Mission-Relevant Criteria for Naval Task Group (NTG) Course of Action (COA) Analysis

Tania Randall, Aren Hunter, Mark Hazen, James Karle

Defence Research and Development Canada – Atlantic Research Centre
Dartmouth, Nova Scotia, Canada
{tania.randall, aren.hunter, mark.hazen, james.karle}
@drdc-rddc.gc.ca

Abstract. Defence Research & Development Canada – Atlantic Research Centre is developing a prototype planning and course of action (COA) testbed for naval command teams. As part of this effort, a study was conducted to understand how criteria for the assessment of Naval Task Group (NTG) courses of action are selected, defined, prioritized and applied. Fourteen naval officers with operational planning experience identified and defined a set of criteria for assessing COAs for four NTG missions. The responses of this first data collection activity were consolidated, categorized, and mapped to the Canadian Forces' Principles of War, to create a short list of relevant criteria for each mission. In a follow-on data collection activity, naval officers were provided the refined criteria lists and, both individually and in groups, completed activities focused on defining the relative importance of the criteria to each mission. The results of the study are presented and framed from the perspective of their application to the COA testbed (COA-T).

Keywords: evaluation criteria, courses of action, mission planning

1 Introduction

Defence Research & Development Canada - Atlantic is developing a prototype planning and course of action (COA) testbed (COA-T) to support shipboard and naval task group (NTG) planning in the 24-72 hour timeframe, across a range of maritime information warfare (MIW) activities. It aims to improve the command team's ability to develop, assess, select and communicate COAs.

From [1], a COA is any potential solution which may result in the accomplishment of the mission. A *viable* COA is a solution that is suitable, feasible, acceptable, compliant, exclusive and complete [2]. Past research [3] indicates that ship and NTG staff will take certain liberties to adapt formal planning processes such as the Canadian Forces Operational Planning Process (OPP) [2], Joint Operational Planning Process (JOPP) [4] or Naval Planning Process (NPP) [1], in order to suit their needs. Nevertheless, the process of assessing the mission requirements, developing feasible COAs, and analyzing the COAs in order to select a COA is common across the various formal and modified methods. The JOPP illustrates the steps of the planning process at a level of detail appropriate for the discussion in this paper, as shown in Figure 1.

The JOPP is the planning method used by the US military to "facilitate interaction between the commander, staff, and subordinate and supporting headquarters throughout planning" ([4], pp. IV-1). It describes a process to "examine a mission; develop, analyze, and compare alternative COAs; select the best COA; and produce a plan or order" ([4], pp. IV-1). In this paper, our primary focus is on Step 4: COA Analysis, and also touches on Step 5: COA comparison.

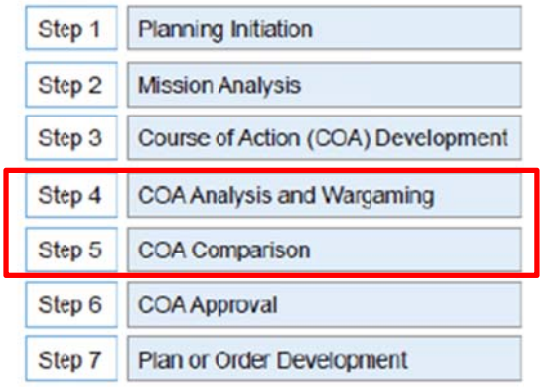

Fig. 1. The Joint Operation Planning Process [4]

1.1 Course of Action Analysis

The COA Analysis step involves reviewing the COAs developed in the previous stage for viability, wargaming of the viable COAs (i.e., exploring how the COA and the situation will evolve over time in order to gain a sense of how the COA will play out and what it's strength and weaknesses may be), and development of the evaluation criteria and their relative importance for use in the COA comparison process. As described in the JOPP, criteria represent aspects of the situation that are critical to accomplishing the mission, and the criteria selected must be sufficient to differentiate the COAs. Some examples of evaluation criteria are: covertness, vessel coverage, fuel usage, and security.

1.2 Course of Action Comparison

In the COA Comparison step, the staff evaluate all COAs against the established evaluation criteria in order to select the COA that best supports the mission. OPP, JOPP, and NPP propose decision matrices as the decision aid that systematically lists the evaluation criteria and helps in differentiating COAs [5]. In our studies, we use weighted comparison matrices, consistent with those found in sample planning briefs obtained from Royal Canadian Navy (RCN) Subject Matter Experts (SMEs) and observed RCN OPP training sessions. In its simplest form, the weighted comparison matrix allows systematic comparison of more than one COA (typically, up to 3 COAs are chosen), by assessing the value of each projected COA outcome with respect to each criteria, and calculating a weighted sum across all criteria to determine an overall value for the COA. This is illustrated in Figure 2.

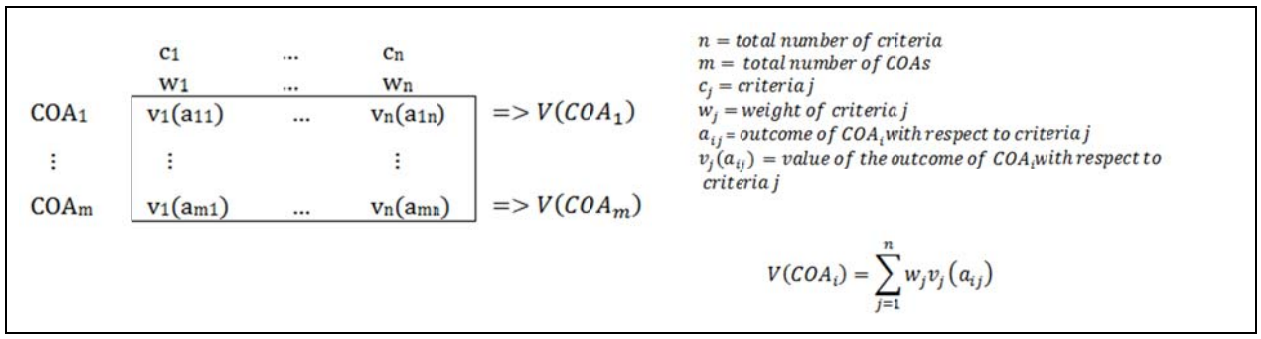

Fig. 2. Weighted Comparison Matrix (Multiple COAs)

In the event that the mission involves some uncertainty about the action of an opposing force (i.e., an 'enemy'), the weighted comparison matrices will also include consideration for possible COAs to be taken by the enemy (referred to as Enemy COAs, or, ECOAs). This more complex matrix is illustrated in Figure 3.

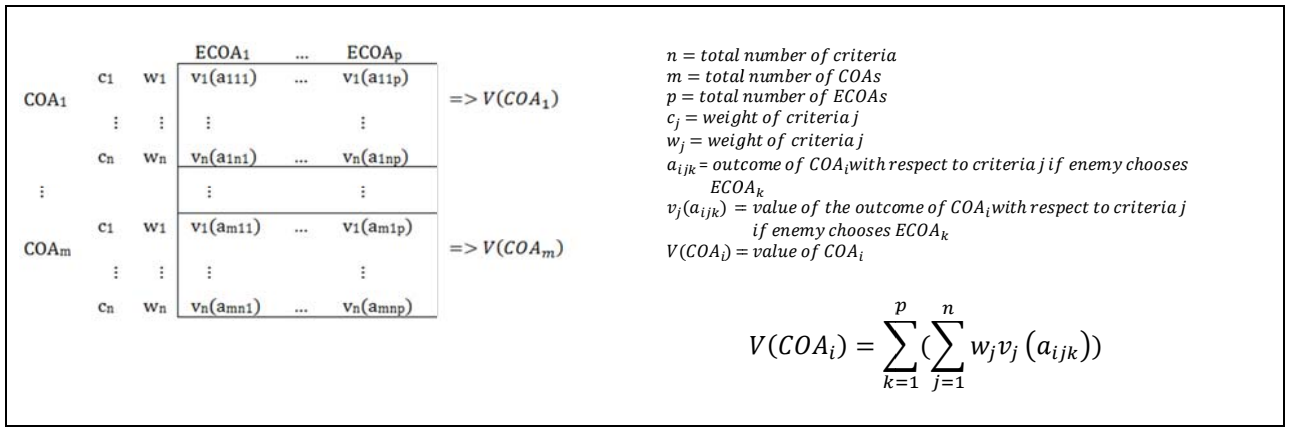

Fig. 3. Weighted Comparison Matrix (Multiple COAs and Enemy COAs)

1.2.1 Course of Action Difference Thresholds

In practice, when $V(COA_a) > V(COA_b)$, COA_a is considered preferred to, or, better able to accomplish the mission than, COA_b. However, it is noted in [6] that there are many uncertainties within the numbers used to calculate an overall COA value. In the simpler case, without ECOAs, the projected outcome, a_{ij}, of COA_i with respect to criterion j involves uncertainty. Further, the planners' assessment of the 'value' of this uncertain outcome, $v_j(a_{ij})$, with respect to criterion j adds further uncertainty. These uncertainties are amplified in the more complex case, where we are also dealing with the uncertainty of the enemy's actions. Thus, it may be more appropriate to define some difference threshold that must be exceeded in order to say that one COA is better than another. While dealing with these uncertainties is not the focus of the research reported here, it is important to keep them in mind when the evaluation criteria and weightings are applied in the COA comparison process.

1.3 Motivation for Work

In June 2016, authors attended an RCN-hosted naval operational planning working group aimed at tailoring the OPP for the naval environment and developing draft Contingency Operation Plans (CONPLANs) and Operations Orders (OPORDs) for a selected set of four tasks that Canadian Fleet Atlantic (CANFLTLANT) must "be prepared to" do [7]. These 'tasks', which we refer to as 'mission types' throughout this paper, were: Maritime Interdiction Operations (MIO), Humanitarian Assistance/Disaster Recovery (HADR), Combat Patrol (CP), and Non-combatant Evacuation Operation (NEO). The fleet staff organizers selected the JOPP, rather than the CF OPP for use by this working group. The JOPP describes the "interplay among the Strategic, Operational and Tactical levels, whereas the OPP resides primarily at the Operational level" [4], and JOPP is thought to provide richer guidance for missions such as those considered here, which were focussed at the Maritime Component Command (MCC) and Naval Task Group (NTG) level, rather than the level of individual ships. The JOPP, like other planning approaches, provides guidance in a generic fashion in order to cover all possible situations. When considered in light of a specific mission type, it is possible to begin weeding out parts of the guide that do not apply, and developing mission-specific briefing templates (i.e., PowerPoint decks, which are commonly used as the outputs of various stages of the planning process) that can then provide a more refined starting point when an operational tasking arrives.

Naval staff were divided into syndicates to take initial steps towards generating these mission-oriented planning templates. A key element of the process is the selection of evaluation criteria and associated weightings. The current practice is for the planning team (or commander) to select criteria and weightings in situ. A concern is that time constraints in situ do not allow for full consideration of the impact of the criteria choices, nor the development of well-considered criteria weightings. Thus, if default sets of COA evaluation criteria and weightings specific to NTG mission sets could be determined, then they could reduce the time and cognitive effort required in the COA analysis, and provide improved guidance to users. The default criteria and weightings would be added to the planning templates in the short-term, and used to guide the content of visualizations and development of decision aids in the longer term.

In order for COA-T to aid the planner in developing COAs (Step 3), analysing COAs (Step 4), and comparing COAs (Step 5), it is necessary for us to understand what things planners are considering when completing these tasks. We anticipated that the COA evaluation criteria and weightings would provide initial insights into some of the foundational requirements of COA-T, while also providing a more immediate output for the CANFLTLANT staff.

1.4 Overview of Studies

This paper describes two studies aimed specifically at exploring mission-specific criteria and weightings, and a third study that was leveraged to explore application of the default criteria. Figure 4, adapted from [5], shows a break-down of the COA analysis phase and how the three studies map to the COA analysis and comparison processes.

Fig. 4. COA Analysis and Comparison (adapted from [5])

Study 1 aimed to identify and define the evaluation criteria for the specific CANFLTLANT missions described above. Study 2 took the evaluation criteria identified in Study 1 for each mission, and explored various individual and group approaches to determining criteria weightings. Study 3 used the default criteria of Study 1 in an applied setting, where two teams fully developed and compared COAs for two of the mission sets.

2 Identification of Mission-Specific Evaluation Criteria

2.1 Approach

The four mission descriptions that were created for the Naval OPP working group were employed in this study. Each mission description included brief paragraphs on the situation, higher command intent, the mission, and the available friendly forces. A brief summary of each mission is shown in Table 1.

Mission	Summary
MIO	Employ NTG to conduct MIO in support of task force anti-narcotics counter-terrorism operations to reduce the flow of illegal drugs and financial support to terrorist groups.
HADR	NTG to transit to hurricane-struck area, to provide humanitarian assistance and disaster response in order to reduce the scale of human suffering and support the local people and government.

CP	NTG to conduct a combat patrol in order to identify, locate, and track vessels transferring enemy fighters in order to provide targeting communication to air assets and destroy enemy targets at sea, should air assets be unavailable.
NEO	NTG to plan and conduct an evacuation of Canadians from their current location to an alternative location, no later than DDMMYYY, ensuring safety of Canadians in the region.

Table 1. Mission Summaries (used in Study 1 and Study 2)

An Excel-based survey was used to provide each participant with a description of the missions and the input that was being sought, including 5 evaluation criteria and associated descriptions. A partial survey is shown in Figure 5.

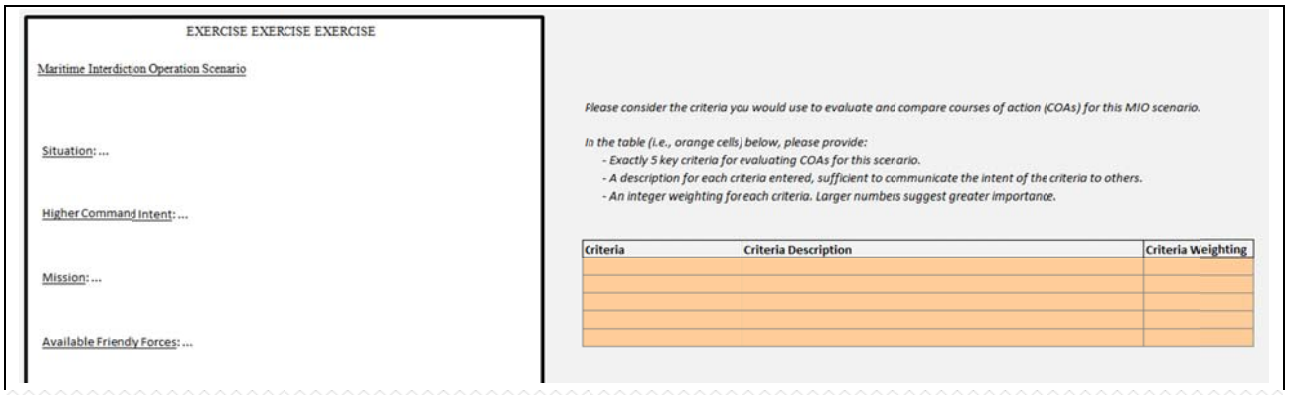

Fig. 5. Partial EXCEL page (scenario details redacted), used for data collection

2.1.1 Participants

Fourteen RCN participants with OPP training and/or experience participated in this study: six Lieutenants, five Lieutenant Commanders, and three Commanders. To limit the time commitment required (and maximize response rate), each participant was asked to complete the survey for only two of the four mission sets: either HADR and Combat Patrol, or MIO and NEO, determined by the initial letter of their surname. We rectified a slight imbalance in the number of responses for each combination when the second (and final) call for participants went out. Table 2 indicates the final break-down of participants. In addition to these, a Lieutenant Commander piloted the survey and provided feedback prior to the broader distribution.

	HADR and CP	MIO and NEO
Lt(N)	3	3
LCdr	2	3
Cdr	2	1

Table 2. Number of participants by rank and mission sets (Study 1)

2.2 Analysis

The responses received were quite varied. In some cases, the criteria indicated were more reflective of a request for additional information (e.g., what are my rules of engagement?) than they were of evaluation criteria. This may indicate that mission descriptions that we re-purposed could have benefited from some additional background information or assumptions for the purposes of our study. Responses that did not fit the definition of evaluation criteria were not used in the analysis. Upon initial visual inspection of the responses, the best method of consolidating them into a set of four (or so) top criteria for each mission type was not evident. The expectation that

participants would provide similar responses that could be easily grouped was not met. Also, the level of detail in the criteria descriptions varied widely, making some more clear than others. In addition, some participants used different terms to describe the same or similar things.

We enlisted the help of a SME in both naval missions and operational planning processes to organize the responses. Planning guidance suggests that an initial set of high level criteria should be derived from the commander's intent. The CF's principles of war (POW) also provide a reasonable starting point for developing evaluation criteria. In Canadian military doctrine, there are ten "principles of war" that govern the application of military power [5]. In brief, the CF's principles of war are [8]: selection and maintenance of the aim, maintenance of morale, offensive action, security, surprise, concentration of force, economy of effort, flexibility, cooperation and administration.

Our RCN SME conducted a careful review of the commander's intent statements, deriving evaluation criteria consistent with both the intent and POWs. The collection of criteria provided by the participants was then examined and each response was either mapped to a POW category or removed from the analysis. If no participant data existed for a POW category, even when expected by our SME based on the commander's intent, that POW was also removed. The number of mappings of participant data to each POW category was used to infer importance of the criteria.

A second RCN SME - a Lieutenant Commander with relevant at-sea experience - reviewed the results and indicated agreement with the criteria that were identified.

2.3 Results

The final sets of criteria and their inferred priority (1=High, 4=Low) for each mission are provided in Table 3, mapped to the Principles of War.

	Criteria by Mission (with Rank)			
Principle of War	**MIO**	**NEO**	**Combat**	**HADR**
Selection and Maintenance of the Aim		Time (3)	Geographical Coverage (1)	
Concentration of force	Effect (1)		Readiness (2)	Resource Allocation (2)
Economy of Effort	Sustainability (2) Force Employment (3)	Economy of Effort (4)		
Security	Risk (4)	Security (1)	Security (2)	
Flexibility		Flexibility		Flexibility (1)
Cooperation		Cooperation (2)	Cooperation (3)	Cooperation (3)
Administration (Logistics)		Logistics (3)		Sustainment (4)
Offensive Action			Offensive Action (3)	

Table 3. Evaluation Criteria by Mission Types (with priorities)

In a related research effort [5], DRDC contractors describe a framework for defining criteria to limit ambiguity and redundancy. It involves systematically defining the criteria in terms of four elements: intent (i.e., the direction of change), effect, subject, and target of the approach or effect (e.g., own force). They illustrate the framework with three examples, read left to right, as shown in Figure 6.

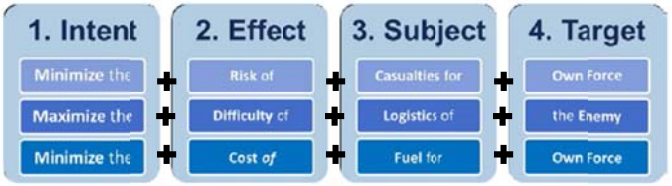

Fig. 6. COA Criteria Definition Framework (modified from [5])

For example, "Minimize the risk of casualties for own force". In this case, for the criterion 'risk', it is important to know who the risk is being posed to (i.e., own forces), that it relates to casualties, and that the intent (in this case, obviously) is to *minimize* risk. The order of the elements is not critical, so long as they are clear in the criteria description.

Table 4 provides definitions for the criteria identified through Study 1 and listed in Table 3. These definitions are rooted in those provided by study participants, adjusted to follow the recommended criteria definition scheme where needed to minimize ambiguities.

	Criteria	Description
MIO	Effect	Which COA will maximize the number of interceptions (of enemy force)?
	Sustainability	Which COA creates the shortest re-supply chain whilst maintaining the desired operational effect?
	Force Employment	Which COA demonstrates the best employment of force (e.g., alignment of ship capabilities to mission requirements)?
	Risk	Which COA will minimize risk to CAF personnel?
NEO	Time	Which COA minimizes Time on Station while Canadians are being evacuated?
	Security	Which COA maximizes security (e.g., range of fire, accommodations in place) for operations and Canadians?
	Economy of Effort	Which COA maximizes organic assets thereby minimizing the need for external support?
	Flexibility	Which COA is most easily modifiable to account for changing conditions and political agreements?
	Cooperation	Which COA maximizes the use of deliberate measures to engage other militaries and relevant government and non-government departments?
	Logistics	Which COA has best procedures in place to support the verification and processing of Canadians?
CP	Geographical Coverage	Which COA maximizes the geographical coverage of traffic in the waters between <<location x>> and <<location y>>?
	Offensive Action	Which COA is most prepared to take offensive action on enemy vessels?
	Security	Which COA maximizes own ship security (e.g., standoff ranges)?
	Readiness	Which COA maximizes crew and ship readiness?
	Cooperation	Which COA maximizes communications with the Combined Air Operations Center?
HADR	Flexibility	Which COA is most prepared for possibility of situation deterioration?
	Resource Allocation	Which COA balances consideration between weapons and HADR supplies?

	Cooperation	Which COA contains most deliberate measures to support the Rapid Response Team and engage with relevant government and non-government departments?
	Sustainment	Which COA contains best logistical support for re-supply?

Table 4. Evaluation Criteria Definitions

3 Exploration of Criteria Weighting Methods

This study takes the criteria identified for each mission in Study 1, and examines various methods of determining weights for these criteria.

3.1 Approach

For each mission type, the top four criteria were carried forward into this second study. However, the NEO mission was not used for further data collection; rather it was used an example for participants. Thus, only the HADR, MIO and CP missions were fully examined in this study. The current practice for determining criteria weightings is simple - the planning team uses their intuition and experience to create, typically as a group, integer-based weightings for each criterion that are appropriate for the mission at hand. This *direct* approach is straightforward and efficient, though lacks the rigor of some weight elicitation techniques described in the multi-criteria decision making (MCDM) literature. The authors of [9] suggest that weighting techniques may be selected based on their perceived validity in the decision making context, the cognitive effort required of the participants, and the complexity of the analysis. In the current time-sensitive planning environment, without tools to support the analysis, it could be difficult to make an argument for a more complex weight elicitation technique. However, given a planning support tool like will be demonstrated using COA-T that could easily convert user inputs into the appropriate set of weightings, the option of using a more sophisticated weighting method becomes more realistic. The cognitive effort required by the planners to create the necessary input data, however, remains a consideration. The first thing to determine is whether or not using a different technique results in criteria weightings that are meaningful and impact the end result (i.e., result in the selection of a different COA).

In this study, we consider two alternative techniques to the current direct, intuitive weighting approach, and also consider the possibility of applying the weighting techniques in a group setting (as is current practice) or as individuals. In addition to the *direct* method, we also used a *points allocation* and *Analytic Hierarchy Process* (AHP) approach [10]. Points allocation (PA) involves splitting 100 points amongst the criteria to indicate their relative importance. It is straightforward, but offers a couple of benefits over the direct method:

1. it forces the evaluator to consider trade-offs between criteria (i.e, more points for one criteria means less points for another; this is not the case when using the direct, ad hoc approach – the sum of weights is unlimited in that case), and,
2. it may produce weights with a finer resolution (e.g., with the direct method, it is common to see users define weights which are simple multiples of each other (for instance, $w_1 = 2$, $w_2=1$, $w_3=1$). With PA, weights such as $w_1 = 55$, $w_2=25$, $w_3=20$ are just as likely).

The Analytical Hierarchy Process was developed by Saaty [11], and involves multiple decision makers individually prioritizing a set of distinct evaluation criteria, and conducting a set of pairwise comparisons between the criteria, using their judgement to indicate each criterion's relative importance to all those above it in the prioritization hierarchy. The method applies matrix algebra to derive an overall set of weights for the criteria. The AHP process is widely accepted and supports the comparison of incommensurable criteria (e.g., mission cost, loss of life) two at a time, which allows for more precise judgements than reviewing the complete criteria set at once (as in the previous

methods discussed). For the AHP method, data was collected only at the individual level, and analyzed for group consensus.

This study took place in a group setting, over the course of 1 hour. Participants were briefed on the results of Study 1 and the intent of Study 2, as well as given a copy of a complete example, using the NEO mission set.

3.1.1 Participants

Fourteen RCN participants with OPP training and/or experience participated in this study: six Lieutenants, four Lieutenant Commanders, and four Commanders. Twelve of these participants had previously participated in Study 1, where they identified criteria for two of the four mission sets. To save time and maintain consistency in Study 2, we aimed to assign participants to a mission that they already studied in Study 1. However, creating similarly-sized teams which were balanced across ranks and mission sets (especially given that we now had only 3 missions instead of 4) was challenging. The participants ultimately assigned to each group are indicated in Table 5.

Participant experience with the OPP varied by rank, and ranged from training only, to implementation in exercises, to application during deployed operations, and teaching the OPP to other officers. Participant positions within the RCN included weapons, operations room, combat, executive, and commanding officers.

	HADR	CP	MIO
Lt(N)	2	2	2
LCdr	1	1	2
Cdr	1	2	1

Table 5. Number of participants by rank and mission (Study 2)

3.1.2 Data Collection Survey

A paper-based survey was used to guide data collection. The survey structure is described in Table 6. In addition to that shown here, a brief demographic and exit survey were conducted.

Section	Description
Reprint of the Mission Description	
Assumptions	Three viable COAs already exist, and evaluation criteria have already been determined. Survey 1 criteria and descriptions for that mission were then provided.
Part A: Individual Exercise	
A1: Ranking	Rank the four criteria provided in order of priority.
A2: Direct Method	Enter a weight (1 to 10) for each criterion to indicate its relative importance. (1=low, 10=high)
A3: Points Allocation Method	Divide 100 points amongst the four criteria to indicate their relative importance.
A4: AHP Method	Enter a number (1, 3, 5, 7, 9) in the third column to indicate the relative importance of the two criteria (X, Y) in the corresponding row. 1 = X is equal in importance to Y 3 = X is moderately more important than Y 5 = X is strongly more important than Y 7 = X is very strongly more important than Y

	9 = X is extremely more important than Y
Part B: Group Exercise	
B1: Ranking	As a group, rank the four criteria provided in order of priority.
B2: Direct Method	As a group, enter a weight (1 to 10) for each criterion to indicate their relative importance.
B3: Points Allocation Method	As a group, divide 100 points amongst the four criteria to indicate their relative importance.

Table 6. Study 2 Data Collection Survey Structure

3.2 Analysis

The calculations applied to the collected data are described in Table 7, where:

Q = Set of all participants
J = Set of all criteria
w_{jq} = Weight of criteria j, assigned by participant q
w_j = Weight of criteria j
w_j' = Normalized weight of criteria j

Method	Analysis Applied
Part A: Individual Exercise	
A1: Ranking	(This was used only to enable application of the AHP method.)
A2: Direct Method	$w_k' = \frac{\sum_Q w_{jq}}{\sum_J \sum_Q w_{jq}}$
A3: Points Allocation Method	$w_j' = \frac{\sum_Q w_{jq}}{\sum_J \sum_Q w_{jq}}, where \sum_J w_{jq} = 100 \; for \; each \; q$
A4: AHP Method	Excel-template applied, see 3.2.1
Part B: Group Exercise	
B1: Ranking	(no analysis needed)
B2: Direct Method	$w_j' = \frac{w_j}{\sum_J w_j}$
B3: Points Allocation Method	$w_j' = \frac{w_j}{\sum_J w_j}, where \sum_J w_j = 100$

Table 7. Study 2 Data Analysis

3.2.1 AHP Excel-Template

To perform the AHP calculations, we leveraged an Excel-based AHP template [12] which allowed for easy entry of participant data, and calculation of the overall criteria weights for each mission set. An example is shown in Figure 7. The first image to the left is an example of an individual's data that was copied into the worksheet. The image to the right is the summary page produced by this AHP Excel template for this example.

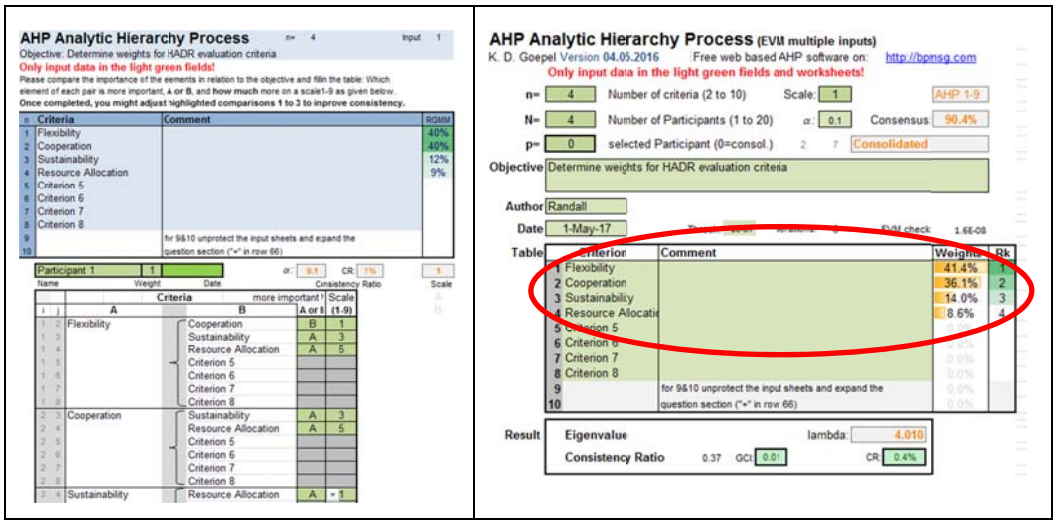

Fig. 7. Sample individual and summary tabs of AHP template [12]

3.2.2 Pairwise Comparison of Elicitation Methods

This study resulted in 5 different sets of weights for the criteria of each mission set. In order to explore how similar the weights derived from each method are, a simple EXCEL tool was later used to compare which of three randomly assessed COAs would be selected, using a weighted comparison matrix such as described in Section 1.2. Mimicking the process observed at an RCN OPP training session, a color of red, yellow or green was selected for each cell, with corresponding values of -1, 0, and 1, respectively. An image of a sample run is shown in Figure 8, with randomly selected values. The value, $V(COA_i)$, for each set of weights is shown on the right-hand side of the image, and the COA with the highest V(COAi) for each set of weights is highlighted in green. In the example shown, all weights result in COA3 being selected.

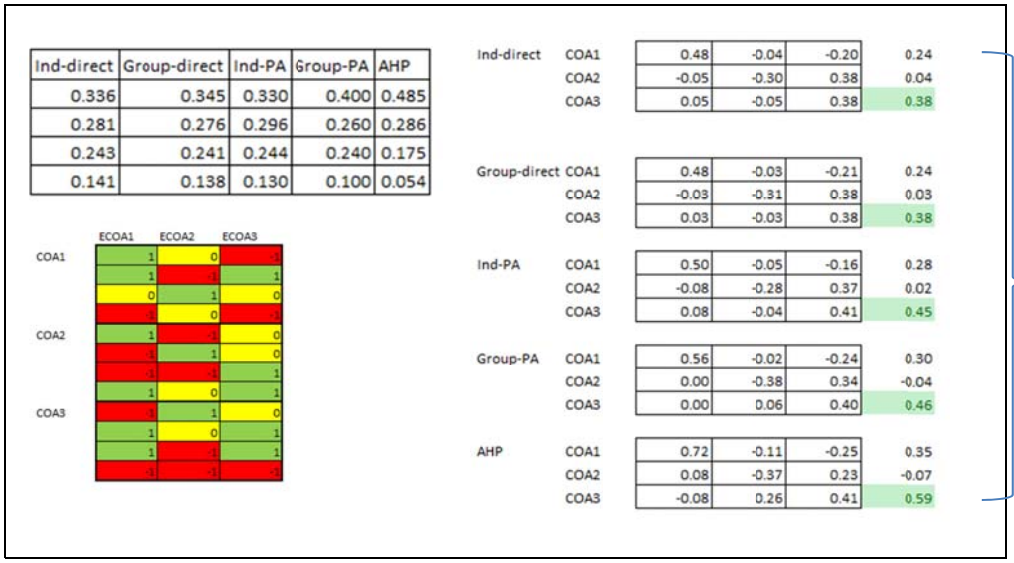

Fig. 8. Sample application of weights with random values for $v_j(a_{ijk})$

For each mission set, 100 such comparisons (runs) were completed and the number of times the results indicated the same COA being selected for the following pairs were tallied:

- Comparisons between individuals and group:
  - Individual-direct vs Group-direct Total # of matches out of 100: _____
  - Individual-PA vs Group-PA Total # of matches out of 100: _____
- Comparisons between methods:
  - Individual-direct vs Individual-PA Total # of matches out of 100: _____
  - Group-direct vs Group-PA Total # of matches out of 100: _____
  - AHP vs Group-direct Total # of matches out of 100: _____
  - AHP vs Group-PA Total # of matches out of 100: _____
  - AHP vs Individual-PA Total # of matches out of 100: _____

For instance, for the run shown in Figure 8, 1 point would be added to each of these comparison categories, as all weighting methods resulted in COA3 being selected.

3.3 Results

3.3.1 Weights Determined by each Elicitation Method

The weights obtained for the criteria provided for each mission using each method are detailed in Table 8. These values were calculated as described in Table 7.

	Criteria	Weights by Method				
		Individual-direct	Group-direct	Individual-PA	Group-PA	Individual-AHP
HADR	Flexibility	0.293	0.265	0.313	0.350	0.414
	Cooperation	0.285	0.265	0.300	0.300	0.361
	Sustainment	0.213	0.235	0.200	0.200	0.140
	Resource Allocation	0.209	0.235	0.188	0.150	0.086
MIO	Effect	0.304	0.250	0.362	0.250	0.409
	Risk	0.254	0.281	0.269	0.300	0.371
	Force Employment	0.241	0.250	0.212	0.250	0.157
	Sustainability	0.200	0.219	0.157	0.200	0.063
Combat Patrol	Geographic Coverage	0.336	0.345	0.330	0.400	0.485
	Security	0.281	0.276	0.296	0.260	0.286
	Offensive Action	0.243	0.241	0.244	0.240	0.175
	Readiness	0.141	0.138	0.130	0.100	0.054

Table 8. Study 2 Results: Criteria weights by method

For both the HADR and Combat Patrol scenario, each method resulted in the criteria being ranked the same way in terms of overall priority. For HADR, all methods agree that Flexibility ≥ Cooperation > Sustainment > Resource Allocation in terms of importance. For Combat Patrol, all methods agree that Geographic Coverage > Security > Offensive Action > Readiness in terms of importance. For the MIO scenario, the three methods based on individual data agree that Effect > Risk > Force Employment > Sustainability, while the two group methods agree that Risk and Effect should be reversed such that Risk > Effect > Force Employment > Sustainability in terms of importance.

To visualize the spread of weights identified for each criterion (Table 8), the method and mean weights are plotted in Figure 9.

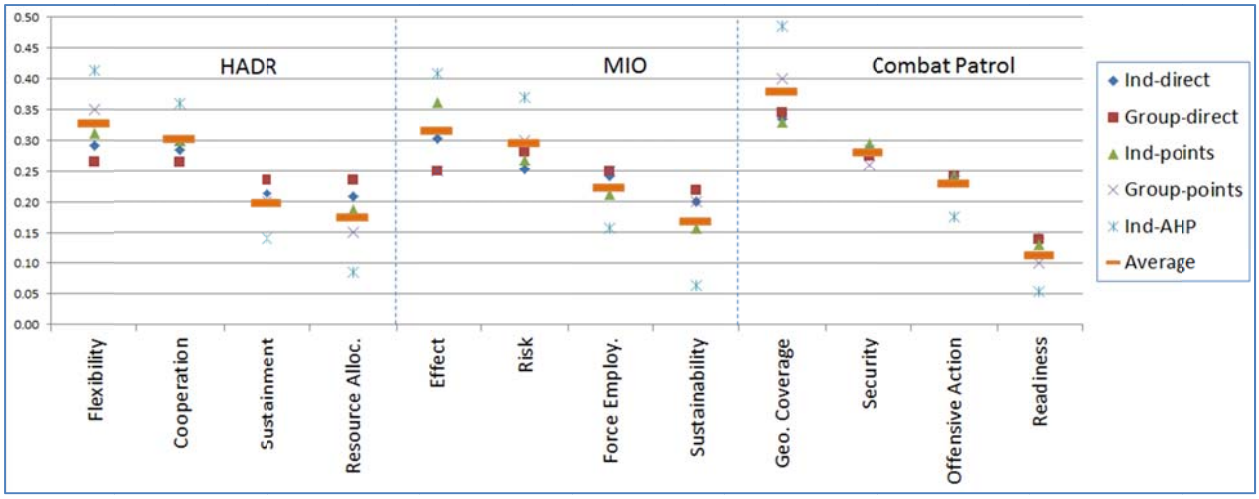

Fig. 9. Study 2 Results: Spread of criterion weights by method

Note that for each mission type, the spread of weights is greatest for the AHP method. In fact, for each mission, the two highest weights and the two lowest weights consistently result from the AHP approach. This type of result is expected from AHP, which is recognized for its ability to differentiate between criteria [10]. It is also noticeable that the Group-direct results give little discrimination between criteria (for HADR and MIO), while the use of PA for those missions resulted in weights that are more dispersed.

3.3.2 Participant Opinions

Table 9 shows the average of individual's opinions about their comfort with the group rankings and weightings, and the appropriateness of the criteria provided.

	Group rankings reflect my opinion	Group weightings reflect my opinion	Criteria appropriateness
HADR	4.25	3.75	3.25
MIO	3	3.5	3
CP	4.4	4.4	4.6

Table 9. Study 2 Results: Individual Acceptances of Criteria as a ranking between 1 (negative) and 5 (positive)

It was hypothesized that the data would show that participants were more comfortable with their group's ranking than they were with their group's weightings, simply because it is easy to agree that one thing is more important than another than it is to agree on how much more important it is. However the actual results are mixed.

With respect to criteria appropriateness, despite the efforts of Study 1 to capture, categorize and define four key criteria for each mission type, the level of participant agreement with the criteria provided in Study 2 averaged 3.6 out of 5, with MIO participants rating a low 3 out of 5 level of agreement, and Combat Patrol participants feeling much more comfortable with the given criteria, and rating them as 4.6 out of 5.

For the HADR mission, participants indicated general agreement with the given criteria, and also offered: time, personnel training, and support for consideration. For the MIO mission, a few participants indicated (albeit, after a group discussion) that Force Employment and Effect are not mutually exclusive, and should not both be used; effect

was the preferred criteria. In addition, participants offered international and constabulary support as possible criteria. For the Combat Patrol mission, participants indicated general agreement with the given criteria, and also offered sustainment, logistics, and communications.

3.3.3 Pairwise Comparison of Elicitation Method Results

Table 10 shows the number of times out of 100, for each mission, that randomly assigned COA assessment values (as described in Section 3.2.2 and illustrated in Figure 8), $v_k(a_{ikj})$, resulted in the same COA being selected by the compared methods. The 'Group-Direct' method is bolded since it is the current method used by planning teams and therefore an important reference point.

	#times out of 100 that same COA is selected		
Methods Compared	**HADR**	**MIO**	**Combat Patrol**
Individual-direct vs **Group-direct**	97	96	100
Individual-PA vs Group-PA	99	92	94
Individual-direct vs Individual-PA	98	95	99
Group-direct vs Group-PA	95	99	95
AHP vs **Group-direct**	90	85	91
AHP vs Group-PA	95	83	96
AHP vs Individual-PA	94	87	89

Table 10. Comparison of weights according to COA selected

Difference thresholds, as discussed in Section 1.2.1, were not considered in these comparisons. If they were included, the number of matches would be expected to increase.

The greatest number of result differences occurs when comparing AHP to any of the other methods (group or individual). This seems straight forward given that we also noted that AHP had the greatest variability across its weights, and its weights are (Figure 9) least consistent with the other methods. AHP comparisons for the MIO mission resulted in the least matches though this is likely attributed to the fact the participants indicated the least agreement with these criteria overall and therefore would naturally find it more difficult to consistently describe their relative merits.

4 Application of Default Criteria

The final study, Study 3, was not designed specifically for the testing of default criteria, rather it had a primary purpose of exploring alternative, more intuition-based, processes for COA development and selection. That experiment involved the application of the current COA analysis and selection processes as its baseline condition, and involved the use of the same MIO and Combat Patrol mission statements as used in our previous studies. We were able to leverage that study to further explore the concept of default criteria for mission types, by providing participants with the criteria identified in Study 1 and used in Study 2 as a starting point for their analysis.

In this study, groups of three Naval Lieutenants participants were provided with one of the MIO or CP mission descriptions and asked to develop a COA, using their existing planning process. This involves developing one or more possible COAs, and developing ECOAs, if applicable, and then applying a weighted decision matrix to arrive at a preferred COA. Each team was given the top four criteria identified in Study 1, along with their definitions. The MIO team was also given an additional default criterion of "public relations", as this had been identified by our RCN SME but previously removed from the list as a result of no matching participant data from Study 1. While Study 2 had not been fully analyzed at this point, observations of the Study 2 data collection session had provided

indications that additional criteria should be considered. Participants in study 3 were not given any default weightings as we did not want previous assessments of criteria importance to bias the participants towards using (or not using) a particular criterion. Participants were told that they could use these criteria, or select only particular ones to use, and/or could add additional criteria as they saw fit. An Excel-template was provided that automatically builds the decision matrix based on the user-selected criteria, and applies the user-supplied weights to calculate COA values, once the criterion assessments have been made.

4.1 Analysis

The weights for the criteria of Study 2 were normalized such that $\sum_J w'_j = 1$, where J is the set of default criteria given to participants of Study 3. Let S be the set of criteria selected by the Study 3 participants. Then J∩S is the set of criteria that Study 2 and Study 3 have in common. The Study 3 participants used their typical, group-direct, method to assign integer weight values (w_s) to each criterion. In order to compare the weights of the criteria in J∩S across studies, the weights of Study 3 are scaled as follows:

$$w'_s = \frac{\sum_{J \cap S} w'_j}{\sum_{J \cap S} w_s} \times w_s$$

4.2 Results

Table 11 shows the default and selected criteria for Study 3, and the associated weights. The weights for the default criteria as determined by the group-direct method in Study 2 are also shown here, but were *not* given to Study 3 participants.

The MIO team used all the default criteria, and added one more: Deterrence. There is little in the way of similarly between the weights applied to the common criteria. In fact, the order of importance is not even consistent. For Study 2, when the group-direct method was applied, Risk > Effect ≥ Force Employment > Sustainability. For Study 3, when the group-direct method was applied, Effect > Sustainability > Risk ≥ Force Employment. However, we also note that the Study 2 group-direct weightings have very small differences, thus, without a better understanding of the underlying variance of the process it is difficult to determine if the ordering is truly accurate. In Study 3 however, it is clear that Effect>Sustainability>(the rest of the criteria). The fact that the team used all six criteria in the decision matrix is interesting; typically fewer criteria are used, however, the Excel-template provided made it easier to manage a larger set of criteria, and may explain the reasoning for this.

		Incl. in default list (from Study 2)?	Group-direct weight (Study 2, w'_j)	Incl. in team list (Study 3)?	Group-direct weight (Study 3, w_s)	Study 3 weights – scaled (w'_s)
MIO	Effect	Y	.250	Y	3	.429
	Risk	Y	.281	Y	1	.143
	Force Employment	Y	.250	Y	1	.143
	Sustainability	Y	.219	Y	2	.286
	Public Relations	Y	-	Y	1	.143
	Deterrence	N	-	Y	1	.143
CP	Geographical Coverage	Y	.345	Y	3	.466

	Security	Y	.276	Y	1	.155
	Offensive Action	Y	.241	N	-	
	Readiness	Y	.138	N	-	
	Cooperation (MPA Support)	N	-	Y	2	.310

Table 11. Comparison of criteria and weights for Study 2 and Study 3

The CP team did not use the default Offensive Action or Readiness criteria, but did use Geographical Coverage and Security. Both groups indicated that Geographical coverage is more important that Security, though the first group suggests it is 1.25x as important, while the second group suggests it is 3x as important. The CP team also added the criterion: Cooperation (Maritime Patrol Aircraft Support). Note that Cooperation was a criterion identified in Study 1 for CP, but was not in the top four criteria, and therefore not carried forward into further studies.

The MIO team developed two COAs and did not use ECOAs. The CP team developed two COAs and three ECOAs.

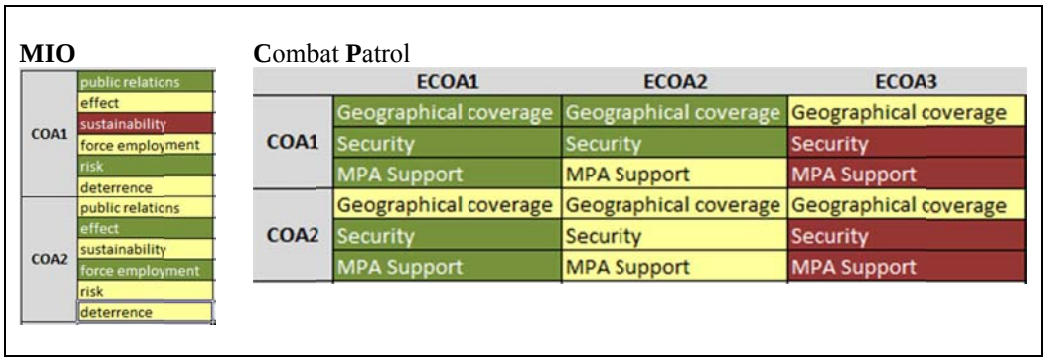

Fig. 10. Study 3 – Completed Weight Comparison Matrices

Using the assessments of Figure 10 for the MIO scenario, and applying the weights of Study 2 for each criterion in J∩S, COA_1 receives a value of 2.8 and COA_2 receives a value of 3.1. Using the Study 3 weights, COA_1 receives a value of 2.6 and COA_2 receives a value of 3.1. In both cases, COA_2 is preferred.

Using the assessments of Figure 10 for the CP scenario, and applying the weights of Study 2 for each criterion in J∩S, COA_1 receives a value of 6.6 and COA_2 receives a value of 5.6. Using the Study 3 weights, COA_1 receives a value of 6.7 and COA_2 continues to receive a value of 5.6. In both cases, COA_1 is preferred.

In a post-experiment questionnaire, Study 3 participants were asked to rate, on a scale of 1 (Strongly Disagree) to 5 (Strongly Agree), whether or not having predetermined criteria was useful. On average, they agreed that it was useful (3.7/5).

4.3 Discussion

4.3.1 Default Criteria

Study 1 aimed to identify a small set of (4 to 6) criteria for each of four mission types: HADR, MIO, CP and NEO. Responses were varied, but organized and categorized to match the principles of war, and subsequently described according to the COA definition framework. Despite the analytical process applied to the Study 1 data, and the SME reviews of proposed criteria, the participants of Study 2 (largely, the same participants as in Study 1), indicated an average agreement with the given criteria of only 3.6 out of 5. This may in part be explained by the participant's bias towards their own responses for Study 1, rather than the group consensus. When these same criteria were

offered as defaults for Study 3, one team used the default set and included an additional criterion, while the other team discounted two of the default criteria and included an additional criterion. Despite not adhering to the exact criteria given, the participants did indicate that they found default criteria useful (3.7/5). The discrepancies in the preferred criteria could exist as a result of: experience levels, which was considerably higher for participants in Study 1; the study methods (Study 3 involved actual COA development which could have helped clarify which criteria were most useful in distinguishing between them); and/or the knowledge of the team members.

In general, the support for providing default criteria through PowerPoint templates or a more sophisticated system like COA-T is somewhat positive, but requires more study. While the planner would always have the option of changing the defaults, there could be a tendency to accept the criteria (e.g., MIO scenario of Study 2) since there is an expectation that the system has based its defaults on something concrete.

From Study 1, we obtained a fairly lengthy list of criteria overall, with associated definitions. A tool such as COA-T should consider providing the planner with the full list of criteria (and definitions) to choose from, with the option of ordering them according to how frequently criteria have been previously used for that mission type. The user would then decide to select the top criteria, select alternatives, or add new criteria as appropriate. When new criteria and definitions are added, these would become available for selection the next time the tool is used.

The sample sizes in our studies were not sufficient to conclusively identify the most appropriate criteria. However, the criteria and preferences identified in this report can be used to seed the system, and the top criteria in the list can evolve over time with system use.

4.3.2 Default Criteria Weights

For the criteria used in both Study 2 and Study 3, the weights assigned by the groups were notably different, and criterion priority was not consistent across the groups. A larger set of data would be needed to confidently recommend specific sets of criteria or weightings. There is a concern, however, that providing defaults weights could result in a cognitive bias. There could be a tendency to accept the default numbers, particularly if they do not differ ‘enough’ from what the planners themselves might have intuitively specified. This concern should be explored through further experimentation.

4.3.3 Weighting Method

The authors of [9] point out that since there is no ‘golden standard’ (i.e., no correct answer) for criteria weighting, the validity of weighting techniques cannot easily be assessed. We saw in Study 2, Table 8 and Figure 9, that the different weighting techniques resulted in similar weightings within the Study 2 groups. When these weights were applied to 100 random decision matrices for each scenario, the end result was the same across select pairs of methods, an average of 94 times out of 100. This suggests that the end result (i.e., the chosen COA) is not overly sensitive to the weighting method chosen, though it is sensitive to the criteria ordering (ranking) due to the integer valuation (i.e., -1/0/1) of the value function. One could argue that there is not enough of a difference to justify changing the current process and adding additional effort in terms of the data entry required to calculate the weighting set. The authors of [9] show in their own trials that when weights were intended to support group decision making, the elicitation method had negligible impact on the result.

So long as these methods are applied manually, the direct method will most likely prevail. However, a simple Excel spreadsheet could make the application of more diverse weighting methods, such as points allocation, much easier. Alternatively, the AHP method is well-accepted in the literature as a premium method for identifying weights, and in particular, has the greatest ability to differentiate amongst the criteria [9]. We see some evidence of this in Figure 9 where the AHP tends to differ the most from the mean. Its strength is also its weakness, however, wherein pairwise comparisons do not require planners to consider the relative merits of all criteria at once, but the many

simple comparisons can become burdensome as the number of criteria increases. For the typical 3 to 5 criteria used by the RCN for COA comparisons, the AHP remains reasonable to apply, barring intense time pressures. The AHP template that was applied in this research also includes a field for scaling the contribution of input from each individual to the total. For the data reported here, all participants were treated equally; however, it is worth considering whether the inputs of more senior and experienced personnel should be given more weight than those with less experience.

4.3.4 The Use of Decision Matrices

This paper has explored the process of selecting and weighting criteria for COA analysis and selection, as supported by a weighted decision matrix. This process is described in the JOPP and OPP and is the process taught to and exercised by the RCN. This is an analytical approach to support decision making which is riddled with approximations, uncertainties and subjective assessments. The JOPP notes that the decision matrix is meant as an aid only, and that its value is in encouraging a systematic process for viewing the strengths and weaknesses of each COA, not for picking the solution. In fact, if the preferred COA, according to the decision matrix, is not consistent with the planner's expectations, they are to critically look at whether or not an important criterion was missed, or some other part of the analysis was incorrect. Thus, getting the weights exactly right is unlikely as important as the planner's experience and ability to recognize when the outcome is amiss.

4.3.5 Implications for COA-T

COA-T is intended to help the planners through the entire planning process. In this paper, we have focused only on the steps of COA Analysis and touched on COA comparison. With respect to criteria selection, we have suggested that an augmentable list of criteria with well-structured definitions be provided and ordered according to historical frequency of use. As we saw in Study 3 for the MIO mission, providing technology to support the analytical process may result in the planner considering a larger number of criteria for a mission, when appropriate, since it is easier to do so. This would result in a more complete analysis that has not left out relevant criteria in the interest of mathematical simplicity. With respect to criteria weightings, we do not have sufficient data to recommend default values, but default ranking of criteria will be implied by the provided criteria list when ordered by frequency of use.

Support for identifying weightings should be provided by the system. The implementation of the three methods described here (direct – i.e., the current method, PA, and AHP) within the system would not be complex, and the best method may be driven by any number of factors, such as time available, command team experience, number of criteria, etc. In addition to these methods, the option to specify and review the weights visually should be considered. As well, the tool should support exploration of the sensitivity of the results to a given criterion weight, allowing the planner to fully consider the impact of the chosen weights and possibly reconsider them. This could be achieved by allowing interactive user adjustments to one weight while holding the proportion of the other weights constant, or, by providing analysis on the delta required in a given weight to produce a different outcome. Finally, as discussed in Section 1.2.1, any indication of a COA preference given by the system should take uncertainty thresholds into account. When the value of multiple COAs cannot be distinguished beyond the boundaries of the uncertainty thresholds, they should be displayed as equal.

4.3.6 Study Design

The studies reported here were not intended as structured experiments, rather as focus groups akin to those that would be involved in the shipboard planning process in the real world. Given the knowledge gained through these studies, it would have been worthwhile having participants develop three COAs for their assigned mission(s) in Study 1, prior to coming up with criteria and definitions. Criteria are meant to differentiate between COAs, not stand on their own, and this facet of criteria selection may not have been well-considered when participants were asked,

due to time constraints, to assume that the COA development step had been successfully completed in this initial study.

For Study 2, participant assignment to focus groups was based firstly on the initial letter of their last names, and secondly a balancing of ranks across groups. This resulted in pseudo-random groups with rank and experience comparable to that of the similar-sized groups that would be involved in shipboard planning at sea. Given the use of paper-based surveys, we did not allow the option of changing or adding criteria for their assigned mission. Participants were asked to use the summarized results of Study 1 directly. However, we saw in our exit survey of Study 2 that participants were not in complete agreement with the criteria provided, meaning that we may have gathered more diagnostic data had we given the groups the full list of criteria from Study 1, and let them complete the consolidation and selection process themselves. As is often the case, we had to balance research effort and available participant time with the quality of results needed for the purpose at hand. While these studies do not lend themselves to statistical analysis, they provide ample insight into the criteria selection and weighting process to move forward with identifying appropriate technological supports to offer through COA-T.

4.4 Future Research

Given time and resources, a structured experiment designed specifically to assess the validity of the weights produced by each of the weighting methods could be constructed. With Study 2, we could only compare the weights across methods, as there is no correct (gold standard) answer available. To determine which method produces the 'best' weights in this naval context, we would need to apply the weights across many scenarios of each mission type and have SMEs assess their agreement with the COA selected as a result of each weighting method. As we have seen, however, the COA selected is not always sensitive to the differences in the weights, so constructing scenarios and COAs that would highlight the differences would require careful consideration.

A simpler experiment could involve a more intuitive assessment of the weighting results (Table 8) by SMEs. For instance, a visual representation of the weights (such as in Figure 11, for the MIO mission) could be provided along with the mission description, and SMEs could be asked to select the weight set (A, B, C, D, or E) which best matches their own perception.

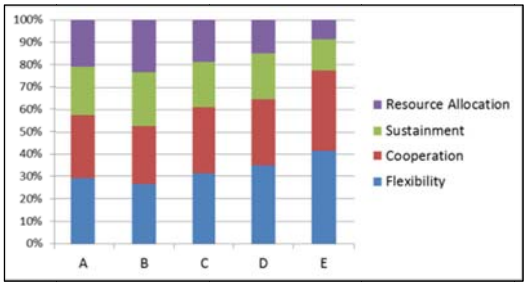

Fig. 11. Future Study

An additional study to examine cognitive biases associated with providing default weights for mission types should be considered. In particular, it should look at the tendency to accept default weights (rather than manipulate them based on intuition or experience), particularly when the default weights do not fall too far from the participant's own judgements.

The criteria described in this paper are meant for distinguishing amongst multiple viable COAs in order to *select* a COA. Other criteria or factors will need to be considered earlier in the planning process to support the *development* of the viable COAs. For example, weather may not be considered important when differentiating COAs, since it is conceivable that all COAs involve encountering the same weather conditions. However, weather could be very

important when building the set of viable COAs. The overall criteria list may be usable for both purposes, but the criteria relevant to building the COAs may differ from those selected to differentiate the COAs. It is expected that map-based visualizations of these criteria to support viable COA-building will be of value, but further exploration of the concept is needed.

In the studies reported here, COA selection was not particularly reliant on the criteria weights. However, there may have been greater discrepancies in the results if a different COA value function, $V(COA_i)$, had been chosen. The weighted sum value function, as used in our studies and by the RCN, can mask negative assessments with positive ones, and some important information can be lost. Alternative value functions should be considered and assessed for their appropriateness and impact on the results. As well, the process for determining appropriate difference thresholds for a given decision matrix should be examined in order to incorporate reasonable thresholds into COA-T. Further, the value functions applied at the criterion level, $v_j(a_{ijk})$, could be studied further. At this time, the possible values for $v_j(a_{ijk})$ are determined 'on-the-fly' (e.g., red/yellow/green = -1/0/1) and their impact on the COA value across all criteria has not been adequately considered.

4.5 References

[1] Department of the Navy, Navy Planning NWP 5-01, Navy Warfare Publication, Norfolk, 2013.

[2] Department of National Defence, The Canadian Forces Operations Planning Process (OPP), Canadian Forces Joint Publication 5.0 (CFJP 5.0), 2008.

[3] Kramer, C. and Baker, K., Maritime Information Warfare: Ship Command Team Task Characterization for Planning, DRDC-RDDC-2017-C116, May 2017.

[4] Department of Defense, Joint Publication (JP) 5-0: Joint Operational Planning (JOP), 2011.

[5] Lawrynczyk, A. and Kramer, C., Criteria Management and Course of Action Sketching, Contract W7707-145734, DRDC Contractor Report (to be published), 2017.

[6] Abi-Zeid, I, Belanger, M, Guitouni, A., Martel, J.-M., The Reformulation of the Family of Criteria in Commander's Expert System Advisor (CESA), DREV-TM-9739, June 1998.

[7] Email from CANFLTLANT, Subject: Naval OPP Working Group, Date: June 17, 2016.

[8] Department of National Defence, Canadian Forces Joint Publication (CFJP-01) Canadian Military Doctrine, Colonel Steven P. Noonan, MSC, CD, 2009.

[9] van TilJ, Groothuis-Oudshoorn CGM, Lieferink M, et al. Does Technique Matter: A Pilot Study Exploring Weighting Techniques for a Multi-Criteria Decision Support Framework. Cost Effective Resource Allocation, 2014.

[10] Konstantinos, P., The Analytic Hierarchy Process, Lesson 1, International Hellenic University, date unknown.

[11] Saaty, T.L. (1980) The Analytic Hierarchy Process, New York: McGraw Hill. International, Paperback (1996, 2000), Pittsburgh: RWS Publications.

[12] Goepel, Klaus D., BPMSG AHP Excel Template with Multiple Inputs, version 2015-06-07 – http://bpmsg.com, Singapore 2013.