Self-identified
Privately acknowledged	Categorical	Inferred	Inferred	Disputable	Clear	Clear	Self-identified
Publicly discoverable	Inferred	Disputable	Disputable	Clear	Clear	Self-identified	Self-identified
Publicly announced	Disputable	Clear	Clear	Clear	Clear	Self-identified	Self-identified

against critical undersea infrastructure (reproduced from Bohnert et al., 2024)

STEP 2. Case generation selects a specific configuration from the uncertainty space discovered in Step 1. The morphological grid (Hofmann, Stockbrügger and Sill Torres, 2026) provides the seven dimensions — strategic goal, operational goal, target, means of impact, degree of state control, temporal structure, and ambiguity — that describe the structural properties of the event; its phase model of escalation defines characteristic configurations of those dimensions at successive levels of conflict intensity. The planning team selects an escalation phase, which fixes the ambient configuration — the prevailing level of ambiguity, the visibility of state involvement, and whether events appear isolated or patterned. For this pilot, the configuration is drawn

from the contemporary threat profiles documented in Gill et al., 2026: a civilian tanker flagged to Sun is detected dragging its anchor across subsea telecommunication cables in Alpha's exclusive economic zone. The cables connect Alpha and Bravo; Charlie and Delta depend on them for transit connectivity. The morphological dimensions read as follows:

Dimension	**Configuration**
Strategic goal	Testing and signalling
Operational goal	Disruption of critical infrastructure services
Target	Subsea cables
Means of impact	Civilian vessel
Degree of state control	Not assignable
Temporal structure	Singular event
Ambiguity	Very high

Albrecht and Harmsen's (2025) six-field heuristic is layered onto this configuration, producing the information frictions under which practitioners will navigate the scenario: contested expertise over whether the anchor was dragged deliberately, jurisdictional ambiguity over enforcement rights in the EEZ under UNCLOS, and diffuse attribution across the vessel operator, the flag state Sun, and any state that may have directed the act.

With L, M, and a specific slice of X in place, the planning team derives $R(T_0)$ — the pre-exercise hypothesis of which system dependencies must hold for the strategy to meet its objectives under these

conditions. The CORE model (Jungwirth et al., 2023) provides the structure from which $R(T_0)$ is read.

CORE model (Jungwirth et al., 2023) maps thirteen domains, including Intelligence, Military Defence, Economy, Infrastructure, Information, and Social/Societal, across three spaces (Governance, Services, Civic) and three layers (local, national, international). Its analytical value for $R(T_0)$ lies in the interconnection web between domains (reproduced below as Figure 8): when a threat enters through one domain, its effects cascade across spaces and layers through dependencies that may or may not hold under stress. The planning team traces the strategy's assumptions onto this web to identify where it is exposed. Three illustrative dependencies for the pilot configuration: Intelligence→Political — the matrix assumes that attribution intelligence produced within Alpha's and Bravo's intelligence domains reaches civilian decision-makers across all four states with sufficient speed and fidelity to support the reveal/conceal choice, but NATO and EU intelligence-sharing channels exclude Charlie and Delta to different degrees, so the four states enter the matrix at different levels of attribution confidence; Economy to Diplomacy, the matrix assumes that economic response actions can be coordinated across four states with different legal frameworks, different economic exposures to Sun, and different institutional memberships, within the timeframe the matrix implies; Information to Social/Societal, the matrix assumes that public communication can be managed in parallel with operational response across four publics with sharply different threat perceptions without one undermining the other. Because R is contingent on the slice of X being tested, the same strategy evaluated against the

same objectives may surface different dependencies under different conditions.

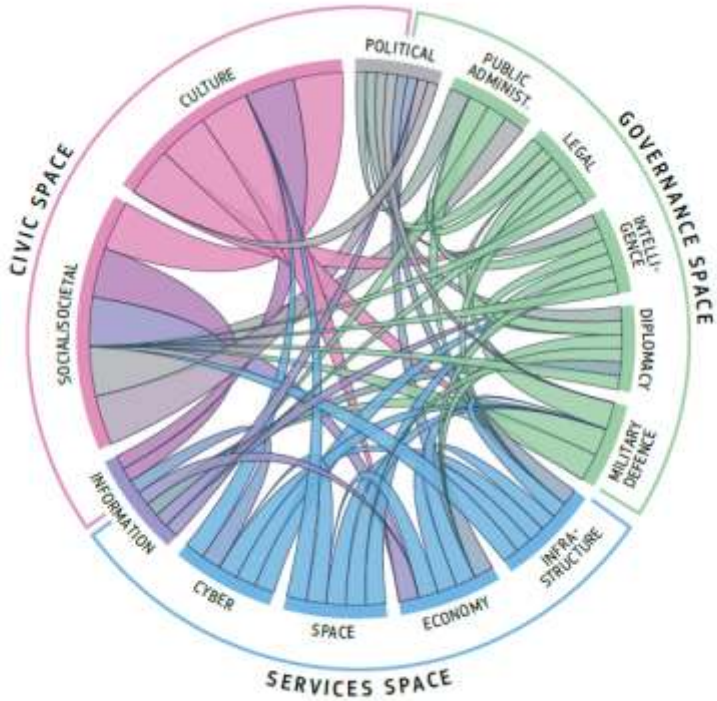


Figure 8. CORE model – connections between domains (reproduced from Jungwirth et al., 2023)

The narrative vignette built from this configuration is a practitioner-friendly wrapper for the underlying morphological structure — what participants experience as a scenario is, for the planning team, a systematic probe of the threat space. The crescendo across successive vignettes is a progression through escalation phases, where several dimensions shift together as the escalation phase model describes.

Phase II — Live play

STEP 3. During live play, syndicates receive the vignette. They respond to the unfolding situation using their own judgment, institutional procedures, and available expertise. Their responses are recorded. No pre-configured triggers are introduced — the vignette carries the morphological configuration and the information environment designed in the pre-exercise phase. The exercise runs on the interaction between practitioners and the conditions they face. Alpha and Bravo, with access to allied intelligence, reach high confidence? and move toward measures against Sun. Charlie, without that access, stays at low confidence and, dependent on the cables for transit, presses for de-escalation. Delta, economically exposed to Sun, declines to join economic measures.

Phase III — Post-exercise analysis

STEP 4. The vulnerability analysis compares $R(T_0)$ with $R(T_1)$. $R(T_1)$ is read from the recorded responses using the same CORE structure: which of the dependencies the strategy assumed held under play and which did not. The gap between the two is the exercise's finding. Under the configuration tested, the Intelligence→Political dependency may have held while Economy→Diplomacy broke — Charlie and Delta never reached the confidence the matrix requires to unlock the actions Alpha and Bravo took. Because R is contingent on the slice of X tested, running additional vignettes, drawing on different configurations from Step 2, may reveal different vulnerabilities in the same strategy. The main finding of this is that the matrix assumes a single defender with one confidence level. In a shared maritime space, the four states occupy different cells at the same time, and the strategy fails its coordination objective.

STEP 5. The tradeoff analysis turns from the dependencies to the objectives themselves. The

thresholds defined in Step 1 may pull against each other: maintaining attribution confidence at the level the matrix requires may demand revealing capabilities that undermine future deterrent capacity; the speed of response practitioners achieved may have come at the cost of allied coordination; public communication choices may have constrained operational flexibility. This step surfaces which tradeoffs practitioners made under pressure and whether the strategy's design anticipated them. Taken together, Steps 4 and 5 complete the analytical cycle: findings feed back into decision structuring — revising L, adjusting thresholds within M, or redefining which regions of X to probe in the next exercise iteration. In this example, Alpha could raise Charlie's and Delta's confidence by releasing intelligence, meeting the coordination objective but breaching the reveal/conceal objective. The matrix gives no guidance on this. Thus the feedback could be that a coalition dimension, or pre-agreed protocols for releasing intelligence to non-allied partners needs to be added.

Discussion

The framework proposed in this article enhances the planning phase of exercises that are, by their nature, a Level 3 pedagogical compromise contending with a Level 4 phenomenon. Any exercise must hold a finite and identifiable future for the duration of play to provide practitioners with a learning context, and since exercises are envisaged here as tools that also enable structured strategy testing, it is equally a question of falsifiability: a strategy can only be tested against conditions that can be specified. What the framework changes is the planning side. The planner is not simulating a specific future but testing properties of the strategy against characteristics that could manifest in many forms, structured through morphological analysis (Hofmann et al., 2026) and drawn, for the pilot in this

article, from pre desgiend scenarios (Gill et al., 2026). Whether the slices, taken iteratively across exercises, accumulate meaningful coverage of the space depends on whether each slice is structurally distinct from those that preceded it. For practitioners, the pedagogical gain is the realisation of unknown unknowns — the scenario becomes a vehicle, rather than a form of prediction. In doing so, the exercise supports decision-making — the very capability that hybrid threats are designed to undermine. What surfaces during play are the implicit assumptions practitioners carry. The post-exercise analysis details and examines them, because it is these assumptions, not the scenario, that are the object of analysis. Addressing them equips practitioners with a clearer picture of what shapes — and distorts — their judgment in the gray zone

In doing so, it embeds an iterative strategy-simulation feedback loop. A strategy is tested, vulnerabilities are discovered, the strategy is revised, and the revised strategy is tested again, against different morphological configurations, under different informational conditions. Each cycle refines the understanding of where the strategy holds and where it does not. In the post-exercise analysis, planners qualitatively identify alternative futures, combinations of conditions they had not considered, while simultaneously priming the simulation architecture for computational RDM. Because the pre-exercise phase encodes every vignette as an XLRM-compatible configuration, the qualitative outputs are structured data.

Structuring tabletop exercises through DMDU does more than improve individual exercises. It infuses the learning process with deep uncertainty thinking. Practitioners who experience characteristic-based play — who are asked not only what they would do but under what conditions their response fails and what it

produces downstream, carry that cognitive framework back to their institutions. The current exercise architecture emerged through learning and has been standardised across the NATO ecosystem. As with any adaptive strategy, standardisation can stifle advancement when it falls out of step with developments in the field, and as section 3 has argued, this mismatch has emerged. But what is currently a mismatch can become a finely tuned diffusion infrastructure: redesign one template and it propagates through existing institutional channels.

Hybrid threats are inherently multi-domain. The framework proposed in this article narrows the field productively by anchoring it in critical maritime infrastructure. CMI is a newly emerging field from the perspective of human-made threats, with a well-defined exercising ecosystem, regulatory momentum through the CER Directive, and a topic that is bounded legally and geographically by the maritime domain. Perhaps most importantly, the methodological building blocks are native to the case: Hofmann et al.'s morphological analysis and Albrecht and Harmsen's heuristic were specifically designed for critical maritime infrastructure and the maritime domain.

This does not mean the approach is confined to CMI. The XLRM grammar, the three phase TTX lifecycle, and the logic of characteristic-based scenario discovery are portable to any hybrid threat domain, cyber operations, information campaigns, political subversion, wherever strategies exist but have not been stress-tested for robustness across conditions. What changes is the content of X, L, and M, rather than the architecture. Nevertheless, the field of critical maritime infrastructure protection is where the tools, the ecosystem, the regulatory momentum, and the author's empirical

exposure converge, making it the natural domain in which to test what this article proposes.

The framework proposed in this article carries limitations that should be stated plainly. Qualitative scenario discovery is lower-resolution than computational: the coverage of the threat space is limited to what is played, and patterns emerge across exercises rather than from a single run. The method depends on the quality of the morphological matrix, the heuristic overlay, and the planning team's judgment — inputs that are themselves subject to bias, blind spots, and institutional constraints. Most fundamentally, this article proposes a framework that it does not validate empirically. The next step is empirical testing — and the exercising ecosystem described in section 3, with its institutional infrastructure, trained practitioners, and expanding regulatory mandate, provides the setting in which that testing can and should take place.[1]

Conclusion:

Adaptation to gray zone threats in the domain of critical maritime infrastructure is well under way. A sizeable exercising ecosystem has emerged, strategies have coalesced from the measures it produced, and regulatory requirements will expand this further. Yet the instrument that proved so central to developing these strategies — the tabletop exercise — has not been restructured to test them. It remains oriented toward familiarisation and solution elicitation, leaving unexamined the question of whether strategies hold across conditions and what they produce downstream.

This article has proposed a reorientation of the TTX planning and analysis phases through the grammar of

[1] The author thanks Mark Workman and Tom Smith for their contributions to the development of the pilot exercise

CC BY NC ND

decision-making under deep uncertainty and robust decision-making. The shift is from scenario-based risk ranking to characteristic-based strategy testing — from asking what we should do to asking under what conditions what we decided to do fails, and what consequences it generates. The tools for doing so are already embedded in the adaptive process and in the dedicated analytical work that has emerged at the intersection of CMI and hybrid threats. The article's contribution resides in the grammar that connects them. The proposed approach awaits empirical testing within the existing exercising ecosystem.

AI-use disclosure statement

The author used Claude Opus 4.6 for drafting support and language refinement during manuscript preparation. All arguments and final editing decisions were made and checked by the author, who takes full responsibility for the integrity and accuracy of the manuscript and its references.

Literature:

1. Albrecht, L. and Harmsen, T. (2025) 'Navigating Diffuse Maritime Information Settings: A Heuristic for Improved Analysis of Hybrid Threats', 5th European Workshop on Maritime Systems Resilience and Security, Rostock, 21 May 2025. doi: 10.5281/zenodo.17120684.
2. Andreassen, N., Elvegård, R., Villanger, R. and Johnsen, B.H. (2025) 'Enhancing cognitive motivation: an evaluation model for emergency preparedness exercises', *The Learning Organization*, 32(7), pp. 35–52. doi: 10.1108/TLO-06-2023-0100.
3. Bakken, B.T., Hærem, T. and Lund-Kordahl, I. (2025) 'Preparing for Hybrid Threats to Security', in Borch, O.J. and Heier, T. (eds.) *Preparing for Hybrid Threats to Security*. Available at: https://library.oapen.org/handle/20.500.12657/93079.
4. Bohnert, M., Kochhar, A.K., Justice, D. and DeLuca, P. (2024) Winning the Undersea Knife Fight: A Deterrence

Framework to Shape Subsea Dilemmas into Competitive Maritime Security Strategies. Santa Monica, CA: RAND Corporation.
5. Bueger, C. and Liebetrau, T. (2023) 'Critical maritime infrastructure protection: What's the trouble?', *Marine Policy*, 155, 105772. doi: 10.1016/j.marpol.2023.105772.
6. Development, Concepts and Doctrine Centre (DCDC), 2017. *Wargaming Handbook*. Shrivenham: Ministry of Defence. Available at: doctrine_uk_wargaming_handbook.pdf
7. Freedman, L. (2013) *Strategy: A History*. Oxford: Oxford University Press
8. Giannopoulos, G., Smith, H. and Theocharidou, M. (2021) *The Landscape of Hybrid Threats: A Conceptual Model*. Luxembourg: Publications Office of the European Union.
9. Gill, T., Kraska, J., Ringbom, H. and Schatz, V. (2026) *Handbook on Maritime Hybrid Threats, Volume II: Contemporary Scenarios and Legal Scans*. Hybrid CoE Paper 32. Helsinki: The European Centre of Excellence for Countering Hybrid Threats.
10. Hofmann, D., Stockbrügger, J. and Sill Torres, F. (2026) 'Assessing hybrid threats to critical infrastructure through morphological analysis and escalation phases' *Proceedings of the 23rd ISCRAM Conference*. The Hague, June 2026.
11. HSEEP (2020) *Homeland Security Exercise and Evaluation Program*. Washington, DC: U.S. Department of Homeland Security.
12. Hybrid CoE (2026) 'Hybrid threat wargaming builds resilience against the unexpected'. Helsinki: European Centre of Excellence for Countering Hybrid Threats. Available at: https://www.hybridcoe.fi/news/hybrid-threat-wargaming-builds-resilience-against-the-unexpected/.
13. Jervis, R. (1976) *Perception and Misperception in International Politics*. Princeton: Princeton University Press.
14. Jungwirth, R., Smith, H., Willkomm, E., Savolainen, J., Alonso Villota, M., Lebrun, M., Aho, A. and Giannopoulos, G. (2023) *Hybrid Threats: A Comprehensive Resilience Ecosystem*. Luxembourg: Publications Office of the European Union. doi:10.2760/37899.
15. Lempert, R.J. (2019) 'Robust Decision Making (RDM)', in Marchau, V.A.W.J. et al. (eds.) *Decision Making under Deep Uncertainty: From Theory to Practice*. Cham: Springer, pp. 23–51.

16. Lempert, R.J., Groves, D.G., Popper, S.W. and Bankes, S.C. (2006) 'A General, Analytic Method for Generating Robust Strategies and Narrative Scenarios', *Management Science*, 52(4), pp. 514–528.
17. Lempert, R.J., Popper, S.W. and Bankes, S.C. (2003) *Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis*. Santa Monica: RAND Corporation
18. Magnussen, L.I., Torgersen, G.-E., Boe, O. and Saeverot, H. (2025) 'Competence for hybrid threats: a strategic competitive development model', in Borch, O.J. and Heier, T. (eds.) *Preparing for Hybrid Threats to Security*. Available at: https://library.oapen.org/handle/20.500.12657/93079.
19. Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M. and Popper, S.W. (eds.) (2019) *Decision Making under Deep Uncertainty: From Theory to Practice*. Cham: Springer.
20. Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M. and Popper, S.W. (2019) 'Introduction', in Marchau, V.A.W.J., Walker, W.E., Bloemen, P.J.T.M. and Popper, S.W. (eds.) *Decision Making under Deep Uncertainty: From Theory to Practice*. Cham: Springer, pp. 1–20.
21. Mitrescu, S. (2025a) 'On Infrastructure', in Scutaru, G. and Margvelashvili, M. (eds.) *Defending Maritime Assets: Hybrid Threats to Critical Maritime Infrastructure*. Dordrecht: Springer. NATO Science for Peace and Security Series C: Environmental Security.
22. Mitrescu, S. (2025b) 'In the Crosshairs: Hybrid Threats and the Challenge to Maritime Infrastructure', in Scutaru, G. and Margvelashvili, M. (eds.) *Defending Maritime Assets: Hybrid Threats to Critical Maritime Infrastructure*. Dordrecht: Springer. NATO Science for Peace and Security Series C: Environmental Security.
23. Monaghan, S. and McDonald, A. (2023) 'Campaigning in the Gray Zone', *CSIS Briefs*. Washington, DC: Center for Strategic and International Studies.
24. Mumford, A. and Carlucci, M. (2022) 'What is the Gray Zone?', in Brister, P. and Feyerabend, L. (eds.) *Winning in the Gray Zone*. Quantico: Marine Corps University Press, pp. 195–208.
25. Neal, A.W. (2026) 'After Nord Stream: Four Security Dilemmas in Critical Maritime Infrastructure Protection', *International Affairs*, 102(1), pp. 207–225. doi: 10.1093/ia/iiaf229.
26. Popper, S.W. (2019) 'Robust Decision Making and Qualitative Analysis', in Marchau, V.A.W.J. et al. (eds.)

Decision Making under Deep Uncertainty: From Theory to Practice. Cham: Springer.
27. Rosenhead, J. (1989) *Rational Analysis for a Problematic World: Problem Structuring Methods for Complexity, Uncertainty and Conflict*. Chichester: Wiley.
28. Rumsfeld, D. (2002) Department of Defense news briefing, 12 February. Washington, DC: U.S. Department of Defense.
29. Sukin, L. and Hedgecock, K. (2026) 'Implausible Deniability and Escalation in the Gray Zone', *Security Studies*, 35(1), pp. 73–103. doi: 10.1080/09636412.2026.2616810.
30. Vassileva, B. and Zwilling, M. (2018) 'Hybrid Warfare Simulation-Based Learning: Challenges and Opportunities', *Information & Security: An International Journal*, 39(3), pp. 220–234. doi: 10.11610/isij.3919.
31. Wirtz, J.J. (2017) 'Life in the "Gray Zone": Observations for Contemporary Strategists', *Defense & Security Analysis*, 33(2), pp. 106–114.
32. Workman, M. (2026) 'All the Gear and No Idea: How to Address the Poor Application of High-Quality Analysis for Long Term Strategy Development and Decision Making in UK Defence', The Boundary Papers, No. 1 (May). Environmental Security Research Group. Series editor: Duraid Jalili. Available at: https://environmental-security.org/boundary-papers.

BY NC ND