SYM-AM-19-065

PROCEEDINGS
OF THE
SIXTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

THURSDAY SESSIONS
VOLUME II

Acquisition Research:
Creating Synergy for Informed Change

May 8–9, 2019

Published: April 30, 2019

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

A Robust Framework for Analyzing Acquisition Alternatives (AoA)

Moshe Kress—Operations Research Department, Naval Postgraduate School. [mkress@nps.edu]

Brian Morgan—Operations Research Department, Naval Postgraduate School. [blmorgan@nps.edu]

Abstract

Analysis of Alternatives (AoA) is a complex multi-dimensional decision procedure used by the U.S. Department of Defense as part of the acquisition process. The four dimensions of the procedure are *alternatives*, *criteria*, *scenarios*, and *stakeholders*. Current AoA studies lack the structural rigor needed from such a complex procedure, which involves measurements, evaluations, analyses, and modeling, as well as social and group-decision aspects—all in a highly uncertain environment. We propose a structured paradigm for conducting AoA, rooted in well-established methods of multi-criteria decision analysis. The core of the methodology comprises the concepts of ratio-scale matrices and the Euclidean norm. The ratio-scale matrices are used to elicit evaluations, preferences, and opinions from individual stakeholders and analysts, and the Euclidean norm is utilized to mitigate possible preference inconsistencies and help form consensus.

Introduction

The U.S. Department of Defense (DoD) Acquisition System comprises three interconnected stages that start with specifying requirements—a procedure called Joint Capabilities integration and Development System (JCIDS). The second stage, called the Acquisition Process, focuses on determining appropriate materiel solutions for the requirements specified in the first stage. The third stage has to do with executing the decision made at the second stage. It includes funding and control activities contained in the Planning, Programming, and Budgeting Execution (PPBE) Process (DoD, 2017).

The first stage comprises tactical and operational analyses based on wargames, simulations combat models, and input from subject matter experts. It identifies gaps in current capabilities and projects future needs based on evolving threats and operational postures. The third stage is the Department of Defense's resource allocation process that includes an annual budget, for presentation to Congress, linking missions to the requested funding.

The second stage—the Acquisition Process—comprises two interrelated phases: a *creative* phase and an *analytic* phase. The outcome of the creative phase is a set of potential materiel solutions to the operational requirements specified in the Initial Capabilities Document (ICD) produced in the first stage. This set comprises acquisition alternatives to be analyzed in the analytic phase. Obviously, the set includes only those alternatives that are evidently reasonable and viable. In other words, no alternative in the set can be without a capability, or violate a clear requirement, specified in the ICD. The analytic phase focuses on evaluating the alternatives with respect to several criteria, while incorporating quantitative analysis with multiple stakeholders' opinions and preferences. The outcome of this phase is a recommendation on the most preferred alternative(s) to be considered for acquisition. This recommendation must be based on multi-criteria evaluations of the alternatives and reflect a consensus among stakeholders' opinions, goals, and preferences.

This study focuses on the analytic phase of the second stage—also called Analysis of Alternatives (AoA; see, for example, RAND Corporation, 2006). Our objective is to propose a comprehensive formal framework for executing AoA and introduce a unified analytic structure into it. The proposed framework is general enough to be easily tuned to any specific AoA study in any branch of the armed services.

The core of the analytic phase (AoA) is a Multi-Criteria Decision Analysis (MCDA) process. In this process, alternatives are evaluated according to a set of criteria, and the resulting evaluations are then aggregated into a rating or a score that represents the relative standing of each alternative. In a DoD acquisition context, the criteria typically include scenario-dependent operational effectiveness, technological feasibility and risk, supportability, compatibility (with existing systems), and cost. While the general spirit of MCDA is indeed present in typical current AoA projects, its actual manifestation varies significantly among studies (see, e.g., DoN, 2006; RAND Corporation, 2006, 2016; Souders et al., 2004; Smith et al., 2011; TRADOC, 2011). Crucially, most of these studies lack the structural or formal rigor that is desired in a critical decision process such as AoA. Typical weaknesses relate to in-context evaluations of alternatives with respect to criteria, determining the weights of criteria, treating uncertainty and risk, and adequately aggregating preferences among stakeholders. These issues are either not addressed in those studies at all or they are treated in inconsistent ways across studies. Moreover, as much as it is an analytic process, AoA is also a social process that involves several (sometimes many) stakeholders. Different stakeholders, representing various DoD branches and organizations, may have different opinions, points of view or preferences regarding the importance (i.e., weights) of criteria. They may also differ in their assessments about the likelihood of future scenarios and disagree about the values of alternatives with respect to qualitative (subjective) criteria, where measures of performance (MOPs) and/or measures of effectiveness (MOEs) are either difficult to compute or do not exist altogether. Even measurable (objective) criteria, such as detection range, velocity, and firing accuracy (say, probability of hit), may be scaled differently by different stakeholders. This important social aspect seems to be ignored in current AoA studies. In our proposed framework, we attempt to remedy these, and other, shortcomings.

The main contributions of this paper are (a) proposing a clear “standard” for conducting AoA in the U.S. Department of Defense, (b) explicitly addressing the role of scenarios and stakeholders in the AoA process, and (c) developing an all-inclusive distance-based model that addresses, simultaneously, all four dimensions of AoA—alternatives, criteria, scenarios, and stakeholders.

MCDA models considering alternatives and criteria are quite abundant (e.g., AHP, ELECTRE, and PROMETHEE; Behzadian et al., 2010; Figueira et al., 2005; Saaty, 1980). There are also MCDA models that consider scenarios (Montibeller, 2006; Stewart, 1997), and those which consider consensus formation among multiple stakeholders (Cook et al., 1996). But, to the best of our knowledge, the model presented here is the first attempt to tackle all four dimensions in a unified and robust fashion in the context of DoD AoA.

The paper is organized as follows. In the next section, we review the four main dimensions of an acquisition AoA: *alternatives*, *criteria*, *scenarios*, and *stakeholders.* Then we describe the set of criteria relevant to a typical DoD acquisition AoA, and their imbedded hierarchy. The next section is the heart of this paper; it formally describes the MCDA methodology underlying the AoA process. Then we address the uncertainty associated with future scenarios and the way it affects the choice of the winning alternative. Finally, we outline the six steps of a robust AoA study.

In the rest of the paper we refer to the subject of the AoA study as *item.* An item can be a Navy fighter, a radar system, a transport vehicle, a supply ship, a command and control system, etc. The objective of the AoA process is to select for acquisition the most appropriate item out of a set of alternative items.

The Four Dimensions of Acquisition AoA

Once the operational needs and/or capability gaps have been identified at the JCIDS stage, an initial set of items—potential materiel or non-materiel solutions to these needs—is generated at the onset of the next stage: the Acquisition Process. Generally speaking, the ultimate goal of the acquisition process is selecting, out of an initial set of possible items, an item that provides the best balance between (in-context) utility or value and potential cost and risk. The members of this set of items are called *alternatives*. The alternatives represent the first main component of the acquisition AoA. The other three main components are *criteria*—the touchstones according to which alternatives are evaluated, *scenarios* that provide the operational backdrop for the evaluation, and *stakeholders* who contribute analytic inputs, as well as preferences, opinions and judgements, into the acquisition decision process.

Alternatives

Generating the initial set of alternatives is the “creative” part in acquisition AoA. The generators of the alternatives are typically defense agencies, who may suggest existing materiel options or off-the-shelf items, and defense contractors who propose either existing products currently produced or items that are at various stages of maturity in the research-and-development stage. The items suggested may range from the mundane (e.g., the current “status quo” alternative) to the daring (e.g., an item based on revolutionary, and perhaps even immature, technological concept).

In some AoA studies, there exists a legacy item (ship, weapon, C2 system, etc.) that either is near its end of life or its capabilities are insufficient for emerging requirements. In such cases, it is important to clearly identify the characteristics of the legacy alternative, which can be considered as a baseline according to which potential upgrades are considered (MITRE, n.d.).

The set of alternatives should be carefully constructed. It must be non-trivial (e.g., just two alternatives where one clearly dominates the other), but also manageable in size. There is hardly an effective and meaningful way of handling the evaluation of dozens of alternatives. One way of reducing the size of the alternatives’ set is eliminating similar alternatives—alternatives that differ marginally or those that are evidently dominated by other alternatives.

The alternatives should also be realistic in the sense that they are technologically feasible and grounded in industry’s capabilities. The set of alternatives should not include idealized items that have no practical basis in industry or government. The set of acquisition alternatives may be divided into categories:

- Modified existing items currently in operation,
- As-is or modified off-the-shelf items available in the market but not yet in operational use,
- Repurposing and/or recombining existing items with new technologies, and
- Newly developed items (USAF, 2016).

The four categories differ in their potential effectiveness, cost, and the risk associated with their acquisition. To modify a legacy item would probably be cheaper and less risky than developing a completely new one, but a new item would most likely be more effective and more attuned to current requirements than the modified legacy item. Roughly and generally speaking, the main thrust of the AoA process is to tradeoff among these three contrasting aspects—effectiveness, cost, and risk.

Criteria

The merits and weaknesses of the alternatives are evaluated by criteria, which represent various aspects related to the operation, functionality and reliability of the alternatives, the risk associated with their selection, and the cost factors related to acquiring, handling, and maintaining them. In general, the set of criteria for evaluating defense (physical) acquisition items such as weapons, sensors, and platforms is divided into four subsets:

- Effectiveness
- Operationability, reliability, maintainability, and logistics (ORML)
- Cost
- Risk

While effectiveness is measured by specially constructed measures of effectiveness (MOE), and cost is typically measured in money spent (and/or to be spent), ORML criteria are measured by both MOEs and cost factors.

Effectiveness

An old adage states that “among all the alternative items that are completely useless for a certain requirement, the best one to be selected is the cheapest.” In other words, the main driver for selecting an alternative is its usefulness or effectiveness with respect to the requirements that generated the acquisition process. The term *effectiveness* may be elusive and may mean many different things. Measuring effectiveness of an alternative is probably the most challenging part of an AoA study. To demonstrate the complexity of this challenge, consider the following simple (in fact, simplistic) example:

> The requirement is for an anti-air (AA) weapon, and the only two criteria are fire-rate and single-shot kill probability (SSKP). There are two alternative weapons for consideration. Weapon A has a higher fire rate than Weapon B, but smaller SSKP than Weapon B. Which weapon is more effective? Weapon A can deliver higher “quantity” of shots while Weapon B has a better “quality” per shot.

One way of measuring the (relative) effectiveness of the two AA weapons is to determine a tactical or operational objective (e.g., maximize number of targets killed), determine an appropriate MOE (e.g., expected number of killed targets within a certain time period) and construct a model (analytic or simulation) that calculates the values of the MOE for the two weapons. Another way to determine the relative effectiveness of the two weapons is to treat each attribute—fire-rate and SSKP—as separate criteria, give a score to each weapon with respect to each criterion, and then combine the scores of the two criteria, via, say, a weighted combination, into a single score—one for each weapon.

The first approach could be considered “objective” in the sense that there is a quantitative model that bridges between the data and assumptions, and the final evaluation of the two weapons. The second approach is “subjective” in the sense that stakeholders and/or subject-matter experts must provide their personal input in determining the scores of the alternatives with respect to the criteria and the weights of the criteria.

In reality, and unlike the above example, *effectiveness* has more than two aspects, and thus measuring effectiveness becomes more challenging. Ideally, there would be one measurable objective (e.g., maximize probability of winning the engagement/battle/campaign) that encompasses all relevant operational aspects of the item and the scenario in which it is to be implemented (see below). The measurable objective would be formalized as an MOE, which could be reliably computed in a comprehensive model. Unfortunately, this ideal setting seldom occurs. Either there are multiple objectives or the scenario and the role of the item in it are so complex that no model can reliably capture all the salient aspects.

The bottom line is that, in reality, effectiveness in an AoA is evaluated by a mixture of the two approaches—the analytic "objective" approach and the opinion- or experience-based, "subjective" approach. The goal is to enhance, as much as possible, the analytic side and thus minimize the possible biases and disagreements (see below) that may be generated from the subjective approach.

Operationability, Reliability, Maintainability, and Logistics (ORML)

During its course of operations, an item must be operated (or controlled) by qualified persons, professionally maintained, and regularly serviced and resupplied. These requirements result in operational, as well as economic, implications. Obviously, ceteris paribus, an item that is more reliable and requires less maintenance, less qualified operators and lighter logistic burden is preferred to an item that is rated worse on any of these aspects. The question is that of trade-offs; how much effectiveness one would be willing to sacrifice for a simpler, more robust, and lower-maintenance system?

Operationability is a criterion that reflects two salient aspects of a newly acquired item: (a) compatibility with existing systems, currently in use, with which the new item has to interact, and (b) human-system integration (HSI). A new radar must interact with existing sensors, command and control systems and weapons, and therefore must be *compatible* with them. However, measuring compatibility is challenging; there is no natural MOE that could be defined and objectively evaluated for measuring how well a certain alternative item interacts with current systems. This is a "subjective" criterion that must be evaluated qualitatively by subject-matter experts (SMEs). Similar restrictions also apply to the other part of Operationability—HSI. While, in principle, one could use the number of operators, classified by technical background, length of service and pay-grade, and estimated length of the training period, as surrogate MOEs for HSI requirements, in practice it would be difficult to do it. Here, once again, evaluating this criterion will most likely be done by qualitative input obtained from SMEs.

Reliability affects the readiness of the item. The more reliable an item is, the less frequently it is unexpectedly down. This criterion is quantitative and is typically measured by the mean time between failures (MTBF). While measuring MTBF of an existing system is a relatively straightforward statistical task, estimating the value of this criterion for items in a design or development phases is challenging because of lack of statistical data. Thus, reliability estimates must depend on engineering-based projections based on the item's design and the technical specifications of its components, and perhaps some statistical data available about similar systems. In many cases, these estimates are provided by the vendors of the items, in which case the projected reliability values must be taken cautiously.

Maintainability is an attribute that describes the technical and physical burden associated with an item. Arguably, a modular item that requires a "plug-and-play"–type service is more maintainable than an item that comprises hard-wired components, which, when failed, result in the need for system-wide service. Similar to reliability, maintainability of an item could be measured by MOEs such as average mean service time over all components of the system, or the maximum mean service time, or other statistical measures of repair and maintenance services. The same challenges that apply to the reliability criterion, when an item is still in the design or development phases, apply here too. Maintainability could be considered a fully quantitative criterion only for existing items, which have accumulated enough maintenance experience and data. Otherwise, maintainability is evaluated by SMEs.

Logistics refers to the operations and cost aspects related to the transshipment of items, and the supply chain of consumables (e.g., ammunition, fuel) and repair parts needed for their operation and maintenance. There are typically two logistic aspects associated with an item: (a) the physical infrastructure needed for storage and maintenance of the item, and (b) transportation and handling equipment for transporting the item and its required supplies. For example, transporting fuel requires specially designated tankers. Certain items may also impose logistic constraints. For example, Vertical Launching Systems (VLS) missiles used by the U.S. Navy cannot currently be resupplied at sea. In order to replenish this type of ammunition, warships must return to port.

Cost

While end-users of a military item—commanders, combat developers, operations officers, etc.—are mostly concerned with effectiveness and operationability of the item, DoD program managers and budget officials may be mostly concerned with its overall cost (see the discussion of *stakeholders* below). Cost comprises several expenditures that vary in their nature (e.g., R&D, production, life-cycle), the time horizon during which they are to be realized, and the certainty regarding their monetary size. Arguably, costs related to future expenditures (e.g., maintenance) are more uncertain then the R&D cost for, say, an item in an advanced development stage, or purchasing price for an off-the-shelf item. The cost criterion can be broken down to sub-criteria representing its various components in order to reflect preferences of immediate versus future expenditures.

Risk

The most complex and elusive criterion in the AoA process is risk, and arguably, it mostly applies to items that are not readily available. The risks are: delays in development, disrupted production schedules, running costs over budget, and difficulties in assimilating the item in the force. Alternatives, which are already existing items or very close to maturity, have relatively little or no risk regarding their availability at the time when they are needed. But other risks, associated with re-production, costs, and assimilation may still exist. For less mature alternatives, the more technologically challenged the item, the higher the probability that something will go wrong during the research and development stages, as well as in the production phase. The problem is that it is extremely difficult to estimate this probability. Therefore, this criterion is essentially “subjective,” where risk assessments are mostly based on inputs from subject matter experts or qualitative projections based on data from past similar experiences.

Scenarios

Scenarios may be considered as “Uber criteria.” They form the settings in which the alternatives are evaluated with respect to the “regular” criteria described earlier. There are two types of scenarios to be considered in an AoA study. The first type refers to the operational setting in which the item is designated to operate. Military conflict scenarios—and in particular, combat scenarios—are used for in-context evaluation of the effectiveness, operationability, and logistics of items such as weapons, C2 systems, sensors, and other defense- and military-related items. For example, the importance of the *range* criterion of a sensor—an Effectiveness criterion—may depend on the typical detection ranges applicable in a certain operational scenario. The importance of the robustness of a vehicle to road conditions may depend on the typical terrain in a scenario. Thus, the designated operational setting of an item is important for evaluating the item's potential effectiveness. An alternative that performs well over a wide range of plausible scenarios may be preferred to an alternative that performs very well on limited operational settings but poorly on other likely settings.

The second type of scenarios applies to AoA of items that do not yet exist and are in various stages of the research and/or development phases. These scenarios describe economic, social, political, and technological factors that may affect the risk associated with selecting a certain alternative. For example, if a certain alternative requires a considerable R&D effort, the Risk, and perhaps the Cost, criteria associated with that alternative will be impacted by the availability of economic resources and technological capabilities.

Both types of scenarios incorporate a fair amount of uncertainty that must be factored in the AoA study. The way scenarios are incorporated in an AoA study is discussed in more detail in the modeling part described below.

Stakeholders

As much as a technical and analytical process, AoA is also a social phenomenon involving a plethora of stakeholders who may represent different interests, viewpoints, agendas, and goals. For example, combatants—the future users of the item—may focus on the effectiveness of the item and its compatibility with existing combat systems currently in use. Combat developers may look at a much wider picture and will be concerned with issues of force structure and other strategic considerations. Technical experts will focus on the scientific and engineering aspects, and in particular on potential technological challenges that may affect the Risk criterion. Finally, budget officials will naturally focus on the programmatic aspects associated with the developing, production, operation, and maintenance of the item. In other words, the Cost criterion plays a major part for these stakeholders.

Because the AoA process is complex and multidimensional, and some criteria (dimensions) may be conflicting, it is important to select a balanced mix of stakeholders for the study—representing all the aforementioned groups of decision makers and experts who represent different aspects of the decision problem.

The Set of Criteria

The criteria are the touchstones that determine the in-context value of an alternative. Obviously, the goal is to select the alternative with the highest overall value when all relevant criteria are considered. The set of criteria should adhere to some structural, as well as content, rules and properties, which are described in the following sections.

Criteria Tree

It is convenient to view the set of criteria as a hierarchical structure. This view is not new; it is manifested, for example, in the Analytic Hierarchy Process (AHP; Saaty, 1980), which is used by the DoD. The idea is to break down the main four criteria—(1) effectiveness, (2) operationability, reliability, maintainability, and logistics (ORML), (3) cost and (4) risk—into sub-criteria, sub-sub-criteria, and so on. This breakdown induces a tree structure whose leaves (lowest hierarchy) are criteria that can either be measured by Measures of Performance (MOPs) or Measures of Effectiveness (MOEs), or can meaningfully be evaluated qualitatively by subject matter experts (SMEs). The aforementioned four criteria constitute the first layer of the criteria tree, as shown in Figure 1.

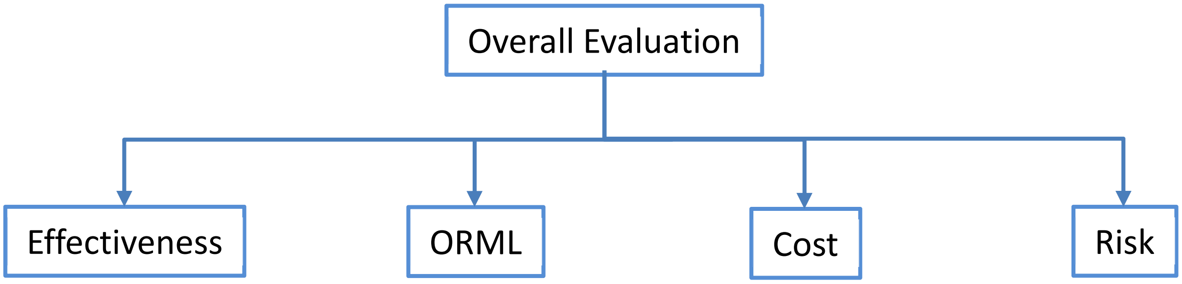

Figure 1. First Layer of a Criteria Tree

If the item to be selected is, for example, some kind of a ground fighting vehicle (e.g., a tank), then possible second and third layers of sub-criteria, which evolve from the Effectiveness criterion, are shown in Figure 2.

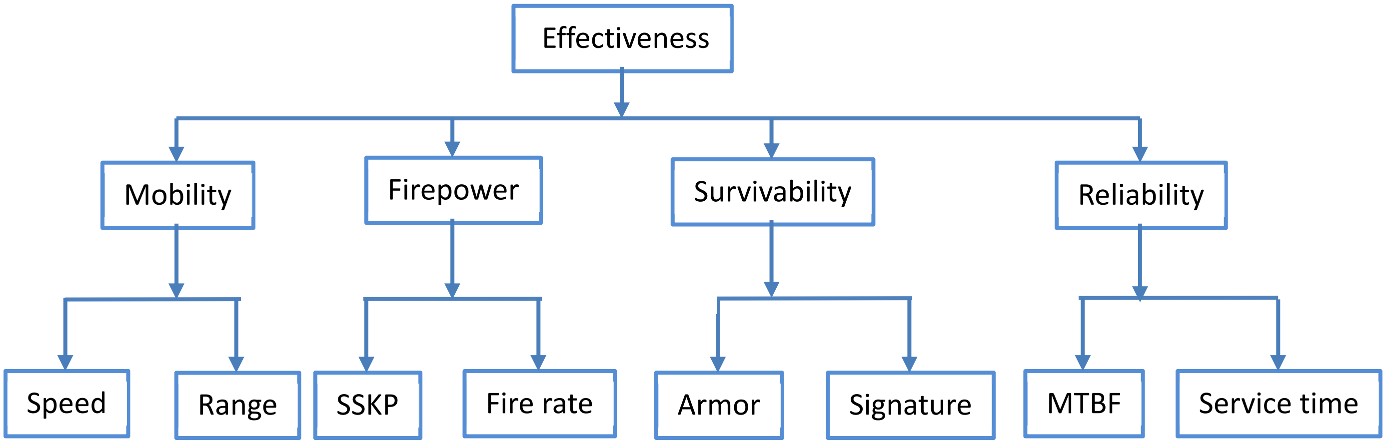

Figure 2. Second and Third Layers of a Criteria Tree

The third layer can be further broken down. For example, *speed* could be divided according to the type of terrain, *single-shot-kill-probability* (*SSKP*) may be separated into *Day-SSKP* and *Night-SSKP*, and so on.

Criteria Properties

A proper design of the criteria tree is crucial for the success of an AoA project. Specifically, the number of layers in the criteria tree and the granularity of each layer depend on the context and thrust of the analysis and on the complexity of the parent criterion in a higher layer. On the one hand, it is important to include all relevant sub-criteria that affect the parent criterion. On the other hand, we need to avoid over-cluttering the criteria tree such that it remains as manageable as possible. Keeney and Raiffa, in their seminal 1976 work, suggested some rules or properties that should guide the way criteria are selected for the analysis. In particular, the set of criteria must be *complete* in the sense that the “leaves” of the criteria tree—the end criteria at the lowest layer—cover all the aspects affecting the choice of the item. The criteria must also be *operational*—they must be relevant to the decision problem and meaningful to the decision-makers. Another important rule is to avoid *redundancy* that can lead to the undesirable effect of double counting. For example, the criterion “range of an aerial platform” may be redundant in the presence of the criterion “endurance of an aerial platform.” Finally, as mentioned above, the set of (end) criteria must be as small as possible, notwithstanding the other properties.

It is noted that in some cases, breaking down a criterion to more refined sub-criteria (in a lower layer of the criteria tree) may be counterproductive when the sub-criteria are interdependent. Two criteria are dependent if the importance or *weight* (see Weights of Criteria section below) of one criterion is affected by the evaluation of the alternative with respect to the other criterion. For example, the speed and maneuverability of a fighter aircraft might be dependent; if the aircraft is slow, the maneuverability may be more important than if the aircraft is fast. In that case the two sub-criteria may be combined into a single criterion such as *flight performance*.

Weights of Criteria

Different criteria may have different levels of importance, or different *weights*. An important fact to remember is that these weights are *subjective*. There is no scientific method that could measure the “true” weight of a criterion. Different stakeholders may have different opinions regarding the impact a certain criterion has on the overall value of an alternative. Moreover, the weight of a criterion may also depend on the scenario; a certain capability of an item may vary according to the scenario in which the item is to be employed. For example, the importance of the criterion “Electro-Optical Signature” of a platform depends on the detection capabilities of the adversary in a conflict scenario. Absent such capabilities, the weight of this criterion will most likely be quite low. Another example is the reliability of equipment. If the system has large redundancy with respect to the availability of this equipment, then the weight of the *reliability* criterion may be lower than in the case where the system relies on a single availability of that equipment. Also, the economic, political, social, and technological scenario may affect the weight to be assigned to the risk-related criteria. In the next section, we describe a method for eliciting weights that take into account the aforementioned factors: multiple stakeholders and multiple scenarios.

Methodology

The AoA process is about comparing the *values* of the alternatives. The best alternative—the one to be selected for acquisition—is the alternative that provides the highest overall value. But how can one combine multiple criteria and opinions into a single value? What is the scalar function that translates measurements and evaluations of the alternatives with respect to the various criteria, and evaluations of criteria weights, into a single value that can be compared among alternative items? The problem is exacerbated in the presence of multiple stakeholders who may provide a plethora of opinions and multiple scenarios that may result in different in-context evaluations.

We start with the basic construct, which is common in many decision settings—a linear value function (see e.g., Saaty, 1980). Simply stated, if w_j denotes the weight of criterion j, and v_{ij} is the value of alternative i with respect to criterion j, $i = 1,...,I, j = 1,...,J,$ then the overall value of alternative i, V_i^*, which is to be compared with the overall values of the other alternatives, is given by

$$V_i^* = \sum_{j=1}^{J} w_j v_{ij}, \; i = 1,...,I \qquad (1).$$

The alternative with the highest V_i^* value is the most likely alternative to be selected. Note that we state “Most likely” and not “definitely.” This distinction is discussed further in the last two sections of this report.

As mentioned earlier, the challenges are to determine the values of w_j, $j = 1,...,J$, and v_{ij}, $i = 1,...,I, j = 1,...,J$, taking into consideration the presence of multiple stakeholders and multiple scenarios. We will construct our value function step-by-step, starting with determining the values of the weights w_j.

Determining Criteria Weights for a Certain Scenario

Consider a certain reference scenario s. Assuming this scenario is realized, we wish to elicit from R stakeholders criteria weights w_{js}, $j = 1,...,J$, that (a) reflect the relative importance of the various criteria if scenario s prevails, and (b) represent a consensus among the stakeholders regarding these weights. An efficient and effective way to elicit preferences from decision-makers is through ratio-scale matrices, similarly to the setup used in AHP (see Saaty, 1980). The idea is as follows: Each stakeholder $r, r = 1,...,R,$ is asked to compare two criteria weights, say w_{js} and w_{ks}, with respect to scenario s. In other words, the stakeholder provides an assessment regarding the extent one criterion is more (or less) important than the other. The comparison is in terms of the ratio between the two weights. That is, p_{jks}^{r} is the assessment of stakeholder r regarding the ratio $\frac{w_{js}}{w_{ks}}$. Different stakeholders may have different opinions regarding the very same ratio. In other words, for two stakeholders r and r', we may have $p_{jks}^{r} \neq p_{jks}^{r'}$. A natural way to mathematically resolve such discrepancies is using least squares. The same way least squares are used to fit a “consensus” line among a clutter of points in statistical linear regression, we can derive a consensus set of weights by minimizing the least-square or L_2 distance. The usefulness and effectiveness of the least-square (L_2) measure as consensus forming method in decision analysis is described in Golany and Kress (1993). Formally, we solve the following non-linear optimization problem:

$$Min \sum_{j<k} \sum_{r=1}^{R} \left(\frac{w_{js}}{w_{ks}} - p_{jks}^{r} \right)^2 \qquad (2)$$

$$st$$

$$\sum_{j=1}^{J} w_{js} = 1, \quad w_{js} \geq 0.$$

The objective function is separable and quadratic, and therefore the optimization problem is easily solvable for real-size problems by tools as simple as the MS Excel Solver. The constraint is just a normalization of the criteria weights, which facilitates simpler computations down the road.

The optimal solution of Problem 2 is a vector $W^*_s = (w^*_{1s},\dots,w^*_{Js})$ of criteria weights that represent an L_2 consensus regarding the criteria importance in the presence of scenario s.

We solve Problem 2 S times—once for each possible scenario. For brevity and simplification, we drop the * sign from future notation.

The model in Problem 2 described above for criteria weights can obviously be applied, sequentially, to the different levels of the criteria tree (see above). For each master criterion at level l, we solve Problem 2 for the "child" criteria at level $l+1$. The weight of the end criterion at the bottom level is the product of the criteria weights leading to that criterion. For example, considering Figures 1 and 2, we first solve Problem 2 for *Effectiveness, ORML, Cost*, and *Risk.* Next, for the master criterion *effectiveness*, we solve 2 for *Mobility, Firepower, Survivability*, and *Mobility*. Similarly, Problem 2 is solved for the children (if any) of *ORML, Cost*, and *Risk.* Finally, we solve 2 for the lowest level (e.g., *Speed* and *Range* for *Mobility, SSKP* and *Fire-rate* for *Firepower,* etc.). The weight of the end-criterion in the value function 1, say *Speed,* is the product $w_{Effectiveness,s} \times w_{Mobility,s} \times w_{Speed,s}$.

Determining Alternatives' Values for a Certain Criterion and Scenario

Once again, we consider a certain scenario s. Let us also consider a certain criterion j. Similarly, to the way criteria's relative weights are elicited from stakeholders, the objective here is to obtain the ratio figure d^r_{iljs} that represents stakeholder's r opinion regarding the ratio between the value of alternative i and alternative l with respect to criterion j, in the presence of scenario s. Similarly to Problem 2, we solve now

$$Min \sum_{i<l} \sum_{r=1}^{R} \left(\frac{v_{ijs}}{v_{ljs}} - d^r_{iljs} \right)^2 \qquad (3)$$

$$st$$

$$\sum_{i=1}^{I} v_{ijs} = 1, \quad v_{ijs} \geq 0.$$

Problem 3 has an identical structure as Problem 2. Here, the optimal values $V^*_{js} = (v^*_{1js},\dots,v^*_{Ijs})$ are the mathematical consensus values of the alternatives with respect to criterion j, under the assumption of scenario s. Problem 3 is solved $J \times S$ times, once for each criterion and each scenario. As before, for brevity and simplification, we drop the * sign from future notation.

Note that in both Problems 2 and 3 we assume a homogeneous or "democratic" set of stakeholders; no stakeholder's opinion is considered more influential, or with higher weight, than others. If this is not the case, and certain stakeholders' opinions weigh more than others, then the objective functions in 2 and 3 are weighted accordingly with stakeholders' *r*-indexed weights. The problems are still easy to solve.

The Alternative's Value Function in a Scenario

Following the solutions of Problems 2 and 3, we can compute the value of an alternative, with respect to a certain scenario. This value reflects the consensus weights of the criteria $W_s = (w_{1s},\dots,w_{Js})$, and the consensus (relative) values of the alternatives

$V_{js} = (v_{1js}, ..., v_{Ijs})$. Formally, the consensus overall value of alternative i in scenario s is given by

$$\overline{v}_{is} = \sum_{j=1}^{J} w_{js} v_{ijs} \qquad (4)$$

In other words, if an "Oracle" could tell us the scenario s to be unfold, then the alternative i to be considered for selection is the one for which $\overline{v}_{is}$ is maximized. Absent such Oracle, the probabilities of the various scenarios must be taken into consideration when trying to identify the best alternative.

Scenarios' Probabilities

The old adage claims that "it is very difficult to forecast, especially the future." Nobody knows for sure which of the possible scenarios will actually be realized. Different stakeholders may have different opinions about the likelihood of the various scenarios. The combined assessment of scenarios' probabilities is obtained using the same methodology as in Problems 2 and 3.

Let a_{st}^{r} denote the assessment of stakeholder r about the relative likelihood of scenarios s and t. That is, a_{st}^{r} is the subjective opinion of stakeholder r regarding the extent scenario s is more (or less) likely than scenario t. The consensus probabilities q_s, $s = 1, ..., S$, of the various scenarios is obtained as the solution of the quadratic optimization problem

$$Min \sum_{s<t} \sum_{r=1}^{R} \left(\frac{q_s}{q_t} - a_{st}^{r} \right)^2 \qquad (5)$$

$$st$$

$$\sum_{s=1}^{S} q_s = 1, \quad q_s \geq 0.$$

The optimal solution $Q^* = (q_1^*, ..., q_S^*)$ is the consensus probability distribution of the scenarios. As before, we drop the * sign from future notation.

Selecting the Winning Alternative

Following the operations described in the previous section, the AoA team has an initial set of parameters that reflect the stakeholders' L_2-consensus regarding (a) the weight of criteria in each scenario, (b) the relative value of each alternative with respect to each criterion in each scenario, and (c) the (subjective) probabilities of the scenarios.

Recall from the previous section that for each scenario s we have now a calculated value $\overline{v}_{is}$ for each alternative i. This value represents the L_2-consensus outcome of the stakeholders' group decision process with respect to the relative standing of alternative i, if scenario s is realized. The L_2-consensus about the likelihood of scenario s is q_S. Thus, we have now a (subjective) probability distribution of alternatives' values over scenarios where each value $\overline{v}_{is}$ is associated with a probability q_S.

There are several ways to proceed from this point and identify the alternative that is most likely to be the best among the set of I alternatives. The most natural measure is the expected value where the “winning” alternative is alternative i for which $\sum_{s=1}^{S} q_s \bar{v}_{is}$ is maximized. Here we choose the alternative that “on-average” over the possible scenarios produces the highest relative value. This linear measure is quite common and easy to explain to decision-makers, but it is not always the right yardstick for choosing an alternative, in particular when the specific likelihoods of scenarios are to be looked at in more detail.

Another possible measure is the mode of the distribution; we simply choose the alternative that performs the best with respect to the *most likely* scenario. That is, if $s' = \arg\max q_s$ then the selected alternative i is the one for which $\bar{v}_{is'}$ is maximal. This measure is appropriate if there is one scenario that stands out as very likely—much more than any other scenario. If the induced (subjective) entropy of the scenarios, as implied from the stakeholders’ projections, is high, then obviously the mode measure will be inappropriate.

Lastly, and probably most appropriately, it would be better to select an alternative that is *good* over a large set of scenarios than an alternative that is *excellent* over a smaller set of scenarios. The goal here is to seek robustness in the choice of the winning alternative. The idea is as follows.

First, we set a probability threshold. This threshold represents the level of confidence, with respect to the realized scenario, which we wish to associate with the winning alternative. Suppose this probability level is α. Reasonable values of α are in the range 0.6–0.9. Next we generate all the minimal subsets of scenarios whose combined probabilities are at least α. For each such subset, we identify the alternative(s) for which the ***minimum*** value across the scenarios in the subset is ***maximal***. Formally, let $T_1,...T_M$ denote the set of all the α-valued subsets of scenarios. Each subset T_m comprises scenarios with combined probabilities of at least α, and any scenario removed from that set reduces the combined probabilities to less than α (hence, minimal subsets).

Consider an α-valued subset $T_m = \{s_1,...,s_{n_m}\},\ m = 1,...,M,$ where we have $\sum_{k=1}^{n_m} q_{s_k} \geq \alpha.$ Note that each α-valued subset of scenarios may contain different number of scenarios. Define $v_i(T_m) = Min_{s\in T_m} \bar{v}_{is}$ and $i_m = \arg\max\{v_1(T_m),...,v_I(T_m)\}$. Alternative i_m is the max-min alternative of the α-valued subset T_m. In other words, alternative i_m provides the highest *guaranteed* value among all alternatives if it is given that one of the scenarios in T_m is realized. Finally, $i^* = \arg\max\{v_{i_1},...v_{i_M}\}$ is the alternative that has the highest value with probability of at least α. Obviously, v_{i^*} is monotone non-increasing in α; the higher the required probability threshold, the smaller the assured alternative value. To demonstrate this procedure, consider the following example:

- Three scenarios, A, B, and C, with probabilities 0.3, 0.3, and 0.4, respectively
- There are 3 alternatives

- The overall values $\overline{v}_{is}$ of the (alternative x scenario) combinations are shown in Table 1.

Table 1: Values of Alternatives

Scenario Alternative	A (0.3)	B (0.3)	C (0.4)
1	0.7	0.5	0.95
2	0.6	0.8	0.6
3	0.9	0.4	0.5

Suppose α =0.6, which implies that we have here three subsets of scenarios that satisfy the minimum probability threshold requirement: $T_1 = \{A, B\}$ with probability 0.3+0.3=0.6, $T_2 = \{A, C\}$ with probability 0.3+0.4=0.7, and $T_3 = \{B, C\}$ with probability 0.3+0.4=0.7.

For the first scenario set, we have: $v_1(T_1) = 0.5$, $v_2(T_1) = 0.6$, $v_3(T_1) = 0.4$, and therefore the max-min alternative i_1 is alternative 2 with value 0.6. For the second scenario set, we have: $v_1(T_2) = 0.7$, $v_2(T_2) = 0.6$, $v_3(T_2) = 0.5$, and therefore the max-min alternative i_2 is alternative 1 with value 0.7. For the third scenario set, we have: $v_1(T_3) = 0.5$, $v_2(T_3) = 0.6$, $v_3(T_3) = 0.4$, and therefore the max-min alternative i_3 is alternative 2 with value 0.6. Thus, alternatives 1 and 2 are candidates for selection. But the maximum value over the eligible α -valued scenario sets is 0.7 and is obtained by alternative 1. Therefore, at confidence level of at least 0.6, the highest valued alternative is alternative 1.

Notice how the likelihoods of the scenarios affect the choice of alternatives. If, instead of the probabilities values in Table 1, the scenario probabilities were 0.6, 0.2, and 0.2 for scenarios A, B, and C, respectively, then it is easily seen that alternative 3 becomes the most preferred one with value 0.9. Going back to the original probabilities, if the threshold α is now 0.8, then we only have one subset (the complete set of scenarios), and the max-min alternative is alternative 2 with min value of 0.6.

To summarize, this quantile-type approach is both flexible, in the sense that one could choose the confidence level for selecting the best alternative, and robust by adopting the max-min measure of alternatives' values. This approach selects an alternative that is good over a wide range of possible scenarios instead of an alternative that is excellent in only limited number of situations.

Implementation

In the last two sections, we described a formal decision process for conducting AoA, in the presence of several uncertain scenarios, by a group of stakeholders who may have different perspectives and opinions regarding the subject matter. Disagreements and inconsistencies in preferences and assessments may occur with respect to criteria weights, alternative valuations and scenario likelihoods. The proposed group-decision model produces a consensus rating of the alternatives based on minimizing disagreements in the L_2 metric sense.

While the model is transparent and relatively simple to implement in a spreadsheet, it **should not** be considered as a "black box" that automatically produces a winning alternative based on stakeholders' and analysts' inputs. The "winning" alternative that emerges from the model may not necessarily be the final choice in situations where the value(s) of the runner-up(s) is (are) not significantly different from the winner's value. The model is a technical tool that, following a properly designed sensitivity analysis, can help guide the AoA process towards a robust decision. A possible paradigm for conducting a well-structured analysis of acquisition alternatives is as follows:

Step 1: Establish an AoA team that is tailored in size and scope to the military problem being considered. The team must comprise a group of stakeholders (e.g., field commanders and end-users), decision makers (e.g., budget managers, defense officials) and analysts (e.g., engineers, cost-estimators, operations-research analysts, and other subject-matter experts).

Step 2: The AoA team reviews documents describing operational setting, requirements, and capability gaps. An initial set of possible acquisition alternatives is generated. The analysts in the team start gathering more detailed data and information about the operational setting and the possible alternatives.

Step 3: Non-starter alternatives are identified and removed from consideration. Such alternatives are items that are rejected up front because of reasons such as not meeting minimum capability thresholds, they are too costly, or they are based on immature technologies. The team defines the sets of alternatives, criteria, and scenarios. This step also includes open discussions that set the stage for the detailed analysis to follow.

Step 4: Each member r in the AoA team provides her/his estimates for p^{r}_{jks} (see Determining Criteria Weights for a Certain Scenario section), d^{r}_{iljs} (see Determining Alternatives' Values for a Certain Criterion and Scenario), and a^{r}_{st} (see Scenarios' Probabilities). This step includes also operations-research and cost-estimation analyses, which provide valuable inputs to the AoA team.

Step 5: Model implementation on data gathered in Step 4. Output: set of alternative ratings.

Step 6: Discussion on the model results (alternative ratings) and performing sensitivity analysis on all three factors: criteria weights, alternatives' values and scenarios' probabilities. Step 5 may be repeated several times based on the discussions in this step.

We can see that the model described in the Methodology section acts as a decision aid and facilitator for discussions among the team members rather than an "Oracle" that crunches numbers and provides a "solution."

References

Behzadian, M., Kazemzadeh, R. B., Albadvi, A., & Aghdasi, M. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research, 200*(1), 198–215.

Cook, W. D., Kress, M., & Seiford, L. M. (1996). A general framework for distance-based consensus in ordinal ranking models. *European Journal of Operational Research, 96*, 392–397.

DoD. (2017, August 10). *Operation of the defense acquisition system* (DoD Instruction 5000.02). Washington, DC: Author.

DoN. (2006). *Seabase to Shore Connector analysis of alternatives final report*. Reston, VA: WBB Consulting

Figueira, J., Mousseau, V., & Roy, B. (2005). ELECTRE methods. In J. Figueira, S. Greco, & M. Ehrgott (Eds.), *Multiple criteria decision analysis: State of the art surveys* (pp. 133–162). New York, NY: Springer-Verlag.

Golany, B., & Kress, M. (1993). A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices. *European Journal of Operational Research*, *69*, 210–220.

Keeney, R. L., & Raiffa H. (1976). *Decisions with multiple objectives*. J. Wiley & Sons.

MITRE. (n.d.). Systems engineering guide: Analyses of alternatives. Retrieved April 23, 2018, from https://www.mitre.org/publications/systems-engineering-guide/acquisition-systems-engineering/acquisition-program-planning/performing-analyses-of-alternatives

Montibeller, G., Gummer, H., & Tumide, D. (2006). Combining scenario planning and multi-criteria decision analysis in practice. *Journal of Multi-Criteria Decision Analysis, 14*, 5–20.

RAND Corporation. (2006). *RAND Project Air Force: Analysis of alternatives (AoA) for KC-135 Recapitalization* (Executive summary). Santa Monica, CA: Author.

RAND Corporation. (2016). *Maritime Tactical Command and Control analysis of alternatives*. Santa Monica, CA: Author.

Saaty, T. L. (1980). *The analytic hierarchy process*. New York, NY: McGraw Hill.

Souders, R. M., Schulze, S., Ginburg, Y., & Goetke, J. (2004). *MPF(F) analysis of alternatives: Final summary report* (CNR D0009814.A2/Final). CNA.

Smith, M. W., Ritz, C., Smith, M., & Braun, C. (2011). *T-AO(X) analysis of alternatives final report* (CRM D0026011.A2/Final). CNA.

Stewart, T. J. (1997). Scenario analysis and multicriteria decision-making. In J. Climaco (Ed.), *Multicriteria analysis* (pp. 519–528). Berlin, Germany: Springer.

TRADOC. (2011). *Joint Light Tactical Vehicle analysis of alternatives final report* (TRAC-F-TR-11-025). TRADOC Analysis Center.

USAF, Office of Aerospace Studies. (2016, July 6). *Analysis of alternatives handbook*. Retrieved from http://afacpo.com/AQDocs/AoAHandbook.pdf

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net