UNIVERSITY OF SPLIT
FACULTY OF ELECTRICAL ENGINEERING, MECHANICAL ENGINEERING AND NAVAL ARCHITECTURE

POSTGRADUATE DOCTORAL PROGRAMME IN ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGY

QUALIFYING EXAM

AI AND MULTI-CRITERIA DECISION SUPPORT FOR COURSE OF ACTION ANALYSIS AND EVALUATION IN MILITARY PLANNING

Ena Sarajlić

Split, September 2026

CONTENTS

1 INTRODUCTION

A commander making decisions about what course of action to take seems like a straightforward problem. There are many different options, each has been analysed, and one needs to recommend. What complicates things is that there is no clear way to measure or compare these options. For example, one course of action may provide advantages in speed but disadvantages in casualty costs. Another course of action may preserve reserves but sacrifice surprise. No formula or exchange rate exists to convert between these different values [1].

Doctrine provides a method or procedure for making the decision but does not provide guidance on the actual judgment being supported [2], [3], [4], [5]. Additionally, three other factors compound the difficulty of selecting a course of action. First, the people who score the options are typically the same individuals who created the options [2]. Second, because the enemy reacts to whatever options were selected, the circumstances under which an option was evaluated have changed prior to executing that option. Third, after having selected a course of action, it must be defendable before headquarters personnel, government officials, and potentially multiple levels of both within an alliance operation [3], [5]. Therefore, a methodology that produces a number without providing insight into how that number was generated will satisfy none of these three requirements.

This study outlines what is currently known in regard to supporting decision making for military commanders during the selection of a course of action and upon what basis it is done. The study evaluates and compares four separate doctrines: the US Military Decision-Making Process (MDMP) [2], [6] and Joint Planning Process (JPP) [3]; the North Atlantic Treaty Organization (NATO) Allied Joint Publication-5 (AJP-5) with the Comprehensive Operations Planning Directive (COPD) [4], [7], [8]; and the Croatian Joint Doctrine Publication 5 (Združeni doktrinarni priručnik, ZDP-5) [5] on the aspects at which they can be differentiated: what constitutes a course of action, when the criteria are set, what format a criterion takes, and how comparisons are calculated. These methodologies are then compared against the body of knowledge developed in operational research specifically designed for this type of problem and against recent literature over the past ten years.

This study aims to answer three primary questions. First, how do these doctrines formally evaluate alternative courses of action and at what juncture do they revert control of decision making to judgment? Second, what are the requirements of multi-criteria decision analysis (MCDA) relative to evaluating alternatives that cannot be provided by the doctrines? Third, how close has recent literature narrowed the gap between requirements of multi-criteria decision analysis and capabilities provided by current methodologies? Also, the study assesses how the proposed approaches have been evaluated and the reported effectiveness.

Beginning with the key objective behind those questions is to develop a decision support system that employs artificial intelligence (AI), machine learning (ML), and/or related techniques to evaluate multiple candidate Courses of Action (COA); rank-order said COA; provide an explanation as to why they were selected; and present a recommended COA to a commander. The goal of developing such a system sets the standard for what the review will need to accomplish. Because a system like this needs to clearly define both its selection criteria and how those selections are aggregated and presented in a format that allows a commander to inspect and defend at headquarters level, the doctrinal definitions in Section 2 and criterion structures in Sections 3 and 5 comprise the specifications for the system rather than its background.

Consequently, the sections follow that order. Section 2 provides a comparative overview of the four doctrines. Section 3 describes the frameworks from which evaluation criteria are developed and identifies an implicit step in all four doctrines whereby an operational consideration is transformed into a criterion. Section 4 examines the types of AI methods currently employed in planning support or prototypes using previously established methodologies to perform any component of the task identified in this paper. Section 5 defines MCDA and describes the comparison machinery that a decision support system employing such methodology would require running; along with the methods used to connect it to the AI described in Section 4. Section 5 compares the doctrinal comparison procedures contained in Sections 2 and 3 against the machinery and methods defined in Section 5. Section 6 summarizes a PRISMA-guided systematic review of 45 studies, conducted to answer four questions presented there, and records for each included study which method family it draws on, from search and optimisation, simulation and probabilistic graphical models, through reinforcement learning and neural networks, to large language models (LLM) and explainable AI (XAI). Sections 7 and 8 integrate the findings of the previous sections and establish the starting point for future work.

2 CONCEPTUAL FOUNDATIONS AND TERMINOLOGY OF COAS

A course of action (COA) is a specified sequence of decisions and operations, selected from a set of feasible alternatives, that moves a system from an initial state toward a defined goal state under given constraints. Everything else in a military planning process either produces a COA or acts on one. Four doctrines define the term, and they define it differently: the differences reflect divergent assumptions about what a plan is for, who produces it, and how much of the answer the commander supplies before staff work begins. The US MDMP treats a COA at the tactical level of the land forces and the JPP at the joint level, NATO's COPD ties it to operational design and the comprehensive approach, and Croatian doctrine adopts the NATO framework with its own adaptations. Settling what a COA is, and where the four doctrines part company, is a precondition for the AI-based comparison and recommendation system this work builds towards. What the doctrines agree a COA contains is the input schema such a system has to accept; every point on which they differ is a place where a model configured against one doctrine will not transfer to another, and a Croatian system operating inside a NATO-led coalition has to sit on both sides of that line.

2.1 MDMP

US Army doctrine, principally Army Doctrine Publication (ADP) 5-0 [6] and Field Manual (FM) 5-0 [2], places COA development third in the seven-step Military Decision-Making Process (MDMP), between mission analysis and COA analysis (wargaming). The definition it inherits from the Department of Defense (DoD) Dictionary [9] is broad: "any sequence of activities that an individual or unit may follow, or a scheme developed to accomplish a mission".

Figure 2.1 shows the seven steps. COA work occupies steps 3 to 6, and step 2 carries a substep in which the COA evaluation criteria are developed.

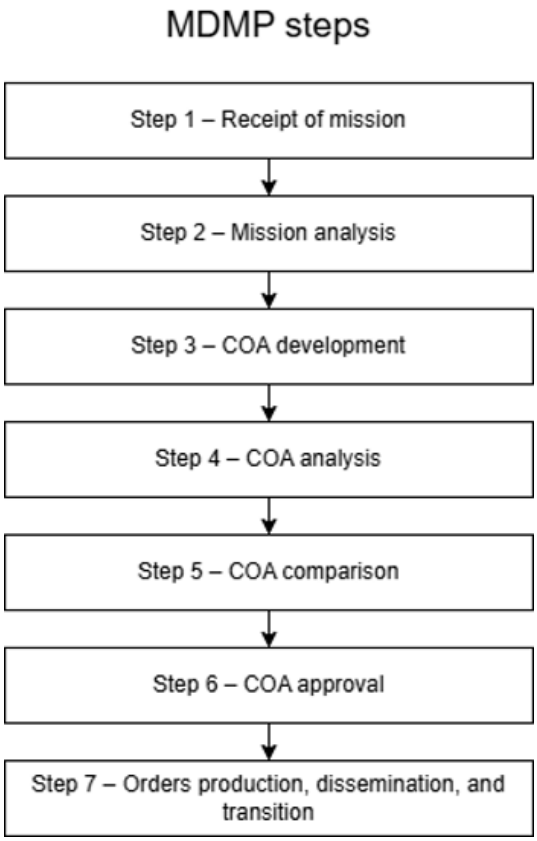


Figure 2.1 MDMP steps (drawn by the author from [2])

Evaluation criteria are developed before any COA exists. Doing so removes a source of bias ahead of COA analysis and comparison, and it fixes what data the staff must capture during wargaming. The criteria themselves are mission specific and, within that constraint, largely arbitrary: casualty limits, speed, opportunity to manoeuvre, risk, logistic supportability, force protection, timing and political considerations are among the examples given. A well-defined criterion has five elements: a short title, a definition, a unit of measure, a benchmark, and a formula stated either in comparative terms (less is better) or in absolute ones (a night movement is better than a day movement) [2]. Table 2.1 gives a worked set. How the weights attaching to those criteria are arrived at is not stated, though leaders must give the rationale for each when recommending a solution to the decision maker [2].

Table 2.1 Example evaluation criteria [2]

Short Title	Definition	Unit of Measure	Benchmark	Formula
Casualties	Casualties taken during the entire operation	Number of casualties	136 casualties	Less than 136 is an advantage. Greater than 136 is a disadvantage. Less is better.
Tempo	How long it will take the enemy forces to reach PL RED	Hours	3 hours	Less than 3 hours is an advantage. Greater than 3 hours is a disadvantage. Longer is better.[1]
Complexity	Number of task organization changes required	Number of task organization changes	7 task organization changes	Less than 7 is an advantage. Greater than 7 is a disadvantage. Less is better.

[1] The Tempo row is reproduced as it stands in FM 5-0. It contradicts itself: an enemy approach time below the three-hour benchmark is called an advantage, and the formula then states that longer is better.

The contradiction in the Tempo row of Table 2.1, recorded in the footnote to that table, is not an isolated slip. The five elements of a criterion are written as free text, and nothing forces them into agreement. The benchmark, the direction of preference and the verbal formula are three separate statements, so a reader who stops at the first will score a COA in the opposite direction to one who stops at the last. Official doctrinal publications carry inconsistencies of this kind, and any system that consumes criterion definitions automatically has to detect them rather than assume them away. Section 5.2 returns to the point: a criterion expressed as a monotone value function carries its direction of preference inside the function, which is exactly the property the textual form lacks.

COA development (step 3 of the MDMP) generates the options that later analysis and comparison will act on, each satisfying the commander’s intent and planning guidance. Where time is short the commander may cap the number of COAs the staff develops or rule ones out in advance. Planners then test each prospective COA against five screening criteria [2]:

1. Feasible. The COA can accomplish the mission within the established time, space, and resources available.
2. Acceptable. The COA must balance cost and risk with the advantage gained.
3. Suitable. The COA can accomplish the mission within the commander's intent and planning guidance.
4. Distinguishable. Each COA must differ significantly from the others, for example in scheme of manoeuvre, lines of effort, phasing, use of the reserve, or task organisation.
5. Complete. The COA incorporates how the main effort contributes to mission accomplishment, accounts for all units and capabilities across the operation's duration, shows how supporting efforts set and preserve conditions for the success of the main effort, shows how available resources enable main and supporting efforts, and shows how actions transform current conditions to the desired end state.

COA development opens with an assessment of relative combat power. Combat power is the total means of destructive and disruptive force a formation can apply at a given time, generated through leadership, firepower, information, mobility and survivability [2]. Planners compare friendly and enemy forces two echelons down, producing force ratios for combat manoeuvre units and, where relevant, for functional and multifunctional units such as field artillery or aviation [2]. The ratios are a starting point rather than a conclusion: FM 5-0 is explicit that COAs must not be built on the arithmetic alone, since the assessment also weighs intangible factors such as morale and training and the mission variables of METT-TC(I) (mission, enemy, terrain and weather, troops and support available, time available, civil considerations and informational considerations), which often matter more than the count of tanks or artillery tubes, a tube being a single gun or launcher and the standard unit in which artillery strength is counted [2]. What the step produces is a picture of relative strengths and weaknesses showing where and when the force can be weighted.

COA analysis (step 4 of the MDMP) confirms that a COA still meets the screening criteria. The staff assesses all COAs against those criteria continually and rejects any that fail, and a COA may be discarded at any point up to comparison [2]. The criteria and their weights come from three hands in turn. The chief of staff or executive officer proposes each criterion and its weight from an assessment of relative importance and the commander’s guidance; the commander adjusts the selection and the weighting according to personal experience and vision; and the

staff presents the result to the commander at the mission analysis brief for approval. The criteria used in the decision matrix at COA comparison are the ones approved there.

FM 5-0 offers three analysis techniques: wargaming, key leader discussion, and modelling and simulation. Which is used depends on time, staff experience, and whether the effort supports a new operation or an ongoing one [2]. Wargaming is the most thorough. It follows an action-reaction-counteraction sequence around a visual representation of friendly and threat forces on terrain, with a red cell (a designated group of staff playing the enemy commander) playing the enemy against the most likely and most dangerous enemy COAs, and the results captured in a synchronisation matrix, a decision support template and a decision support matrix [2]. The staff records advantages and disadvantages as they emerge and is instructed not to compare one friendly COA with another during the game [2]. To structure the analysis, planners use one of three methods, belt, avenue-in-depth or box, each framing the area of interest differently: the belt divides it into horizontal bands across the width of the sector and plays each in turn, the avenue-in-depth follows one avenue of approach from the line of departure to the objective, and the box takes a critical area and plays it in detail while the rest of the force is assumed to perform as directed[MB14.1], each framing the area of interest differently [2].

COA comparison (step 5 of the MDMP) runs an advantages and disadvantages analysis, shown in Table 2.2. Every staff member evaluates each COA from their own functional perspective, using the evaluation criteria developed in step 2, and sets out what tells for and against it. Comparing those lists identifies the benefits of each COA and the risks that come with it, relative to the others.

Table 2.2 Example COA advantages and disadvantages analysis [2]

COA	Advantages	Disadvantages
COA 1	Main effort avoids major terrain obstacles. Adequate manoeuvre space available for units executing the essential task and the reserve.	Main effort faces stronger resistance at the start of the operation. Limited resources available to establishing civil control to town X.
COA 2	Supporting efforts provide excellent flank protection of the main effort. Upon completion of the main effort's objective, supporting efforts can quickly transition to establish civil control and provide civil security to the population in town X.	Operation may require the early employment of the division's reserve.

FM 5-0 leaves the technique open, requiring only that whatever the staff applies produce the outcomes and the recommendation the commander needs [2]. Techniques commonly include using a decision matrix, which uses the evaluation criteria developed and expanded upon during mission analysis and COA development to determine which COA is most effective and efficient. Table 2.3 illustrates this.

Table 2.3 Example decision matrix [2]2

Weight	**1**	**2**	**1**	**1**	**2**	**-**
Criteria	Simplicity	Maneuver	Fires	Civil control	Mass	Total
COA 1	2	2 (4)	2	1	1 (2)	8 (11)
COA 2	1	1 (2)	1	2	2 (4)	7 (10)

[2] Criteria and their respective weights and values in the table are example values as stated in [2]

The weights reflect the relative importance of each criterion. Rankings run from 1 to the number of COAs, lower being better, and are summed in the rightmost column; the same rankings are then multiplied by the weights in the topmost row and summed again, the weighted totals shown in parentheses. The preferred COA is the one with the lowest total in both the weighted and the unweighted column [2]. Estimating relative rankings & relative weights can be highly subjective. Doctrine is explicit that the matrix is not to be treated as the decision [2]. Planners should avoid drawing conclusions from quantitative comparisons of subjective weights. Evaluating each COA functionally (one criterion at a time) usually produces more useful data than ranking all COAs against one another. A review of the situation will often change the relative weights applied to each criterion, which may also change the preferred COA. Ultimately, the commander may accept the matrix result, select a different COA, combine elements into a hybrid option, or direct additional COA development with re-focused guidance [2].

COA approval (step 6 of the MDMP) is a formal step. The commander receives the results of the comparison/analysis, selects a COA to execute, and may modify that COA prior to execution. The final product of this step includes a COA decision brief and a warning order.

2.2 JPP

Joint doctrine [3] defines a COA as a potential way, a solution or method, of accomplishing the assigned mission. The joint planning process (JPP), the joint counterpart of the MDMP, applies to supported and supporting joint force commanders and to component commanders in joint planning [3]. A joint COA is a broad but clear concept for employing the joint force across the physical domains, the information environment and the electromagnetic spectrum, stating the forces required, the time in which capabilities must be brought to bear, and the associated risk. It can reach theatre-wide, and it is built on operational design rather than on a scheme of manoeuvre in a single sector [3].

Figure 2.2 shows the seven steps of the JPP. COA work again occupies steps 3 to 6 of the JPP, with substeps in step 2 covering the development of COA evaluation criteria, the initial risk assessment and the commander’s critical information requirements. Steps or tasks may be run concurrently, truncated or modified as the situation demands. In a crisis they may be conducted simultaneously to speed the process, with steps 4 to 7 repeated as often as new requirements have to be integrated into the plan [3].

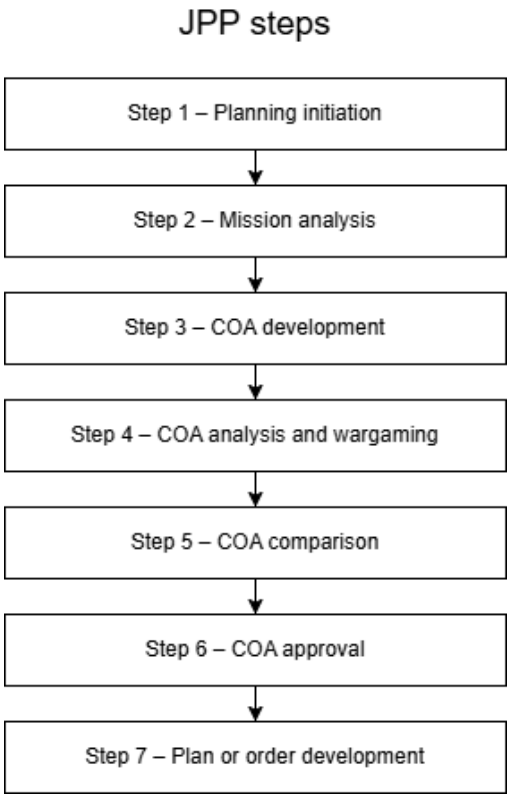


Figure 2.2 JPP steps (drawn by the author from [3])

Evaluation criteria are the standards against which the commander and staff measure the relative effectiveness of one COA against the others. They are set during mission analysis, before any COA exists, for two reasons: to remove bias from the later analysis and comparison, and to fix what data the wargame must collect. They change from mission to mission and must be clearly defined and understood by every staff member before the wargame [3]. Where the commander does not give them directly, the staff derives them from the commander's intent statement [3]. Figure 2.3 gives examples.

Figure 2.3 Example of evaluation criteria [3]

The MDMP fixes a five-element form: short title, definition, unit of measure, benchmark and formula. Joint Publication (JP) 5-0 prescribes nothing beyond a clear definition and a standard against which each COA is assessed [3]. Scoring is distributed: the staff member responsible for a functional area scores each COA on the criteria relating to that area, and commanders adjust the selection and weighting according to their own experience and vision [3]. As in the MDMP, the staff presents the proposed criteria at the mission analysis brief, and the criteria used in the comparison at step 5 are the ones approved there.

COA development (step 3 of the JPP) generates the options for later analysis and comparison. The operational approach already holds the commander's broad approach, so each COA expands it with who acts, what type of action occurs, and when, where, why and how the force is employed, the essential tasks being common to all COAs [3]. Planners vary the COAs across the domains, the information environment and the spectrum, and at the strategic level may vary the military end states themselves, so that senior leaders can weigh options at different levels of cost and risk [3]. Time and resources normally allow only two or three distinct COAs per operational approach, and where personnel permit, different teams may develop different COAs in parallel [3]. The JPP also sets out a seven-step sequence for building a COA that the MDMP does not, and it works backwards from the end state [3]. The first step fixes, within the limits of available forces, how much force the theatre needs at the end of the operation or campaign, what those forces will be doing and how they will be postured geographically, using troop-to-task analysis and a sketch. The second works back from those final positions to a base in friendly territory to give the desired basing plan. The third takes the mission statement and determines the tasks the force must accomplish in the physical domains, the information environment including cyberspace, and the electromagnetic spectrum to reach the desired military end state, sketched as a manoeuvre plan and checked against the tasks the Secretary of Defense has directed. The fourth fixes the basing required to posture the force in friendly territory and the tasks needed to reach those bases, sketched as the deployment plan. The fifth tests whether the planned force is sufficient for every task the Secretary of Defense has given the commander. The sixth sequences deployment into theatre by force category, covering combat, protection, sustainment, theatre enablers and theatre opening. The seventh draws the

preceding steps together into force employment, major tasks and their sequencing, sustainment and command relationships.

Each prospective COA is tested for validity against five screening criteria [3]:

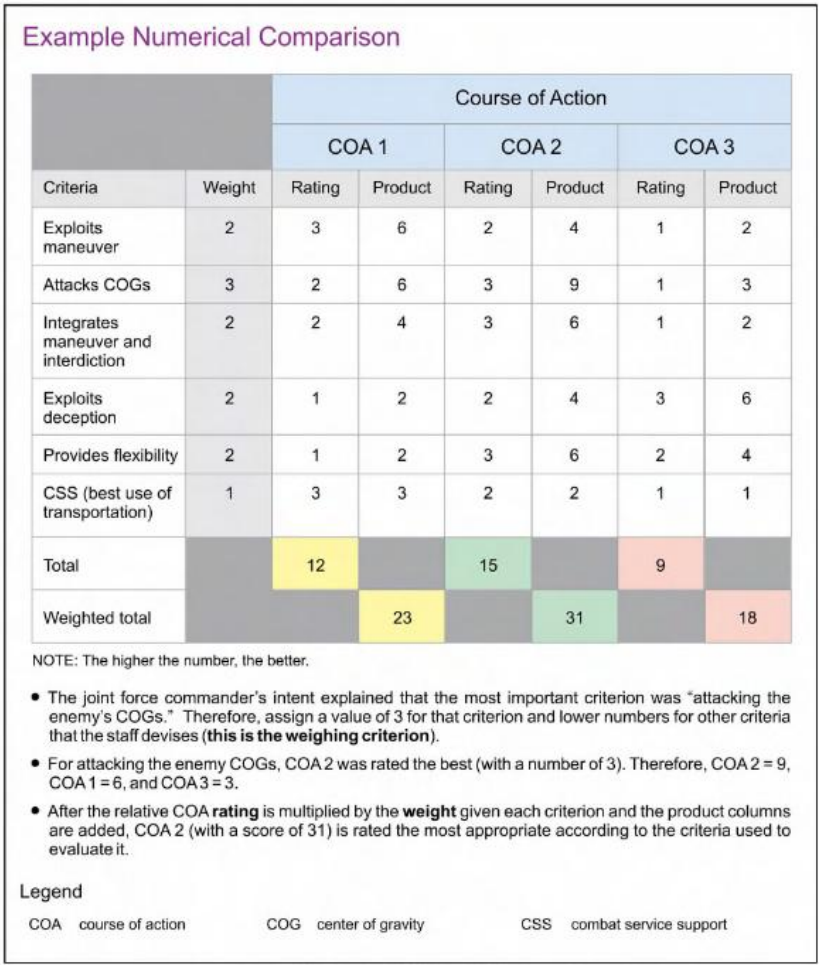
1. Suitable. Accomplishes the mission within the commander's guidance and intent, achieves all essential tasks and the end-state conditions, and accounts for the enemy and friendly centres of gravity.
2. Feasible. Accomplishes the mission within the available time, space, and resources, given the force structure, posture, transportation, and logistics on hand and the expected enemy opposition.
3. Acceptable. Balances cost and risk against the advantage gained and is reconciled with the limitations on the commander and with external limits such as US and international law, policy, and rules of engagement.
4. Distinguishable. Differs significantly from the others in the focus of the main effort, the scheme of manoeuvre and sequencing across domains, the primary mechanism for mission accomplishment, the task organisation, or the use of the reserve.
5. Complete. Answers who, what, where, when, how, and why, covering objectives and effects, the forces required including those of partners, the deployment, employment, and sustainment concepts, the time estimates, and the military end state with its success criteria.

Distinguishability demands more here than in the MDMP. The alternatives must pose the adversary genuinely different problems across domains, not variations on one scheme of manoeuvre. Any COA that fails a criterion is rejected [3].

Where the MDMP opens with relative combat power, the joint equivalent is the centre of gravity analysis carried forward from operational design. A centre of gravity is the source of power that provides strength, freedom of action, or will to act, and analysing its critical vulnerabilities is what carries the comparison of friendly and enemy strengths [3], [10]. Force availability is then established by backward planning: troop-to-task analysis fixes the force required in theatre at the end of the operation and works back to the deployment and basing that put it there [3]. Three further requirements attach to the step. A COA integrates the instruments of national power alongside multinational and interagency partners [3], [11]; none is complete without a sustainment and a deployment concept [3]; and each carries a risk assessment whose consequences reach the strategic level [3]. The step closes with initial COA sketches and statements, a validity test of each, and a development brief after which the commander approves COAs for analysis, directs revisions, or orders another developed [3].

COA analysis and wargaming (step 4 of the JPP) examines each COA separately, according to the commander's guidance, to confirm its validity and expose its advantages and disadvantages [3]. Each retained COA is wargamed against the most likely and the most dangerous enemy COAs, and every friendly COA must face the same threat COAs for the later comparison to be valid [3]. Manual wargaming runs through three variants: deliberate timeline analysis, phasing, and critical events [3]. Results are recorded in a synchronisation matrix with decision points, branches, and sequels, alongside a refined event template, an initial decision support template, and refined CCIRs [3]. A COA whose suitability, feasibility, or acceptability becomes questionable is modified or discarded rather than carried into comparison [3].

COA comparison (step 5 of the JPP) determines which COA performs best against the evaluation criteria. Each COA is judged independently against the criteria and COAs are not compared directly until each has been considered on its own [3]. Appendix F [3] offers several techniques that differ in how far they quantify. The most common is the weighted numerical comparison: the staff scores each COA on each criterion, multiplies each score by the criterion's weight and totals the products, the highest weighted total being the preferred COA. Figure 2.5 gives an example.



Example Numerical Comparison

		Course of Action					
		COA 1		COA 2		COA 3	
Criteria	Weight	Rating	Product	Rating	Product	Rating	Product
Exploits maneuver	2	3	6	2	4	1	2
Attacks COGs	3	2	6	3	9	1	3
Integrates maneuver and interdiction	2	2	4	3	6	1	2
Exploits deception	2	1	2	2	4	3	6
Provides flexibility	2	1	2	3	6	2	4
CSS (best use of transportation)	1	3	3	2	2	1	1
Total		12		15		9	
Weighted total			23		31		18

NOTE: The higher the number, the better.

- The joint force commander's intent explained that the most important criterion was "attacking the enemy's COGs." Therefore, assign a value of 3 for that criterion and lower numbers for other criteria that the staff devises (**this is the weighing criterion**).
- For attacking the enemy COGs, COA 2 was rated the best (with a number of 3). Therefore, COA 2 = 9, COA 1 = 6, and COA 3 = 3.
- After the relative COA **rating** is multiplied by the **weight** given each criterion and the product columns are added, COA 2 (with a score of 31) is rated the most appropriate according to the criteria used to evaluate it.

Legend

COA course of action COG center of gravity CSS combat service support

Figure 2.4 Example numerical comparison [3]

A non-weighted variant applies the same arithmetic without weights, suiting a problem in which the criteria are taken to matter equally. A narrative or descriptive comparison drops the numbers altogether and tabulates the strengths and weaknesses, or advantages and disadvantages, of each COA by criterion. That is the technique closest to the MDMP advantages-and-disadvantages analysis of Table 2.2, and Figure 2.6 gives an example.

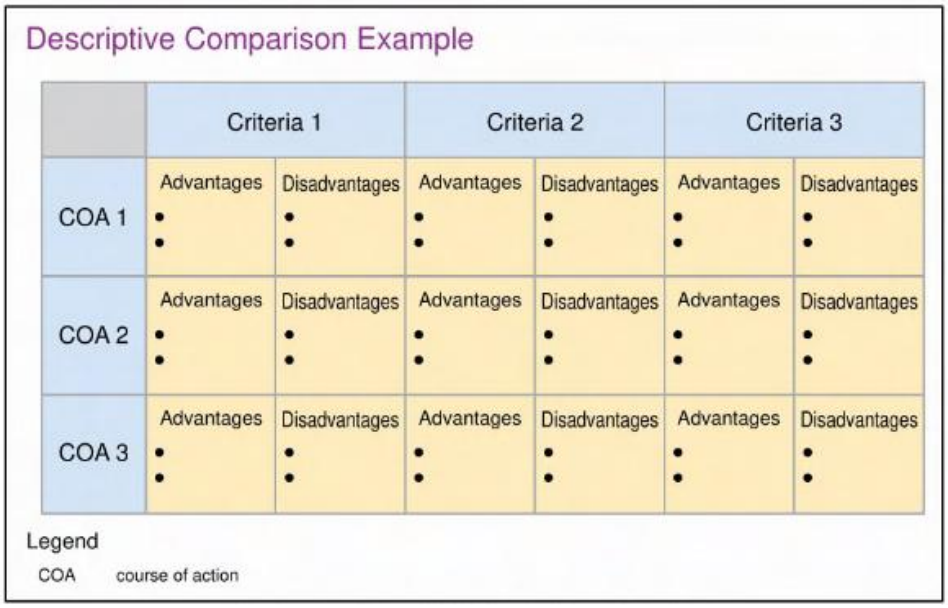


Figure 2.5 Descriptive comparison analysis [3]

A plus, minus and neutral comparison sits between the two, recording each criterion's influence on a COA as positive, neutral or negative, as in Figure 2.7. The COA with the most positive score is the one chosen.

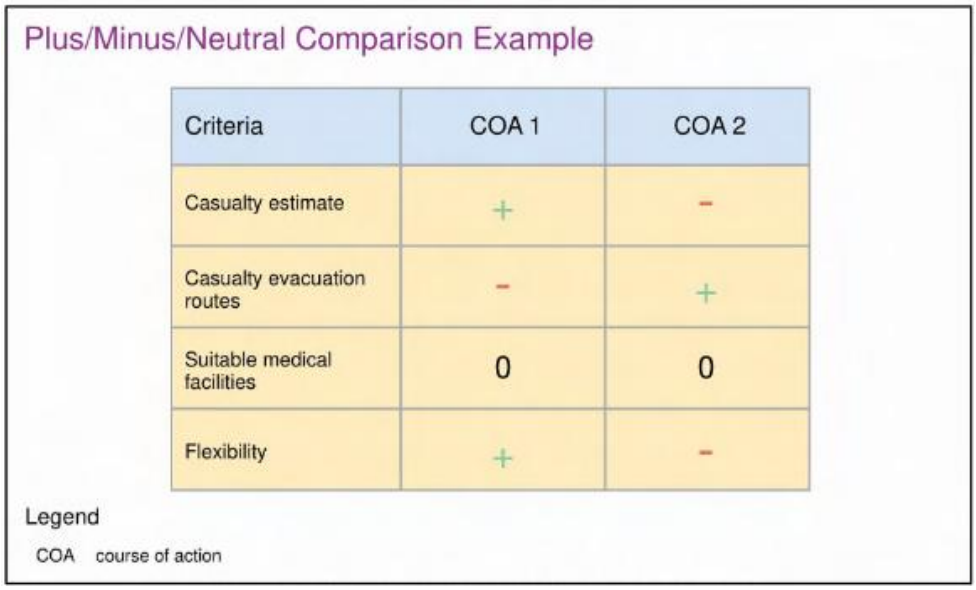


Figure 2.6 Plus/minus/neutral comparison [3]

Each staff section compares separately and may use different criteria, after which the staff reaches consensus on the criteria and weights and has the chief of staff or deputy commander approve them for consistency. The doctrine attaches three warnings to the numerical techniques: they rest on subjective values and weights, the simplified method must not be presented as rigorous mathematical analysis, and comparing COAs by criterion is more accurate than comparing totals [3]. Where COAs were developed for significantly different options with differing end states, side-by-side comparison may not be appropriate at all. The output is a recommended COA with its rationale, together with refined CCIRs and updated staff estimates.

COA approval (step 6 of the JPP) begins with a decision briefing at which the staff presents the comparison and wargaming results and recommends a COA, with all principal staff directors and component commanders attending [3]. The commander combines an independent analysis with the staff recommendation and may concur, concur with modifications, select a different COA, combine COAs, reject all and return to COA development or mission analysis, or defer pending consultation [3]. One case has no MDMP counterpart. Where the end state cannot be determined before the crisis occurs, two or more valid COAs may go to higher authority, and a single one is approved once circumstances become clear [3]. The staff then refines the selection into a clear decision statement and applies a final acceptability check, weighing acceptable risk against the desired objectives. Approval produces the commander's estimate, a concise narrative of how the commander intends to accomplish the mission [3].

2.3 COPD

NATO frames operational-level planning through Allied Joint Publication AJP-5, the keystone doctrine for the planning of operations beneath the capstone AJP-01, applied at the strategic and operational levels through the Comprehensive Operations Planning Directive (COPD) [4], [7], [8]. AJP-5 views the COA as a particular sequence of force and capability employment that occurs within an assigned area [4]. The difference in emphasis between this definition and those used by the US Army is the use of an operational design process to generate COAs. Thus, AJP-5 has established that COAs are generated as a result of a design process; they do not stand alone from their parent process. Rather, the COA represents the sequence of action or effect that leads to achieving the desired conditions for the operation [4].

The planning activities comprise seven sequential steps, shown in Figure 2.8, designed to examine a mission, develop, analyse and compare COAs, select the best, and produce a plan or order [4]. COA work again occupies steps 3 to 6 of the COPD: development, analysis, validation and comparison, and the commander’s decision. The selection criteria are set at the close of step 2, mission analysis. Unlike the MDMP and JPP, AJP-5 assigns the work throughout to the joint operations planning group (JOPG), and steps 3 and 4 are collaborative efforts with the higher and tactical levels [4]. Risk management runs through all the activities, with risk evaluation in development and risk treatment in plan development.

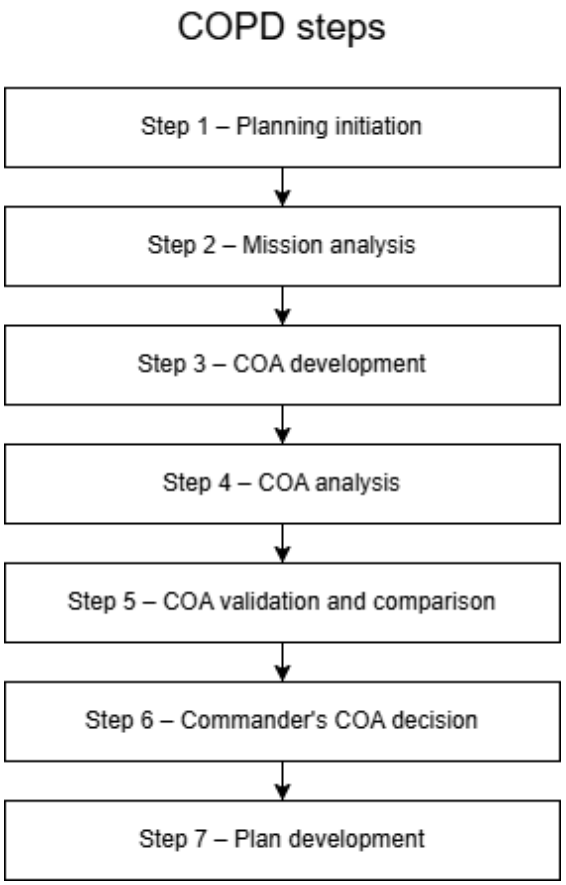


Figure 2.7 COPD steps (drawn by the author from [4])

Selection criteria are set before any COA exists. The commander issues them in the planning guidance at the end of mission analysis, alongside the adversarial COAs to be considered and a broad description of the COAs wanted, drawn from the higher commander's directive, the lines of operation and the decisive conditions [4]. They are not fixed once issued: the commander may modify them before the JOPG develops the COAs in detail [4]. AJP-5 prescribes no form for a criterion and gives no worked example of the kind the MDMP supplies. It does name the methods for applying them at comparison, namely narrative free text, one-word descriptors, numerical rating with a cardinal value, and rank ordering with an ordinal number or a plus, zero and minus qualifier, with the choice left to the commander [4].

COA development (step 3 of the COPD) produces a set of tentative COAs from the restated mission and the initial operations design, each able to accomplish the mission within the commander's initial intent and planning guidance [4]. Development begins with the adversary. Before drafting its own options the JOPG identifies the COAs open to the adversary, the intelligence staff refining the most likely and the most dangerous adversary COA for each, which sharpens the understanding of adversary capabilities and of the risks to the mission [4]. The JOPG also analyses the actions and centres of gravity of relevant national and international actors, so that friendly options neither disrupt them nor forgo interaction [4]. Teams then develop the COAs in outline, with tasks two levels down, phases and sequencing, initial subordinate missions and command arrangements, each option illustrating the main actions and effects, the system elements and centres of gravity targeted, the forces required across the joint functions, the complementary non-military actions and the information activities [4], [12]. Every tentative COA is tested against five viability criteria, and any that fails one is adjusted or rejected [4]:

1. Feasibility. Is the COA possible given the likely time, space and resources, and does it fit the operating environment?
2. Acceptability. Are the likely achievements worth the expected costs in forces, resources, casualties, collateral effects, societal impact and likely media reaction?
3. Completeness and compliance. Is the COA complete and does it conform to Allied joint doctrine?
4. Exclusivity. Is the COA sufficiently varied from the others to let its comparative advantages and disadvantages be told apart?
5. Suitability. Does the COA accomplish the mission and comply with the planning guidance?

Three features distinguish this set from the MDMP and JPP screening criteria. Acceptability is defined more broadly, weighing societal impact and likely media reaction, which reflects the political exposure of an alliance operation. Completeness adds a doctrinal-compliance requirement absent from the American doctrines, since conformity is what makes a multinational force interoperable. And exclusivity replaces distinguishability, framed not as tactical difference for its own sake but as the condition that keeps the later comparison meaningful. Surviving COAs are then evaluated against the identified risks by likelihood and impact across all phases. The step closes with a review at which the commander may rule out or add COAs, modify one, or adjust the criteria [4].

COA analysis (step 4 of the COPD) evaluates the COAs against the commander's guidance, reaffirms their viability and refines them through a cross-functional evaluation and synchronisation [4]. It produces four outputs: an outline concept of operations giving the sequence and purpose of operations by phase, the main and supporting efforts, the effects, the reserve and the strategic communications narrative; the missions and objectives for subordinate commands; the task organisation two levels down from a troops-to-actions analysis; and the operational graphics and timelines [4]. Coherence across forces and functions is plotted in a synchronisation matrix, refined during plan development and included in the OPLAN [4]. Wargaming is the principal analytical tool, but it occupies a more modest position than in the American doctrines. It is used where time permits, ideally playing each COA against the most likely and most dangerous adversarial COAs, and its value lies in letting the staff visualise the operation. AJP-5 is explicit that wargaming results do not validate or invalidate a COA; they are data points showing where risk and opportunity lie [4].

COA validation and comparison (step 5 of the COPD) validates and compares the COAs and recommends one against the commander's selection criteria [4]. The comparison runs in four contexts: consolidating each COA’s advantages and disadvantages from analysis and wargaming using the same set and weight of criteria; comparing the COAs against the selection criteria; rating from the wargame how each coped with the most likely and most dangerous adversarial COAs, with the expected effectiveness, costs and risks of each; and updating the risk description matrix [4]. All COAs should meet the selection criteria, but they will differ in how well they satisfy them, and it is those differences the JOPG compares. The method is the commander’s choice among narrative free text, one-word descriptors, numerical rating with a cardinal value, rank ordering, and plus, minus and zero. The doctrine gives no example of any of them. On that basis the JOPG recommends the COA with the highest probability of mission success within the commander’s risk attitude and acceptable human, materiel and financial costs [4]. Pairing the comparison with an explicit risk assessment against a declared risk

attitude is the clearest structural difference from the MDMP decision matrix and the JPP weighted comparison.

The commander's COA decision (step 6 of the COPD) obtains a decision and refines the chosen COA as the core of the concept of operations [4]. The JOPG presents the COAs with a coordinated recommendation. On that recommendation and their own judgement, the commander may accept a COA in full, accept it with modifications, merge COAs, or order a new one [4]. The decision produces direction on the COA and its branches and sequels, guidance for the concept of operations, issues for the higher commander, the coordination required in theatre and with national and international actors, and the refined intent [4]. The staff then refines the COA into a brief statement and applies a final acceptability check. The activity concludes with the operational planning directive, which promulgates the refined COA, intent, operations design, synchronisation matrix and subordinate missions, and tasks the components to begin planning [4].

Two characteristics shape the process throughout. The first is the comprehensive approach, adopted as Alliance policy at the 2010 Lisbon Summit and carried by AJP-5 as a central tenet [4], [7], [13]. Planning assumes actors beyond the alliance, among them host nations, international and non-governmental organisations and other coalition members, and the assumption is procedural rather than rhetorical: the JOPG confirms their actions before developing COAs [4], each COA states the complementary non-military actions it requires [4], and the coordination needed with them is among the results of the decision. The military effort is one contribution to a wider civil-military response.

The second is political direction. NATO action proceeds under the North Atlantic Council, which sets the end state and strategic objectives, and Allied doctrine treats policy as leading planning rather than entering it once as an input [4], [7]. Military options are developed at the strategic level under the COPD, and operations planning begins only once one is selected. The operational level can influence this only through advice to the strategic commander, which AJP-5 identifies as the single opportunity to affect the process from below [4]. A COA must therefore stay within the political authority the Council has granted, and options that exceed it are unavailable however sound in military terms [7]. What is carried forward is accordingly a design construct, made of objectives, decisive conditions, effects, actions and lines of operation, developed into the concept of operations and then the operation plan.

2.4 Croatian doctrine

Croatian joint doctrine was formed prior to joining NATO. The 2006 Joint Doctrine of the Armed Forces of the Republic of Croatia listed among its foundations the National Security Strategy and Defense Strategy of the Republic of Croatia, NATO doctrinal documents (AJP-01(B), Allied Tactical Publication (ATP)-3.2 and the US Army's FM 3-0), the experience of the Homeland War, the experience of other countries that have become part of Euro-Atlantic organizations, and developments in military science and technology [14]. This doctrine incorporated a term used today for a course of action, *inačica djelovanja*, defined as "an option that will accomplish or support the accomplishment of the mission or task from which detailed plans will be developed." That is a nearly verbatim translation of the NATO Allied Administrative Publication (AAP)-6 definition [14]. The COA construct therefore entered Croatian doctrine together with the NATO alignment, not after it.

Two revisions followed accession in 2009 and restate the relationship to NATO doctrine more tightly. ZDP-1, from 2011, declares harmonisation with AJP-01 specifically, and introduces a precedence rule: in NATO-led operations, Allied doctrine ratified by the Republic of Croatia takes precedence over Croatian doctrine [15]. It fixes the operations planning process in five

stages, initiation (*iniciranje*), orientation (*orijentacija*), concept development (*razvoj koncepata*), plan development (*razvoj plana*) and plan review (*revizija plana*), with COA development and selection falling in the third stage and the fifth providing for revision of the selected COA. It gives no formal COA definition and no detailed development procedure [15].

The current ZDP-1(A), from 2016, moves the harmonisation statement forward, drops the enumeration of foundations, states harmonisation with NATO doctrinal publications, and retains the precedence rule [16]. COAs and operations planning appear only in passing, the detailed treatment having moved to the subordinate planning publication ZDP-5 [5]. The alignment extends beyond doctrine into planning, since the Long-Term Development Plan 2025-2036 [17] ties the national defence planning process to the NATO Defence Planning Process, through which national objectives and plans are harmonised with the Alliance's collective needs and, indirectly, with those of the European Union.

ZDP-5 presents the operations planning process in eight steps, with COA work in steps 3 to 6. Figure 2.9 shows them, with the original Croatian names in brackets.

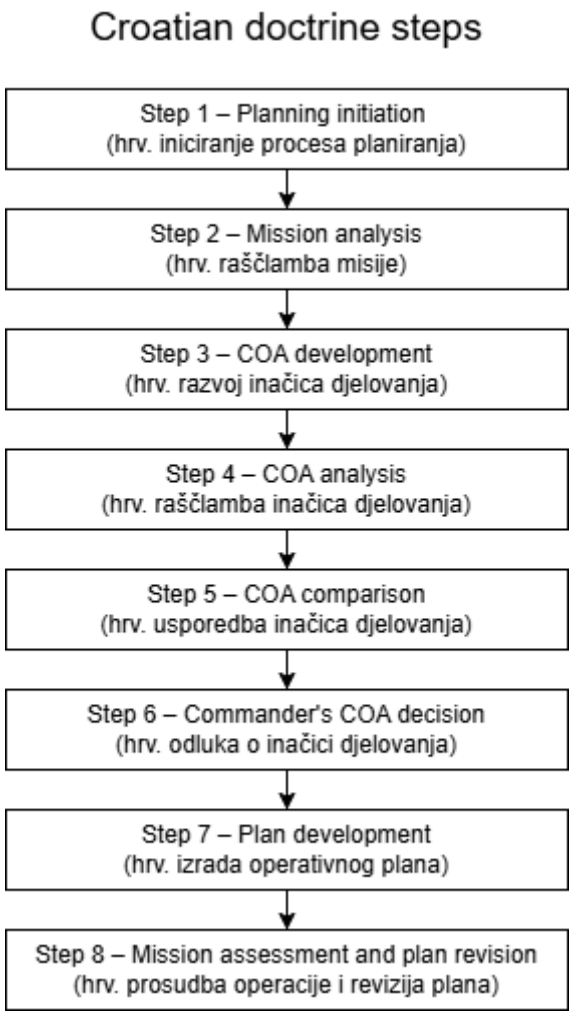


Figure 2.8 Croatian doctrine steps (drawn by the author from [5])

The work on COAs is carried out by the operational planning group (*Operativna planska skupina*, OPS), formed within the operation command after receipt of the preparatory order [5]. It begins from the initial operational design and the operational planning guidelines produced in step 2, in which the commander also approves the most likely and the most dangerous

adversarial COA as the basis of the development phase [5]. Unlike the doctrines described above, ZDP-5 does not state that this step produces explicit evaluation criteria for use in later steps.

Development of COAs (step 3 of the ZDP-5 process) produces several potential options that accomplish the assigned mission in accordance with the commander's intent, worked up in outline only as far as is needed to establish that they are realistic and executable [5]. An option states who will conduct which kind of action, when the operation will begin, where it will take place, to what end and in what manner [5]. Before drafting its own options, the group assesses the adversarial ones, the intelligence staff preparing an estimate of adversary procedures and actions [5]. Working groups then develop each option as a general idea supported by a sketch, showing the sequence of the main activities of the force elements and the resulting effects [5]. Options must differ from one another, and similar solutions to the same problem are not to be developed [5]. Each must describe the sequence and purpose of the main joint activities needed to reach the decisive points, the activities that create the planned effects, the systems and system elements they are directed against, etc. Options carried into further analysis must satisfy five credibility criteria (*kriteriji vjerodostojnosti*) [5]:

1. *Prikladnost* (suitability): the option accomplishes the mission within the commander's guidelines and intent, supports the execution of all important tasks and satisfies the conditions for the desired end state.
2. *Provedivost* (feasibility): the option accomplishes the mission within the determined time and space and under the resource constraints, with the forces and support allocated.
3. *Prihvatljivost* (acceptability): the option balances costs and risks against the benefit obtained and conforms to the constraints imposed, the rules of engagement in particular.
4. *Različitost* (distinguishability): the option differs sufficiently from the others in main effort, scheme of manoeuvre, force organisation and reserve.
5. *Cjelovitost* (completeness): the option answers who, what, where, when, how and why, and contains the objectives, desired effects and tasks, the bulk of the forces required, the plans for deployment, employment and support, and a time estimate.

The commander may adapt these criteria to the requirements and constraints received from the military strategic level, and the planning group presents them to the commander for discussion [5]. The step closes with a briefing on the proposed options. There the commander approves them for further analysis, orders revision, combination or the development of additional ones, and determines which option will be played against which adversarial option; if all are rejected, step 3 begins again [5].

Analysis of COAs (step 4 of the ZDP-5 process) reviews each option from different functional perspectives to establish its advantages and disadvantages, the sufficiency of forces, the branches and sequels of the main activities and the adversary's possibilities. The wargame is its primary instrument [5]. The criteria for evaluating the options during the wargame are developed in this step, not earlier, and they vary with the type of operation: they may include the joint (combat) functions, the principles of war, doctrinal principles, the elements of operational design, the level of risk and the commander's intent [5]. Three wargaming methods are prescribed, playing the main activities by phase against the desired outcome of each phase, playing them to define the decisive conditions, and playing them in defined parts of the area of operations [5]. Play proceeds in cycles of action, reaction and counteraction, each assessed before the initial conditions of the next are set [5]. The outputs are the refined options: the

concept of each by phase, with the main and supporting efforts, the effects and the reserve; the missions and objectives of subordinate commands; the force organisation two levels down; and the operational graphics and timeline [5].

Evaluation and comparison of COAs (step 5 of the ZDP-5 process) proceeds by four methods [5]. The first tabulates the advantages and disadvantages of each option, established during analysis and extended with what the wargame revealed [5]. The second rates each option's effectiveness, expected resource expenditure and risk against the most likely and the most dangerous adversarial options [5]. The third, described as the most frequently applied, is a numerical comparison against the commander's criteria, in which the best option is determined from the sum of the products of the importance of each criterion and the points denoting the degree to which the option realises it [5]. The fourth is a risk assessment: risks are traced from their sources to what they bear on, rated by consequence and likelihood, and the option's risk is classified as unacceptable, conditionally acceptable or acceptable [5]. Three cautions follow. Comparison and selection are a subjective process, conclusions drawn from mathematical methods are to be treated with care, and the decisive factor is the ability of the head of the planning group to articulate why a particular option is proposed [5].

The next step (step 6 of the ZDP-5 process) is the COA selection, where the commander chooses the preferred course among other alternatives [5]. The commander makes the decision based on the staff recommendation, the advice of subordinate commanders and their own judgement and experience, and may take the recommended option or another, with or without modification, or require additional options to be developed [5]. The selected option is then refined and the commander's intent elaborated [5]. The step closes with the operational directive for planning (*operativna direktiva za planiranje*), which carries the refined option, the commander's intent, the final operational design, the synchronisation matrix and the missions of subordinate components and units [5].

What Croatian doctrine does not replicate is the political and material context in which NATO doctrine assumes it will run. Three adaptations follow. None is a doctrinal disagreement; all three are functions of constitutional structure and scale.

The first is the position of political authorisation. Article 7 of the Constitution, as amended in 2010, permits the Armed Forces to cross or operate beyond the state border only on a decision of the Sabor, proposed by the Government with the prior consent of the President of the Republic and adopted by a majority of all members, or by a two-thirds majority if the President withholds consent [18]. The Defence Act operationalises this mechanism and permits the transfer of operational and tactical command over Croatian units to international commands, with or without national caveats on the use of forces (*izuzeća u uporabi snaga*) [19]. A COA that is valid within the COPD framework may therefore be constitutionally unavailable to Croatian forces, so political feasibility screening enters COA development earlier and binds harder than the NATO process itself requires. The mechanism is not theoretical. In October 2024 the President withheld prior consent from the proposed contribution of up to five members of the Armed Forces to the NATO Security Assistance and Training for Ukraine activity (NSATU), which raised the adoption threshold to two-thirds; the decision was not adopted, while the twelve decisions that received consent passed [20].

The second is the contributing-nation posture. Deployment abroad is regulated by a dedicated statute [21], and the Sabor's implementing decisions set numerical and geographic limits in advance. The KFOR decision is representative: for 2025 and 2026 it authorises up to 200 members of the Armed Forces with up to two helicopters, rotation permitted [22]. The posture extends to command arrangements. Croatia plans part of its land contribution through the

Multinational Division Command Centre established with Hungary and Slovakia, which reached full operational capability in 2025 within NATO's military structure [17]. Much national COA development is therefore force-generation planning conducted inside pre-set limits, a national contribution to an operation planned elsewhere, rather than campaign design from a blank sheet.

The third is scale. The Defence Strategy and the Long-Term Development Plan commit to a force whose size will not decrease, but the structure they set out is narrow: land-force development rests on the guards brigades, with the first planning period devoted to completing a single medium infantry brigade and the following period to a single heavy brigade built around the existing armoured-mechanised brigade [17], [23]. A force whose land-manoeuvre development fits within two brigades supports fewer genuinely distinguishable options than the doctrinal framework implies. Feasibility screening therefore does more work in Croatian planning than in NATO planning generally, and COA development in practice generates fewer real alternatives.

2.5 Comparison of COA across doctrines

All four doctrines place COA work in steps 3 to 6 of a process of seven or eight steps, screen options against five criteria, and put a recommendation to the commander for decision. The differences lie elsewhere: in what a COA is taken to be, when the criteria against which options are judged are fixed, how far the comparison is quantified, and where risk enters. The tables below set the four side by side on each of those points.

Table 2.4 Doctrinal frame of the COA process [2], [3], [4], [5], [6], [8]

Attribute	MDMP	JPP	AJP-5 / COPD	ZDP-5
Keystone publications	ADP 5-0, FM 5-0	JP 5-0	AJP-5, applied through the COPD	ZDP-5, under ZDP-1(A)
Level and force	Tactical, land forces	Joint, operational to strategic	Operational, alliance	Operational, national
Planning body	Staff, chief of staff or executive officer leading	Staff, teams may develop COAs in parallel	Joint operations planning group (JOPG)	Operational planning group (OPS)
Steps in the process	7	7	7	8
Steps carrying COA work	3 to 6	3 to 6	3 to 6	3 to 6
What a COA is	A scheme of manoeuvre against an enemy in terrain	A concept for employing the joint force across domains, with forces, timing and risk	An employment of forces in a sequence of actions along the lines of operation, following from operations design	A sequence of actions across domains within the commander's intent
Where development opens	Assessment of relative combat power, force ratios two echelons down	Centre of gravity analysis, troop-to-task and backward planning	Operations design, adversarial COAs and actor analysis	Initial operational design and the adversarial COAs approved in step 2
Number of COAs	Commander may cap the number	Normally two or three per operational approach	A set of tentative COAs	Several potential options

The four processes name the COA steps almost identically, which is what makes them comparable at all. Only the fifth step differs in scope. Both American doctrines compare; NATO and Croatian doctrine validate and compare in the same step.

Table 2.5 Names of the COA steps [2], [3], [4], [5]

Step	MDMP	JPP	AJP-5 / COPD	ZDP-5
3	COA development	COA development	COA development	COA development (*razvoj inačica djelovanja*)
4	COA analysis (wargaming)	COA analysis and wargaming	COA analysis	COA analysis (*raščlamba inačica djelovanja*)
5	COA comparison	COA comparison	COA validation and comparison	COA evaluation and comparison (*vrednovanje i usporedba inačica djelovanja*)
6	COA approval	COA approval	Commander's COA decision	Decision on the COA (*odluka o inačici djelovanja*)

The screening criteria are the strongest point of convergence. All four use five, and three carry the same names and the same tests everywhere. The remaining two vary in label and in what they demand.

Table 2.6 Screening criteria applied to tentative COAs [2], [3], [4], [5]

Concept	MDMP	JPP	AJP-5 / COPD	ZDP-5
Suitability	Suitable: accomplishes the mission within intent and planning guidance	Suitable: adds essential tasks, end-state conditions and centres of gravity	Suitability: accomplishes the mission and complies with the planning guidance	*Prikladnost*: adds the conditions for the end state and consideration of centres of gravity
Feasibility	Feasible: time, space and resources available	Feasible: adds force structure, posture, transportation, logistics and expected opposition	Feasibility: time, space, resources and fit to the operating environment	*Provedivost*: time, space and resource limits with the forces and support allocated
Acceptability	Acceptable: balances cost and risk against the advantage gained	Acceptable: adds law, policy and rules of engagement	Acceptability: adds casualties, collateral effects, societal impact and likely media reaction	*Prihvatljivost*: balances cost and risk against benefit, with external constraints and rules of engagement
Difference between options	Distinguishable: scheme of manoeuvre, lines of effort, phasing, reserve, task organisation	Distinguishable: main effort, sequencing across domains, mechanism, task organisation, reserve	Exclusivity: varied enough for comparative advantages to be told apart	*Različitost*: main effort, scheme of manoeuvre, force organisation, reserve
Completeness	Complete: main and supporting efforts, all units, transformation to the end state	Complete: the six questions, forces, deployment, employment, sustainment, time, end state	Completeness and compliance: the six questions, with how to a limited extent	*Cjelovitost*: the six questions, objectives, effects, tasks, forces, plans and time estimate
Doctrinal compliance	Not required	Not required	Required, inside completeness: conformity to Allied joint doctrine	Not required
When applied	Step 3, re-tested continually through step 4	Step 3, revisited in step 4	Step 3, reviewed at the close of the step	Step 3, before an option enters further analysis
Failure of a criterion	Rejected; a COA may be discarded at any point up to comparison	Rejected	Adjusted or rejected	Not carried forward; if all are rejected, step 3 begins again

Evaluation criteria are a separate instrument from the screening criteria. The rows of Table 2.7 read as follows. Fixed when records the step at which the criteria are settled, and stated rationale the reason each doctrine gives for settling them there. Prescribed form records whether the doctrine specifies what a criterion must contain, and typical criteria named the examples it offers. Who proposes and the commander's role separate the staff work from the commander's, and weighting method records what, if anything, the doctrine says about where the numbers come from. Screening asks whether an option is admissible; evaluation asks how well it performs against the others. Here the four diverge, and the divergence is what any comparison method has to work with.

Table 2.7 Evaluation criteria: provenance and form [2], [3], [4], [5]

Attribute	MDMP	JPP	AJP-5 / COPD	ZDP-5
Fixed when	Step 2, before any COA exists	Step 2, before any COA exists	End of step 2, issued in the planning guidance	Step 4, after the options exist
Stated rationale	Eliminates bias before analysis and fixes what the wargame must capture	Removes bias from analysis and comparison and fixes what the wargame must collect	Issued with the adversarial COAs and a description of the COAs wanted	None stated
Prescribed form	Five elements: short title, definition, unit of measure, benchmark, formula	A clear definition and a standard; no fixed form	No form prescribed	No form prescribed
Worked example	Yes (Table 2.1)	Yes (Figure 2.3)	None	None
Typical criteria named	Casualties, speed, manoeuvre, risk, logistic supportability, force protection, timing, political factors	Derived from the commander's intent where not given directly	Drawn from the higher directive, the lines of operation and the decisive conditions	Joint functions, principles of war, doctrinal principles, elements of operational design, level of risk, commander's intent
Who proposes	Chief of staff or executive officer proposes criteria and weights	Staff; functional-area officers score their own areas	Commander issues them	The planning group develops them
Commander's role	Adjusts selection and weighting, approves at the mission analysis brief	Adjusts selection and weighting, approves at the mission analysis brief	Issues them and may modify them before detailed development	May adapt them to military strategic requirements, discusses them with the planning group
Weighting method	Not stated; rationale must be given for each	Not stated; staff reaches consensus, approved by the chief of staff or deputy commander	Not stated; the same set and weight of criteria applies to all COAs	Not stated; importance 1 to 4 in the example, the priority criterion weighted 1

The doctrines in step 4 are closest in method and furthest apart in the authority they give the result. The American doctrines treat the wargame as the test an option must survive. AJP-5 treats it as a source of data points that neither validates nor invalidates an option.

Table 2.8 COA analysis and wargaming [2], [3], [4], [5]

Attribute	MDMP	JPP	AJP-5 / COPD	ZDP-5
Principal technique	Wargaming, the most thorough of three	Wargaming	Wargaming, used where time permits	Wargaming, the primary instrument
Alternatives named	Key leader discussion, modelling and simulation	None	None	None
Structuring methods	Belt, avenue-in-depth, box	Deliberate timeline analysis, phasing, critical events	None named	By phase against desired outcomes, by decisive conditions, by parts of the area of operations
Adversarial COAs played	Most likely and most dangerous, played by a red cell	Most likely and most dangerous; every friendly COA faces the same threat COAs	Most likely and most dangerous, ideally	Most likely and most dangerous, with red, green and blue teams, arbiters and a recorder
Standing of the results	Advantages and disadvantages recorded as they emerge; a COA may be discarded	Confirms validity; a questionable COA is modified or discarded rather than compared	Data points highlighting risk and opportunity; they do not validate or invalidate a COA	Refined options; observations pass into step 5
Recording products	Synchronisation matrix, decision support template and matrix	Synchronisation matrix, refined event template, initial decision support template, refined CCIRs	Synchronisation matrix, carried into the OPLAN	Synchronisation matrix, operational graphics and timeline
Constraint on the method	COAs are not compared with one another during the game	Each COA is analysed separately under the commander's guidance	Coherence across forces and functions plotted before comparison	Political and military strategic constraints must be built into the options, or the comparison is irrelevant

The comparison methods offered are the point at which the four doctrines visibly part company. Table 2.9 maps which methods each one names.

Table 2.9 Comparison methods named by each doctrine [2], [3], [4], [5]

Method	MDMP	JPP	AJP-5 / COPD	ZDP-5
Advantages and disadvantages by COA	●	●	●	●
Narrative or free text by criterion	–	●	○	–
One-word descriptors	–	–	○	○
Plus, minus and neutral	–	○	○	○
Unweighted numerical sum	○	○	–	–
Weighted numerical sum	●	●	○	●
Rank ordering across COAs	●	–	○	–
Effectiveness against the adversarial COAs	–	–	●	●
Explicit risk assessment of each COA	–	○	●	●

● named as the principal or prescribed method; ○ named as an option; – not named.

Three of the four supply an arithmetic. Table 2.10 sets out what each one computes, which matters because the same operation is performed on scales of different types and read in opposite directions.

Table 2.10 Mechanics of the quantitative comparison methods [2], [3], [4], [5]

Doctrine	Scale	Weights	Aggregation	Preferred COA	Example	Stated caveat
MDMP	Rank from 1 to the number of COAs, lower better	Relative importance, proposed by the chief of staff, adjusted by the commander	Sum of ranks and sum of weighted ranks	Lowest total in both columns	Yes	Avoid drawing conclusions from a quantitative analysis of subjective weights; compare by criterion
JPP	Score of each COA against a standard on each criterion	Agreed by the staff, approved by the chief of staff or deputy commander	Sum of score × weight	Highest weighted total	Yes	Must not be presented as rigorous mathematical analysis
AJP-5 / COPD	Cardinal rating, ordinal rank, plus/zero/minus, or words	The same set and weight of criteria for all COAs; no scale given	Not specified	Highest probability of mission success within the commander's risk attitude	No	Method left to the commander's preference
ZDP-5	Points for the degree to which a criterion is realised; 1 to 3 with ties in the example, fewer better	Importance from 1 to 4, the priority criterion weighted 1	Sum of importance × points	Lowest total	Yes	Comparison and selection remain subjective; treat conclusions from mathematical methods with caution

The decision step converges again. In all four the commander is free to take the recommendation, change it, combine options or send the staff back, and in all four the step ends by promulgating the choice. Only the weight of the closing document differs.

Table 2.11 The commander's decision and its outputs [2], [3], [4], [5]

Attribute	MDMP	JPP	AJP-5 / COPD	ZDP-5
Recommendation from	Staff, at the decision briefing	Staff, at the decision briefing with principal directors and component commanders	JOPG, with a coordinated recommendation	Chief of staff, on the planning group's proposal
Options open to the commander	Select, modify, or direct further development	Concur, concur with modifications, select another, combine, reject all and return, or defer	Accept, accept with modifications, merge COAs, or order a new one	Take the recommended option or another, with or without modification, or require additional options
Named outputs	Refined intent and mission-specific CCIRs	Decision statement, final acceptability check, refined CCIRs	Direction on the COA with branches and sequels, guidance for the concept of operations, issues for the higher commander, coordination required, refined intent	Selected option with branches and sequels, guidance and deadlines, issues for the military strategic level, liaison and coordination priorities, authorities for non-military coordination
Closing document	COA decision briefing and warning order	Commander's estimate	Operational planning directive	Operational directive for planning

Four differences carry through the tables. The first is the object being compared, which ascends from a scheme of manoeuvre in terrain to a design construct expressed in lines of operation and decisive conditions; a criterion that discriminates between schemes of manoeuvre will not necessarily discriminate between designs. The second is the timing of the evaluation criteria. The MDMP, the JPP and AJP-5 fix them before any option exists, and the first two say why, while ZDP-5 develops them in step 4 with the options already on the table. The third is the form of a criterion, which only the MDMP specifies, leaving the other three to apply criteria with no stated unit, benchmark or direction. The fourth is the place of risk, which the American doctrines fold into the criteria set and which NATO and Croatian doctrine keep outside the aggregate as a separate judgement against a declared risk attitude or acceptability class.

The methods themselves converge on a narrow set: an advantages and disadvantages table everywhere, a weighted sum in three of the four, and an ordinal or verbal fallback where numbers are not wanted. Every doctrine that supplies the arithmetic also warns against trusting it. None states how weights are to be elicited, how scores are to be assigned, or what the totals mean when the criteria measure incommensurable things.

3 OPERATIONAL FACTORS

Operational factors are things a commander needs to know about a particular scenario, so planners can assess options based upon those factors. This term is used in a very specific way. It does not relate to the operational art's space, time and force, but to doctrine-defined lists of things that affect a mission from which criteria for judging and comparing COAs will come. Doctrine uses different terms such as "variables", "considerations" and sometimes none at all, for these lists; the ways they organize their lists vary depending on the process being used.

They appear during the planning process in three separate instances. During mission analysis, they serve to establish an understanding of the problem area and create the facts, assumptions and constraints on which all future decisions are made. When developing COAs, they define what can be done because each COA either uses or accommodates them and the screening criteria outlined in Section 2.5 are applied against them. Finally, when evaluating COAs, they represent the pool of items from which the evaluation criteria are selected. It is the last of these applications that this study relies upon. A decision matrix consists of columns, and every column represents an operational factor that was selected to be quantifiable.

No doctrine states how that choice is made. All four fix the shape of the process and two fix the form of a criterion, but none gives a procedure for deriving evaluation criteria from the factor analysis that precedes them. FM 5-0 supplies a five-element form and a worked example, JP 5-0 a clear definition and a standard, AJP-5 and ZDP-5 neither [2], [3], [4], [5]. The pool is described in detail; the drawing from it is left to judgement. This section sets out the pools framework by framework, and Section 3.3 compares them.

3.1 Factors in MDMP

US Army doctrine gives the most explicit treatment, and it separates the pool into two tiers. The operational variables, political, military, economic, social, information, infrastructure, physical environment and time, known as PMESII-PT, describe an operational environment in general terms and without reference to any particular unit. The mission variables then select from that description what bears on the mission at hand [2], [6]. That two-tier structure is the general form of the problem. Every framework below performs both operations, even where it names neither.

The mission variables are METT-TC, with current doctrine appending informational considerations through all six, as METT-TC(I): mission, enemy, terrain and weather, troops and support available, time available, and civil considerations [2], [6]. Each is a category for organising analysis rather than a quantity, and each produces attributes of a different kind.

- Mission supplies the essential tasks against which the others are judged [2].
- Enemy supplies composition, disposition and strength, the most likely and most dangerous enemy COAs, and relative combat power computed during COA development as force ratios two echelons down [2].
- Terrain is analysed through its five military aspects: Observation and fields of fire, Avenues of approach, Key terrain, Obstacles, and Cover and concealment (OAKOC). Weather is treated alongside it through its effects on mobility, sensors and illumination [24]. Making these aspects computable typically means representing each as a raster layer and combining the layers into a single threat or trafficability surface by weighted sum, which turns a doctrinal category into a

scored quantity at the price of fixing the trade-offs between the aspects before any option is examined [25].

- Troops and support available supplies strength, readiness, training state and sustainment reach.
- Time available supplies planning and execution time, tempo relative to the enemy, and the width of the window before a decision must be taken.
- Civil considerations covers the influence of man-made infrastructure, civilian institutions, and the attitudes and activities of civilian leaders, populations and organisations on the conduct of operations [24].

The civil variable is the thinnest of the six and the only one with a framework of its own. ASCOPE [MB22.1]is the acronym for areas, structures, capabilities, organisations, people and events, and it decomposes the civil variable into those six headings. The framework is most often crossed with PMESII-PT so that each cell prompts a specific civil-by-system observation [24]. The crosswalk reaches beyond the Army, since NATO builds its operations design on the same PMESII view of the operating environment and its interdependencies [4]. PMESII is on that account the one framework in this section shared between American and Allied doctrine.

Because METT-TC is tied to the MDMP, it is tactical and land-centric, and its terrain variable has no counterpart above the level at which a scheme of manoeuvre is drawn on ground; none of the three processes treated below adopts it. Another limit appears where an operation turns on the civil environment rather than on an opposing force. In the civil protection and multi-agency response cases of Section 4 the enemy variable is absent or replaced by a hazard, and the weight of the analysis falls on ASCOPE-PMESII instead.

3.2 Factors in JPP, COPD and Croatian doctrine

JP 5-0 names no mission-variable set. Its pool is the operational environment across the physical domains, the information environment and the electromagnetic spectrum, together with the elements of operational design [3]. Two substitutions change what a criterion can be about. It is the critical vulnerabilities of a centre of gravity, rather than a force ratio at a place and a time, that carry the comparison of friendly and enemy strengths [3], [10]. And force availability is produced rather than given, since troop-to-task analysis works backward from the force required in theatre at the end of the operation to the deployment and basing that place it there [3]. Two COAs making different demands on transportation have different forces at the decisive moment, so the factor belongs partly to the option rather than wholly to the situation. Risk is carried through development as a factor in its own right, with consequences reaching the strategic level [3].

Because AJP-5 derives COAs from operations design [4], its pool is the design vocabulary laid over a PMESII reading of the operating environment. Two features have no counterpart in either American framework. Civil actors are analysed with the instrument used on the adversary rather than treated as a description of the environment, the JOPG examining the actions and centres of gravity of relevant national and international actors so that friendly options neither disrupt them nor forgo interaction [4], and the comprehensive approach makes their involvement procedural rather than advisory [4], [7], [13]. Acceptability, second, weighs societal impact and likely media reaction [4], so reputational attributes belong to the formal factor set rather than to the commander’s private judgement. The authority granted by the North Atlantic Council bounds the option space from outside [4].

ZDP-5 names no factor framework. Its pool comes from the initial operational design and the adversarial COAs approved in step 2 [5], and from the required content of a COA [5], every item of which is a NATO operations-design construct and none a METT-TC variable. Terrain reaches the process only through the decisive points and the parts of the area of operations in which the wargame is played [5]. To these it adds the menu from which the evaluation criteria are drawn during analysis, which varies with the type of operation and admits the joint (combat) functions, the principles of war, doctrinal principles, the elements of operational design, the level of risk and the commander's intent [5]. That is the broadest and least prescriptive of the four.

Three further factors are implicit in the comparison methods themselves. The second method rates effectiveness, expected resource expenditure and risk against the adversarial options [5], the third aggregates the commander's criteria as a weighted sum [5], and the fourth rates risk by consequence and likelihood [5]. Effectiveness, cost and risk are on that account the minimum factor set of Croatian comparison. The national conditions of Section 2.4, the Sabor authorisation and the contribution caps set in advance, belong to the pool as well, but they bound the feasible set rather than distinguishing between the options within it.

3.3 Comparison of factor frameworks

Table 3.1 sets the four pools side by side. Table 2.7 has already compared the evaluation criteria themselves, their provenance, form and timing; what follows is the material from which they are drawn.

Table 3.1 Operational factor frameworks by doctrine [2], [3], [4], [5]

Factor	MDMP	JPP	AJP-5 / COPD	ZDP-5
Named factor set	METT-TC(I) over PMESII-PT	None; the elements of operational design	None; the operations design constructs	None; assembled per operation
Environment	IPB (intelligence preparation of the battlefield), with PMESII-PT filtered into the mission variables	The operational environment across the physical domains, the information environment and the spectrum	A PMESII system view of the operating environment and its interdependencies	The initial operational design produced in step 2
Adversary	Relative combat power and force ratios two echelons down; most likely and most dangerous enemy COA	Centre of gravity and its critical vulnerabilities	Adversarial COAs and centres of gravity; the system elements each COA targets	Most likely and most dangerous adversarial COA, approved by the commander in step 2
Civil and political actors	Civil considerations, expanded through ASCOPE	Instruments of national power; multinational and interagency partners	Actor analysis with centres of gravity; societal impact and media reaction inside acceptability	Complementary non-military activities and strategic communications required in every COA
Force and resources	Troops and support available	Troop-to-task and backward planning; deployment and sustainment concepts	Forces required across the joint functions	Forces and support allocated; force organisation two levels down
Space	Terrain and weather through the five military aspects	No terrain factor; variation across domains	The area of operations within the lines of operation	Decisive points, and defined parts of the area of operations in the wargame
Conditions rather than criteria	Constraints and restraints from the higher headquarters	Law, policy and rules of engagement	The political authority granted by the North Atlantic Council	Sabor authorisation, contribution caps, force scale

MDMP is the only doctrine that names a mission-variable set, and the other three organise understanding through operational design. PMESII is common to US intelligence doctrine and to AJP-5, and the constructs of objectives, effects, and decisive conditions or decisive points are common to the JPP, AJP-5 and ZDP-5. A criteria structure meant to serve more than one process should be built on that shared material, with METT-TC and ASCOPE treated as one doctrine's decomposition of it rather than as the general case.

The last row of Table 3.1 records a distinction that matters more in a formal method than in a staff process. Every pool holds factors of two kinds: attributes on which options can be scored against one another, and conditions an option must satisfy to be admissible at all. Doctrine keeps the two apart procedurally, through the screening criteria of step 3 and the evaluation criteria of step 5. An additive model keeps nothing apart. It reduces a violated condition to a low score on one term; a strong score on another term then offsets it, and the offset is never declared. Where the conditions are constitutional, as in the Croatian case, the result is not merely inaccurate: the model recommends an option the commander has no authority to

execute. The same conflation follows wherever operational factors are optimised directly. Unit placement scored as a weighted sum over five terrain factors admits the identical silent trade, and where two scenarios differ only in their weights the preferred placement changes with no statement of which trade the change represents [26].

Every framework above is military, and three of the four assume an opposing force. In the civil protection, search and rescue and multi-agency crisis response cases of Section 4 the enemy factor is absent or replaced by a hazard, and what remains of the pool is closest to ASCOPE-PMESII. Whether a single criteria structure can serve both cases is a precondition for the transferable decision support this work proposes, and decision making in a chain of command under a rapid time frame is common to military and emergency contexts alike [27]. The review in Section 6 covers work framed in course-of-action vocabulary, and the civil-security literature on option selection is not necessarily framed that way. The question is therefore carried into the future work rather than answered here. However, combined civil and military training for disaster and crisis management is already conducted on shared modelling and simulation standards [28], which establishes interoperability at the level of the tools without settling whether the criteria structures behind them can be shared.

4 DECISION SUPPORT FOR THE COA PROCESS IN COMMAND AND CONTROL

Command and control (C2) is the exercise of authority and direction by a commander over assigned forces in the accomplishment of the mission, and the planning processes of Section 2 are executed inside it: by a staff, on a battle rhythm, with whatever tool support the headquarters has [6], [29]. Sections 2 and 3 established what a COA is and where its evaluation criteria come from, and Section 5 sets out the machinery for comparing alternatives. Between them sits a practical question: which systems carry COA evaluation, and how much of the doctrinal process do they implement? What follows is organised by class of system rather than by study. Each class is illustrated by fielded or programme-level exemplars, and individual pieces of research appear only as instances of a class; how the research literature of the past decade is distributed is a question for Section 6.

4.1 Decision support system

The term decision support system (DSS) carries a specific meaning that this survey trades on. In the framework of [30], such systems address semi-structured decisions, problems in which part of the task can be formalised and computed while the judgement that completes it cannot be specified in advance, so the system supports the decision maker rather than replacing them [30]. [31] adds the architecture, a data component, a model component and an interactive dialogue between user and system, and the orientation, effectiveness of the decision rather than efficiency of the process. COA selection is semi-structured in exactly this sense: the performance of each alternative can be computed or simulated, but the evaluation criteria, their weights, the risk attitude and the final choice are the commander's. That supplies the yardstick for what follows. A system supports the COA decision to the degree that it reaches the comparison and selection step with the commander in the loop, not merely the documentation, simulation or generation that precedes it.

C2 doctrine treats planning as continuous rather than episodic. The operations process runs planning, preparation, execution and assessment concurrently, and a headquarters sequences its recurring meetings, briefings and reports on a battle rhythm that ties the planning cycle to the rhythm of execution [6], [24]. Boyd's observe–orient–decide–act (OODA) loop [32], [33], the most widely borrowed model of the decision cycle, makes the same point at the level of the single decision: orientation is revised as observation changes, and a decision holds only until the situation that produced it has moved [32], [33]. The consequence for the COA construct is direct. Doctrine supplies branches, sequels and decision points precisely because a selected COA is expected to be revisited during execution, yet the selection procedures of Section 2 and the comparison methods of Section 5 are both described as one-shot exercises. How far tool support closes that gap is the question this section carries into the review.

4.2 Fielded planning suites and constructive simulation

The first class is the planning suite embedded in doctrine, and NATO's Tools for Operations Planning Functional Area Services (TOPFAS) is the clearest example. As described at its introduction, it supports the operations planning process end to end, from systems analysis of the operating environment through operational design capture to plan development and force activation, on a common database from which the major planning documents are generated [34]. Current AJP-5 still lists it as the tool support for all phases of the process [4]. Its character should be stated precisely. TOPFAS supports the recording, sharing and documentation of planning products, the operational design, the order of battle and the synchronisation of tasks;

the COA comparison it supports is the doctrinal staff procedure of Section 2, captured and briefed, not a computational evaluation of alternatives. For Croatian planning conducted within the Alliance the same suite is the reference environment, since ZDP-1 gives ratified Allied doctrine precedence in NATO-led operations [15] and ZDP-5 mirrors the process the suite implements [5]. What the Croatian Armed Forces (Oružane snage Republike Hrvatske, OSRH) use for purely national planning is not documented in open sources.

Between the documentation suite and the research prototypes sits the fielded constructive simulation. MASA SWORD is taken here as the exemplar of that class because it is the one documented in the academic literature rather than only in vendor material: an aggregated constructive simulation with scenario-generation and analysis tools, in which platoon- and company-level units execute orders autonomously through the vendor's Direct AI behaviour engine, used for staff training, planning support and operations research up to division level [35]. For the COA process its significance is that it gives the analysis step of Section 2 a computational form. Candidate COAs can be played in silico against an automated opponent and compared on simulated outcomes, which is the modelling-and-simulation technique FM 5-0 lists alongside manual wargaming [2]. What the simulation does not supply is the comparison machinery itself; simulated outcomes still enter a decision matrix, or a judgement, through the staff procedure of Section 2. Two qualifications attach to this class. Interoperating national C2 systems with national simulations is a standards problem in its own right, addressed through command and control systems to simulation systems interoperation (C2SIM) [36] with the Coalition Battle Management Language (C-BML) [37] and Military Scenario Definition Language (MSDL) [38] representations and the successive NATO Modelling and Simulation Group (MSG) activities MSG-048 [37], MSG-085 [39] and MSG-145 [40], and the longevity such environments require is visible in country-scale simulation toolsets maintained across two decades [41]. The second qualification weighs more. The outcomes a constructive simulation returns carry a measurable error against the reality they stand for: manoeuvre fidelity measured against real exercise trajectories diverges by amounts that can be quantified [42], and those divergences correlate with the error of the decision support built on top of them [43]. A performance table filled from simulation therefore inherits an error term that none of the comparison methods of Section 5 represents.

4.3 Research prototypes and AI-based planners

AI planning techniques have been applied to the tactical process since the mid-1990s. CADET [44], evaluated with the US Army's Battle Command Battle Laboratory, took a COA sketch and statement and expanded it into a synchronised plan: task decomposition, allocation and scheduling against available assets, predicted enemy actions and counteractions, and estimates of timing, logistics consumption, attrition and risk [45]. Controlled experiments comparing brigade planning with and without such aids found that aided planning produced products of comparable quality in substantially less time [46]. DARPA's Deep Green [47], launched in 2008, aimed further forward. Rather than accelerating deliberate planning it sought to generate and analyse options ahead of need during execution, so that when the situation reached a decision point the commander already held evaluated alternatives. The programme ended without fielding a system. It matters here less for what it delivered than for what it shows, that execution-time COA revision has been recognised as the hard problem at system level for nearly two decades.

Two more recent systems show where the class now sits. An AI-enabled wargaming prototype recommends COA improvements during COA analysis under the constraints practitioners face: sparse training data, limited tactical computing, and the need for user interaction throughout

[48]. COA-GPT, a large-language-model planner, generates candidate COAs from mission information and doctrine excerpts within seconds and refines them against commander feedback, evaluated in a militarised StarCraft II environment [49], [50]. What both compute is a set of options and an estimate of how each would fare. Where comparison appears it is by simulated outcome rather than through the multi-criteria structure of Section 5, and the commander acts on the result as a reviewer of machine-produced options. The class has widened since. LLMs have been assembled into hierarchical agent architectures mirroring the OODA loop, returning a scheme of objectives and tasks from a mission statement [51], and the broader question of what such models contribute to command, control, communications, computers, intelligence, surveillance and reconnaissance (C4ISR) has attracted survey treatment [52]. Naval work on human-autonomy teaming states the problem in its sharpest form: where several agents each propose a course of action, an arbitration function must resolve the conflict, which places the machine rather than the commander at the point of choice [53]. The same tendency appears in mission planning for autonomous systems, where a scheduler decomposes an issued tasking into one optimal plan and, on the developers' own account, showed operators only the best plan rather than the set [54]. How such systems should be organised and what data they require has been treated separately [55], as has the interface through which candidate options are put in front of a commander [56].

4.4 Emergency management systems

The civil security domain has its own systems, and Section 3.3 left open whether one criteria structure can serve it and the military case alike. Emergency management information systems address the same decision problem under different names: response options rather than COAs, incident commanders rather than military ones, and a multi-agency setting in which information sharing is the binding constraint. What they deliver is typically a common operational picture assembled from dispersed reporting, resource and asset tracking, incident logging, and messaging between agencies that do not share a command structure. Support of this kind is characteristically dynamic, the decision being made, evaluated and revised as the event unfolds rather than optimised once [57]. That makes the temporal character of the problem more explicit than in any of the military classes above.

Three differences change what a criterion can be about. The first is the absence of an adversary. A hazard does not react to the plan chosen against it, which removes the action-reaction-counteraction structure on which the wargame of all four doctrines rests, and with it the most likely and most dangerous adversarial COA against which options are scored throughout Section 2. What replaces it is a forecast or a hazard model, so the uncertainty attaching to an option is uncertainty about the event rather than about an opponent's choice. The second is the absence of a single commander. Multi-agency response distributes authority across organisations with separate statutory mandates, whereas the doctrinal processes of Section 2 all terminate in one commander's decision and the methods of Section 5 assume a single decision maker, or a group whose preferences have been reconciled into one structure. That is the sharpest obstacle to a single criteria structure serving both domains. The third is that statutory and life-safety obligations behave in the civil case as the national conditions of Section 3.2 behave in the Croatian one: they bound the set of admissible responses rather than distinguishing between the responses within it.

At the level of tools the boundary between the two domains is less sharp than the doctrinal separation suggests. The constructive simulation of Section 4.2 is fielded for emergency preparedness as well as for defence, and its vendor sells a separate crisis-management simulation built on the same behavioural engine [58]. That system is described as testing

alternative response strategies during an event and computing the outcome of each, after which the crisis team selects one and executes it [59], and it has been used to validate flood contingency plans for the Greater Paris Region and to train the crisis staff who would execute them [60]. The description is the COA process in civil vocabulary. The point at which that system stops is the point at which its military counterpart stops: it computes outcomes for a set of alternatives and returns the choice among them to a person. Two classes, in two domains, on the same architecture, stopping in the same place.

What these systems typically lack is what the military classes lack: a structured comparison layer. The operational picture and the communications are served first, and the choice among response options remains a human procedure conducted outside the tool. What the two domains share is therefore not a solution but an obstacle, which is the more useful basis on which to ask, as Section 3.3 does, whether one criteria structure can serve both. The civil literature supports each of these claims on its own terms. Collaborative crisis management has been reviewed as an information and knowledge problem [61], the decision making structure of cyber incident response has been examined for what it implies about crisis management generally [62], and what-if analysis and scenario planning are both established devices for exploring alternative responses ahead of an event [63], [64]. Where the civil work does reach comparison it does so on the same architecture as the military case: game-theoretic selection among defensive strategies [65], comparison of intervention packages on diffusion and belief-shift metrics in the information environment [66], and, closest to the machinery of Section 5, multiobjective contrastive explanation of why one response option is preferred to another across conflicting objectives [67].

4.5 Comparison across class systems

Section 4.1 stated the test: a system supports the COA decision to the degree that it reaches comparison and selection with the commander in the loop. Table 4.1 applies it to the four classes.

Table 4.1 Classes of decision support and the point at which each stops

Class	Exemplar	What it computes	Where it stops	Commander's role
Doctrinal planning suite	TOPFAS	Nothing; it records, shares and documents the planning products	Before comparison: the comparison it carries is the staff procedure of Section 2, captured and briefed	Author and approver of the products
Constructive simulation	MASA SWORD	Simulated outcomes for each COA played against an automated opponent	At analysis: the outcomes still enter a matrix, or a judgement, outside the tool	Interpreter of simulated outcomes
AI planner or generator	CADET, Deep Green, COA-GPT	Task decomposition, scheduling, timing, logistics and attrition estimates; candidate COAs	At generation and analysis: comparison, where present, is by simulated outcome	Reviewer of machine-produced options
Emergency management system	MASA SYNERGY	The operational picture and the communications; simulated outcomes for alternative response strategies	Before comparison: the choice among response options is made outside the tool	Decision maker without tool support at the point of choice

Two findings follow, both about the tools rather than about the research on them. The first is that no class reaches the comparison step. Documentation stops before it, simulation stops at

the analysis that feeds it, generation stops at producing the alternatives to be compared, and the civil systems stop at the picture. The step doctrine describes in most detail, and the step at which the commander's preferences enter, is the one no class implements. The second concerns the question carried forward from Section 4.1: how far tool support closes the gap between one-shot selection and the continuous revision that branches, sequels and decision points presuppose. Nothing fielded closes it. Deep Green named execution-time COA revision as the problem in 2008 and ended without fielding a system [47], and the only class treating the decision as continuous rather than episodic is the civil one [57]. Section 6 tests whether the past decade has moved either.

4.6 Human control and responsible use

In this section every system supports a decision-making process for the use of force and the position assigned to humans within those systems are both matters of legal/philosophical doctrine prior to being an issue of design/engineering. There are three frameworks applicable to the assignment of positions to humans, and they do not all provide the same direction.

The terms "meaningful" and "human control" have become the standard terminology used for what was formerly referred to as "human-in-the-loop". Meaningful human control can be defined as two distinct requirements. First, the human must possess sufficient context to understand the decision the system is suggesting. Second, accountability for the outcome of the decision must be traced back to the individual responsible for making that decision [68]. Neither requirement will be met by a system that recommends a singular choice without also providing the rationale for that recommendation; specifically, the criteria used, the weights applied and the trade-offs made in arriving at the recommendation. If a reviewer is simply shown a single rank ordered list of options with no explanation of how each option was determined, then the reviewer is not making a decision, but rather validating a decision that has already been made.

NATO adopted six Principles of Responsible Use for artificial intelligence in defence in October 2021: lawfulness, responsibility and accountability, explainability and traceability, reliability, governability, and bias mitigation [69]. Three of these affect how we do the comparison part of the process. Explainability and traceability require us to be able to understand what the output says and to be able to audit how the methodology arrived at those results. This is required for both the aggregation rule (as explained in section 4.2) and not just the way you choose to display things. Responsibility and accountability will also require an individual who has been responsible for making the decision, to have had the ability to make that decision differently. Governability will require the behaviour of the system to be interruptible, as with recommendations, this can be achieved through revising recommendations based upon changing situations. As outlined in Section 4.5, there were no fields of data collected where any fielded class was defined.

The European Union's Artificial Intelligence Act excludes all systems that fall within one of three categories; systems that are "placed on the market", systems that are "put into service" or systems that are 'used exclusively for military, defence or national security purposes'. Therefore, a decision support system such as the one this research aims to build would therefore be excluded from the risk assessment and transparency obligations as described in [70].The exclusion removes a regulatory floor; it does not remove a design constraint. What remains binding is the law of armed conflict, which governs the decision the system supports however that decision was reached, and the national authorisation framework of Section 2.4, which for Croatian forces bounds the option space constitutionally before any comparison begins.

For the design of a comparison layer, the implications are specific rather than exhortatory. A system that identifies its criteria and provides explanations for its weighting and their origin, maintains non-compensatory judgments separate from the compensatory total, and provides a report detailing the range of assumptions under which the recommendation provided still holds good will meet the explainability and accountability requirements as a natural result of being a defensible multi-criteria model. On the other hand, a system that simply provides scores cannot claim compliance and adding an interface after the fact will not rectify this issue.

5 MCDA

Multi-criteria decision analysis (MCDA) is the branch of operational research concerned with choosing, ranking or classifying alternatives that are judged on several criteria at once, where the criteria conflict and share no common natural unit [1]. Its subject is the ordinary case in which no alternative is best on every criterion. The decision cannot then be read off the data; it rests instead on an explicit statement of how the criteria trade against one another. Buying a car on price, safety and fuel economy is the everyday version, since the cheapest car is rarely the safest. Every method in this section is a disciplined way of making that statement and applying it consistently.

5.1 The multi-criteria decision problem

A multi-criteria problem has a small number of fixed parts. There is a set of alternatives, the objects among which a decision is to be made. There is a family of criteria, each a viewpoint against which the alternatives are judged and each carrying a scale on which performance is measured. And there is a performance table, whose rows are the alternatives, whose columns are the criteria, and whose entries record how each alternative scores on each criterion [1]. Everything a method does begins from this table.

A distinction is drawn at the outset. In some problems the alternatives are finite and listed explicitly, a shortlist of suppliers or a set of drafted plans; this is multi-attribute decision making, and it is the case treated here. In others the alternatives are defined implicitly by constraints and are effectively infinite, as when the best mix of outputs is sought within a production limit. That is multi-objective optimisation, and it needs different machinery [71].

What counts as a solution depends on the question asked. [72] set out four forms the recommendation can take: choosing a single best alternative, sorting the alternatives into predefined ordered categories such as accept, review and reject, ranking them from best to worst, and simply describing them so that a decision maker can judge for themselves. Most methods are built for one or two of these. The problematic is therefore settled first, before any method is chosen.

Three questions then run through every method, and the differences between methods are largely differences in how they answer them. The first is how the criteria are weighted, since some matter more than others. A weight can mean two different things, and the two are easy to confuse: it can express the plain importance of a criterion, or it can express a trade-off rate, how much performance on one criterion the decision maker will give up for a unit of performance on another. Value-measurement methods need the second, stricter reading. Treating an importance weight as a trade-off rate is a common source of error [73].

The second question is how the criterion scales are made comparable. Performance measured in casualties, in hours and in numbers of units cannot be combined while it stays in those units, so the scores are first normalised onto a common scale, usually running from zero to one. The rule chosen for this is not a formality. Scores may be divided by the column maximum, stretched between the observed best and worst, or scaled by the length of the column, and different normalisations can change which alternative comes out on top.

The third is compensation. A compensatory method lets strong performance on one criterion redeem weak performance on another; a non-compensatory method refuses this beyond a point and can rule an alternative out for a single bad score whatever its other merits. Neither is right in general. Trading comfort for a lower price is reasonable; trading away a safety minimum is

not. Matching the method's stance on compensation to the problem is the main decision in method selection.

MCDA methods fall into three broad approaches shown in Figure 5.1 [1]. Value-measurement methods give each alternative a single score for its overall worth and rank by it. Outranking methods compare alternatives in pairs and build a relation of relative preference, which may leave some pairs incomparable rather than force an order. Reference-level methods judge alternatives by how close they come to an ideal or to stated targets. The three rest on different assumptions about preference and are willing to conclude different things, which is why none is right for every problem.

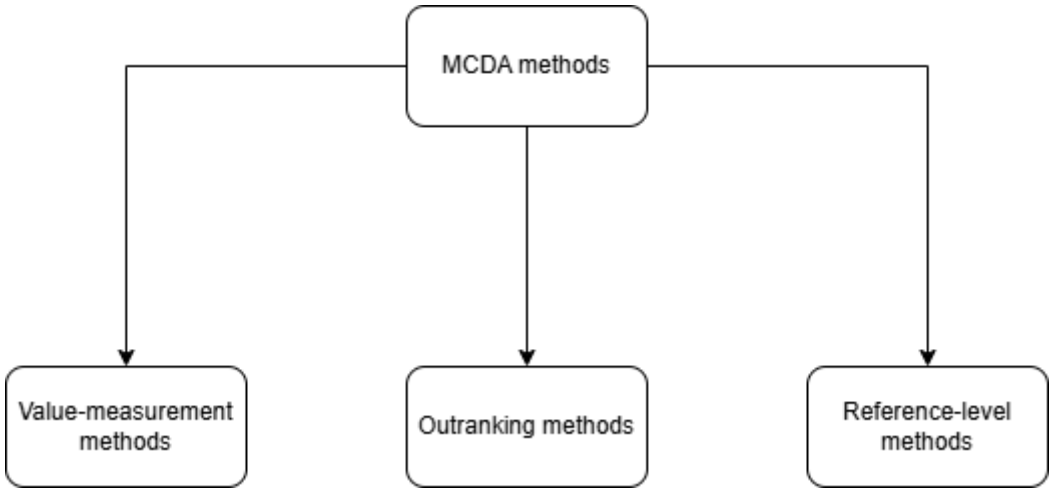


Figure 5.1 Three main MCDA subcategories (drawn by the author)

5.2 Value-measurement methods

Value-measurement methods share one idea: convert each alternative into a single number standing for its overall worth, then rank by that number. The theory behind them is multi-attribute value theory, which sets out when such a number can legitimately be built [73].

The usual construction is additive. Each criterion gets a value function mapping raw performance onto a value between zero and one, zero the worst level considered and one the best, and the overall value is the weighted sum of these single-criterion values, the weights summing to one. The weights are the trade-off constants. The value functions do the real work, turning performances measured in different units into comparable value, and they need not be straight lines: a first hour of delay may cost more value than a fifth.

The additive form is not free. It is valid only when the criteria are mutually preferentially independent, meaning that how much the decision maker cares about one criterion does not depend on the levels of the others, and that has to be checked rather than assumed. Take the value functions as linear and apply the weights to normalised scores, and the method collapses into the weighted sum, or simple additive weighting. It is the most used method in MCDA and the most often misused, because its two strong assumptions, linear value and full compensation, are convenient enough to adopt without noticing.

Where preferential independence fails there is an alternative to abandoning the aggregation. The Choquet integral [71] replaces the additive weights with a capacity, a set function assigning a weight to every subset of criteria rather than to each criterion alone, subject only to monotonicity and normalisation. Two criteria that overlap can then be given a joint weight lower than the sum of their separate weights, so the evidence they share is not counted twice,

and two that reinforce one another a joint weight higher than the sum, so an option strong on both is rewarded for the combination. The weighted sum is the special case in which the capacity is additive. The price is the number of parameters, since a capacity on n criteria has 2^n minus 2 free values, which is why practice restricts it to 2-additive capacities defined by the single-criterion weights and the pairwise interaction indices alone, and reads the result through the Shapley value, which reports each criterion's average contribution across all subsets [74]. This matters here because the criteria families ZDP-5 admits, the joint functions, the principles of war and the elements of operational design, overlap by construction, which is the condition the additive model forbids and the capacity is built to absorb. Section 5.10 records the one military application in the literature reviewed here, and Section 5.9 the option of learning such a capacity from examples rather than eliciting it.

Multi-attribute utility theory [73] extends the approach to decisions under uncertainty, where an alternative delivers not a fixed performance but a range of outcomes with probabilities. Value functions become utility functions whose curvature captures attitude to risk: a cautious decision maker's utility rises quickly and then flattens, so avoiding a bad outcome counts for more than securing an equally large gain. Alternatives are ranked by expected utility. This is the correct treatment when outcomes are genuinely uncertain, at the price of eliciting utilities rather than plain values.

The analytic hierarchy process (AHP) [75] reaches the same kind of overall score by a different route, one many decision makers find easier. Rather than asking for weights and values directly, it asks only for comparisons between two things at a time, with the problem laid out as a hierarchy: the goal at the top, the criteria beneath it, the alternatives at the bottom.

At each level the elements are compared in pairs on a nine-point scale, from equal importance to extreme dominance, producing a matrix of judgements from which the priorities are recovered as the principal eigenvector. Since nobody making many pairwise judgements stays perfectly consistent, the method measures the inconsistency: a consistency index computed from the largest eigenvalue, divided by a tabulated random index, gives a consistency ratio, with values below about 0.1 taken as acceptable. Judgements that fail are revisited before the priorities are used.

The appeal is real: pairwise judgements are easy to make, to explain and to audit. So are the weaknesses. The nine-point scale is a convention with no deeper justification, and the method can suffer rank reversal, in which adding or removing an alternative flips the order of two others that did not change. AHP is best treated as a structured way of eliciting judgement rather than a machine that settles the decision.

5.3 Outranking methods

Outranking methods, developed mainly in the European tradition, refuse the central move of the value-measurement school: they do not build a single score for each alternative [72]. The refusal rests on a doubt: collapsing many criteria into one number hides how the number was reached and forces comparisons the evidence may not support. They compare alternatives in pairs and ask a more modest question of each pair.

The question is whether one alternative is at least as good as another, a relation called outranking, and two tests decide it. The concordance test asks whether a large enough weighted majority of the criteria agree. The discordance test asks whether any single criterion objects so strongly that the claim should be blocked despite the majority. An alternative outranks another only with enough support and no strong objection. The weights here are voting strengths in the

concordance test, not trade-off rates, so they mean something different from value-measurement weights and cannot be carried over unchanged.

Thresholds soften the crisp comparisons of the value-measurement school, acknowledging that small differences in performance are often not meaningful. An indifference threshold sets a difference below which two alternatives count as equivalent on a criterion; a preference threshold sets a larger difference above which one is clearly preferred, with a graded zone between; and a veto threshold sets a difference so large on a single criterion that it blocks an outranking outright, however well the alternative does elsewhere. The veto is what makes the approach non-compensatory.

Outranking methods also allow two alternatives to be left incomparable. Where the evidence favours neither, they say so rather than inventing an order, which is more honest than a forced ranking and less convenient for a decision maker who wants one answer. Two families dominate. ELECTRE (ELimination Et Choix Traduisant la REalité) [72] covers choice, ranking and sorting through a series of variants built on concordance and discordance. PROMETHEE (Preference Ranking Organisation Method for Enrichment Evaluations) [76] gives each criterion a preference function turning every performance difference into a degree of preference between zero and one, then aggregates these into flows: a positive flow measuring how strongly an alternative outranks the others, a negative flow measuring how strongly it is outranked, and a net flow between them. PROMETHEE I [76] ranks on the two flows separately and may leave incomparabilities standing; PROMETHEE II [76] ranks on the net flow and produces a complete order.

5.4 Reference-level and distance-based methods

A third group judges alternatives against a fixed point of reference rather than against each other. The oldest is goal programming, which sets a target level for every criterion and looks for the alternative that comes closest to meeting all of them at once by minimising total shortfall. It suits problems framed in terms of aspirations, where the decision maker can say what would be good enough.

TOPSIS (Technique for order preference by similarity to ideal solution) [77] is the most widely used method of this kind. It builds two reference points out of the alternatives themselves. The positive-ideal solution takes the best observed score on every criterion, a best-of-all-worlds alternative that usually does not exist; the negative-ideal takes the worst. Each real alternative is measured by its straight-line distance to both and scored by relative closeness, so one close to the positive ideal and far from the negative scores near one. The idea is intuitive and the computation light, which explains its popularity. It is fully compensatory and, like AHP, can show rank reversal when the set of alternatives changes.

VIKOR (Višekriterijumska optimizacija i kompromisno rešenje) [78] follows the same compromise logic but measures distance differently, and the difference matters. Where TOPSIS combines the criteria into one straight-line distance, VIKOR balances the overall distance from the ideal against the worst single-criterion shortfall, favouring a solution that is good overall without being badly deficient anywhere. It also reports whether its leading alternative holds a stable and sufficient advantage over the runner-up rather than naming a winner regardless of how close the contest was [78]. That two methods built on the same compromise idea can rank the same alternatives differently, purely because they measure distance in different ways, is a useful reminder that the aggregation rule shapes the result as much as the data do.

5.5 Fuzzy methods

Every method so far assumes a performance table of precise numbers. Real judgements are often not precise: an assessor may say that a cost is about three units, or that risk is high, without committing to a figure, and forcing such judgements into crisp numbers throws away the vagueness that was part of the information. Fuzzy methods keep it and carry it through the calculation [79].

The tool is the fuzzy set, in which an element belongs to a set not yes or no but to a degree between zero and one, given by a membership function [80]. A vague quantity such as about three becomes a fuzzy number, most often triangular, defined by a lower bound, a most likely value and an upper bound, with membership climbing from zero to one and falling back to zero. A trapezoidal number, with a flat top, represents a vague range rather than a vague point. Linguistic terms such as low, medium and high are handled by assigning each a fuzzy number in advance, so words entered by an assessor become quantities the method can compute with.

With performances, and often weights, expressed as fuzzy numbers, the aggregation runs in fuzzy arithmetic, which propagates the imprecision: each alternative ends with a fuzzy overall score that is a range rather than a point. To rank them, these scores are usually turned back into ordinary numbers by defuzzification, the simplest rule being the centroid, which for a triangular number is the average of its three values.

Value-measurement, outranking and distance-based approaches all have fuzzy versions. Fuzzy AHP [81] is among the best known, with the pairwise judgements made in fuzzy numbers and the priorities extracted by a technique such as extent analysis; fuzzy TOPSIS [82] computes the ideal and anti-ideal points and the distances to them in fuzzy arithmetic. The gain is a more faithful treatment of imprecise or linguistic input and a natural fit for group decisions where assessors disagree within a range. The cost is added complexity and fresh modelling choices, the shape of the fuzzy numbers and the defuzzification rule among them, each of which can affect the ranking.

Interval, grey and intuitionistic-fuzzy extensions push the same reasoning further, recording a degree of doubt or non-membership alongside membership and trading more complexity for a finer account of uncertainty. For most practical purposes triangular fuzzy numbers with a simple defuzzification are enough.

5.6 Eliciting weights and value functions

Sections 5.2 to 5.5 take the weights and the value functions as given. Where they come from is a separate body of work, and it is precisely the part the doctrines of Section 2 leave blank: all four require the commander to approve criterion weights and none states how the numbers are produced.

Direct Rating (DR) [83] is the most basic method and the most commonly employed. SMART (Simple Multi-Attribute Rating Technique) [84] uses a similar concept to DR, however SMART requests the decision maker assign the least important criterion a score of 10 and then request the decision maker to evaluate the other criteria relative to the least important criterion. Once these scores are determined, SMART normalizes the scores so that they sum to one. SMARTER (SMART Exploiting Ranks) [85] is an improvement to SMART that replaces numerical scoring with ranking. In SMARTER, the decision maker provides a ranking of the criteria, and SMARTER computes the weights using this rank-ordering as rank-order centroid weights. Rank-order centroid weights are the centroid of all of the possible sets of weights that could have been generated from a particular rank-ordering of the criteria [85]. SMARTER

purchases robustness at the expense of precision, and it is suitable for use within environments where commanders will agree to an ordering, but not agree to a ratio of one ordering to another

SWING [86] weighting is designed to address the ambiguity discussed in section 5.1, between importance and trade-off rate. SWING begins by asking the decision maker to identify a hypothetical alternative at the worst level for each criterion and ask them which criterion they would want to improve to its best level first; once the decision maker identifies which criterion they wish to move first, it receives a 100 point score, while each remaining criterion is evaluated against this criterion to determine what their own "swing" from worst to best level is worth. This makes the resultant values trade-off constants, instead of expressions of importance. Since this question involves a specified range over which the decision maker evaluates swings, the values obtained through this question represent trade-off rates, whereas an expression of importance is needed in order to satisfy the additive model [66]. Trade-off questions explicitly inquire how much of one criterion the decision maker will trade off for a stated increase in another. As such, Trade-off is the most difficult type of elicitation information to obtain and is supported by the most substantial theoretical basis. Therefore, whether or not to use a particular method is not simply a matter of administrative detail since a decision-maker may generate substantially different weights using DR and using SWING questions [63], [80].

MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) [87] exists between AHP and Direct Rating. Similar to AHP, MACBETH uses only pairwise judgments; however, unlike AHP, MACBETH uses only qualitative judgments on a seven-point scale (from No difference to extremely different in terms of attractiveness), and linear programming translates these into an interval scale (MACBETH checks for consistency during this process and indicates where inconsistencies occur). MACBETH derives both value functions and weights from the same type of questions, making it useful for conducting an elicitation under time pressure with a decision maker who prefers not to use numbers.

The disaggregation approach reverses the direction of the exercise. Rather than eliciting the parameters and computing the ranking, UTA (UTilités Additives) [88] takes a ranking the decision maker has already given over a small set of reference alternatives and infers, by linear programming, additive value functions that reproduce it. ROR (Robust Ordinal Regression) [89] generalizes this by not selecting one such model when multiple models are compatible with the stated ranking; ROR exploits all of the compatible value function models to report necessary preference (where one option is ranked higher than another option by every compatible model) and possible preference (where one option is ranked higher than another option by at least one compatible model) [89]. The distinction between necessary preference and possible preference is exactly what a commander needs in order to recommend actions that hold under every model consistent with his/her stated priorities, as opposed to recommending action based upon a single weighted total.

The consequence for the COA process is that a commander required to declare weights before the options exist, which the MDMP, the JPP and AJP-5 all require, is being asked the hardest question in the procedure at the point of least information [2], [3], [4]. A rank order represents a significantly less complicated task for a decision-maker to perform, revise or defend.

5.7 Group decision making

Every method above assumes there is a single decision maker, while doctrine does not. Doctrine does not. FM 5-0 employs a process that requires all members of the Staff to assess each COA based upon their respective functional requirements. JP 5-0 follows a similar procedure. Each staff section compares each COA separately (and may employ different

evaluation criteria) until the staff collectively reaches a consensus; then, the chief of staff/deputy commander reviews and formally approves the final assessment. ZDP-5 utilizes the planning group to perform the task [2], [3], [5]. Section 4.4 demonstrated the same framework applies to civilian cases; however, civilian authority is dispersed throughout various agencies that have separate statutory obligations. As such, what doctrine refers to as "consensus" is essentially an aggregation phase that occurs without any guidelines.

There are two opportunities for group multi-criteria analysis to aggregate. Selecting either opportunity will produce different outcomes [71]. Combining the judgments of multiple assessors produces a single set of input parameters and/or performance scores or pairwise comparisons that can be used to run the methodology once. However, aggregating the priorities of individual assessors allows for running the methodology separately for each assessor and producing a ranked list of options.

In the case of AHP, the conventional approach utilized to combine individual assessments is the geometric mean. This provides preservation of the reciprocal relationship inherent within the comparison matrix. Conversely, the arithmetic mean does not preserve these relationships.

When utilizing the second approach, the issue transitions to being a matter of social choice and applicable results exist. Specifically, Arrow's theorem stipulates that no mechanism exists to combine individual rankings into a group ranking that simultaneously meets a relatively modest set of rational expectations; therefore, any aggregation of staff rankings inherently represents a selection that cannot be defended as neutral [90]. Borda counts, which score each option by the number of alternatives it outranks and sum those scores across assessors, majority rules and rank aggregation each surrender a different one of those conditions. The practical consequence for a headquarters is that the rule has to be declared before it is known which option it favours.

A third route treats the disagreement as information rather than noise. Fuzzy and interval group methods of the kind Section 5.5 describes keep the spread of the assessors' judgements in the calculation instead of averaging it away, so an option on which the staff is divided ends with a wider score than one on which it agrees. Consensus-reaching processes go further, running the aggregation iteratively and feeding back to the assessors whose positions lie furthest from the group until a consensus measure passes a declared threshold [91].That is a formal version of what a chief of staff does when the sections disagree.

Neither of these aspects displace the commander's ultimate decision-making responsibility nor were they intended to. Rather, they provide what doctrine previously left unaddressed: a clear articulated rule governing how to convert multiple functional assessments into a single performance table as well as documentation illustrating how much weight was placed upon each assessor's assessment versus how much weight was placed upon the rule that ultimately converted the assessments.

5.8 Decision making and deep uncertainty

The use of uncertainty to support decision-making is treated in sections 5.2 through the attachment of probabilities to outcomes and section 5.5 through the allowance of vagueness within assessment. Neither treatment is satisfied by the performance table used as part of the COA process. Entries in this table result from a wargame, therefore represent a structured application of judgment against an adversary whose actions are simulated (and thus observable) rather than simply observed, and AJP-5 refuses to allow these results to either validate or invalidate a particular option [4]. Thus, there exists no frequency distribution upon which to base a probability and no mutually accepted distribution upon which to be vague. This

represents "deep" uncertainty; i.e., the situation in which the parties involved cannot agree on a description of the system, nor on the appropriate probability distributions for its input parameters, nor on how to evaluate the potential outcomes [92]. Three approaches exist to treat such deep uncertainty, and each addresses a different aspect of the problem.

Info-gap decision theory drops probability and asks instead how wrong the best estimate can be before a decision fails [93]. Around a nominal estimate it defines a nested family of uncertainty sets indexed by a horizon of uncertainty, and for each option computes the robustness, the largest horizon at which the option still meets a stated minimum performance. Options are compared on robustness rather than on expected performance, and the two orderings frequently disagree, the option with the best nominal score often being the least robust. Applied to a performance table filled from a wargame, the question becomes how far the wargame estimates would have to be wrong before the recommended COA stopped meeting the commander's minimum. That is a quantity a staff can compute, and a commander can be told.

Robust Decision Making inverts the usual order of analysis [92]. Rather than first making predictions and then acting, Robust Decision-Making subjects each candidate alternative to extensive testing using many possible future states. Statistical clustering is used to identify those combinations of circumstances under which each candidate alternative performs poorly. Then those "vulnerability areas" are presented to the decision-maker for assessment of whether the adverse circumstances identified as having the potential to undermine each alternative are sufficiently realistic as to warrant consideration. Unlike traditional decision-making processes that provide a ranked list of preferred actions, Robust Decision-Making produces a graphical representation of the range of circumstances under which each alternative recommendation is valid or not, along with indicators for determining whether the assumptions underlying each alternative are changing.

Scenario robustness is the least demanding of the three and the closest to existing practice. A small set of scenarios is fixed, each option is scored under each, and the options are compared on a robustness criterion across scenarios: maximin, which ranks by worst case, or minimax regret, which ranks by the largest shortfall against the best option available in each scenario. Doctrine already supplies the scenarios. The most likely and the most dangerous adversarial COA are exactly a two-scenario set, every friendly option is required to be played against both, and the results are recorded separately [3]. Doctrine does not however supply a method for resolving conflicting information resulting from evaluations conducted across both scenarios, therefore if an evaluation demonstrates an option performs well against the most likely threat but poorly against the most dangerous threat, an individual will have to make an adjudication regarding how much weight should be assigned to each type of outcome.

5.9 Preference learning and inverse reinforcement learning

Every method above is based on decision makers specifying something: weights, thresholds, rankings of reference alternatives, minimum acceptable performance levels. Preference learning, instead, infers the preference model from behaviour and that where multi-criteria analysis meets machine learning.

Preference learning is the machine learning of orderings (rather than labels), with its standard tasks being object ranking, instance ranking and label ranking [94]. The methods used in preference learning are pairwise decomposition, learning to rank and monotone classifiers which are used when a preference will increase with respect to a criterion, these statistical techniques represent the disaggregation approaches covered in Section 5.6. UTA and robust

ordinal regression solve a small, tightly structured version of the same problem using linear programming and an assumption about additive form; preference learning solves a larger and loose version with more data and weaker assumptions about the shape of the model. Fitting the parameters of an outranking relation or Choquet capacity of Section 5.2 from examples is now routine, which places the two literatures on the same objects as opposed to mere neighbouring them.

Inverse reinforcement learning carries this idea into sequential decisions. While reinforcement learning is given a reward function and finds a policy; inverse reinforcement learning is given demonstrated behaviour and infers the reward function that would make that behaviour very close to optima [95], [96]. The problem is undetermined since many reward functions can rationalise the same behaviour. The methods that address this problem, maximum entropy formulations, Bayesian approaches and adversarial imitation differ mainly by how they choose amongst the compatible explanations. This is the same move robust ordinal regression makes when it works with all the possible value functions instead of one.

For the COA process both the attraction and objection are obvious. Headquarters generates exactly the type of data that these types of methods used, options developed, scored and rejected, with the commander's selected option recorded after each cycle. A modelling approach that uses that record to determine the criteria and trade-offs applied by a commander in practice provides a framework against which criteria stated by a commander could be compared. The gap between what the commander declared and applies is also a finding. However, there is also an objection. A learned reward function describes previous actions but does not warrant future action, while doctrinal requirements for a recommendation must be defensible beforehand. If used to generate a hypothesis about a commander's preferences to be put back to the commander for confirmation, then it preserves the position doctrine assign to the commander. If used to directly score options then it moves the choices into the model and returns a number without indicating why that was generated, which is also failure Section 1 started from.

Reward specification has the same danger as an additive model with no screen gate. A reward function is a scalarisation. Where goals conflict, the weights implied by the reward function have the same impact as the weights of Section 5.6 and are far less visible. Multi-objective reinforcement learning keeps objectives separate and return sets of policies instead of single policies, which is the sequential equivalent of the pareto treatments covered in Section 6.5 [97].

5.10 MCDA in the military domain and the COA process

MCDA has a long record in defence, but the problems it is usually applied to are not operations planning problems. Procurement, capability planning, siting and equipment assessment share a shape that suits the methods: the alternatives exist before the analysis begins, the attributes are largely technical and measurable, the data can be gathered at leisure, and a criteria set can be reused across programmes [98]. One line of that work bears directly on the methods of Sections 5.2 to 5.9. Threat evaluation in air defence has been treated with a Choquet integral [71] in place of the weighted sum, so that the interaction between criteria described in Section 5.2 is modelled rather than assumed away [99].

The distinction that separates the two cases is between rational decision making, which assumes full information and no time constraint, and naturalistic decision making, in which an experienced decision maker recognises a workable option under pressure without enumerating alternatives [98]. COA selection sits between them. Doctrine mandates the rational apparatus, several options developed, analysed and compared against declared criteria, while the conditions of the decision belong to the naturalistic case: compressed timelines, a reacting

adversary, and criteria written for one mission and discarded after it. The alternatives are created by the staff that will score them, and the choice belongs to a commander who must be able to defend it afterwards. Methods proven on a procurement problem therefore do not transfer unexamined. The naturalistic account has been reviewed with the military case explicitly in view [100].

How far the methods of this section have in fact been applied to operational decisions is a question for Section 6. What can be said from the tools themselves is shown in Table 4.1: the systems of Section 4 create alternatives and estimate their consequences, and none models the commander's preferences over the results, which is the step doctrine leaves to a matrix and a caveat.

Selecting a COA is a discrete multi-attribute problem of the kind these sections describe, and Section 2.5 shows all four doctrines modelling it the same way. The alternatives are the options that survive screening, the criteria are the evaluation criteria, the performance table is filled from the wargame, and the weights are the commander's. Table 5.1 sets out this correspondence in full.

Table 5.1 The COA process read as a multi-criteria problem [2], [3], [4], [5]

MCDA construct	Doctrinal counterpart	Where the four differ
Alternatives	COAs developed in step 3	Two or three in the JPP, capped at the commander's discretion in the MDMP; the object ranges from a scheme of manoeuvre to a design construct
Admissibility filter	The five screening criteria	Same five concepts everywhere; only AJP-5 adds doctrinal compliance, and only the MDMP keeps retesting through step 4
Criteria family	Evaluation or selection criteria	Fixed before the options exist in the MDMP, the JPP and AJP-5; developed in step 4, after them, in ZDP-5
Criterion scale	Unit of measure, benchmark and formula	Prescribed only by the MDMP; the other three leave scale and direction unstated
Performance table	Wargame results recorded by criterion	Binding in the MDMP and the JPP; data points that neither validate nor invalidate an option in AJP-5
Weights	Criterion importance approved by the commander	No elicitation procedure in any of the four; ZDP-5 alone fixes a range, 1 to 4 with the priority criterion weighted 1
Aggregation	Decision matrix or numerical comparison	Weighted sum in three of the four; AJP-5 names rating methods without an arithmetic
Recommendation	Recommended COA with its rationale	Ranking in step 5, choice in step 6, with the commander free to modify or merge options

The three doctrines that supply an arithmetic all supply the same one. The MDMP decision matrix, the JPP weighted numerical comparison and the ZDP-5 numerical comparison are weighted sums, the simplest value-measurement model, and AJP-5's numerical rating is the same model with the aggregation left to the staff [2], [3], [4], [5]. Read against the conditions that model requires, the differences catalogued in Table 2.10 stop being presentational. Additive aggregation presumes values on an interval scale and weights that express trade-off rates, the amount of one criterion the commander will give up for a unit of another [73]. Neither doctrine that scores ordinally meets that condition. The MDMP scores by rank order within each criterion, so the numbers entering the products carry no interval information. ZDP-5 scores on a three-point rating of how far a criterion is realised and weights by an importance ranking in which the priority criterion carries 1, so the product multiplies one ordinal index by another and the totals run in the opposite direction from the JPP. The doctrinal caveats follow

from this rather than from institutional modesty: FM 5-0 warning against conclusions drawn from a quantitative analysis of subjective weights, and JP 5-0 forbidding the arithmetic to be presented as rigorous mathematical analysis, both state the consequence of aggregating ordinal quantities [2], [3]. Table 5.2 records what each rule can and cannot support.

Table 5.2 The doctrinal aggregation rules as value-measurement models [2], [3], [4], [5]

Doctrine	What is aggregated	Model	What the model will not support
MDMP	Ranks from 1 to the number of COAs, weighted by criterion importance	Weighted sum over ordinal ranks, lowest total preferred	Ranks are not interval quantities; the totals change with the option set, since a rank depends on which options are present
JPP	Scores of each COA against a standard, weighted by criterion importance	Weighted sum over criterion scores, highest total preferred	Scores are assigned against a standard with no stated scale, so the weights cannot be read as trade-off rates
AJP-5 / COPD	Cardinal ratings, ordinal ranks, plus/zero/minus, or words	Aggregation unspecified; the same criteria set and weights applied to all options	Nothing is aggregated formally, so the comparison rests entirely on the judgement it was meant to structure
ZDP-5	Points for degree of realisation, weighted by an importance rank from 1 to 4	Weighted sum over ordinal points, lowest total preferred	Two ordinal indices multiplied together; ties in the points scale make differences between options smaller than any meaningful threshold

The weighted sum is fully compensatory, yet several of the things a commander weighs are not tradable. Casualties beyond a level, a breach of the rules of engagement, or a risk outside the commander's stated attitude are meant to remove an option, not to be bought back by speed or economy of force. Each doctrine answers by placing a non-compensatory filter ahead of the compensatory rule, the five screening criteria, with acceptability carrying the load [2], [3], [4], [5]. NATO and Croatian doctrine add a second gate behind the comparison, keeping risk out of the aggregate and judging it separately against a declared risk attitude or acceptability class, so a well-scoring option can still be ruled out after the totals are in. Outranking formalises the arrangement: the veto threshold refuses compensation inside the comparison, and the sorting problematic reproduces the screening step.

The theory is the least ambiguous in the timing of the criteria. Building criteria before the alternatives exist is what value-focused thinking prescribes, and the MDMP and the JPP give the reason in the same terms MCDA does, that criteria drawn up in the presence of the options are shaped by them [2], [3], [101]. ZDP-5 develops its evaluation criteria in step 4, with the options already on the table, and offers no rationale. Its candidate families compound the exposure. The joint functions, the principles of war and the elements of operational design overlap heavily, and a family that is not non-redundant double-counts the same evidence and inflates the weight of whatever it counts twice [1], [73]. None of the four states the requirement MCDA treats as prior to weighting, that the family be exhaustive, non-redundant and preference independent.

The performance table is built from a wargame, so its entries are estimates rather than measurements, and AJP-5 says as much in refusing to let wargame results validate or invalidate an option [4]. Multi-attribute utility theory answers the version of this in which outcomes are random with assessable odds, and the fuzzy methods the version in which the assessments themselves are vague. The insistence that every friendly option be played against the same adversarial options is a procedural attempt at the comparability both treatments assume [3].

Wargaming also serves purposes other than filling a performance table, including the study of allied courses of action as a way of developing the analysts' own understanding, a use in which no option is scored and none is selected [102].

All four doctrines note that a change in the weighting can change the preferred option, and all four treat it as a caution to record. MCDA treats it as a quantity to measure, asking over what range of weights a recommendation survives [1], [103]. An option preferred across the plausible range can be put to the commander with a confidence no single total carries, and it answers the demand ZDP-5 makes explicitly, that the head of the planning group articulate why a particular option is proposed [5].

6 LITERATURE REVIEW

This section reports a systematic review of the research literature on computational and methodological support for the development, analysis, comparison and selection of courses of action. It follows the guidelines for systematic literature reviews in software engineering [104] and is reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [105]. Sections 2 to 5 supply the instruments it uses: Section 2 the phases of the COA process, Section 3 the operational factor frameworks from which criteria are drawn, Section 4 the classes of decision support system, and Section 5 the families of multi-criteria method.

Its scope is military operations planning. Section 4 treated civil security systems as context and Section 3.3 left open whether one criteria structure can serve both domains, but this review does not answer that question. The reason is vocabulary rather than evidence. Course of action is a doctrinal term of art, and the civil-security literature on selecting between response options, which is substantial, does not use it. A search built on that phrase retrieves the military literature and only the fraction of the civil literature that happens to share the term, which is too small and too unrepresentative to compare against.

The review addresses four questions.

- RQ1. Which part of the COA lifecycle does the study address: generation, analysis, comparison, selection, or revision of a selected COA during execution?
- RQ2. Which computational methods are applied, and to which phase of the lifecycle?
- RQ3. Where a study addresses comparison or selection, where do the criteria come from, how are they weighted and reconciled, and is the doctrinal distinction between screening criteria and evaluation criteria preserved?
- RQ4. How have the proposed approaches been evaluated, and what evidence of effectiveness is reported?

RQ1 and RQ2 chart the corpus against the instruments of Sections 2 and 5. RQ3 tests the observation of Section 3, that doctrine describes the pool of operational factors in detail but gives no procedure for deriving evaluation criteria from it. RQ4 establishes how far any of the proposed approaches has been tried against the staff procedure it would replace.

6.1 Review method

Three databases were searched on 29 July 2026. Each search combined the phrase “course of action” and its plural with a military planning context and with terms for computational or methodological support. Table 6.1 records the fields, filters and yield of each.

Table 6.1 Search strategy by database, run 29 July 2026

Database	Fields searched	Filters applied	Records
Web of Science Core Collection	TS=(("course of action" OR "courses of action" OR "course-of-action") AND (military OR defen?e OR "command and control" OR wargam* OR "mission planning" OR "operations planning" OR "crisis management" OR "emergency response" OR "emergency management" OR "civil protection" OR "disaster management" OR "search and rescue"))	PY=(2016-2026); LA=(English); DT=(Article OR Proceedings Paper)	181
Scopus	TITLE-ABS-KEY (("course of action" OR "courses of action" OR "course-of-action") AND (military OR defen?e OR "command and control" OR wargam* OR "mission planning" OR "operations planning" OR "crisis management" OR "emergency response" OR "emergency management" OR "civil protection" OR "disaster management" OR "search and rescue"))	PUBYEAR 2016-2026; DOCTYPE ar, cp; LANGUAGE English	280
IEEE Xplore	("All Metadata":"course of action" OR "All Metadata":"courses of action" OR "All Metadata":"course-of-action") AND ("All Metadata":military OR "All Metadata":defense OR "All Metadata":defence OR "All Metadata":"command and control" OR "All Metadata":wargam* OR "All Metadata":"mission planning" OR "All Metadata":"operations planning" OR "All Metadata":"crisis management" OR "All Metadata":"emergency response" OR "All Metadata":"emergency management" OR "All Metadata":"civil protection" OR "All Metadata":"disaster management" OR "All Metadata":"search and rescue")	Year 2016-2026; Content Type: Journals, Conferences	196

Three decisions about the search vocabulary bound what the corpus can contain, and each was deliberate. The three field scopes are not equivalent: the Web of Science topic field and the Scopus title-abstract-keyword field cover roughly the same ground, while IEEE Xplore's all-metadata mode reaches indexing terms the other two do not. That is one reason IEEE contributes the largest block of records found by a single database. The acronym COA was not searched, because it collides with unrelated senses across engineering and medicine at a volume no title and abstract screen could absorb. Civil-security vocabulary other than "crisis management" was excluded, for the reason given at the start of this section.

The cyber-defence sense of the term, in which a course of action is a sequence of attack or response actions across a network, is documented in [106]. It is the clearest instance of the collision the unexpanded acronym would have introduced.

Table 6.2 shows the eligibility criteria. The two criteria that filtered the most are the requirement that a course of action be the object of study rather than the setting, and the requirement that it be one candidate among alternatives rather than a single plan issued for execution; between them they account for 15 of the 20 full-text exclusions.

Table 6.2 Eligibility criteria

Criterion	Included	Excluded
Object of study	A course of action is the object of the work	The course of action is the setting, or names the output of some other process
Alternatives	The course of action is one candidate among alternatives	A single plan is issued, refined or executed with no alternative
Setting	Military operations planning	Exclusively civil security, cyber defence, business or any non-military planning setting
Contribution	Proposes, applies or evaluates computational or methodological support	Doctrinal publication, secondary review, or ordinary-language use of the phrase
Period	Published 2016 to 2026	Outside the window
Language	Written in English	Any other language
Type	Journal article or conference paper	Book, thesis, editorial, tutorial, keynote abstract

Records were merged and deduplicated in Rayyan [107] before screening, and the databases that returned each record were recorded so that overlap could be reported. Title and abstract screening used three coarse exclusion codes, distinguishing the ordinary-language use of the phrase, the wrong setting and the wrong publication type. At full text the reason was written out in full for each excluded report, as PRISMA requires.

Screening and extraction were carried out by a single reviewer. The exclusion codes used at title and abstract were fixed before screening began and applied uniformly, the reason for every full-text exclusion was written out and is reported in Table 6.5, and the extraction form was piloted on twenty studies before the remainder were charted. The review covers peer-reviewed scientific publications only, journal articles and conference papers, as Table 6.2 records. Literature that was not searched includes: theses and dissertations, technical reports, doctrinal publications, and vendor documentation etc. Material of that kind cited in Sections 2 to 4 was gathered separately and forms no part of the corpus. A substantial part of the work on operations planning support is done inside defence organisations and reported in technical reports and at conferences whose proceedings are not indexed in the three databases searched, and a national-language thesis would be the most likely place to find Croatian work on this problem. What follows is therefore a sample of the openly published, indexed and English-language literature rather than of the work being done.

The extraction form has 18 fields, listed in Table 6.3, and was piloted on twenty studies before the remainder were extracted. Two fields admit more than one value per study, the lifecycle phase of RQ1 and the method family of RQ2, so counts reported against them exceed the number of studies. On the criteria fields “not stated” is recorded as a value rather than left blank.

Table 6.3 Data extraction form

Field	Question	Value set
Setting	Domain of the work	military; dual-use; unclear
RQ1 Phase(s)	Which part of the lifecycle	generation; analysis; comparison; selection; revision (multiple allowed)
RQ2 Method family	Class of computational method	17 families or none (multiple allowed)
RQ2 Specific method	Named technique or system	free text
Human position	Where the decision maker sits	absent; reviewer of machine output; at comparison and selection
RQ3 Criteria source	Where the criteria come from	doctrinal framework; ad hoc set; elicited from stakeholders; not stated; not applicable
RQ3 Doctrine named	Which planning doctrine is cited	MDMP; JPP; COPD/AJP-5; national; other; none
RQ3 Weighting	How criteria are weighted in the loop	elicited; explicit numeric; equal; unweighted; not stated; not applicable
RQ3 Aggregation	How criterion scores are reconciled	weighted sum; Pareto dominance; probabilistic inference; qualitative synthesis; distance to reference point; not stated; not applicable
RQ3 Screening vs evaluation	Whether the doctrinal separation survives	distinguished; conflated; screening only; evaluation only; not applicable
RQ4 Research type	Wieringa facet	solution proposal; validation research; evaluation research; experience paper; opinion paper
RQ4 Practitioners	Were practitioners involved	yes; no
RQ4 Baseline	What the work is compared against	none; vs another method; vs unaided human judgement
QA1 to QA5	Quality items	1; 0.5; 0

Each study was assessed on five quality items scored one, one half or zero, giving a maximum of five. Table 6.4 gives the instrument together with the distribution of scores across the corpus. The scores qualify the findings for RQ4 and were not used to exclude any study.

Table 6.4 Quality assessment instrument and results across the 45 included studies

Item	Question	Met	Partly	Not met	Mean
QA1	Are the aims or objectives clearly stated?	44	1	0	0.99
QA2	Is the approach described in enough detail to reproduce?	6	33	6	0.50
QA3	Is an empirical evaluation present?	19	19	7	0.63
QA4	Is the evaluation method described and appropriate to the claim?	2	18	25	0.24
QA5	Are limitations or threats to validity discussed?	14	16	15	0.49

6.2 Search outcome and corpus

The three searches returned 657 records. Removing 195 duplicates left 462 unique records, of which 367 were excluded at title and abstract: 272 for the wrong setting, 79 for the ordinary-language use of the phrase and 16 for the wrong publication type. Of the 95 reports sought, 30 could not be obtained: no full text was available through the institutional subscriptions accessible for this review and none was reachable in open access, through the publisher or through a repository. That is close to a third of the reports reaching the retrieval stage, and the loss is unlikely to be random, since the proceedings of defence-oriented societies are the least

likely to fall inside a general university subscription and the corpus is conference-heavy. Sixty-five reports were assessed at full text, of which 20 were excluded. 45 studies were included. Figure 6.1 shows the complete selection workflow.

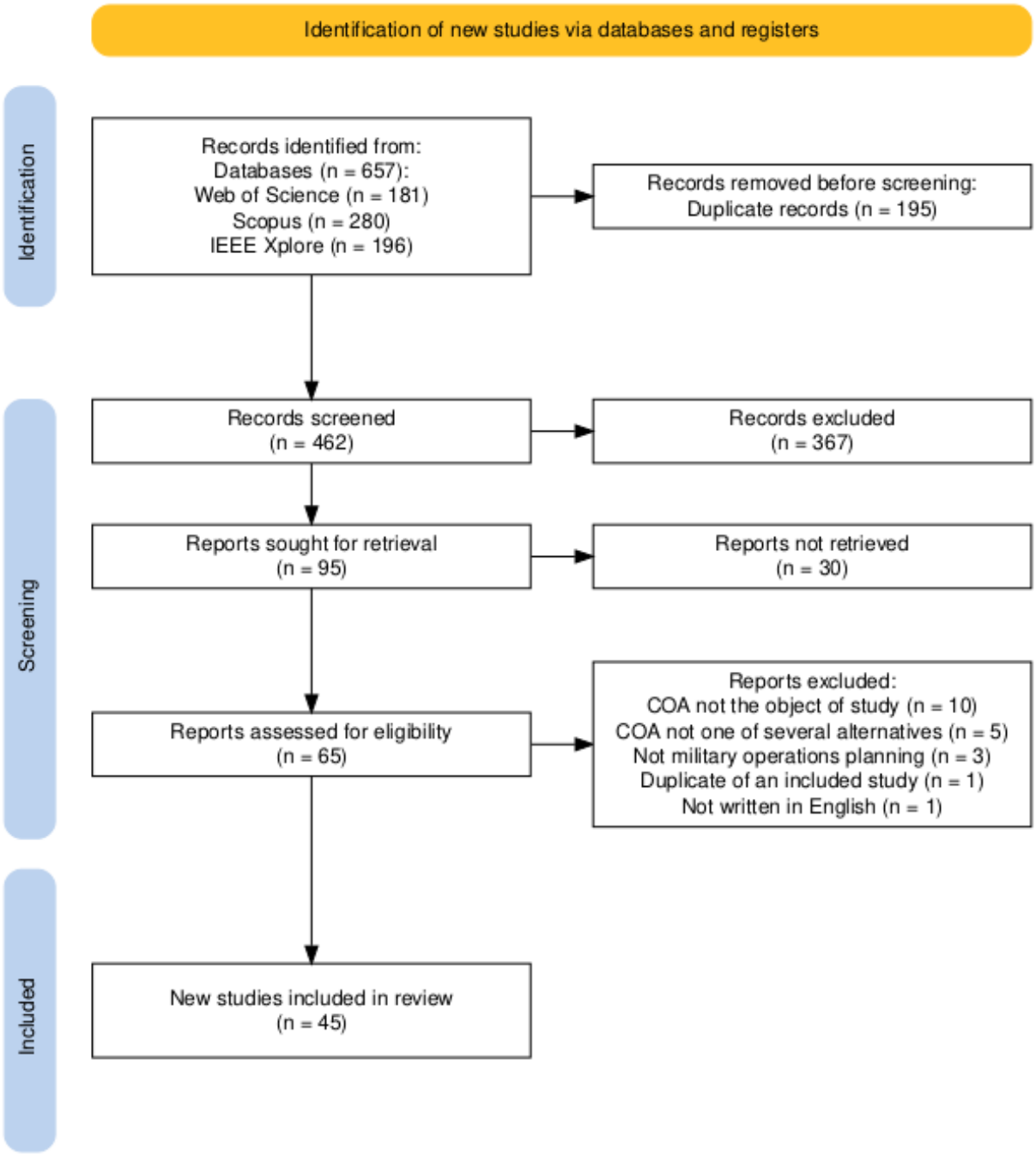


Figure 6.1 PRISMA flow diagram ([108] used to create the diagram)

Table 6.5 groups the full-text exclusions. Five reports were removed because the method returns a single plan rather than a set of alternatives, ten because the course of action was the setting rather than the object of study, and three because the setting was not military operations planning. One is a conference predecessor of an included study and was counted once with it.

Table 6.5 Reports excluded at full text, by reason

Reason for exclusion	n	Reports
COA not the object of study	10	[25], [36], [42], [43], [53], [55], [56], [102], [109], [110]
COA not a candidate among alternatives	5	[26], [51], [54], [111], [112]
Not military operations planning	3	[65], [66], [67]
Duplicate publication of an included study	1	[113]
Not written in English	1	[114]

One of these deserves separate notice. [114] is an evaluation framework for courses of action based on wargaming, eligible on every criterion except that it is written in Chinese with only its metadata in English. Scopus indexes that metadata in English, so the language filter did not catch it. It is the single substantive loss the English-language restriction imposed on this review.

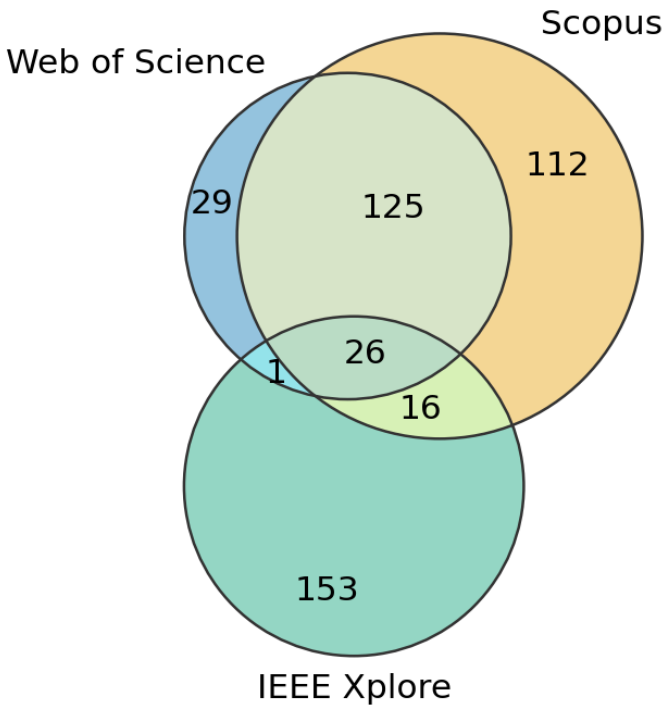


Figure 6.2 Database overlap across the 462 unique records

The overlap in Figure 6.2 justifies the three-database design. Only 26 records were returned by all three services and 294 of 462 by one alone, so any single-database search would have missed the majority of the pool. IEEE Xplore contributes the largest exclusive block, which follows from its all-metadata search mode and from the conference-heavy character of the literature.

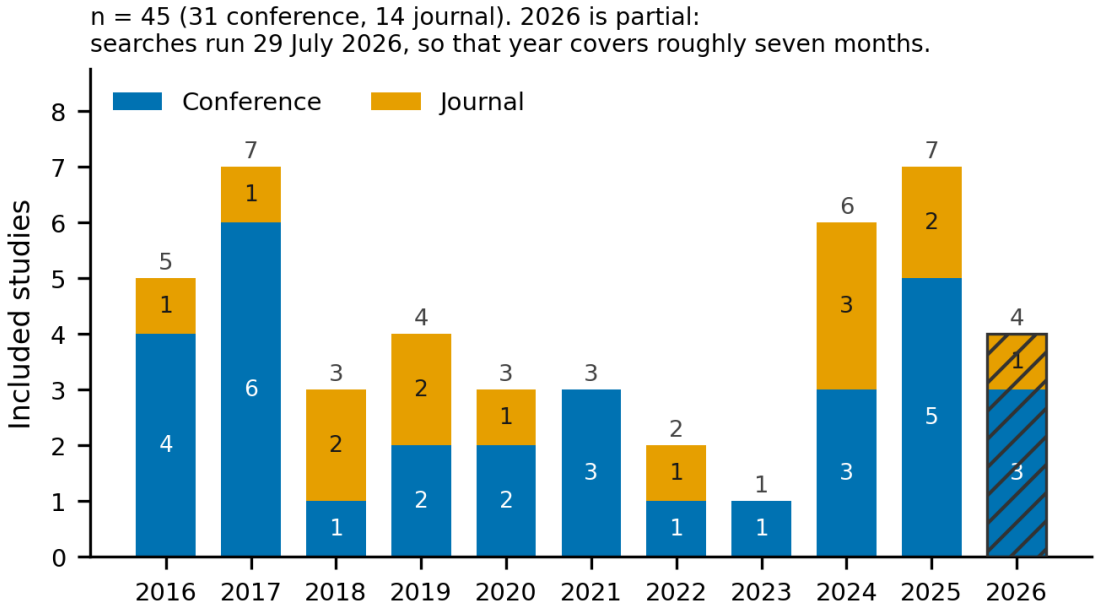


Figure 6.3 Included studies by year of publication and publication type

31 of the 45 studies are conference papers and 14 are journal articles. 42 studies are set in military operations planning proper. Three are dual-use, naming military application among several: a generator of manoeuvre options for unmanned platforms [115], a mobile common operational picture built for defence and civil protection alike [116] and a state-actor model applied to intervention analysis [117].

The distribution across time falls from seven studies in 2017 to one in 2023 and recovers to six and seven in 2024 and 2025. The recovery coincides with a change of method rather than a return of the earlier work, as Section 6.4 shows.

Table 6.6 gives the corpus in full: each study with its phases, method families, the position it gives the human decision maker, its research type, what it compares itself against and its quality total. The sections that follow read that table one question at a time.

Table 6.6 The 45 included studies. Phases: G generation, A analysis, C comparison, S selection, R revision. Method families: SO search-optimisation, SIM simulation, PGM probabilistic graphical model, RL reinforcement learning, AP automated planning, KR knowledge, CBR-case-based reasoning, NLP-natural language processing, MCDA-multi-criteria decision analysis, VIS-visualisation, GT-game theory, IF-information fusion, MAS-multi-agent system, LLM-large-language model, NN-neural network, ML-machine learning, XAI-explainable AI – Part I

Ref	Yr	T	Phases	Method families	Human	Research type	Baseline	Pr	QA
[118]	2016	C	G	KR	In loop	Proposal	—	y	3
[119]	2016	C	AC	SIM	In loop	Validation	—	—	3.5
[120]	2016	C	GAC	SIM	In loop	Proposal	—	—	2
[121]	2016	C	GAC	SO, CBR, NLP	In loop	Validation	Method	—	3
[122]	2016	J	CS	MCDA	In loop	Validation	Human	y	4.5

Table 6.7 The 45 included studies – Part II (phase and method-family legend as for Part I)

Ref	Yr	T	Phases	Method families	Human	Research type	Baseline	Pr	QA
[123]	2017	C	GC	SO	Absent	Validation	Method	—	3
[124]	2017	C	GC	SO, SIM	Absent	Proposal	—	—	2
[125]	2017	C	A	SIM	In loop	Experience	—	y	2.5
[126]	2017	C	A	SIM	In loop	Opinion	—	—	2.5
[127]	2017	C	AC	SIM	In loop	Validation	—	—	3
[113]	2017	J	GACSR	PGM, SO, ML	In loop	Proposal	Method	—	2
[128]	2017	C	AC	PGM, SO	Absent	Validation	Method	—	3.5
[117]	2018	J	GAC	SIM	In loop	Validation	—	—	4.5
[129]	2018	C	GCS	KR	Absent	Proposal	—	—	1
[130]	2018	J	GCS	—	In loop	Proposal	—	—	2.5
[131]	2019	C	CS	VIS	In loop	Validation	Method	y	5
[132]	2019	J	ACS	GT	Reviewer	Proposal	—	—	1.5
[133]	2019	J	G	—	In loop	Evaluation	—	y	3.5
[134]	2019	C	GA	—	In loop	Proposal	—	—	2.5
[116]	2020	C	G	CBR, IF	Reviewer	Proposal	—	—	1.5
[135]	2020	C	GC	GT	In loop	Proposal	—	—	2.5
[136]	2020	J	GCS	SO	Absent	Validation	Method	—	3.5
[137]	2021	C	CSR	PGM	Reviewer	Proposal	—	—	1.5
[138]	2021	C	ACS	IF	Reviewer	Proposal	—	—	2
[139]	2021	C	AC	SIM, RL	In loop	Validation	—	—	3
[140]	2022	C	GCS	MAS, AP, SO	Absent	Proposal	Human	y	2
[141]	2022	J	GC	NLP, KR, SO	Reviewer	Validation	—	y	4
[142]	2023	C	GC	SO, AP	Absent	Validation	Method	—	3.5
[143]	2024	C	GAC	RL, SO	Absent	Validation	Method	—	4
[49]	2024	C	GAR	LLM	Reviewer	Validation	Method	—	4
[144]	2024	J	GA	AP, PGM, NN	Absent	Validation	Method	—	3
[145]	2024	J	CS	MCDA	In loop	Proposal	—	y	3.5
[146]	2024	J	GCS	SO	Reviewer	Validation	Method	—	3
[147]	2024	C	GAC	SO, SIM, XAI	Reviewer	Proposal	—	—	3
[148]	2025	C	AC	RL, SO	Absent	Validation	Method	—	4
[149]	2025	C	GCSR	PGM, MCDA, SO	Reviewer	Validation	Method	y	1.5

Table 6.8 The 45 included studies – Part III (phase and method-family legend as for Part I)

Ref	Yr	T	Phases	Method families	Human	Research type	Baseline	Pr	QA
[150]	2025	C	CSR	NN, ML	Absent	Validation	Method	—	3.5
[151]	2025	C	CS	MCDA, MAS	Reviewer	Validation	—	—	2
[115]	2025	C	G	SO, RL	In loop	Validation	Method	—	4
[152]	2025	J	CS	MCDA	In loop	Validation	—	y	2
[153]	2025	J	ACS	GT, PGM, AP, KR	Reviewer	Proposal	—	—	3.5
[154]	2026	C	A	RL, MCDA	Reviewer	Proposal	—	—	1.5
[155]	2026	J	CSR	PGM	Reviewer	Proposal	—	—	1.5
[156]	2026	C	AC	SIM, PGM	Reviewer	Proposal	—	—	2.5
[157]	2026	C	GC	SO	In loop	Validation	—	—	3.5

6.3 Coverage of the COA lifecycle (RQ1)

Comparison is addressed by 35 of the 45 studies, generation by 25, analysis by 21, selection by 18 and revision of a selected course of action during execution by six.

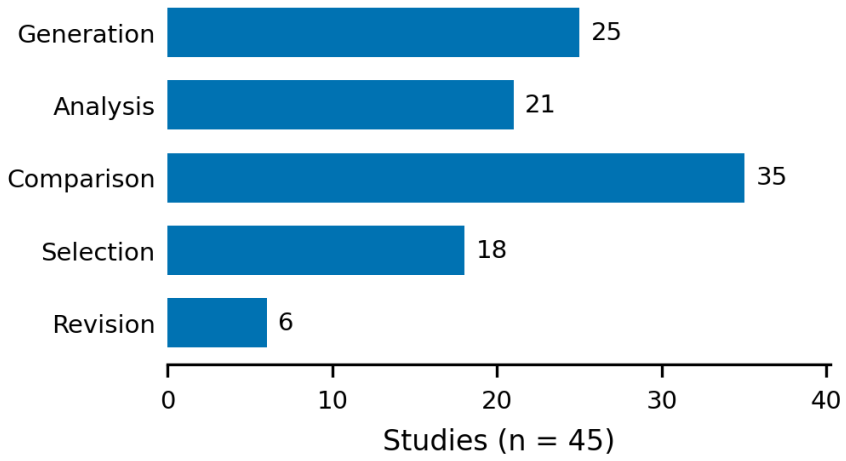


Figure 6.4 Number of included studies addressing each phase of the COA lifecycle

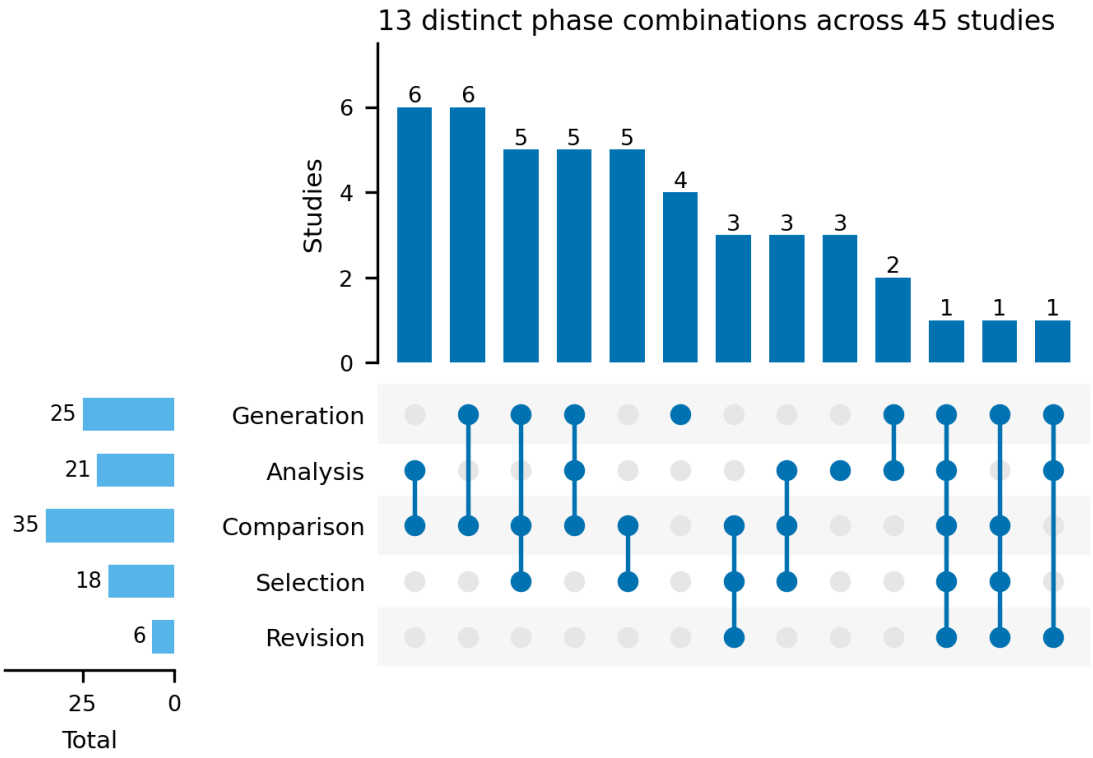


Figure 6.5 Combinations of lifecycle phases addressed, ordered by frequency

The proposition carried forward from Section 4.5, that work concentrates before the comparison step, is not supported in the form it was stated. Comparison is the most-covered phase in the corpus, not the least. The discontinuity falls one step later: 27 of the 45 studies stop short of selection. Typical of the pattern is a search that enumerates manoeuvre options and ranks them on a fitness function [123], [136], [142], [157], or a simulation that plays each option against an automated opponent and reports the outcomes [119], [124], [127], [139]. Both produce a scored set and stop.

Where selection is addressed it is usually because the method is explicitly a decision method rather than a generator. Five of the six studies applying multi-criteria decision analysis reach selection [122], [145], [149], [151], [152]; the sixth applies the machinery at analysis, scoring assets rather than plans [154]. The same holds for the game-theoretic treatment that solves for an equilibrium choice [132] and the information-fusion approach that returns a single recommended option with a confidence measure [138].

Revision is the genuine gap. Six studies address what happens to a selected course of action once execution begins. Four of the six track a changing situation with a probabilistic graphical model and re-derive the recommendation rather than re-planning: the SCOAR (Self -Healing Course of Action Revision) method for revising an executing course of action, presented first as a standalone method [137], and then as one of three orchestrated services [155], the proactive decision-support framework built on a 6R architecture [113] and a rolling optimisation over a dynamic Bayesian network [149]. The remaining two re-generate options directly, one through a language model prompted with the changed situation [49] and one through a Bayesian neural network whose predictive uncertainty triggers reconsideration [150]. The doctrines of Section 2 all describe execution as a phase in which the plan is revisited; the research literature has barely begun to support it.

The position given to the human decision maker follows the same boundary. Twenty studies place a human at comparison and selection, fourteen cast the human as a reviewer of machine output and eleven have no human in the loop at all. The reviewer position in recent work is as follows: it covers the language-model planner [49], the multi-agent negotiation of military, legal and ethical perspectives [151], the theory-of-mind proportionality model [156], the protection decision support that scores assets rather than plans [154] and the organisational-structure optimiser that returns a preferred configuration [146].

The fourteen studies casting the human as a reviewer and the eleven with no human at all together account for more than half the corpus, and the reviewer position is the one meaningful human control is least able to accommodate: a reviewer handed a ranked list with no account of its construction has the form of control without its substance, and the traceability of responsibility that the NATO principles require has nothing to attach to. Only the twenty studies that place a person at comparison and selection put the human where both the concept and the doctrines of Section 2 assume the human to be, and even there the assumption holds only if the comparison is reported in terms the person can interrogate.

6.4 Computational methods (RQ2)

Seventeen method families appear across the corpus and 20 of the 45 studies combine more than one. Search and optimisation methods lead with 16 studies, followed by simulation with 10, probabilistic graphical models with 8, multi-criteria decision analysis with 6, reinforcement learning with 5, and automated planning and knowledge representation with 4 each. Three studies propose no computational method, being process or position studies [130], [133], [134].

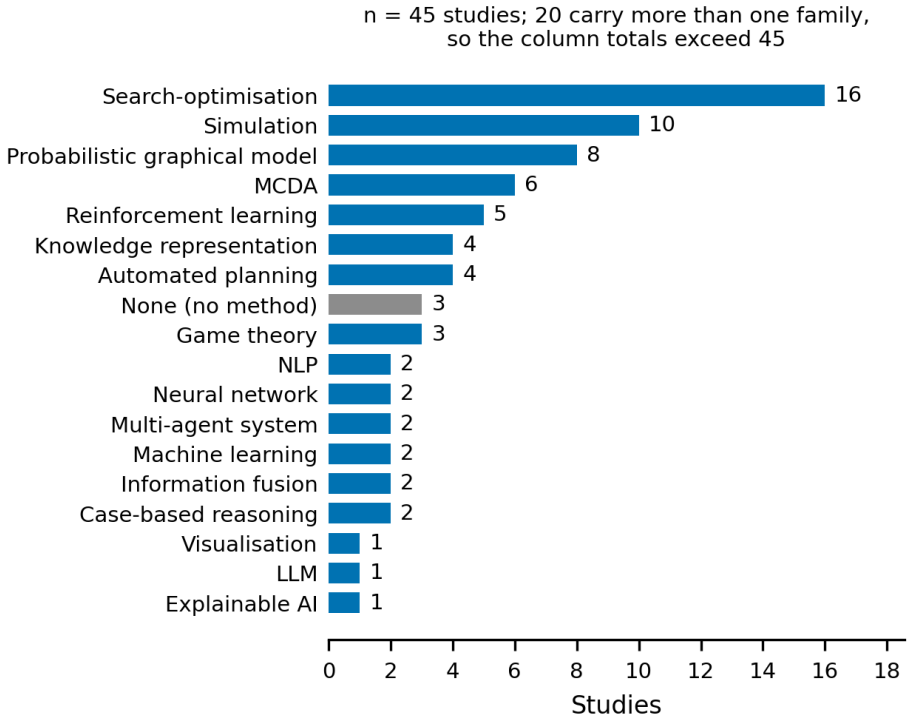


Figure 6.6 Method families across the 45 included studies.

The search and optimisation group is heterogeneous in objective rather than in mechanism. Genetic and evolutionary algorithms dominate, applied to anti-aircraft force distribution [123], manoeuvre endpoints [147], platform-to-task assignment [136], and multi-objective option

generation with crowding preserved through k-nearest neighbours [142]. Two studies pair the search with a learned heuristic: a beam search guided by a proximal-policy-optimisation policy [148] and a graph reinforcement learner over network interdiction [143]. One casts the whole problem as bi-objective constrained optimisation with a normative gate ahead of it [157], and one generates a pool of candidates before pruning it [115].

Simulation is used to fill the performance table rather than to choose. It appears as data-farming experiments coupling two simulation systems [119], as an ensemble of pre-existing models assembled for a campaign question [127], as an open-source naval combat simulator used in a confrontation exercise [125], as hybrid-intelligence multi-branch wargaming in which an agent plays the branches [139] and as a campaign-model mission planner [124]. Two further studies argue for agent-based modelling in the analysis step without building the system [120], [126].

Execution-time revision is mostly in the probabilistic group and it also supplies the influence-network treatments of effect propagation [128], [144] and the theory-of-mind model of adversary reasoning [156]. Knowledge representation covers ontology-based option representation with a two-level reasoner [129], the recognition-primed decision agent [118] and the translation of free-text orders into a formal representation [141]. Game theory appears as a two-player zero-sum formulation of option development [132], as a graph model for conflict resolution [135] and inside a three-phase pipeline with hierarchical task networks [153]. Case-based reasoning [116], [121] and information fusion [116], [138] complete the tail, and one study contributes the corpus's only explicit treatment of how options should be displayed rather than computed [131].

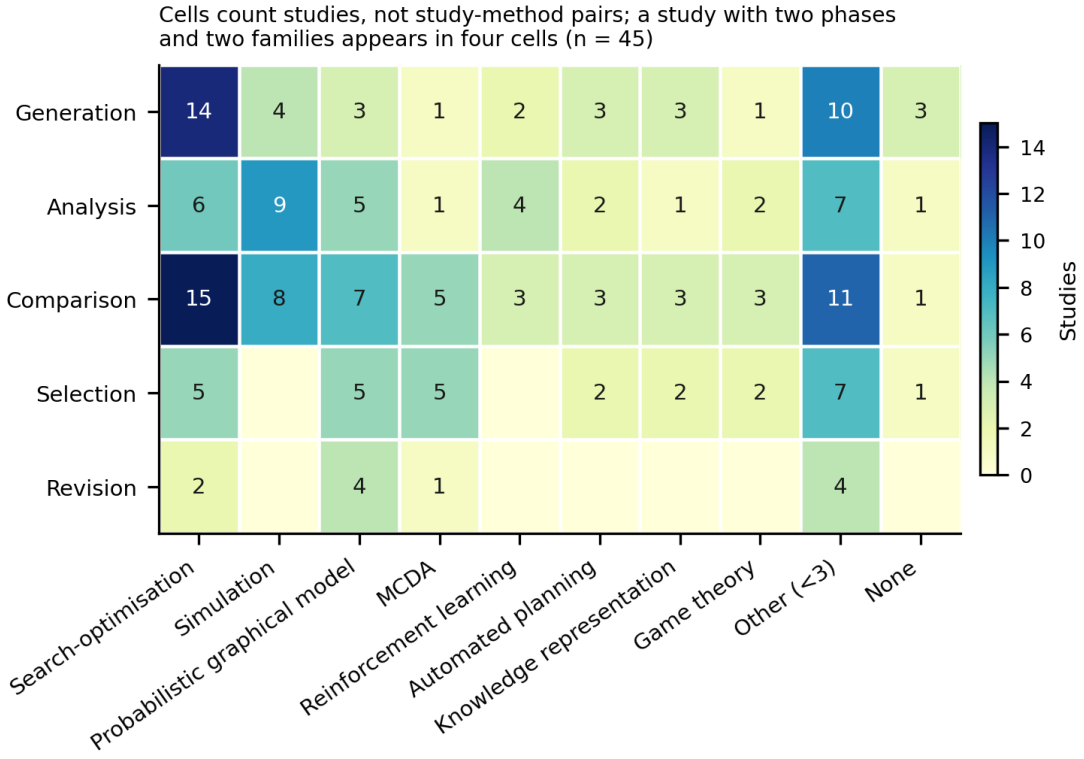


Figure 6.7 Method family by lifecycle phase

Figure 6.7 shows which methods are brought to which part of the process, and three patterns are legible. Search and optimisation methods spread across generation and comparison in almost equal measure, 14 and 15 studies, consistent with their use to enumerate a space of alternatives and rank what is enumerated by the same objective function. Simulation

concentrates at analysis, where 9 of its 10 studies sit, and appears in no study addressing selection or revision; it produces the estimates a comparison then consumes, which is the division of labour the wargame occupies in doctrine. Multi-criteria decision analysis appears in all five stages with comparison and selection having the most.

The composition of the corpus changes over the period. In 2016 to 2019 simulation leads with 7 of the 19 studies in that window and search and optimisation follows with 5. In 2023 to 2026 the leading families are search and optimisation with 8 of 18, probabilistic graphical models and multi-criteria decision analysis with 5 each, and reinforcement learning with 4. Reinforcement learning [115], [139], [143], [148], [154], neural networks [144], [150] and the single LLM study [49] are all from 2021 onward and concentrated after 2023.

6.5 Criteria, weighting and aggregation (RQ3)

RQ3 asks, of studies that compare or select, where the criteria come from, how they are weighted and reconciled, and whether the doctrinal distinction between screening criteria and evaluation criteria survives. Table 6.7 gives the answer study by study for the 35 studies addressing comparison or selection, and Figure 6.8 gives the totals across the whole corpus.

Table 6.9 Criteria, weighting and aggregation in the 35 studies addressing comparison or selection – Part I

Ref	Criteria source	Doctrine	Weighting	Aggregation	Criteria use
[119]	Ad hoc	COPD/AJP-5	Numeric	Weighted sum	Evaluation
[120]	Doctrinal	MDMP	n/a	n/a	Screening
[121]	Ad hoc	none	Unweighted	Pareto	Evaluation
[122]	Doctrinal	other	Elicited	Weighted sum	Evaluation
[123]	Ad hoc	none	Not stated	Weighted sum	Evaluation
[124]	Ad hoc	none	Not stated	Not stated	Evaluation
[127]	Doctrinal	other	Not stated	Qualitative	Evaluation
[113]	Ad hoc	none	Not stated	Not stated	Evaluation
[128]	Ad hoc	none	Not stated	Probabilistic	Evaluation
[117]	Ad hoc	none	Not stated	Qualitative	Evaluation
[129]	Ad hoc	none	Not stated	Not stated	Evaluation
[130]	n/a	MDMP	n/a	n/a	n/a
[131]	Ad hoc	MDMP	Equal	Not stated	Evaluation
[132]	Not stated	none	Not stated	Probabilistic	Evaluation
[135]	Ad hoc	none	Not stated	Weighted sum	Evaluation
[136]	Ad hoc	none	Numeric	Weighted sum	Evaluation
[137]	Ad hoc	MDMP	Numeric	Weighted sum	Evaluation
[138]	Doctrinal	MDMP	Equal	Probabilistic	Evaluation
[139]	Ad hoc	none	Not stated	Not stated	Evaluation

Table 6.10 Criteria, weighting and aggregation in the 35 studies addressing comparison or selection – Part II

Ref	Criteria source	Doctrine	Weighting	Aggregation	Criteria use
[140]	Doctrinal	other	Not stated	Pareto	Evaluation
[141]	Elicited	MDMP	Elicited	Weighted sum	Evaluation
[142]	Doctrinal	other	Unweighted	Pareto	Evaluation
[143]	Ad hoc	none	n/a	n/a	Evaluation
[145]	Doctrinal	national	Elicited	Weighted sum	Evaluation
[146]	Ad hoc	none	Elicited	Weighted sum	Evaluation
[147]	Doctrinal	national	Equal	Weighted sum	Evaluation
[148]	Ad hoc	none	Numeric	Weighted sum	Evaluation
[149]	Ad hoc	none	Elicited	Weighted sum	Evaluation
[150]	Ad hoc	none	Not stated	Not stated	Evaluation
[151]	Doctrinal	MDMP	Elicited	Weighted sum	Evaluation
[152]	Doctrinal	national	Elicited	Distance to ref.	Evaluation
[153]	Ad hoc	JPP	Equal	Weighted sum	Evaluation
[155]	Ad hoc	MDMP	Numeric	Weighted sum	Evaluation
[156]	Doctrinal	other	Not stated	Pareto	Evaluation
[157]	Doctrinal	other	Not stated	Pareto	Distinguished

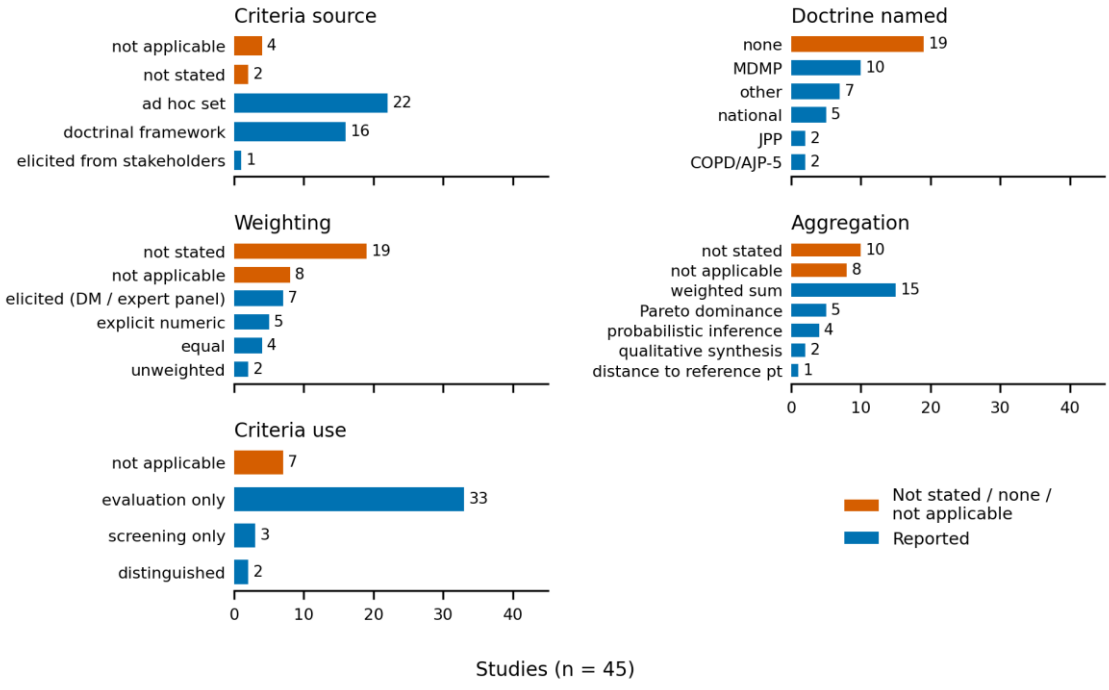


Figure 6.8 Criteria source, doctrine named, weighting, aggregation and criteria use

Across the whole corpus of 45 studies, 22 build their criteria as an ad hoc set assembled for the problem at hand, 16 draw on a doctrinal framework and one elicits them from stakeholders in a structured session with practitioners [141]. 19 studies name no planning doctrine at all. The remaining six are the studies for which the field does not arise or was not reported: two do not say where their criteria came from and four propose no comparison procedure for the field to apply to. Within the 35 studies that address comparison or selection, listed in Table 6.7, the split is 20 ad hoc, 12 doctrinal, one elicited, one not stated and one not applicable. Where doctrine is named it is most often the MDMP, in 10 studies Where doctrine is named it is most often the MDMP, in 10 studies [49], [120], [126], [130], [131], [137], [138], [141], [151], [155], followed by a national doctrine in 5 [125], [134], [145], [147], [152], the Joint Planning Process in 2 [144], [153], and COPD or AJP-5 in 2 [119], [133]. The remaining seven name a framework that is not a planning doctrine at all, most often the law of armed conflict where the criteria are ethical or legal rather than operational [122], [156], [157]. The two fields move together: of the 19 studies naming no doctrine, 17 use an ad hoc criteria set and two do not say where their criteria came from [118], [132].

Weighting is unstated in 19 of the 45 studies, 14 of them among the 35 in Table 6.7. Seven derive weights from a decision maker or expert panel [122], [141], [145], [146], [149], [151], [152], five fix weights numerically and declare the values [119], [136], [137], [148], [155], four weigh equally [131], [138], [147], [153], and two state explicitly that no weighting is applied [121], [142], all of them within the 35, and three record the field as not applicable.

Of the 45 studies, 15 aggregate by weighted sum, five by Pareto dominance [121], [140], [142], [156], [157], four by probabilistic inference [128], [132], [138], [144], two by qualitative synthesis [117], [127] and one by distance to a reference point [152]; 10 aggregate without specifying the method used and eight record the field as not applicable. All but one of the counts above fall inside the 35 studies of Table 6.7, the exception being the fourth probabilistic treatment, which addresses analysis rather than comparison. The dominance of the weighted sum is the substantive point rather than its frequency. Section 5.6 set out the device by which every doctrine keeps non-compensatory judgements out of a compensatory total, placing the five screening criteria ahead of the comparison so that an unacceptable option is removed rather than compensated. A weighted sum applied without that gate reinstates exactly the trade the doctrine forbids. The five Pareto treatments are the exception: refusing to aggregate at all, they return a non-dominated set and leave the trade-off to the decision maker, which is the refusal an outranking veto threshold encodes.

33 of the 45 studies apply criteria to evaluate alternatives with no screening step. Two preserve the distinction: [157] screens candidates through a proportionality gate and passes only the survivors to a Pareto comparison, and [134] puts each alternative through suitability, feasibility and acceptability tests before evaluation. Three name doctrinal adequacy criteria with no evaluation criteria for them to be conflated with [49], [120], [125], and the remaining seven propose no comparison procedure to which the field applies [115], [116], [118], [126], [130], [133], [154].

Four studies name doctrinal screening criteria. In [49] the four criteria are written into the system prompt and every generated option must justify itself against them, and in [134] each alternative passes suitability, feasibility and acceptability tests before evaluation begins; in both the criteria do work. In [120] the feasibility, acceptability and suitability tests are quoted with their definitions and never applied, and in [125] the wargame is described as producing information from which those tests could be developed rather than as applying them, with no alternative screened out. Of the 45 studies in this review, two make a doctrinal screening criterion operative.

The subgroup using multi-criteria decision analysis is the exception that measures the rest. All six state their aggregation rule and the five performing a comparison elicit their weights rather than leaving them implicit: an additive model of ethical violation validated against expert raters [122], a hybrid model pairing Defining Interrelationships Between Ranked criteria (DIBR) II with Evaluation based on Distance from Average Solution (EDAS), deriving criterion weights from a stated ordering [152], a fuzzy Decision-Making Trial and Evaluation Laboratory (DEMATEL) and Complex Proportional Assessment (COPRAS) treatment on triangular numbers [145], a multi-agent system reconciling military, legal and ethical perspectives into one ranking [151] and a rolling optimisation that re-weights as the situation changes [149]. Adopting a named method from the multi-criteria literature evidently carries a reporting discipline with it, and the difference between these six and the other 39 studies is less a difference of rigour in the underlying work than of what the method obliges the author to write down.

Doctrine describes the pool of operational factors in detail and gives no procedure for deriving evaluation criteria from it; the research literature has not filled that gap but has routed around it, assembling criteria for each problem and, in a majority of cases, declining to say how those criteria are traded against one another.

6.6 Evaluation and evidence of effectiveness (RQ4)

Classified on the Wieringa facet [158], 23 studies are validation research evaluating a proposal in the laboratory or in simulation, 19 are solution proposals without such an evaluation, and three remain: one evaluation research study conducted in practice, a qualitative study of the option-development session as staff actually run it [133], one experience paper reporting a naval confrontation exercise [125] and one opinion paper setting out a roadmap for agent-based modelling in the analysis step [126].

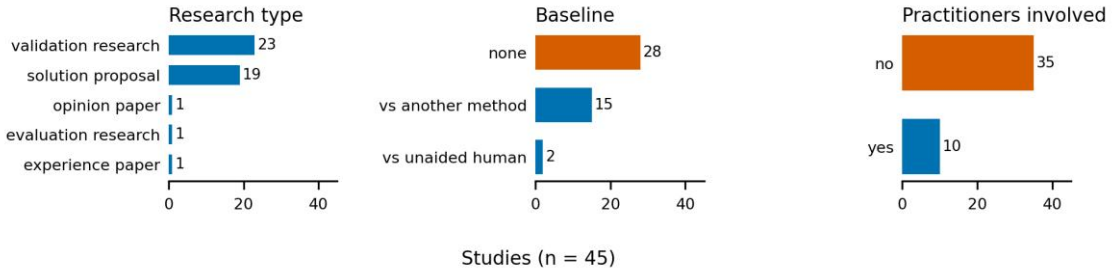


Figure 6.9 Research type, comparison baseline and practitioner involvement

What the studies compare themselves against is the more informative field. Twenty-eight report no baseline of any kind, 15 compare against another computational method, and 2 compare against unaided human judgement: an ethical-violation model tested against the assessments of expert raters [122] and a multi-agent planner compared with plans produced by human participants [140]. The distribution is almost entirely explained by research type. Of the 19 solution proposals, 17 offer no baseline. Of the 23 validation research studies, 14 compare against another method. The literature therefore does contain comparative evaluation, but it is comparison of one algorithm against another. Two studies of 45 test a proposed method against the staff procedure it is meant to support, which is the comparison that would establish whether any of this work improves on current practice.

Practitioners were involved in 10 studies. However, their presence does not track quality: the group spans quality totals from 1.5 to 5.0, includes the highest-scoring study in the corpus [131]

and the lowest-scoring members of the multi-criteria group [149], [154], and in several cases the involvement amounts to subject-matter experts supplying parameter values rather than to evaluation with users.

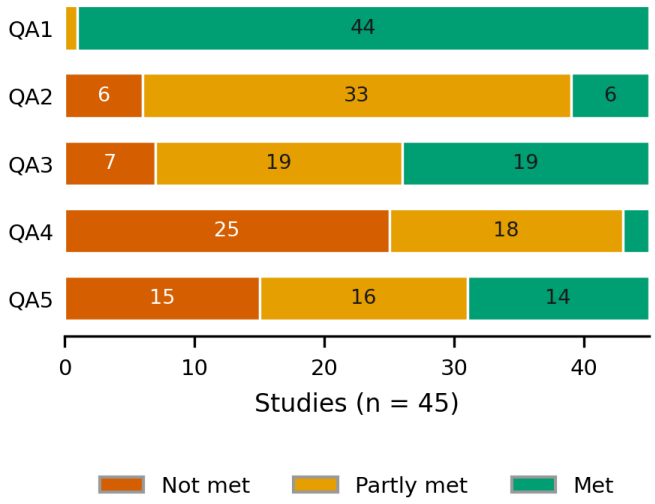


Figure 6.10 Quality assessment across the five items

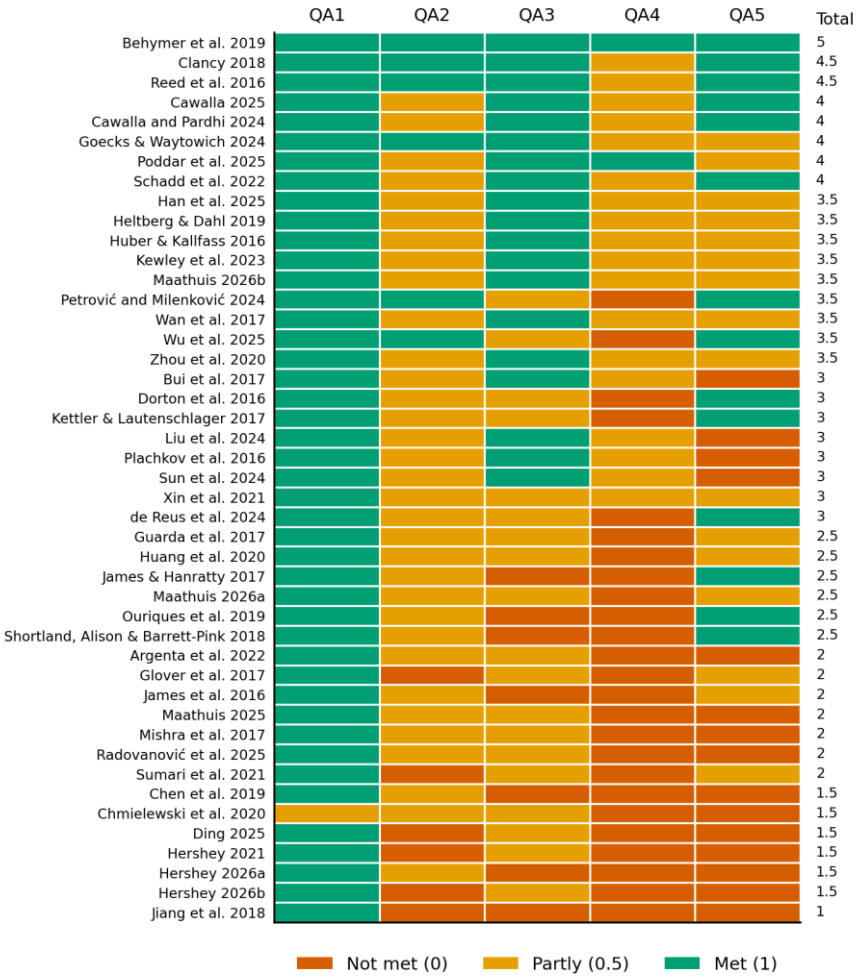


Figure 6.11 Quality assessment by study, ordered by total score

The quality profile in Figures 6.10 and 6.11 is uneven in an instructive way. Aims are stated clearly in effectively every study, giving a mean of 0.99 on the first item. The mean falls to 0.63 on the presence of an empirical evaluation, 0.50 on reproducibility of the described approach, 0.49 on the discussion of limitations, and 0.24 on whether the evaluation method is described and appropriate to the claim made. Twenty-five of the 45 studies score zero on that last item. The mean total is 2.86 of 5, with one study at 5.0 [131] and one at 1.0 [129].

The deficit is concentrated where it does most damage to any attempt to synthesise findings across the corpus. A literature can be read for its ideas when reproducibility is weak; it cannot be read for evidence of effectiveness when 25 of 45 studies either omit an evaluation or describe one that does not test the claim being made, and only 2 describe one that does. The conference format accounts for part of this, since 31 of the 45 studies are conference papers with a page budget an evaluation section competes for, but it does not account for the pattern within the journal subset, where the same item is the weakest of the five.

The answer to RQ4 is therefore narrow. The corpus reports a great deal of activity and very little evidence. Where effectiveness is demonstrated it is nearly always effectiveness relative to another computational method on a constructed problem, and whether any of these approaches improves on the staff procedure described in Section 2 remains, on this evidence, open.

7 DISCUSSION AND FUTURE WORK

On RQ1, the majority of the studies in the corpus focus on comparing courses of action (COAs) and very few studies address executing a selected COA while continuing to revise. Additionally, the greatest disparity between two methods occurs when moving from comparison of COAs to selecting one COA. Of the 45 studies reviewed, 27 studies identify alternative solutions to a problem, however none of the studies evaluate how to select among those alternatives. On RQ2, there are 17 method families used to evaluate COAs, with search and optimization being the most commonly used method families. Simulation was primarily applied during the analysis phase of evaluating COAs. Multi-criteria decision analysis (MCDA) was only applied to the phases of COA comparison and selection. Reinforcement learning and language-based machine learning methods were introduced in the last three years of research. On RQ3, the majority of studies create their criteria for evaluating COAs based upon an ad-hoc basis, and in 19 of the 45 studies referenced in this review, no methodology was provided regarding how to plan for creating criteria for evaluating COAs. In addition, in 19 of the 45 studies, no weightings were assigned to any criteria that were created. The weighted sum was found to be the primary method of aggregating criteria for evaluating COAs. Finally, although some studies referenced separating screening from evaluation processes, such separation was only referenced in 2 of the 45 studies. On RQ4, in 28 of the 45 studies referenced in this review, no baseline information was included in order to provide context for evaluating COAs. In 2 of the studies referenced in this review, comparisons were made with respect to unassisted human judgment. Finally, based on the literature referenced within this review, the lowest quality element related to using methods for evaluating COAs relates to determining if the method for evaluating COAs is applicable to supporting the claims that are intended to be supported by those evaluations.

The gap established in Section 3 between a detailed doctrinal description of operational factors and a lack of procedures for transforming those operational factors into evaluation criteria for COAs is not addressed by the research literature referenced in this review. Similarly, the mechanism developed in Section 2.5 that allows doctrine to keep non-compensatory judgments from being incorporated into a compensatory total is missing from all but two of the studies that would require that mechanism.

Doctrines provide lengthy descriptions of pools of operational factors and then allow judgment to determine how those operational factors should be transformed into criteria for evaluating COAs. As such, the research literature has largely circumvented this gap rather than filling it. Twenty-two studies develop a set of criteria specific to the problems that are being studied and then abandon that set of criteria. There is no knowledge exchange between any two studies that independently develop their respective sets of criteria. Furthermore, neither set of criteria can be audited against the commander's expressed priorities. Only FM 5-0 describes what constitutes a valid criterion: a title, a definition, a unit of measure, a benchmark and a direction. The remaining three documents do not describe any format for developing a criterion.

Acceptability appears to carry the heaviest burden in terms of establishing suitability, feasibility and acceptability for each potential COA. While acceptability is consistently referred to as an important consideration in each document, acceptability is defined and described least clearly of all the five considerations listed above. For example, in AJP-5, acceptability encompasses both societal impact and anticipated media reaction [7], however, there is no discussion in any study referencing AJP-5 regarding how societal impact or anticipated media reaction are determined or validated. Two studies [49], [134] out of the 45 studies referencing four U.S. Army Doctrines explicitly reference a screening criterion prescribed by doctrine and apply it.

The instrument that forms the foundation for non-compensatory decisions is least clearly defined of all elements involved in making decisions regarding which COA to select. This is significant because of what lies beyond the screen. A weighted sum trades everything off against everything else. Therefore, the five screening criteria are placed in front of it: an option that fails acceptability is eliminated and not compensated for by anything else. Fifteen studies utilize a weighted sum to combine criteria and 33 studies reference criteria without applying a screening process prior to aggregating them.

Selecting a COA is a textbook discrete multi-attribute problem. A finite number of alternatives exists, multiple competing criteria are present in disparate units, weights represent prioritized preferences established by the commander, and a single decision-maker is responsible for defending their subsequent choices. However, three doctrines utilize a weighted sum as a means for combining competing criteria without referring to it as such or providing either scales required to establish it or references to alternative approaches developed since the 1970s specifically designed for solving this type of problem.

Similarly, six studies referenced in this review apply a named multi-criteria method. Those studies also specify both their weights and their aggregation rule. The values that are input into any of these rules are generated from the wargame. As discussed earlier in this section, the wargame is perhaps the least investigated tool utilized throughout this process. This is because it serves as the venue for evaluating options and; in practice; refining options into formats that are suitable for comparative evaluation; and on each doctrine's own account; it is a structured application of judgment rather than measurement. AJP-5 further limits the utility of wargame outcomes in that it prohibits wargame results from validating or invalidating any option and instead views wargame results as data points [7]. Therefore; any performance table generated based on wargame results will include an amount of uncertainty that no comparison method represented in Section 5 addresses; whereas; any constructive simulations replacing the wargame will produce estimates of that uncertainty themselves. None of the 45 studies analysed herein quantify either aspect of uncertainty associated with performance tables generated from wargames.

There are five areas of research activity that could potentially contribute towards closing gaps identified in Section 3: The first area is building a set of criteria based upon shared constructs existing within each doctrine (objectives/effects/decisive conditions). That set must be shown to be exhaustive, non-redundant, or preference-independent. Screening criteria must remain formally separate from evaluation criteria to which they are currently collapsed. The second area is defining each screening criterion utilizing standards similar to those identified within FM 5-0 for defining evaluation criteria (definition/unit/benchmark/direction), thereby allowing staff members to record and reviewers to audit acceptability assessments as staff record them. Each depends on a step that neither doctrine nor body of literature references directly: converting qualitative judgments into values usable by methods for evaluation purposes; which is typically where errors arise. The third area is actually identifying appropriate methods for evaluating COAs: relating compensation procedures to items designated as non-negotiable by commanders and reporting ranges of weights over which recommendations survive. The fourth area is examining aspects related to conducting wargames: what portion of variability within a performance table stems from variations due to option characteristics versus variations caused by players' interpretations thereof; which can be answered by repeatedly running the same option through different cells. The fifth area is the one that leads to previous four activities; i.e., each method identified in Section 5 evaluates a collection of options at a point-in-time; and each doctrine identified in Section 2 executes selected options over time via branching, sequencing and decision-making nodes that implicitly assume each node will be revisited at

some future point-in-time. However; neither doctrine nor literature provides guidance concerning durations for maintaining recommendations or availability of rejected options when circumstances have changed or resources have been allocated. Deep Green named this in 2008 and six studies of 45 have addressed it since [47]. Recommendations that possess no expiration date constitute less effective instruments than they seem to be; and treating COAs as trajectories rather than as singular points would be necessary to assign duration(s) to. In addition to concepts identified previously in this review; this review introduces a new requirement: addressing "transfer" issues not resolved in Section 3.3 necessitates a second review conducted using civil-security terminology; as civilian literature concerning selection among response options utilizes none of the terms used in this review.

The survey reported here is the first stage of a doctoral project, and the direction it sets is the third and the fifth of those lines taken together. The next stage is a criteria and comparison layer for the COA process: built on the constructs the four doctrines already share, with the screening criteria formalised as a sorting problem in the sense of Section 5.10 rather than folded into a weighted total, the weights elicited by a procedure a commander can complete before the options exist, and the recommendation reported together with the range of weights and of wargame estimates over which it survives rather than as a single score. The generative and estimating work the corpus already does well is taken as given and treated as the source of the performance table; the contribution is the layer above it, which no class of fielded system in Section 4 implements. A prototype of that layer will be built against a Croatian planning scenario constructed for the purpose and evaluated with planning staff against the matrix procedure it is meant to support, which is the comparison two studies of 45 currently make.

8 CONCLUSION

The four doctrines treat COA selection as a discrete multi-attribute decision making process; they leave the part which has the answer to be determined by the planner's judgment. None of the doctrines provide a methodology for developing evaluation criteria from the detailed descriptions of operational variables included in their doctrines; converting an operational variable to a criterion is left as an un-stated step in each of the four doctrines. All but one of the doctrines include a weighted sum; and each of those that provide arithmetic for calculating the weighted sum warn against using it. Every doctrine provides a non-compensatory screen prior to applying a compensatory rule, therefore eliminating a suboptimal alternative, rather than compensating for it.

Across 45 studies from 2016 to 2026, the literature routes around these gaps rather than closing them. Criteria are most often assembled ad hoc, no planning doctrine is named in 19 studies, weighting is unstated in 19, aggregation is dominated by the weighted sum, and the separation of screening from evaluation survives in two. The corpus covers comparison most heavily and revision during execution barely at all, with its sharpest discontinuity between the two: 27 of 45 studies score alternatives without addressing the choice among them. Only in the past three years have reinforcement learning, neural networks, and a single LLM been used as tools in decision support, and they are primarily being used to generate alternatives and estimate how well each will perform. A total of twenty studies place humans at both comparing alternatives and selecting alternatives; fourteen studies place humans at reviewing machine generated output as compared to machine generated output; and eleven studies do not include humans in either comparing or selecting alternatives.

Future decision support systems must combine the generative speed of artificial intelligence with comparison methods that are transparent, defensible and open to revision during execution. By consolidating doctrine, decision theory and a decade of research into a single synthesis, this study contributes a foundation for researchers and planners bringing artificial intelligence into operations planning.

REFERENCES

[1] V. Belton and T. J. Stewart, *Multiple Criteria Decision Analysis*. Boston, MA: Springer US, 2002. doi: 10.1007/978-1-4615-1495-4.

[2] Headquarters, Department of the Army, "Planning and Orders Production," Department of the Army, Washington, DC, USA, Field Manual FM 5-0, 2024. Accessed: Jul. 17, 2026. [Online]. Available: https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN44590-FM_5-0-001-WEB-3.pdf

[3] Joint Chiefs of Staff, "Joint Planning," Department of Defense, Washington, DC, USA, Joint Doctrine Publication JP 5-0, 2020. Accessed: Jul. 17, 2026. [Online]. Available: https://www.esd.whs.mil/Portals/54/Documents/FOID/Reading%20Room/Joint_Staff/18-F-1152_JP_5-0_Joint_Planning_2020.pdf

[4] North Atlantic Treaty Organization (NATO), "Allied Joint Doctrine for the Planning of Operations," NATO Standardization Office, Brussels, Belgium, Allied Joint Publication (AJP) AJP-5, Edition B, Version 1, 2026. Accessed: Jul. 20, 2026. [Online]. Available: https://assets.publishing.service.gov.uk/media/6a43a8dbfa72f76532a4e3bd/AJP_5_EdB_NATO.pdf

[5] Glavni stožer OSRH, "Planiranje operacija," Glavni stožer Oružanih snaga Republike Hrvatske, Zagreb, Združeni doktrinarni priručnik ZDP-5, 2019.

[6] Headquarters, Department of the Army, "The Operations Process," Department of the Army, Washington, DC, USA, Army Doctrine Publication ADP 5-0, 2019. Accessed: Jul. 17, 2026. [Online]. Available: https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN18126-ADP_5-0-000-WEB-3.pdf

[7] North Atlantic Treaty Organization (NATO), "Allied Joint Doctrine," NATO Standardization Office, Brussels, Belgium, Allied Joint Publication (AJP) AJP-01, 2022.

[8] Allied Command Operations, "Comprehensive Operations Planning Directive," Supreme Headquarters Allied Powers Europe (SHAPE), Mons, Belgium, Version 3.1, 2023.

[9] Joint Chiefs of Staff, *Department of Defense Dictionary of Military and Associated Terms*. Washington, DC, USA: Department of Defense, 2025. Accessed: Jul. 17, 2026. [Online]. Available: https://www.jcs.mil/doctrine/dod-terminology-program/

[10] Joint Chiefs of Staff, "Joint Campaigns and Operations," Department of Defense, Washington, DC, USA, Joint Doctrine Publication JP 3-0, 2022. Accessed: Jul. 17, 2026. [Online]. Available: https://www.jcs.mil/doctrine/joint-doctrine-pubs/3-0-operations-series/

[11] J. Kelly and M. Brennan, *Alien: how operational art devoured strategy*. Carlisle Barracks/Pa: US Army War College Press, 2009.

[12] J. Strange, *Centers of Gravity & Critical Vulnerabilities: Building on the Clausewitzian Foundation So That We Can All Speak the Same Language*. in Perspectives on Warfighting, no. 4. Quantico, VA, USA: Marine Corps University, 1996. Accessed: Jul. 20, 2026. [Online]. Available: https://jfsc.ndu.edu/Portals/72/Documents/JC2IOS/Additional_Reading/3B_COG_and_Critical_Vulnerabilities.pdf

[13] North Atlantic Treaty Organization (NATO), "Lisbon Summit Declaration," Official texts and resources | NATO. Accessed: Jul. 20, 2026. [Online]. Available: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2010/11/20/lisbon-summit-declaration

[14] Glavni stožer OSRH, "Združena doktrina oružanih snaga Republike Hrvatske," Glavni stožer Oružanih snaga Republike Hrvatske, Zagreb, 2006.

[15] Glavni stožer OSRH, "Doktrina Oružanih snaga Republike Hrvatske," Glavni stožer Oružanih snaga Republike Hrvatske, Zagreb, Združeni doktrinarni priručnik ZDP-1, 2011.

[16] Glavni stožer OSRH, “Doktrina Oružanih snaga Republike Hrvatske,” Glavni stožer Oružanih snaga Republike Hrvatske, Zagreb, Združeni doktrinarni priručnik ZDP-1(A), 2016.
[17] Hrvatski sabor, *Dugoročni plan razvoja Oružanih snaga Republike Hrvatske 2025.-2036.*, vol. akt br. 1434. 2025. Accessed: Jul. 23, 2026. [Online]. Available: https://narodne-novine.nn.hr/clanci/sluzbeni/2025_07_104_1434.html
[18] Hrvatski sabor, *Ustav Republike Hrvatske*. 1990. Accessed: Jul. 23, 2026. [Online]. Available: https://narodne-novine.nn.hr/clanci/sluzbeni/2010_07_85_2422.html
[19] Hrvatski sabor, *Zakon o obrani*. 2013. Accessed: Jul. 23, 2026. [Online]. Available: https://narodne-novine.nn.hr/clanci/sluzbeni/2013_06_73_1452.html
[20] Vlada Republike Hrvatske, *Prijedlog odluke o sudjelovanju pripadnika Oružanih snaga Republike Hrvatske u NATO aktivnosti za sigurnosnu i obučnu potporu Ukrajini (NSATU)*. 2024. Accessed: Jul. 23, 2026. [Online]. Available: https://www.sabor.hr/sites/default/files/uploads/sabor/2024-10-03/181302/PO_OSRH_NSATU.pdf
[21] Hrvatski sabor, *Zakon o sudjelovanju pripadnika Oružanih snaga Republike Hrvatske, policije, civilne zaštite te državnih službenika i namještenika u mirovnim operacijama i drugim aktivnostima u inozemstvu*. 2002. Accessed: Jul. 23, 2026. [Online]. Available: https://narodne-novine.nn.hr/clanci/sluzbeni/2002_03_33_711.html
[22] Vlada Republike Hrvatske, *Prijedlog odluke o sudjelovanju Oružanih snaga Republike Hrvatske u operaciji potpore miru KFOR na Kosovu*. 2024. Accessed: Jul. 23, 2026. [Online]. Available: https://www.sabor.hr/sites/default/files/uploads/sabor/2024-10-03/181404/PO_OSRH_KFOR_KOSOVO.pdf
[23] Hrvatski sabor, *Strategija obrane Republike Hrvatske*, vol. akt br. 1433. 2025. Accessed: Jul. 23, 2026. [Online]. Available: https://narodne-novine.nn.hr/clanci/sluzbeni/2025_07_104_1433.html
[24] Headquarters, Department of the Army, “Intelligence Preparation of the Battlefield,” Department of the Army, Washington, DC, USA, Army Techniques Publication ATP 2-01.3, 2019. Accessed: Jul. 22, 2026. [Online]. Available: https://home.army.mil/wood/application/files/8915/5751/8365/ATP_2-01.3_Intelligence_Preparation_of_the_Battlefield.pdf
[25] M. A. Frey and A. Schulte, “Tactical Situation Modelling of MUM-T Helicopter Mission Scenarios using Influence Maps,” in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy: IEEE, Oct. 2019, pp. 552–558. doi: 10.1109/SMC.2019.8914650.
[26] S. Bisht, S. B. Taneja, V. Jindal, and P. Bedi, “APSO based automated planning in Constructive Simulation,” *Def. Sci. J.*, vol. 73, no. 5, pp. 564–571, Aug. 2023, doi: 10.14429/dsj.73.18497.
[27] K. Virtanen, H. Mansikka, M. Kankaisto, and R. P. Hämäläinen, “Communicating Commander’s Intent in a Chain of Command With Multi-Criteria Decision Analysis—An Experiment in Air Combat,” *J. Multi-Criteria Decis. Anal.*, vol. 33, no. 1, p. e70030, 2026, doi: 10.1002/mcda.70030.
[28] M. Asprusten *et al.*, “Using M&S Standards in Combined Civil and Military Training for Disaster and Crisis Management,” presented at the Virtual Simulation Innovation Workshop 2022, SIW 2022, 2022. [Online]. Available: https://www.scopus.com/pages/publications/85186957667?origin=resultslist
[29] Headquarters, Department of the Army, “Mission Command: Command and Control of Army Forces,” Department of the Army, Washington, DC, USA, Army Doctrine Publication ADP 6-0, Jan. 2020.

[30] G. A. Gorry and M. S. Scott Morton, “A framework for management information systems,” *Sloan Manage. Rev.*, vol. 13, no. 1, pp. 55–70, 1971.
[31] R. H. Sprague, “A Framework for the Development of Decision Support Systems,” *MIS Q.*, vol. 4, no. 4, pp. 1–26, Dec. 1980, doi: 10.2307/248957.
[32] J. A. Boyd, *A Discourse on Winning and Losing*. Maxwell AFB, Alabama: Air University Press, 2018. [Online]. Available: https://media.defense.gov/2018/May/22/2001920668/-1/-1/0/B_0151_Boyd_Discourse_Winning_Losing.pdf
[33] F. Osinga, *Science, strategy and war: the strategic theory of John Boyd*. Delft: Eburon Academic Publishers, 2005.
[34] H. Thuve, “TOPFAS (Tool for Operational Planning, Force Activation and Simulation),” in *Proc. 6th Int. Command and Control Research and Technology Symp. (ICCRTS)*, Annapolis, MD, USA, 2001.
[35] T. Havlík, M. Blaha, L. Potužák, O. Pekař, and V. Šlouf, “Wargaming Simulator MASA SWORD for Training and Education of Czech Army Officers,” *Eur. Conf. Games Based Learn.*, vol. 16, pp. 811–815, Oct. 2022, doi: 10.34190/ecgbl.16.1.914.
[36] J. M. Pullen and O. M. Mevassvik, “Coalition command and control — Simulation interoperation as a system of systems,” in *2016 11th System of Systems Engineering Conference (SoSE)*, Kongsberg, Norway: IEEE, Jun. 2016, pp. 1–6. doi: 10.1109/SYSOSE.2016.7542896.
[37] North Atlantic Treaty Organization (NATO), “Coalition Battle Management Language (C-BML),” NATO Research and Technology Organisation, Neuilly-sur-Seine Cedex, France, RTO Technical Report TR-MSG-048, Feb. 2012. Accessed: Aug. 27, 2026. [Online]. Available: https://publications.sto.nato.int/publications/STO%20Technical%20Reports/RTO-TR-MSG-048/$$TR-MSG-048-ALL.pdf
[38] D. R. Wittman and NATO Science and Technology Organization, “An Introduction to the Simulation Interoperability Standards,” NATO Science and Technology Organisation, Neuilly-sur-Seine, France, STO Educational Notes STO-EN-MSG-141, 2015. Accessed: Aug. 27, 2026. [Online]. Available: https://publications.sto.nato.int/publications/STO%20Educational%20Notes/STO-EN-MSG-141/EN-MSG-141-06A.pdf
[39] North Atlantic Treaty Organization (NATO), *Standardisation for C2-simulation interoperation*. Neuilly-sur-Seine Cedex, France, 2015.
[40] North Atlantic Treaty Organization (NATO), “Modelling and Simulation Group 145: Operationalization of Standardized C2-Simulation Interoperability,” NATO Science and Technology Organisation, Neuilly-sur-Seine, France, STO Technical Report TR-MSG-145, Apr. 2021.
[41] B. G. Silverman, D. M. Silverman, G. Bharathy, N. Weyer, and W. R. Tam, “StateSim: lessons learned from 20 years of a country modeling and simulation toolset,” *Comput. Math. Organ. Theory*, vol. 27, no. 3, pp. 231–263, Sep. 2021, doi: 10.1007/s10588-021-09324-1.
[42] L. F. Weyland, M. P. D. Schadd, and H. C. Henderson, “Exploring the Fidelity of Synthetic Data for Decision Support Systems in Military Applications,” in *2024 International Conference on Military Communication and Information Systems (ICMCIS)*, Koblenz, Germany: IEEE, Apr. 2024, pp. 1–8. doi: 10.1109/ICMCIS61231.2024.10540934.
[43] L. F. Weyland, M. P. D. Schadd, and H. C. Henderson, “How a Simulator’s Fidelity Impacts the Performance of a Decision Support System: A Military Case Study,” in *2025*

International Conference on Military Communication and Information Systems (ICMCIS), Oerias, Portugal: IEEE, May 2025, pp. 1–10. doi: 10.1109/ICMCIS64378.2025.11047897.
[44] L. Ground and A. Kott, "Course of Action Display and Evaluation Tool (CADET) Enhancements:," Defense Technical Information Center, Fort Belvoir, VA, Mar. 2000. doi: 10.21236/ADA378107.
[45] A. Kott, L. Ground, R. Budd, and J. Langston, "Toward Practical Knowledge-Based Tools for Battle Planning and Scheduling," in *Proceedings of the Eighteenth National Conference on Artificial Intelligence and Fourteenth Conference on Innovative Applications of Artificial Intelligence*, Edmonton, Alberta, Canada, Aug. 2002.
[46] R. Rasch, A. Kott, and K. D. Forbus, "Incorporating AI into military decision making: an experiment," *IEEE Intell. Syst.*, vol. 18, no. 4, pp. 18–26, Jul. 2003, doi: 10.1109/MIS.2003.1217624.
[47] J. R. Surdu and K. Kittka, "The Deep Green Concept," in *Proceedings of the 2008 Spring Simulation Multiconference (SpringSim '08)*, Ottawa, Canada: Military Modelling and Simulation Symposium (MMS), 2008. Accessed: Jul. 22, 2026. [Online]. Available: https://www.bucksurdu.com/Professional/Documents/Case10307_TheDeepGreenConceptRevised.pdf
[48] P. J. Schwartz *et al.*, "AI-enabled wargaming in the military decision making process," in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II*, SPIE, Apr. 2020, p. 118. doi: 10.1117/12.2560494.
[49] V. G. Goecks and N. Waytowich, "COA-GPT: Generative Pre-trained Transformers for Accelerated Course of Action Development in Military Operations," in *2024 International Conference on Military Communication and Information Systems (ICMCIS)*, Koblenz, Germany: IEEE, Mar. 2024. doi: 10.1109/ICMCIS61231.2024.10540749.
[50] O. Vinyals *et al.*, "StarCraft II: A New Challenge for Reinforcement Learning," Aug. 16, 2017, *arXiv*: arXiv:1708.04782. doi: 10.48550/arXiv.1708.04782.
[51] X. Kuang, D. Zeng, G. Li, F. Yang, and X. Li, "HOLA: Hierarchical OODA-LLM Agent Architecture with Commander-like Cognition," *IEEE Trans. Games*, pp. 1–12, 2026, doi: 10.1109/TG.2026.3689572.
[52] D. Verdejo and E. M. Laurent, "How can LLM improve efficiency of C4ISR?," in *IFIP Advances in Information and Communication Technology*, Paris, France: Springer Nature Link, Nov. 2024, pp. 1–17.
[53] A. J. Tate and C. J. Cooke, "Human Autonomy Teaming in Naval C2 – Insights from Dstl's Intelligent Ship Project," in *Proceedings of the International Ship Control Systems Symposium*, Online. doi: 10.24868/11145.
[54] S. Jackson, C. Golsteijn, R. Johnson, and A. Munafo, "Scheduling and Tasking of Autonomous Systems for Collaborative Missions," in *OCEANS 2022, Hampton Roads*, Hampton Roads, VA, USA: IEEE, Oct. 2022, pp. 1–5. doi: 10.1109/OCEANS47191.2022.9976994.
[55] S. A. Sarcia' and G. Colo', "Organizing Structures and Information for Developing AI-enabled Military Decision-Making Systems," in *2023 IEEE International Workshop on Technologies for Defense and Security (TechDefense)*, Rome, Italy: IEEE, Nov. 2023, pp. 455–460. doi: 10.1109/TechDefense59795.2023.10380925.
[56] N. Wahab, N. Hasbullah, S. Ramli, and N. Z. Zainuddin, "Verification of a Battlefield Visualization Framework in Military Decision Making using Holograms (3D) and multi-touch technology," in *2016 International Conference on Information and Communication Technology (ICICTM)*, Kuala Lumpur, Malaysia: IEEE, 2016, pp. 23–26. doi: 10.1109/ICICTM.2016.7890770.
[57] B. Van de Walle and M. Turoff, "Decision support for emergency situations," *Inf. Syst. E-Bus. Manag.*, vol. 6, no. 3, pp. 295–316, Jun. 2008, doi: 10.1007/s10257-008-0087-z.

[58] “SYNERGY,” MASA Group. Accessed: Jul. 28, 2026. [Online]. Available: https://www.masasim.com/en/synergy
[59] “MASA Group | PreventionWeb.” Accessed: Jul. 28, 2026. [Online]. Available: https://www.preventionweb.net/organization/masa-group
[60] “Simulation using artificial intelligence in training senior crisis staff members and in validating contingency plans | UNDRR.” Accessed: Jul. 28, 2026. [Online]. Available: https://www.undrr.org/publication/simulation-using-artificial-intelligence-training-senior-crisis-staff-members-and
[61] M. Benali and A. R. Ghomari, “Information and knowledge driven collaborative crisis management: A literature review,” in *2016 3rd International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, Dec. 2016, pp. 1–3. doi: 10.1109/ICT-DM.2016.7857229.
[62] J. Groenendaal, I. Helsloot, and C. Reuter, “Towards More Insight into Cyber Incident Response Decision Making and its Implications for Cyber Crisis Management,” 2022.
[63] E. Rome, T. Doll, S. Rilling, B. Sojeva, N. Voß, and J. Xie, “The Use of What-If Analysis to Improve the Management of Crisis Situations,” in *Managing the Complexity of Critical Infrastructures: A Modelling and Simulation Approach*, R. Setola, V. Rosato, E. Kyriakides, and E. Rome, Eds., Cham: Springer International Publishing, 2016, pp. 233–277. doi: 10.1007/978-3-319-51043-9_10.
[64] J. Wolbers, M. Luesink, M. Van Duin, and S. Kuipers, “Scenario planning to enable foresight in crisis management,” *Proc. Int. ISCRAM Conf.*, vol. 21, May 2024, doi: 10.59297/zjzq2w57.
[65] E. Van Den Berg and S. Robertson, “Game-Theoretic Planning to Counter DDoS in NEMESIS,” in *MILCOM 2019 - 2019 IEEE Military Communications Conference (MILCOM)*, Norfolk, VA, USA: IEEE, Nov. 2019, pp. 1–6. doi: 10.1109/MILCOM47813.2019.9020844.
[66] A. F. Olivieri, R. E. Guadagno, S. Solari, and E. Russo, “EWACS as a Backbone for Wargaming and Decision Support in the Information Environment,” in *CEUR Workshop Proceedings*, Cagliari, Italy, Feb. 2026.
[67] M. Blom *et al.*, “Improving Tactical Decision-Making Through Multiobjective Contrastive Explanations,” *IEEE Trans. Hum.-Mach. Syst.*, vol. 55, no. 5, pp. 855–864, Oct. 2025, doi: 10.1109/THMS.2025.3591163.
[68] F. Santoni de Sio and J. van den Hoven, “Meaningful Human Control over Autonomous Systems: A Philosophical Account,” *Front. Robot. AI*, vol. 5, p. 15, Feb. 2018, doi: 10.3389/frobt.2018.00015.
[69] “Summary of the NATO Artificial Intelligence Strategy,” Official texts and resources | NATO. Accessed: Aug. 28, 2026. [Online]. Available: https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2021/10/22/summary-of-the-nato-artificial-intelligence-strategy
[70] “Regulation (EU) 2024/1689 of the European Parliament and of ...” Accessed: Aug. 28, 2026. [Online]. Available: https://eur-lex.europa.eu/legal-content/CA/LSU/?uri=oj%3AL_202401689
[71] S. Greco, M. Ehrgott, and J. R. Figueira, Eds., *Multiple Criteria Decision Analysis: State of the Art Surveys*, vol. 233. in International Series in Operations Research & Management Science, vol. 233. New York, NY: Springer, 2016. doi: 10.1007/978-1-4939-3094-4.
[72] B. Roy, *Multicriteria Methodology for Decision Aiding*, vol. 12. in Nonconvex Optimization and Its Applications, vol. 12. Boston, MA: Springer US, 1996. doi: 10.1007/978-1-4757-2500-1.

[73] R. L. Keeney and H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press, 1993.
[74] M. Grabisch and C. Labreuche, “A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid,” *4OR*, vol. 6, no. 1, pp. 1–44, Mar. 2008, doi: 10.1007/s10288-007-0064-2.
[75] T. L. Saaty, *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. McGraw-Hill International Book Company, 1980.
[76] J. P. Brans and Ph. Vincke, “A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making),” *Manag. Sci.*, vol. 31, no. 6, pp. 647–656, 1985.
[77] C.-L. Hwang and K. Yoon, “Methods for Multiple Attribute Decision Making,” in *Multiple Attribute Decision Making: Methods and Applications A State-of-the-Art Survey*, C.-L. Hwang and K. Yoon, Eds., Berlin, Heidelberg: Springer, 1981, pp. 58–191. doi: 10.1007/978-3-642-48318-9_3.
[78] S. Opricovic and G.-H. Tzeng, “Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS,” *Eur. J. Oper. Res.*, vol. 156, no. 2, pp. 445–455, Jul. 2004, doi: 10.1016/S0377-2217(03)00020-1.
[79] R. E. Bellman and L. A. Zadeh, “Decision-Making in a Fuzzy Environment,” *Manag. Sci.*, vol. 17, no. 4, pp. B141–B164, 1970.
[80] L. A. Zadeh, “Fuzzy sets,” *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965, doi: 10.1016/S0019-9958(65)90241-X.
[81] D.-Y. Chang, “Applications of the extent analysis method on fuzzy AHP,” *Eur. J. Oper. Res.*, vol. 95, no. 3, pp. 649–655, Dec. 1996, doi: 10.1016/0377-2217(95)00300-2.
[82] C.-T. Chen, “Extensions of the TOPSIS for group decision-making under fuzzy environment,” *Fuzzy Sets Syst.*, vol. 114, no. 1, pp. 1–9, Aug. 2000, doi: 10.1016/S0165-0114(97)00377-1.
[83] D. von Winterfeldt and W. Edwards, *Decision Analysis and Behavioral Research*. Cambridge University Press, 1986.
[84] W. Edwards, “How to Use Multiattribute Utility Measurement for Social Decisionmaking,” *IEEE Trans. Syst. Man Cybern.*, vol. 7, no. 5, pp. 326–340, May 1977, doi: 10.1109/TSMC.1977.4309720.
[85] W. Edwards and F. H. Barron, “SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement,” *Organ. Behav. Hum. Decis. Process.*, vol. 60, no. 3, pp. 306–325, Dec. 1994, doi: 10.1006/obhd.1994.1087.
[86] J. Mustajoki, R. P. Hämäläinen, and A. Salo, “Decision Support by Interval SMART/SWING—Incorporating Imprecision in the SMART and SWING Methods,” *Decis. Sci.*, vol. 36, no. 2, pp. 317–339, 2005, doi: 10.1111/j.1540-5414.2005.00075.x.
[87] C. A. Bana a Costa, J.-M. De Corte, and J.-C. Vansnick, “MACBETH,” *Int. J. Inf. Technol. Decis. Mak.*, vol. 11, no. 2, pp. 359–387, 2012, doi: https://doi.org/10.1142/S0219622012400068.
[88] E. Jacquet-Lagreze and J. Siskos, “Assessing a set of additive utility functions for multicriteria decision-making, the UTA method,” *Eur. J. Oper. Res.*, vol. 10, no. 2, pp. 151–164, 1982.
[89] S. Greco, V. Mousseau, and R. Slowinski, “Ordinal regression revisited: Multiple criteria ranking using a set of additive value functions,” *Eur. J. Oper. Res.*, vol. 191, no. 2, pp. 416–436, 2008.
[90] K. J. Arrow, *Social Choice and Individual Values*. Yale University Press, 2012. Accessed: Aug. 27, 2026. [Online]. Available: https://www.jstor.org/stable/j.ctt1nqb90

[91] I. Palomares, L. Martínez, and F. Herrera, “A Consensus Model to Detect and Manage Noncooperative Behaviors in Large-Scale Group Decision Making,” *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 3, pp. 516–530, Jun. 2014, doi: 10.1109/TFUZZ.2013.2262769.
[92] V. A. W. J. Marchau, W. E. Walker, P. J. T. M. Bloemen, and S. W. Popper, Eds., *Decision Making under Deep Uncertainty: From Theory to Practice*. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-05252-2.
[93] *Info-Gap Decision Theory*. 2006. Accessed: Aug. 27, 2026. [Online]. Available: https://shop.elsevier.com/books/info-gap-decision-theory/ben-haim/978-0-12-373552-2
[94] J. Fürnkranz and E. Hüllermeier, “Preference Learning,” in *Encyclopedia of Machine Learning and Data Mining*, Springer, Boston, MA, 2017, pp. 1000–1005. doi: 10.1007/978-1-4899-7687-1_667.
[95] A. Y. Ng and S. J. Russell, “Algorithms for Inverse Reinforcement Learning,” in *Proceedings of the Seventeenth International Conference on Machine Learning*, in ICML ’00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Jun. 2000, pp. 663–670.
[96] S. Arora and P. Doshi, “A survey of inverse reinforcement learning: Challenges, methods and progress,” *Artif. Intell.*, vol. 297, p. 103500, Aug. 2021, doi: 10.1016/j.artint.2021.103500.
[97] C. F. Hayes *et al.*, “A practical guide to multi-objective reinforcement learning and planning,” *Auton. Agents Multi-Agent Syst.*, vol. 36, no. 1, p. 26, Apr. 2022, doi: 10.1007/s10458-022-09552-y.
[98] J. P. Caylor and T. P. Hanratty, “Survey of Multi-Criteria Decision-Making Methods for Complex Environments,” CCDC Army Research Laboratory, Aberdeen Proving Ground, MD, USA, Technical Report ARL-TR-9054, 2020.
[99] X. Liu, J. Yao, X. Lu, H. Guo, and W. Wu, “Threat Evaluation of Air Targets Based on the Generalized lambda-Shapley Choquet Integral of GIFSS,” *Aerospace*, vol. 8, no. 5, p. 144, 2021, doi: 10.3390/aerospace8050144.
[100] N. Li, J. Huang, Y. Feng, K. Huang, and G. Cheng, “A Review of Naturalistic Decision-Making and Its Applications to the Future Military,” *IEEE Access*, vol. 8, pp. 38276–38284, 2020, doi: 10.1109/ACCESS.2020.2974317.
[101] R. L. Keeney, *Value-Focused Thinking: A Path to Creative Decisionmaking*. Harvard University Press, 1996.
[102] D. Vatne, M. Guttelvik, A. C. Hennum, and S. Malerud, “Wargaming for the Purpose of Knowledge Development: Lessons Learned from Studying Allied Courses Of Action,” *Scand. J. Mil. Stud.*, vol. 5, no. 1, pp. 297–308, Sep. 2022, doi: 10.31374/sjms.122.
[103] R. P. Hämäläinen and S. Alaja, “The threat of weighting biases in environmental decision analysis,” *Ecol. Econ.*, vol. 68, no. 1, pp. 556–569, Dec. 2008, doi: 10.1016/j.ecolecon.2008.05.025.
[104] B. Kitchenham and S. Charters, “Guidelines for performing Systematic Literature Reviews in Software Engineering,” Keele University and University of Durham, Durham, UK, Technical Report EBSE-2007-01, 2007. Accessed: Aug. 12, 2026. [Online]. Available: https://www.researchgate.net/publication/302924724_Guidelines_for_performing_Systematic_Literature_Reviews_in_Software_Engineering
[105] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.
[106] S. B. Son, S. Park, H. Lee, J. Kim, S. Jung, and D. H. Kim, “Tutorial on Course-of-Action (COA) Attack Search Methods in Computer Networks,” in *2022 13th International Conference on Information and Communication Technology Convergence (ICTC)*, Oct. 2022, pp. 1972–1975. doi: 10.1109/ICTC55196.2022.9952533.

[107] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, “Rayyan—a web and mobile app for systematic reviews,” *Syst. Rev.*, vol. 5, no. 1, p. 210, Dec. 2016, doi: 10.1186/s13643-016-0384-4.
[108] N. R. Haddaway, M. J. Page, C. C. Pritchard, and L. A. McGuinness, “PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis,” *Campbell Syst. Rev.*, vol. 18, no. 2, p. cl2.1230, Jun. 2022, doi: 10.1002/cl2.1230.
[109] C. H. Rinaudo, W. B. Leonard, J. E. Hopson, T. R. Coumbe, J. A. Pettitt, and C. Darken, “Applying Deep Reinforcement Learning to Train AI Agents in a Wargaming Framework,” in *SoutheastCon 2024*, Atlanta, GA, USA: IEEE, Mar. 2024, pp. 1131–1136. doi: 10.1109/SoutheastCon52093.2024.10500249.
[110] L. Ouriques, C. Eduardo Barbosa, and G. Xexéo, “Understanding Military Collaboration in Wargames,” in *2023 26th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, Rio de Janeiro, Brazil: IEEE, May 2023, pp. 1920–1925. doi: 10.1109/CSCWD57460.2023.10152624.
[111] R. Nelson *et al.*, “Air movement operations planning heuristic improvement,” *J. Def. Anal. Logist.*, Jan. 2025, doi: 10.1108/JDAL-02-2024-0003.
[112] L. Sebastian and S. Axel, “Enhancing Tactical Military Mission Execution through Human-AI Collaboration: A View on Air Battle Management Systems,” in *2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS)*, Toronto, ON, Canada: IEEE, May 2024, pp. 1–7. doi: 10.1109/ICHMS59971.2024.10555606.
[113] M. Mishra *et al.*, “A Context-Driven Framework for Proactive Decision Support With Applications,” *IEEE Access*, vol. 5, pp. 12475–12495, 2017, doi: 10.1109/ACCESS.2017.2707091.
[114] Liu H., Tang Y., Hu X., and Qiao G., “Research on Evaluation Framework of COA Based on Wargaming,” *J. Syst. Simul.*, vol. 30, no. 11, 2018.
[115] P. Poddar, E. T. Esfahani, K. Dantu, and S. Chowdhury, “Automated Generation of Diverse Courses of Actions for Multi-Agent Operations using Binary Optimization and Graph Learning,” in *2025 IEEE 21st International Conference on Automation Science and Engineering (CASE)*, Los Angeles, CA, USA: IEEE, Aug. 2025, pp. 2901–2908. doi: 10.1109/CASE58245.2025.11163864.
[116] M. Chmielewski, M. Kukiełka, P. Pieczonka, and T. Gutowski, “Methods and analytical tools for assessing tactical situation in military operations using potential approach and sensor data fusion,” *Procedia Manuf.*, vol. 44, pp. 559–566, 2020, doi: 10.1016/j.promfg.2020.02.255.
[117] T. Clancy, “Application of Emerging-State Actor Theory: Analysis of Intervention and Containment Policies †,” *Systems*, vol. 6, no. 2, p. 17, May 2018, doi: 10.3390/systems6020017.
[118] S. Dorton, B. Terry, B. Jaeger, and P. B. Shearer, “Development of a Recognition Primed Decision Agent for supervisory control of autonomy,” in *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, San Diego, CA, USA: IEEE, Mar. 2016, pp. 173–179. doi: 10.1109/COGSIMA.2016.7497806.
[119] D. Huber and D. Kallfass, “Applying data farming for military operation planning in nato MSG-124 using the interoperation of two simulations of different resolution,” in *2015 Winter Simulation Conference (WSC)*, Huntington Beach, CA, USA: IEEE, Jan. 2016, pp. 2535–2546. doi: 10.1109/WSC.2015.7408363.
[120] A. James, T. P. Hanratty, D. C. Tuttle, and J. B. Coles, “Agent based modeling in tactical wargaming,” presented at the SPIE Defense + Security, B. D. Broome, T. P. Hanratty, D.

L. Hall, and J. Llinas, Eds., Baltimore, Maryland, United States, May 2016, p. 985106. doi: 10.1117/12.2230916.

[121] A. Plachkov *et al.*, "Automatic Course of Action Generation using Soft Data for Maritime Domain Awareness," in *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*, Denver Colorado USA: ACM, Jul. 2016, pp. 1071–1078. doi: 10.1145/2908961.2931678.

[122] G. S. Reed, M. D. Petty, N. J. Jones, A. W. Morris, J. P. Ballenger, and H. S. Delugach, "A principles-based model of ethical considerations in military decision making," *J. Def. Model. Simul. Appl. Methodol. Technol.*, vol. 13, no. 2, pp. 195–211, Apr. 2016, doi: 10.1177/1548512915581213.

[123] T. Q. Bui, T. T. Nguyen, B. Q. Nguyen, and S. T. Le, "On the mathematical program in theater anti-aircraft distribution problem," in *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore: IEEE, Dec. 2017, pp. 999–1003. doi: 10.1109/IEEM.2017.8290042.

[124] P. Glover, S. Collander-Brown, and S. J. E. Taylor, "Using a genetic programming approach to mission planning to deliver more agile campaign level modelling for military operational research," in *2017 Winter Simulation Conference (WSC)*, Las Vegas, NV: IEEE, Dec. 2017, pp. 4465–4468. doi: 10.1109/WSC.2017.8248165.

[125] T. Guarda *et al.*, "Wargames Applied to Naval Decision-Making Process," in *Recent Advances in Information Systems and Technologies*, Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, and S. Costanzo, Eds., in Advances in Intelligent Systems and Computing, vol. 571. Cham: Springer International Publishing, 2017, pp. 399–406. doi: 10.1007/978-3-319-56541-5_41.

[126] A. James and T. P. Hanratty, "A technology path to tactical agent-based modeling," presented at the SPIE Defense + Security, T. P. Hanratty and J. Llinas, Eds., Anaheim, California, United States, May 2017, p. 1020707. doi: 10.1117/12.2266018.

[127] B. Kettler and J. Lautenschlager, "Expeditionary Modeling for Population-Centric Operations in Megacities: Some Initial Experiments," in *Advances in Cross-Cultural Decision Making*, S. Schatz and M. Hoffman, Eds., in Advances in Intelligent Systems and Computing, vol. 480. Cham: Springer International Publishing, 2017, pp. 3–15. doi: 10.1007/978-3-319-41636-6_1.

[128] L. Wan, W. Li, J. Dai, and A. Huang, "The modeling and solution of course of action generation based on IN-DEA," in *Journal of Physics: Conference Series*, China: IOP Publishing, 2017.

[129] Z. Jiang, W. ChuanHua, X. X. Wei, and D. Wei, "An Architecture Design for Fighting Method Decision Based on Course of Action," in *Recent Developments in Mechatronics and Intelligent Robotics*, F. Qiao, S. Patnaik, and J. Wang, Eds., in Advances in Intelligent Systems and Computing, vol. 690. Cham: Springer International Publishing, 2018, pp. 21–25. doi: 10.1007/978-3-319-65978-7_4.

[130] N. Shortland, L. Alison, and C. Barrett-Pink, "Military (in)decision-making process: a psychological framework to examine decision inertia in military operations," *Theor. Issues Ergon. Sci.*, vol. 19, no. 6, pp. 752–772, Nov. 2018, doi: 10.1080/1463922X.2018.1497726.

[131] K. Behymer, H. Ruff, G. Calhoun, J. Bartik, and E. Frost, "Presentation of Autonomy-Generated Plans: Determining Ideal Number and Extent Differ," in *Advances in Human Factors in Robots and Unmanned Systems*, J. Chen, Ed., in Advances in Intelligent Systems and Computing, vol. 784. Cham: Springer International Publishing, 2019, pp. 90–101. doi: 10.1007/978-3-319-94346-6_9.

[132] C. Chen, Z. Du, X. Liang, J. Shi, and H. Zhang, "Modeling and solution based on stochastic games for development of COA under uncertainty#br#," *J. Syst. Eng. Electron.*, vol. 30, no. 2, p. 288, 2019, doi: 10.21629/JSEE.2019.02.08.

[133] T. Heltberg and K. S. Dahl, “Course of Action Development – Brainstorm or Brickstorm?,” *Scand. J. Mil. Stud.*, vol. 2, no. 1, pp. 165–177, Oct. 2019, doi: 10.31374/sjms.30.
[134] L. Ouriques, C. E. Barbosa, G. Xexeo, and J. D. Souza, “Analyzing Knowledge Codification for Planning Military Operations,” in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy: IEEE, Oct. 2019, pp. 2620–2625. doi: 10.1109/SMC.2019.8914150.
[135] Y. Huang, B. Ge, B. Zhao, and K. Yang, “Course of Action Generation Using Graph Model for Conflict Resolution,” in *2020 IEEE 15th International Conference of System of Systems Engineering (SoSE)*, Budapest, Hungary: IEEE, Jun. 2020, pp. 000249–000254. doi: 10.1109/SoSE50414.2020.9130479.
[136] Y. Zhou, H. Zhao, J. Chen, and Y. Jia, “A novel mission planning method for UAVs’ course of action,” *Comput. Commun.*, vol. 152, pp. 345–356, Feb. 2020, doi: 10.1016/j.comcom.2020.01.006.
[137] P. C. Hershey, “Method for Self-Healing Course of Action Revision (SCOAR),” in *2021 IEEE International Systems Conference (SysCon)*, Vancouver, BC, Canada: IEEE, Apr. 2021, pp. 1–6. doi: 10.1109/SysCon48628.2021.9447122.
[138] A. D. W. Sumari, K. B. D. R. N. Widyasari, and V. A. Lestari, “A Knowledge Growing System-based Decision Making-Support System Application for Forces Command and Control in Military Operations,” in *2021 International Conference on ICT for Smart Society (ICISS)*, Bandung, Indonesia: IEEE, Aug. 2021, pp. 1–6. doi: 10.1109/ICISS53185.2021.9533245.
[139] J. Xin, W. Xinnian, Z. Fang, and D. Ran, “Uncertainty Based Hybrid-intelligence Multi-branch Wargaming,” in *2021 7th International Conference on Computer and Communications (ICCC)*, Chengdu, China: IEEE, Dec. 2021, pp. 1664–1669. doi: 10.1109/ICCC54389.2021.9674397.
[140] R. Kewley, C. Argenta, and K. Brawner, “Behaving Like Soldiers: A Multi-Agent System Approach to Course of Action Planning for Simulated Military Units,” in *The International FLAIRS Conference Proceedings*, May 2022. doi: 10.32473/flairs.v35i.130684.
[141] M. Schadd, A. M. Sternheim, R. Blankendaal, M. Van Der Kaaij, and O. Visker, “How a machine can understand the command intent,” *J. Def. Model. Simul. Appl. Methodol. Technol.*, vol. 22, no. 1, pp. 41–58, Jan. 2025, doi: 10.1177/15485129221115736.
[142] R. Kewley, C. Argenta, and C. McGoarty, “Tradeoffs within Representational Behaviors: Multi-Objective Optimization Strategies for Simulated Military Units,” in *The International FLAIRS Conference Proceedings*, May 2023. doi: 10.32473/flairs.36.133631.
[143] J. Cawalla and R. Pardhi, “Graph Reinforcement Learning for Courses of Action Analysis,” in *2024 International Conference on Military Communication and Information Systems (ICMCIS)*, Koblenz, Germany: IEEE, Apr. 2024, pp. 01–09. doi: 10.1109/ICMCIS61231.2024.10540763.
[144] Y. Liu, Y. Wang, H. Li, G. Wang, and J. Ai, “An adaptive operation planning and EBO-BPNN optimization method for decision support systems,” *Sci. Rep.*, vol. 14, no. 1, p. 21838, Sep. 2024, doi: 10.1038/s41598-024-72808-y.
[145] I. Petrović and M. Milenković, “Improvement of the operations planning process using a hybridized fuzzy-multi-criteria decision-making approach,” *Vojnoteh. Glas.*, vol. 72, no. 3, pp. 1093–1119, 2024, doi: 10.5937/vojtehg72-51473.
[146] L. Sun, Y. Zhou, H. You, C. Zhu, and W. Zhang, “An Efficient Planning Method to Recommend COA Based on New Command and Control Organizational Structure Model,”

IEEE Trans. Comput. Soc. Syst., vol. 11, no. 6, pp. 7359–7372, Dec. 2024, doi: 10.1109/TCSS.2024.3408653.

[147] N. De Reus *et al.*, "Generating Explainable Military Courses of Action," in *2024 International Conference on Military Communication and Information Systems (ICMCIS)*, Koblenz, Germany: IEEE, Apr. 2024, pp. 1–10. doi: 10.1109/ICMCIS61231.2024.10540875.

[148] J. Cawalla, "Learning Heuristics for Course of Action Analysis with Reinforcement Learning," in *2025 International Conference on Military Communication and Information Systems (ICMCIS)*, Oerias, Portugal: IEEE, May 2025, pp. 1–9. doi: 10.1109/ICMCIS64378.2025.11047783.

[149] N. Ding, "The Optimal Decision for Course of Action under Dynamic Warfare Environment," in *2025 IEEE 2nd International Conference on Electronics, Communications and Intelligent Science (ECIS)*, Yueyang, China: IEEE, May 2025, pp. 1–6. doi: 10.1109/ECIS65594.2025.11086810.

[150] J. W. Han, N. L. Ywet, A. A. Maw, and J.-W. Lee, "Bayesian Adaptive Tactical Decision Support System for Manned-Unmanned Teaming Operations,"

[151] C. Maathuis, "Multi-Agent System for Courses of Action Comparison in Military Operations," presented at the European Conference on Cyber Warfare and Security, Jun. 2025, pp. 384–392. doi: 10.34190/eccws.24.1.3318.

[152] M. Radovanović, M. Živković, and M. Crnogorac, "Application of Decision-Making Support Model in the Operations Planning Process at the Tactical Level," *Vojen. Rozhl.*, vol. 106, no. 1, pp. 85–103, Jan. 2026, doi: 10.3849/2336-2995.34.2025.01.085-103.

[153] J. Wu, Y. Lu, D. Li, Y. Ren, W. Zhou, and S. Yuan, "Modeling and Solution for Course of Action Planning Driven by Operational Task Network," *IEEE Access*, vol. 13, pp. 41169–41193, 2025, doi: 10.1109/ACCESS.2025.3532979.

[154] P. C. Hershey, "Raytheon Company - Approved for Public Disclosure: System for Adaptable, Scalable, and Autonomous Protection Verification and Decision Support (ASAP VDS) During Mission Planning and Execution," in *2026 IEEE International Systems Conference (SysCon)*, Halifax, NS, Canada: IEEE, Apr. 2026, pp. 1–6. doi: 10.1109/SysCon66367.2026.11503529.

[155] P. C. Hershey, "System of Systems Engineering Approach for Applying AI/ML to Decision Support," *IEEE Syst. J.*, pp. 1–10, 2026, doi: 10.1109/JSYST.2026.3654847.

[156] C. Maathuis, "A Theory of Mind Model for Proportionality Assessment in Military Operations," in *2026 International Conference on Unmanned Aircraft Systems (ICUAS)*, Corfu, Greece: IEEE, Jun. 2026, pp. 913–918. doi: 10.1109/ICUAS69441.2026.11598631.

[157] C. Maathuis, "Proportionality Assessment Optimization Model in Military Operations," in *2026 International Conference on Development and Application Systems (DAS)*, Suceava, Romania: IEEE, May 2026, pp. 1–6. doi: 10.1109/DAS69882.2026.11553312.

[158] R. Wieringa, N. Maiden, N. Mead, and C. Rolland, "Requirements engineering paper classification and evaluation criteria: a proposal and a discussion," *Requir. Eng.*, vol. 11, no. 1, pp. 102–107, Mar. 2006, doi: 10.1007/s00766-005-0021-6.

LIST OF ABBREVIATIONS AND ACRONYMS

AAP	Allied Administrative Publication
ADP	Army Doctrine Publication
AHP	Analytic Hierarchy Process
AI	Artificial Intelligence
AJP	Allied Joint Publication
AP	Automated Planning
ASCOPE	Areas, Structures, Capabilities, Organisations, People, and Events
ATP	Allied Tactical Publication
C-BML	Coalition Battle Management Language
C2	Command and Control
C2SIM	Command and Control Systems to Simulation Systems Interoperation
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance, and Reconnaissance
CADET	Course of Action Display and Evaluation Tool
CBR	Case-Based Reasoning
CCIR	Commander's Critical Information Requirement
COA	Course of Action
COPD	Comprehensive Operations Planning Directive
COPRAS	COmplex PRoportional Assessment
DARPA	Defense Advanced Research Projects Agency
DEMATEL	Decision-Making Trial and Evaluation Laboratory
DIBR	Defining Interrelationships Between Ranked criteria
DoD	Department of Defense
DR	Direct Rating
DSS	Decision Support System
EDAS	Evaluation based on Distance from Average Solution
ELECTRE	ELimination Et Choix Traduisant la REalité (Elimination and choice expressing reality)
FM	Field Manual
GT	Game Theory
IF	Information Fusion
IEEE	Institute of Electrical and Electronics Engineers
IPB	Intelligence Preparation of the Battlefield
JOPG	Joint Operations Planning Group
JP	Joint Publication
JPP	Joint Planning Process
KR	Knowledge Representation
LLM	Large Language Model
MAS	Multi-Agent System
MACBETH	Measuring Attractiveness by a Categorical Based Evaluation Technique

MCDA	Multi-Criteria Decision Analysis
MDMP	Military Decision-Making Process
METT-TC	Mission, Enemy, Terrain and weather, Troops and support available, Time available, Civil considerations
METT-TC(I)	METT-TC with Informational considerations
ML	Machine Learning
MSDL	Military Scenario Definition Language
MSG	Modelling and Simulation Group
NATO	North Atlantic Treaty Organization
NLP	Natural Language Processing
NN	Neural Network
NSATU	NATO Security Assistance and Training for Ukraine
OAKOC	Observation and fields of fire, Avenues of approach, Key terrain, Obstacles, and Cover and concealment
OODA	Observe–Orient–Decide–Act
OPLAN	Operation Plan
OPS	Operativna planska skupina (Operational Planning Group)
OSRH	Oružane snage Republike Hrvatske (Armed Forces of the Republic of Croatia)
PGM	Probabilistic Graphical Model
PL	Phase Line
PMESII	Political, Military, Economic, Social, Information, Infrastructure
PMESII-PT	PMESII plus Physical environment and Time
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
PROMETHEE	Preference Ranking Organisation Method for Enrichment Evaluations
QA	Quality Assessment
ROR	Robust Ordinal Regression[
RL	Reinforcement Learning
SCOAR	Self -Healing Course of Action Revision
SIM	Simulation
SMART	Simple Multi-Attribute Rating Technique
SMARTER	Simple Multi-Attribute Rating Technique Exploiting Ranks
SO	Search and Optimisation
TOPFAS	Tools for Operations Planning Functional Area Services
TOPSIS	Technique for Order Preference by Similarity to Ideal Solution
US	United States
UTA	UTilités Additives (Additive Utilities)
VIKOR	Višekriterijumska Optimizacija i Kompromisno Rešenje (Multi-criteria optimisation and compromise solution)
VIS	Visualisation
XAI	Explainable Artificial Intelligence
ZDP	Združeni doktrinarni priručnik (Joint Doctrine Publication)

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

This qualification paper examines how the four doctrines governing Croatian operations planning support the selection of a course of action, what multi-criteria decision analysis requires of that support, and how far the research literature of the past decade has supplied what doctrine leaves out. It covers the conceptual foundations and terminology of courses of action, the operational factor frameworks from which evaluation criteria are drawn, decision support for the course of action process in command and control, and the methods of multi-criteria decision analysis. It then reports a systematic review of 45 studies published between 2016 and 2026.

Each of the four doctrines represents course of action selection as a discrete multi-attribute problem, and each leaves to judgment the parts that carry the solution. No doctrine supplies a method for deriving evaluation criteria from the operational variables it describes in detail, and none specifies how weights are to be obtained. Three of the four supply a weighted sum, and every doctrine that supplies arithmetic also warns against relying on it. Each nonetheless places a non-compensatory filter ahead of the compensatory rule, so that an inadmissible option is removed rather than compensated.

The review finds the literature routing around these gaps rather than closing them. Criteria are most often assembled ad hoc, weighting is unstated in 19 of the 45 studies, aggregation is dominated by the weighted sum, and the separation of screening from evaluation survives in two. The corpus covers comparison most heavily and revision during execution barely at all, with its sharpest discontinuity between the two: 27 studies score alternatives without addressing the choice among them. Reinforcement learning, neural networks and a single large language model study enter only in the last three years, applied to generating options and estimating outcomes rather than to comparison, and the commander is most often absent or cast as a reviewer of machine output. The gap is at the comparison step, which needs to be transparent, defensible and open to revision as the operation unfolds.

Keywords: course of action; military operations planning; multi-criteria decision analysis; decision support systems; artificial intelligence; systematic literature review

PROŠIRENI SAŽETAK

Ovaj kvalifikacijski rad ispituje na koji način četiri doktrine koje uređuju planiranje operacija podupiru odabir inačice djelovanja, što višekriterijska analiza odlučivanja traži od takve potpore te u kojoj je mjeri znanstvena literatura posljednjeg desetljeća popunila ono što doktrina ostavlja neodređenim. U radu su obuhvaćeni pojmovni temelji i terminologija inačica djelovanja, okviri operativnih čimbenika iz kojih se izvode kriteriji vrednovanja, potpora odlučivanju u procesu zapovijedanja i nadzora te metode višekriterijske analize odlučivanja. Nakon toga izložen je sustavni pregled 45 studija objavljenih u razdoblju od 2016. do 2026. godine.

Sve četiri doktrine prikazuju odabir inačice djelovanja kao diskretan više atributni problem i sve prosudbi prepuštaju upravo one korake o kojima rješenje ovisi. Nijedna doktrina ne propisuje postupak izvođenja kriterija vrednovanja iz operativnih varijabli koje inače potanko opisuje, pa je pretvorba operativnog razmatranja u kriterij neizrečen korak u sve četiri. Nijedna ne određuje kako se dolazi do težina. Tri od četiri doktrine navode ponderirani zbroj, a svaka doktrina koja daje računski postupak ujedno upozorava da se na njegov rezultat ne treba oslanjati. Unatoč tome, svaka doktrina postavlja nekompenzacijski filtar ispred kompenzacijskog pravila, tako da se neprihvatljiva inačica odbacuje, a ne nadoknađuje dobrim ocjenama drugdje.

Pregled literature pokazuje da istraživanja te praznine zaobilaze, a ne zatvaraju. Kriteriji se najčešće sastavljaju neposredno za konkretni slučaj, u 19 od 45 studija nije naveden nijedan doktrinarni okvir planiranja, u njih 19 težine nisu obrazložene, u agregaciji prevladava ponderirani zbroj, a razdvajanje probira od vrednovanja zadržano je u dva rada. Korpus najviše obuhvaća usporedbu inačica, dok je njihova izmjena tijekom izvođenja gotovo neobrađena, a najoštriji prekid nalazi se upravo između te dvije faze: 27 studija ocjenjuje alternative bez razmatranja izbora među njima.

Najizrazitija promjena u korpusu odnosi se na uvođenje umjetne inteligencije. Podržano učenje, neuronske mreže i jedna studija s velikim jezičnim modelom pojavljuju se tek u posljednje tri godine, i to u generiranju inačica i procjeni njihovih ishoda, a ne u usporedbi i odabiru. Dvadeset studija smješta čovjeka u fazu usporedbe i odabira, četrnaest mu dodjeljuje ulogu ocjenjivača strojno proizvedenih inačica, a jedanaest ga uopće ne uključuje. Dvadeset i osam studija ne navodi nikakvu osnovicu za usporedbu, a samo dvije uspoređuju svoj pristup s neposredovanom ljudskom prosudbom, pa ostaje neodgovoreno unapređuje li ijedan predloženi pristup postojeći stožerni postupak.

Ovim je radom doktrina, teorija odlučivanja i desetljeće istraživanja objedinjeno u jedan pregled, iz kojega proizlazi da nedostatak nije u brzini generiranja inačica, nego u koraku usporedbe koji bi bio provjerljiv, obranjiv i otvoren za izmjenu tijekom izvođenja operacije.

Ključne riječi: inačica djelovanja; planiranje vojnih operacija; višekriterijska analiza; sustavi za podršku odlučivanju; umjetna inteligencija; sustavni pregled literature