COMBAT ASSESSMENT FOR THE SIMULATION OF WARFARE

THESIS

Benjamin Finch (Civ)

AFIT-ENS-MS-21-D-028

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-21-D-028

COMBAT ASSESSMENT FOR THE SIMULATION OF WARFARE

THESIS

Presented to the Faculty

Department of Operations Research

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Benjamin Finch (Civ), B.S. Mathematics

6 December 2021

AFIT-ENS-MS-21-D-028

COMBAT ASSESSMENT FOR THE SIMULATION OF WARFARE

THESIS

Benjamin Finch (Civ), B.S. Mathematics

Committee Membership:

Mark M. Gallagher, Ph.D
Chair

Maj Phillip R. Jenkins, Ph.D
Member

Stephen Sturgeon, Ph.D
Member

AFIT-ENS-MS-21-D-028

Abstract

Simulated combat requires knowledge of how both friendly and enemy forces are progressing relative to the stated friendly objectives and the believed enemy objectives. In the Department of Defense (DoD), the structure of these objectives is hierarchical, from the national strategic level down to the tactical level. Military assessment seeks to answer two primary questions: 1) are we creating the effects that we desire? and 2) are we accomplishing tasks to standard? Little research has been conducted in assessment methodologies for simulated combat. Some predominant assessment application arenas are education and gaming, which provide useful lessons for military combat assessment within a simulation. This work steps through several desirable characteristics for a simulated combat assessment methodology gleaned from DoD policy and these areas of research. After developing a value hierarchy from these characteristics, this thesis provides and evaluates several candidate methodologies for use within a combat simulation – the existing Combat Effectiveness & Combat Vulnerability methodology within the Bayesian Enterprise Analysis Model (BEAM), Bayesian Networks, Value-Focused Thinking, and Linear Programming. Each alternative's evaluation is informed by its application to a small combat simulation. We then create an alternative from the Value-Focused Thinking and Linear Programming alternatives with a better evaluation that the other four. The thesis terminates with some conclusory thoughts on the Linear Programming and ideas for future research.

To those who have taught me that "here" and "there" are one; and to Aristotle, whose meticulous and ethereal journey of observation illuminates the hidden catholic places of truth

Acknowledgements

I would like to thank my research advisor, Dr. Mark Gallagher, for his unfailing commitment to higher and inner learning. This work is made possible through his dedication to knowledge free from bias and judgement.

I am also deeply indebted to Dr. Alex McAllister, whose selfless investment in my person has continued to buttress my multi-faceted growth.

Benjamin L. Finch

Table of Contents

Page

Abstract . . . iv

Acknowledgements . . . vi

List of Figures . . . viii

List of Tables . . . ix

I. Introduction . . . 1

II. Predominant Contributions to Assessment . . . 7

III. Value Hierarchy for Selecting an Assessment Methodology . . . 14

3.1 The Value Hierarchy . . . 14
3.2 Potential Simulated Combat Assessment Methodologies . . . 19
3.2.1 Combat Effectiveness and Combat Vulnerability . . . 19
3.2.2 Bayesian Networks . . . 23
3.2.3 Value-Focused Thinking: Objectives Hierarchy . . . 27
3.2.4 Linear Program . . . 30

IV. Test Application . . . 36

4.1 Dresher's Game . . . 37
4.1.1 Dresher's Game 1^{st} Time-Step Adjudication . . . 40
4.2 Combat Effectiveness and Combat Vulnerability . . . 42
4.3 Bayesian Networks . . . 47
4.4 Objectives Hierarchy . . . 52
4.5 Linear Program . . . 56

V. Evaluation of Potential Methodologies Using the Value Hierarchy . . . 65

5.1 Realism . . . 65
5.2 Efficiency . . . 70
5.3 Robusticity . . . 74
5.4 Methodology Creation and Finalization . . . 78

VI. Conclusion . . . 86

Appendix A. Dresher's Game Conditional Probability Tables for the Bayesian Network Methodology . . . 89

Bibliography . . . 93

List of Figures

Figure Page

1. JCS Assessment Interaction 4
2. Value Hierarchy for Selecting an Assessment Methodology 15
3. Current Assessment Approach within BEAM 22
4. Basic Bayesian Networks and their Associated Joint Distributions 24
5. Creative Problem Solving Bayesian Network 26
6. Example Means-Ends Objectives Network 29
7. Example Objectives Hierarchy 30
8. Dresher's Game Objectives Hierarchy 39
9. Dresher's Game Bayesian Network 50
10. Dresher's Game Objectives Hierarchy 53

List of Tables

Table Page

1. Notional Conditional Probability Table 21
2. Initialization and First Time-Step Parameters for Dresher's Game Applications 42
3. CPT for Enemy Bombers Defending Against Friendly Bombers in a Friendly Offensive Mission 43
4. CPT for Enemy Fighters Defending Against Friendly Bombers in a Friendly Offensive Mission 43
5. CPT for Enemy Bombers Defending Against Friendly Fighters in a Friendly Defensive Mission 44
6. Red-Blue Def:Off Ratio and Resulting Health Distributions 45
7. Discrete-Uniform Quartile Bins for Red and Blue Assets 45
8. CE $\times$ CV States at End of First Time-Step 46
9. Dresher's Game LP Initialization Parameters 60
10. Evaluation of Alternatives - Categorical Labels 65
11. Evaluation Distributions by Objective 79
12. VFT-LP Evaluation – Categorical Labels 83
13. Evaluation Distributions by Objective for Top 3 Methodologies 84
14. Distribution of $S_{f,R,2}$ 89
15. Distribution of $S_{bb,R,2}$ 89
16. Distribution of $S_{bf,R,2}$ 89
17. Distribution of $S_{f,B,2}$ 89
18. Distribution of $S_{bb,B,2}$ 89
19. Distribution of $S_{bf,B,2}$ 89

Table Page

20. Distribution of $S_{bf,B,2}|S_{f,R,2}$ . . . 90

21. Distribution of $S_{bf,R,2}|S_{bb,R,2}$ . . . 90

22. Distribution of $S_{bb,R,2}|S_{f,B,2}$ . . . 91

23. Distribution of $S_{bf,B,2}|S_{bb,B,2}$ . . . 91

24. Distribution of $S_{f,B,2}|S_{bb,R,2}, S_{bf,R,2}$ . . . 91

25. Distribution of $S_{bf,B,2}|S_{bb,B,2}, S_{f,R,2}$ . . . 92

26. Distribution of $S_{bf,R,2}|S_{bb,R,2}, S_{f,B,2}$ . . . 92

27. Distribution of $S_{f,R,2}|S_{bb,B,2}, S_{bf,B,2}$ . . . 92

COMBAT ASSESSMENT FOR THE SIMULATION OF WARFARE

I. Introduction

Similar to real war, combat simulations require knowledge of how friendly and enemy forces are progressing. This knowledge extends potential simulation stopping conditions beyond the temporal and allows for the analysis of progress in relation to time, money, and asset posture. While recording the destruction of assets and utilization of consumables may provide worthy analytical results, commanders are often interested in the broader question, *did this scenario result in a win or a loss?* To answer this question, a simulation must contain some definition of winning or losing. This is most simply attained via stated objectives to complete. Combat simulation must be able to assess the operational environment and report progress toward or from these objectives. From the concrete (e.g., destroy all enemy ports) to the more abstract (e.g., achieve naval superiority) objectives, a singular combat assessment methodology inside of a simulation should be versatile enough to provide an answer to the win-loss question in the face of many different definitions of winning. This thesis provides an answer to such a methodology to be used in *campaign-level* combat assessment for simulated war.

A straightforward definition of assessment is "the process of using data to demonstrate that stated goals and objectives are actually being met" (1, p. 554). In the defense realm, the United States Joint Chiefs of Staff (JCS) define assessment as "a continuous process that measures the overall effectiveness of employing joint force capabilities during military operations" (2). In general, *assessment* is a word used for continual or ongoing feedback intended to improve a process. In more formal settings, assessment is divided into two categories: summative and formative. *Summative assessment*, or what some may term as evaluation, performs post-hoc reviews of performance. Examples of summative assessment include simple grading and comparisons to a benchmark and statistical or other analytical methods for obtaining comparative results from process outputs. The "external" (3, p. 19) characteristic of summative assessment naturally leads to its presentation of results as distant and/or static. Some examples of summative assessment include yearly personnel reviews and appraisals, student exams, and business metrics reporting. Alternatively, *formative assessment* focuses on continuous learning processes, intended to provide feedback during a given process (4). In education, formative assessment includes a collaboration of the student and the assessor to "actively produce [the student's] best performance" (5, p. 242). Interactive in nature, formative assessment intends to enhance performance before a process is ended, leveraging data (or experience) to customize the aid given to the assessed party. In this way, formative assessment may also include progress tracking or reporting. This thesis focuses on formative assessment methodologies within the context of campaign-level simulated war. We provide a suggestion to the question, "how should one approach campaign-level combat assessment within a computational simulation?"

The approach to assessment within a simulation of war, particularly if used for military training or analysis, should be to mimic actual decision-makers' assessment of the war effort. In actual combat contexts, military subordinates and analysts prepare an assessment for their commanders' situational awareness. The commanders use the assessment to provide further direction to manipulate the operational environment. As we model a war effort within a simulation of combat, the assessment portion should effectually be a model of the combined subordinate/analyst assessment and commander feedback. We include this concept in our value hierarchy for combat simulation assessment approaches in Section 3.1. Adjacently, we include the necessary characteristic that the methodology should be simple in its communication, allowing for streamlined presentation to commanders and other decision-makers. Useful to simple communication is the distilling of assessment outcomes into categories (e.g., win or loss), which is discussed in more detail in Section 3.1.

As the authoritative source on joint force operations, the United States JCS's "Joint Publication 3-0" (2) provides insight into the type of decision-making to be mimicked. The JCS concern themselves most with "operation assessment," which is the process used to "measure progress toward accomplishing tasks, creating conditions or effects, and achieving objectives" (2, p. II-9). The JCS assert that *operation assessment* should "begin during mission analysis when the commander and staff consider what to measure and how to measure it" (2, p. II-9). For a computational model employed in the field, this step would occur before the model is run, in which analysts set the preliminary objectives and any initial parameters. Furthermore, an "objective" in this definition is a goal that directs a course of action. An objective may be a phrase, as given in the National Strategic Objective for the DoD, or it may be the goal of an individual military task. In any case, the objectives within an assessment methodology for a simulation of war dictate where the proxy commanders' attention lie and in which direction they suggest movement.

Figure 1 provides the nested relationship between different levels of warfare with corresponding objectives. Within military applications, the assessment framework given in the purple arrow of Figure 1 is often called a "strategy-to-task" framework (6; 7; 8). Beneath objectives in the framework are effects that may be assessed. An "effect" is "the result, outcome or consequence of an action" (2, p. GL-9). A "task" is considered to be the smallest unit of military operation, ranging from destruction of an enemy asset to shipment of materiel. Some tasks are the objectives themselves, and some tasks support corresponding objectives with no intermediary effects. However, as Figure 1 illustrates, higher-level objectives inform lower-level objectives, which in turn direct military operations at all levels. Meanwhile, assessment is the mechanism used to provide bottom-up feedback. In the provision of feedback, assessment should answer two key questions: 1) "are we creating the effect(s) or condition(s) in the [operational environment] that we desire?" and 2) "are we accomplishing tasks to standard?" (2, p. II-11).

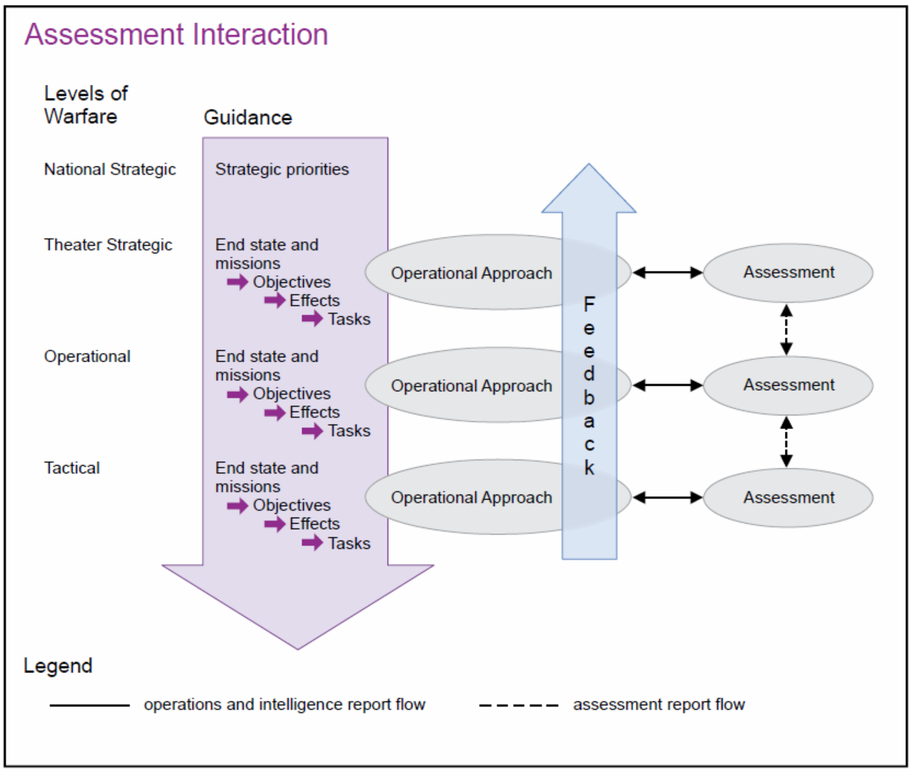

Figure 1: Interaction of Assessment as Defined by the JCS (2)

The focus of these two central assessment questions naturally leads combat assessment of simulated war to the formative realm. Although analysts have traditionally answered these questions retroactively using summative assessment methods, formative assessment allows for a reactive simulated combat environment when paired with the intent of answering the two key questions. With formative assessment methods, the simulation (or commanders) may receive real-time updates of the operational environment status in relation to specific objectives. Rather than performing post-hoc analyses, formative assessment more closely mimics the real-time war effort of assessing battle damage against friendly forces along with mission debriefs and intelligence reporting on adversary units. As such, this thesis focuses on formative assessment methodologies.

The remainder of this thesis is devoted to synthesizing JCS operational assessment requirements into a coherent structure for effective deployment of an assessment methodology within a combat modeling simulation environment. Gallagher et al (9) use resolution to define different levels of combat: *system/engineering, engagement, mission, campaign, defense enterprise*, and *whole government.* JCS doctrine establishes that each level of combat resolution be hierarchical in its assessment. This thesis does not attempt to narrow the scope of assessment to any particular engagement resolution; rather, we present general results in the attempt to apply to the broadest set of DoD applications possible. This research is focused on answering two research questions:

Research Questions

1) What are the desired characteristics of a combat assessment methodology for a programmatic/computerized simulation?
2) How should one conduct combat assessment within a programmatic simulation of warfare?

The next chapter provides an introduction to assessment through other predominant application areas. Taking the lessons learned from Chapter II and JCS assessment guidance, we construct a value hierarchy in Chapter III. Presenting several alternative methodologies for evaluation, we investigate their mechanics in a small combat simulation in Chapter IV. We then evaluate these alternative assessment methodologies in Chapter V for use in simulated warfare. Chapter V ends with a recommendation for a methodology for simulated combat assessment. We provide some concluding remarks and suggestions for further research in chapter VI.

II. Predominant Contributions to Assessment

In this section, we summarize and comment on the efficacy of several assessment applications. We utilize the various assessment approaches presented to develop a hierarchy of values in Section 3.1 that may be used to determine the worthiness of assessment approaches for combat simulations. Commencing with the readily recalled application area of education, we demonstrate how, historically, these assessment frameworks lack applicability in the combat assessment realm. Continuing with an example of assessment in games, we comment on the automation of assessment in chess. In this case, although formative assessment is tackled marvelously, the implementation of artificial intelligence (AI) is not reasonable for a combat simulation due to technological constraints. A combination of education and gaming is subsequently presented as a modern and promising area of research that may contribute to combat assessment within simulated warfare.

The National Center for Biotechnology Information (NCBI) defines an assessment framework as a "structured conceptual map of the learning outcomes of a programme of study along with details of how achievement of the outcomes can be measured" (10). Education is severely lacking in assessment frameworks as the NCBI defines. Many education assessment methodologies that provide conceptual maps or learning objectives do not explicitly link student progress to these objectives. Others provide links between student learning and achievement of objectives via a Likert scale, but then do not connect these scores to a larger framework. The International Association for the Evaluation of Educational Achievement (IAEEA) provides an archetypal educational assessment framework (11). Since the IAEEA reaches schools across 55 countries and in every continent (p. 78 – 81), the assessment framework is also widely accepted. However, the framework is mostly qualitative in nature. As such, it does not provide a structured conceptual map. While the authors graphically display learning outcomes (p. 61), they do not clearly state the mechanisms through which learning outcome achievement should be measured.

Some in the educational realm have developed assessment frameworks that have both the requisite conceptual map and details connecting student learning to achievement of objectives (12; 13; 14). Even so, such frameworks are largely untranslatable into the context of simulated combat, as educational assessment is historically summative. Standardized testing is a frequently recalled example of the summative nature of most educational assessment (15). The assessment framework in the context of these tests includes a calculation of an overall grade-point average (GPA) from individual course grades, which themselves are the composition of individual assignment and exam grades. Whether the GPA framework is a summative or formative type of assessment depends on the time context one considers. Over an assignment-level or lesson-level time frame, the assessment seems more summative. In contrast, a longer time frame, say a semester or year (or longer), provides ample opportunity for substantial formative assessment. On an assignment level, the minimum requirement for completion of the assessment of the assignment is typically the assignation of either a letter (e.g. A, B, C, etc.) or a percentage (0% – 100%). This minimum effort is clearly summative assessment. A more involved summative feedback could also look like correction of punctuation or addressing a calculation error. These examples provide the student with knowledge of the error, but do not address the learning deficits that caused the error. Individualized feedback addressing learning gaps or tips for how to progress toward a better assessment introduces formative assessment into this time context. However, with a longer time frame, the assessment takes on a more complex structure, as individual assignments may be graded dependently, referencing prior feedback or student errors. In this case, the student learns, perhaps with external encouragement over time, why they are making the mistakes they are making.

The time-dependent nature of assessment within education may naturally produce formative assessment.

The GPA framework is, in part, an analogue of some simulated combat methodologies. Most distinct from the GPA framework applied to education is that simulated combat lacks the student. The presence of the student obscures the influence of the GPA framework's formative assessment, as human psychology intervenes. Computerized simulations do not have this benefit of organic processing and learning. However, assessment methodologies may be able to address performance gaps within the operational environment. Current methodologies could, for example, explicitly search and reference past behavior in their feedback mechanisms. While some methodologies grade or otherwise utilize minute combat interactions, others only assess the aggregate effect of combat maneuvers. A singular value for progress towards scenario objectives, which is analogous to the overall GPA, is useful in many cases. The GPA framework does not perfectly translate its formative assessment characteristics into the assessment of simulated combat. While formative assessment is not explicitly part of the GPA framework, we recommended that the assessment of simulated combat incorporates formative assessment explicitly. One way an assessment framework may instantiate formative assessment is by addressing resource gaps with the achievement of combat objectives. This concept is related to instructors providing their student(s) individualized feedback of their learning deficits in order to progress toward classroom objectives.

The GPA framework also contains another measure of influence for combat assessment. The GPA framework inadvertently implies that its metric, given on a scale of 0.0 - 4.0, accurately and succinctly summarizes student achievement. As a instantiation of the *Law of Large Numbers*,[1] a student's GPA approximates their actual (theoretical) average achievement. Translating this into the assessment within simulated combat, we should consider multiple possible next time-steps within a scenario, rather than a single possible next time-step. We may do this by considering cases where friendly or enemy assets, or both, are strong, weak, or of medium strength in the next time-step. By sampling in this fashion, we can form a distribution of performance across the next time step. In this way, we have a two-dimensional application of the Law of Large Numbers, where having more possible futures sampled gives us a clearer picture of the theoretical future distribution. It may be of interest to instead obtain sample means of these levels of performance. In this case, the Law of Large Numbers becomes two-dimensional, otherwise known as the *Central Limit Theorem*. In either case, modeling variability within a simulation – a highly volatile context – is advantageous. Therefore, distributional results would enhance an assessment framework for simulated combat. In building the value hierarchy in Section 3.1, we consider that the distributional output of an assessment framework for simulated combat is one approach to recording the partial completion of objectives and categorization into *win, loss,* or *unresolved*.

Simulated war also requires a well-defined assessment framework capable of quantitative integration with a feedback loop within a scenario. Research into games such as chess has more recently involved applying search algorithms via AI to master the game. A prominent example and breakthrough in AI technology is IBM's *AlphaZero*, which uses Monte Carlo Decision Trees as a basis for its in-game play (17). Rather than an explicit assessment framework, however, machine learning applications create models by training them on data. The model attempts to select the best move pro-

[1]First proved by Jakob Bernoulli in 1713, the Law of Large Numbers states that a large collection of independent and identically distributed random variables has a sample mean approximately that of their theoretical mean (16).

vided any board configuration. Such a method for assessing the quality of any given move is rather straightforward in that the desired end-state of the game is known: end on a move in whose reply the opponent's king is in direct attack and cannot escape. Capturing this algorithmically is non-trivial, but at least the end-state is known. Combat simulation is not as well-defined, as each combat scenario may have different objectives. An additional complication of mirroring AI training techniques in the simulated combat sphere is the sheer size of the decision space. In chess, the decision space is moderate, as the number of legal moves is limited within each turn. However, Shannon demonstrated that at there are roughly 7×10^{14} possible games within just the first 5 turns (18). The difficulty with simulating chess lies in deciding which of these games to play to maximize the probability of a win.

In simulated combat, the difficulty lies also in such a decision, although the decision space is much larger. Unhindered by the constraint of an 8×8 board, simulated combat has a decision space at least as large as the number of assets-to-locations assignments. In modeling a country's military assets, this lower bound becomes astronomical. Therefore, training algorithms located in AI research and practice are impractical with current technological constraints. Still, this area of research has captured a way to assess progress toward a stated goal formatively, which is desired in combat simulation assessment. In Section 3.1, we incorporate the need for computational efficiency into the decision for an assessment methodology for simulated warfare. The driving factor in this decision is the decision space's size.

In the forefront of modern assessment research is *serious games* – a combination of education and games as a way to measure student characteristics, rather than solely their knowledge. Serious games are broadly defined as "digital games created not with the primary purpose of pure entertainment, but [rather] with the intention of serious use as in training, education, and health care" (19). Concerning the characteristics, the multiple-choice-question-type of student assessment fails miserably, specifically when dealing with assessments of processes, such as in mathematics (20), "complex problem solving, communication, and reasoning skills" (21). Serious games have been used to assess levels of systematic thinking, inductive reasoning, creative design skills, self-esteem, collaboration, interpersonal skills, and aggressiveness (22). Assessment frameworks for the measurement of a user's abstract characteristics can be quite extensive. One such example is in the assessment framework of the game *Elder Scrolls IV: Oblivion* (23). In addition to addressing the issue of statistical dependence within a sequence of actions (23, p. 300), the evidence-centered Bayesian model for assessing a player's actions is capable of assessing both direct and abstract goals. Section 3.2 explores this approach for combat simulation. However, combat assessment could benefit from being able to distill abstract concepts into quantifiable connections with individual operating environment actions. We consider these ideas in our value hierarchy in Section 3.1.

A key component of assessment executed within serious games is assessment without the user's knowledge of being assessed. *Stealth assessment* – assessments that are "embedded deeply within games to unobtrusively, accurately, and dynamically measure how players are progressing relative to targeted competencies" (24) – is the focus of recent research tasked with this exact problem: observation of a process disturbs the process itself.[2] Fortunately, a computational simulation does not encounter the same difficulty as observation of other processes, since the process being observed is a programmatic model. A separate module may handle observations to gather the information, and the simulation may resume afterwards, undisturbed. Nevertheless, stealth assessment research may be tweaked to apply to combat modeling. First and

[2] The central problem with observational studies, sometimes referenced via the Heisenberg Uncertainty Principle, is the inability to precisely measure quantum particles' momentum and position simultaneously. For more information, see (25).

foremost, this research has tackled the issue of statistical dependence between actions within a serious game and its consequence on assessment (24). That is, the actions early on in a game (or in a series of combat actions) impacts the potential actions later on in the game (or combat sequence). Second, stealth assessment has been implemented in multiple scenarios to measure different types of learning and learner attributes within serious games (26). Combat assessment can learn from this research by utilizing the robust and flexible capabilities of stealth assessment in serious games. We consider the statistical dependence of actions within the operational environment as we incorporate multiple combat domains.

In this chapter, we reviewed and commented on assessment contributions outside of simulated warfare applications. Education assessment delineates between summative and formative assessment, reflecting upon how providing resource gaps as feedback can be instrumental to success. The GPA framework also naturally points to the need for a distributional characterization of the operational environment. We can categorize this distribution in terms of *win, loss,* and *unresolved*. Assessment within chess demonstrated that even an enclosed-space game with scarce components requires an efficient assessment methodology. Within serious games, we see the benefit of incorporating abstract goals into an assessment methodology, as the simulated combat application may utilize more ethereal objectives. Lastly, stealth assessment delved deeper into serious games, revealing the significance of dependent actions as relates to assessment within a multi-domain combat context. The next chapter synthesizes these lessons learned with the JCS requirements from Chapter I to construct a value hierarchy capable of evaluating potential methodologies for combat assessment within simulated war.

III. Value Hierarchy for Selecting an Assessment Methodology

In this chapter, we develop a value hierarchy for the evaluation of assessment methodologies within the simulated combat model context. We develop a set of measurable and specific values for the hierarchy from the characteristics gleaned from Chapters I and II. We end by providing several candidate assessment methodologies. In the subsequent chapter, we demonstrate on a small application how these methodologies would function in a simulation of combat. We then utilize the information from this chapter and Chapter IV to provide a detailed evaluation of the methodologies in Chapter V and to evaluate our created alternative.

3.1 The Value Hierarchy

In this section, and throughout the rest of this chapter, we utilize the value hierarchy concepts as prescribed by Ralph L. Keeney (27). We start this section by providing and describing the devised value hierarchy for evaluating simulated combat assessment methodologies. We then briefly discuss four methodologies to be considered as alternatives for evaluation by our hierarchy. The subsequent chapter presents the mechanics of each methodology in a small combat simulation before evaluating the methodologies in Chapter V, which ends by providing a suggested methodology for use within a campaign-level simulation of combat.

A *value hierarchy* is an effective tool for quantitatively evaluating partial achievement of criteria. The value hierarchy proposed to evaluate the candidate assessment frameworks is shown in Figure 2. The *strategic objective* of the framework is "Maximize the Suitability of the Assessment Framework for a Campaign-Level Combat Modeling." The fundamental objectives are to maximize realism, maximize efficiency, and maximize robusticity. The fundamental objectives are colored green in Figure 2 and are the first tier under the strategic objective.

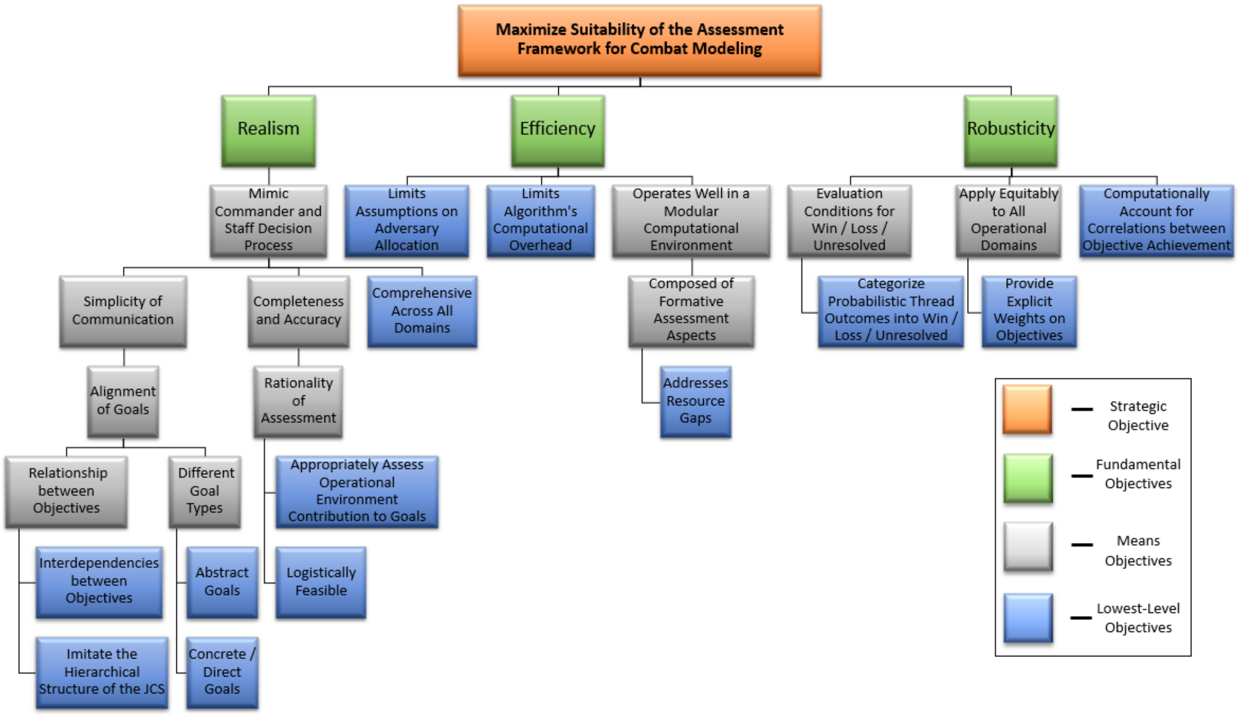

Figure 2: Value Hierarchy for Selecting an Assessment Methodology

Several subsequent objectives are important to this hierarchy for connecting the measurable objectives of the hierarchy to the desired aspects of an assessment methodology within a simulation of warfare.

The Realism portion of the value hierarchy addresses several key components of a desired methodology: realistically mimic commander and staff decisions, simplify communication, and completely and accurately assess the operational environment. An assessment methodology can become more realistic by simplifying communication, maximizing the completeness and accuracy of the assessment methodology, and by ensuring that the methodology includes all possible operational domains. By simplifying communication, we mean that the structure of the methodology should be reasonably similar to that of the strategy-to-task framework of Figure 1. An assessment methodology can create this effect by having an integrated alignment of goals within the framework, connecting the lowest-level objectives with the highest-level ones in a similar fashion to the strategy-to-task framework. However, to get to measurable objectives, we break the alignment of goals concept down into the modeling of relationships and the incorporation of different goal types. Within the relationship-between-objectives concept, the assessment framework should model the interdependencies between objectives and imitate the strategy-to-task structure. Within the JCS framework, objectives will not necessarily be independent. The framework should therefore capture this aspect of the JCS hierarchy to maximize the accuracy of the approximation to commander and staff decision processes. Under "Address Different Goal Types," a combat assessment methodology should also be able to include both abstract and concrete, or direct, goals. Within the JCS strategy-to-task structure, both types of objectives will be utilized in real decision-making. An assessment methodology for simulated combat should also be flexible enough to handle multiple types of objectives to better mimic this decision process.

The second aspect of mimicking commander and staff decision-making is the completeness and accuracy of the assessment provided by a methodology. This objective can be more technically phrased as the rationality of assessment. Two key aspects of an assessment methodology's rationality are whether or not, and to what degree, it can appropriately assess the operational environment's contribution to objectives and whether or not its methodology provides a logistically feasible conceptualization of the operational environment. The former aspect ensures both completeness and accuracy by targeting the connection between the operational environment and the assessment. The latter aspect contributes to the accuracy of the methodology. If the methodology evaluates based on worst-case or best-case scenarios, then it may over-allocate assets to the possible set of combat actions, and therefore assess a combat posture that is logistically infeasible for that moment in the simulated war.

The last aspect of realism addresses whether or not the methodology can incorporate multiple combat domains. In order to fully mimic commander and staff decision processes, the assessment methodology cannot be blind to any part of the operational environment.

The second fundamental objective in the value hierarchy given in Figure 2 is to maximize efficiency. Here, we target the desired efficiency of an assessment methodology. The methodology must work efficiently within a modular programmatic environment, as well as be internally efficient. We can increase the efficiency of an assessment framework via three lines of effort. First of all, limiting assumptions on the adversary's allocation limits the amount of data or computation necessary. If only friendly asset information is required for the assessment, the speed of computation will increase. Moreover, we can also limit the computational overhead to increase the assessment methodology implementation's efficiency. This includes larges quantities of data as well as long algorithmic processes. These two objectives contribute directly to technological efficiency. By being more efficient internally as a methodology, its implementation should mirror this quality.

A combat assessment methodology should also fit well in a modular computational environment, streamlining the coding structure. The key way for an assessment module to communicate with other modules is by providing feedback to other modules. This is the cornerstone characteristic of formative assessment. The major role that combat assessment can play in providing a formative assessment for the simulation is to analyze and report resource gaps. Here, the framework yields a personalized, continual report of the potential improvements that the friendly side can make to obtain a more positive assessment result. While addressing these gaps may result in a heavier computational load, an assessment module improves its efficiency in communicating with other modules by minimizing its output data size. However, providing feedback increases the efficacy of assessment and improves the quality of either side's allocation.

The last fundamental objective is to maximize robusticity. Robusticity here is a catch-all for the inclusion technical characteristics. Specifically, we break robusticity into three components: allow for the partial completion of objectives, be capable of including multiple combat domains, and address objective interactions. A robust assessment methodology is able to account for and include several niche aspects of simulated combat. For example, every assessment methodology should correspond to at least one output – the overall rating. However, when subjected to quantitative assessment, as we are in this case, we may want to distill the numerical rating into categories. A simple ternary option is win/loss/unresolved. Assessing a certain portion of the simulation as one of these three labels can ease communication, simplify stopping conditions, and lays the foundation for one way to improve simulation accuracy. In the education portion of Chapter II, we saw the need to compile multiple outcomes in an assessment. In combat assessment, this may look like splitting a scenario into potential futures utilizing slightly different enemy/friendly allocations. Evaluating all of these potential futures creates a probability distribution of assessment outcomes, say on a domain of [0,1]. An assessment methodology capable of categorically separating this distribution into the aforementioned ternary can then output a percentage of the scenario won, lost, and unresolved. By dwindling the unresolved portion, we can become increasingly accurate in our assessment.

Another facet of robusticity is the equitable application of the assessment methodology to all operational environment domains. While being able to *include* all domains is an aspect of realism, an equitable methodology attributes proportional (or otherwise weighted) significance to the domains. For example, it may not be reasonable to assess the space domain as equal to the ground domain if the size of the space operational environment is significantly smaller than that of the ground domain. However, the quantity of assets may not be the sole indicator of import. No matter the motivating factor, attribution of weights to different objectives allows for an equitable policy according to stakeholder or analyst input.

The third component of increased robusticity is to account for correlations between objective achievement. One example of the statistical dependence of achievement to objectives is a situation where we have the following two objectives: 1) achieve air superiority, and 2) destroy enemy Integrated Air Defense Systems (IADS). Clearly, destroying enemy IADS is one way that we can achieve air superiority. Achievement of these two objectives is therefore coupled. An assessment methodology that can account for these dependencies will improve fidelity.

3.2 Potential Simulated Combat Assessment Methodologies

This section introduces the list of four alternatives to be evaluated through value hierarchy described above. All assessment methodologies have only two potential data inputs to consider: the perceived resulting assets from the previous time step's combat resolution and the current perceived forces.

3.2.1 Combat Effectiveness and Combat Vulnerability

The following explanation of the Bayesian Enterprise Analysis Model's (BEAM's) current methodology is summarized from the BEAM overview article, *Bayesian Analysis of Complex Combat Scenarios* (28).

To understand the assessment methodology *Combat Effectiveness and Combat Vulnerability* (CE/CV), we first explain some of the mechanics of the simulation and of the adjudication algorithm. A scenario within BEAM is broken into phases, which progress linearly in time. Each phase has a set of user-defined goals, which are either abstract (e.g. air superiority) or concrete. The simulation progresses discretely, by a user-defined interval, called a *time-step*. Each time-step has 16 sets of combat, called *threads*. These threads have associated weights summing to unity, as each is relatively more or less likely to occur than others. This construction adds variability to the probabilistic outcomes from the adjudication algorithm.

The adjudication algorithm is centered on data, fed from the Joint Wargaming Analysis Model (JWAM) (28, p. 15). These data help to construct an "extensive set of Conditional Probability Tables (CPTs)" (28, p. 15), tailored to specific missions and asset types – targeted, defending, and attacking assets. CPTs represent the stochasticity of adjudication, and function to incorporate multiple concurrent runs (28, p. 14). An example of a CPT is taken from (28) in Table 1. The outcomes are the proportion of the targeted asset that remains functional, represented in probability of occurrence. Note that each row of Table 1 sums to 1, since one of the outcomes must occur. In addition, all columns, save the first and last, sum to an equal probability – 0.6. This is because, "within BEAM, all of the CPT outcomes are represented as discrete-uniform distributions of specified assets" (28, p. 16), denoted by $\mathbf{U}(a)$. These distributions are not homogeneous, as the width of each bin differs with the asset and the time-step. For example, one asset may have a discrete-uniform distribution of $\mathbf{U}(\alpha) = [0.25, 0.25, 0.25, 0.25]$ with bin boundaries $b_\alpha = [0, 0.2, 0.5, 0.7, 1]$, while another asset has a discrete-uniform distribution of $\mathbf{U}(\beta) = \mathbf{U}(\alpha)$ but with bin boundaries $b_\beta = [0, 0.5, 0.6, 0.8, 1]$.

To integrate asset entities with probabilistic outcomes of the CPTs, BEAM represents assets with their probability distribution in state vectors (28, p. 8). The interpretation is the remaining capability or resource of the asset from the original value, given within the continuous interval [0,1]. For example, if a scenario starts off with 10 F-16s, but this time-step sees only three remaining, then the F-16's remaining

Table 1: A Notional Conditional Probability Table (CPT) of targeted enemy assets distribution based on the ratio of defensive to offensive missions (28, p. 16).

Def:Off	Targeted Asset Quantity Bins						
Ratio	[0.0]	(0,0.2]	(0.2,0.5]	(0.5,0.6]	(0.6,0.9]	(0.9,1.0]	[1.0]
0:1	1	0	0	0	0	0	0
1:2	0	0.40	0.28	0.17	0.10	0.05	0
1:1	0	0.16	0.26	0.30	0.19	0.09	0
2:1	0	0.04	0.06	0.13	0.31	0.46	0
1:0	0	0	0	0	0	0	1

capability is 0.3. The probability distributions on this interval are discrete-uniform, with four equal-probability bins. The quartiles are referred to as low, fair, moderate, and high. Combining friendly and enemy assets in the 16 total combinations of these quartiles results in the 16 threads per time-step (28, p. 8–9).

BEAM's adjudication algorithm evaluates each of the 16 threads and "categorizes its outcome into three possible states: (1) one side [has] achieved all of their phase goals, (2) one side loses the war, or (3) the [combat] is not yet resolved" (28, p. 22). Threads falling into the first category are stored until the proceeding phase. Those in the second category are stored until the end of the scenario, as they have terminated. Threads in the third category proceed into subsequent time steps until one side's phase goals have been achieved, or until the proportion of category three threads is insignificant in comparison to its complement. (Thread proportions are calculated via the product of their respective thread weights). Phase goals are "stated in terms of achieving one of five states of effectiveness relative to the enemy" (28, p. 23): Enemy Supremacy, Enemy Superiority, Contested, Friendly Superiority, and Friendly Supremacy. These regions can be visually demonstrated by two-dimensional cross sections between intervals in the axes Combat Vulnerability (CV) and Combat Effectiveness (CE), as shown in Figure 3. Additionally, the user inputs a positive probability, say ρ, as well as a parameter defining the size of the regions, Δ. Once the thread achieves the defined probability ρ of being in one of the regions, the thread has achieved the phase goal sufficiently. This approach therefore assumes monotonic movement toward either Friendly Supremacy or Enemy Supremacy after the probability ρ has been surpassed.

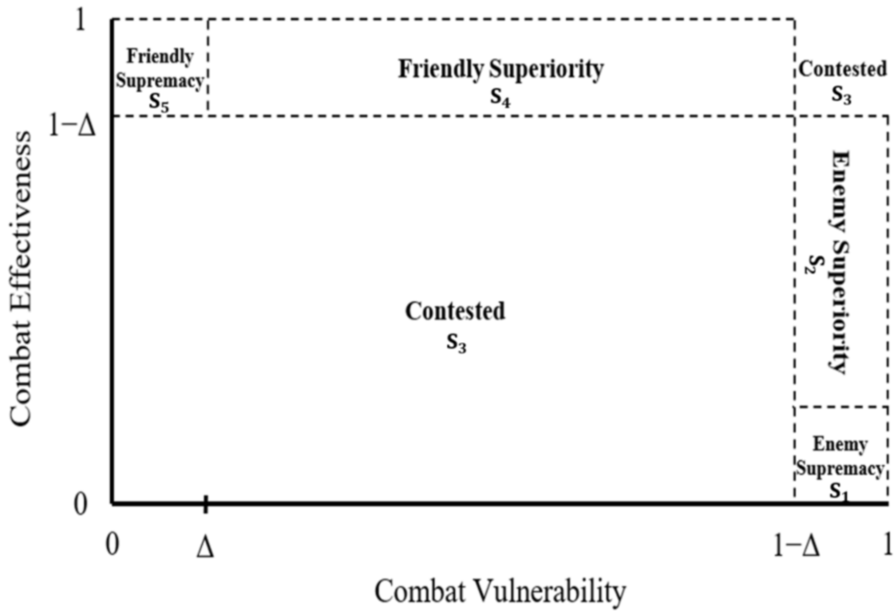

Figure 3: Current Assessment Approach within BEAM

Mathematically, CV is the "probability distribution of average losses of friendly action assets for [a] mission group G" (28, p. 24). Let a_A denote an action asset, so that the CPTs provide the discrete-uniform distribution of these assets after combat, $\mathbf{U}(a_A)$. Then let δ_{a_A} be the "ratio of the resulting number of assets to the allocated number of assets" (28, p. 24), which are found in the ranges in the column headers of each CPT. The body of each CPT contains the probability $Pr\{\delta_{a_A}\}$ of being assigned to the ratio δ_{a_A}. To obtain the distribution of CV, we average the asset distributions across all missions in the mission group G

$$\mathbf{U}(CV_G) = \frac{1}{|G|} \sum_{a_A \in G} \mathbf{U}(\delta_{a_A \in G}) \tag{1}$$

as on page 25 of (28).

CE is calculated similarly. We now consider attacking enemy assets, $a_E \in E$, which may be destroyed, and defending friendly assets, $a_D \in F$. Now, δ_{a_E} represents the ratio of enemy assets destroyed and δ_{a_D} is the ratio of surviving friendly defensive assets. Letting $G = E \cap F$, we have the distribution of CE as

$$\mathbf{U}(CE_G) = \frac{1}{|G|}\left(\sum_{a_E \in E} \mathbf{U}(\delta_{a_E \in E}) + \sum_{a_D \in F} \mathbf{U}(\delta_{a_D \in F})\right) \tag{2}$$

as on page 25 of (28).

The distributions $\mathbf{U}(CV_G)$ and $\mathbf{U}(CE_G)$ are given by the probability of being in each of the following intervals: $[0, \Delta], (\Delta, 1 - \Delta)$, and $[1 - \Delta, 1]$. The joint probability distribution of CE and CV yields probability of being in each region of Figure 3.

Lastly, consider the states $S_i, i = 1, 2, 3, 4, 5$, associated with the regions in Figure 3. Achieving a particular phase goal is defined by meeting or exceeding the threshold ρ of being in a state $S_j, j = 1, 2, 3, 4, 5$, such that $j \geq i$ and S_i is the goal state of the phase goal. A loss of a phase goal occurs when the probability of being worse than that phase goal $S_j, j < i$, meets or exceeds ρ. If any single phase goal is lost, the thread is considered lost. However, all phase goals must be won in a thread for the thread to be considered won. Any thread not meeting one of these two criteria is unresolved (28, p. 26).

3.2.2 Bayesian Networks

A Bayesian Network is a graphical model used "for representing multivariate probability distributions" in a directed acyclic graph, oft exploited for its ability to convey and calculate conditional probabilities. Having unidirectional arcs, a Bayesian network explicitly defines dependencies; nodes (variables) at the tail of the arc are dependent upon those at the head of the arc.

To build a Bayesian network by hand, there are a few basic rules. Consider random variables A, B, and C. We can position these variables in a directed graph using various combinations of arcs. Figure 4 shows four foundational examples. In network 1 of the figure, we have placed no arcs. In this case, all three random variables are marginally independent of each other. Therefore, the joint distribution is simply the cross product of the three univariate distributions.

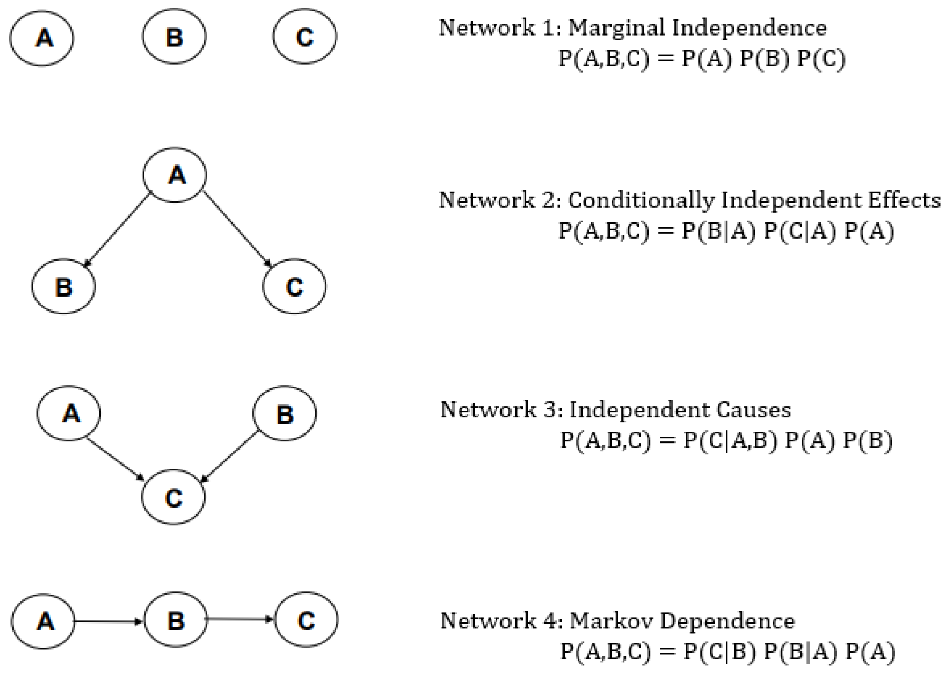

Figure 4: Basic Bayesian Networks and their Associated Joint Distributions

Network 2, however, introduces dependence. In this case, variable A as an influence on variables B and C. B and C are considered the *effects* of the *cause*, A. We consider that B and C are conditionally independent of each other, in the presence of A; that is, they are dependent only through their conditional dependence upon A. We can find the joint distribution representing this Bayesian network by crossing the two conditional distributions with each other and with the distribution of A.

Flipping Network 2 on its head yields Network 3 from Figure 4. In this case, we have independent causes influencing a single effect. Since C is dependent upon both

A and B, we cross the conditional distribution $f(C|A, B)$ with the two univariate distributions of A and B.

The last foundational example is Network 4. Here, the network is based on Markovian principles. The Bayesian network here is similar to considering an entry $a_{1,1}$'s distribution in a transition matrix after two iterations. With this linear dependence network, we can find the joint probability distribution defining this Bayesian network by crossing the conditional distributions of the variables from the right side of the network to the left.

Shute et. al. (23) explicate the usage of a Bayesian network within an assessment context. The authors provide a tangible example of dynamic stealth assessment in immersive games, which they wish to expand to serious games for educational purposes in the future. In their article, the authors focus on the game *Elder Scrolls IV: Oblivion* as an adaptive immersive game that could allow for a Bayesian network integrated with individual player data. Specifically, Shute et. al. break down the concept of "Creative Problem Solving" into a Bayesian network and some notional player data to instantiate an example of dynamic stealth assessment of abstract attributes. Their main result is shown in Figure 5.

The joint probability distribution defining the Bayesian network in Figure 5 is easily deduced from the basic conditional probability rules established above. Let CPS, PS, C, E, N, OE, and ON stand for *CreativeProblemSolving, ProblemSolving, Creativity, Efficiency, Novelty, ObservedEfficiency,* and *ObservedNovelty,* respectively. Then the joint distribution defining this Bayesian network is

$$f(CPS, PS, C, E, N, OE, ON) =$$
$$g(OE|N) \cdot h(OE|E) \cdot i(N|C) \cdot j(E|PS, C) \cdot k(C|CPS) \cdot l(PS|CPS) \cdot m(CPS)$$

where each of g, h, i, j, k, l, m are functions defining the corresponding distribution.

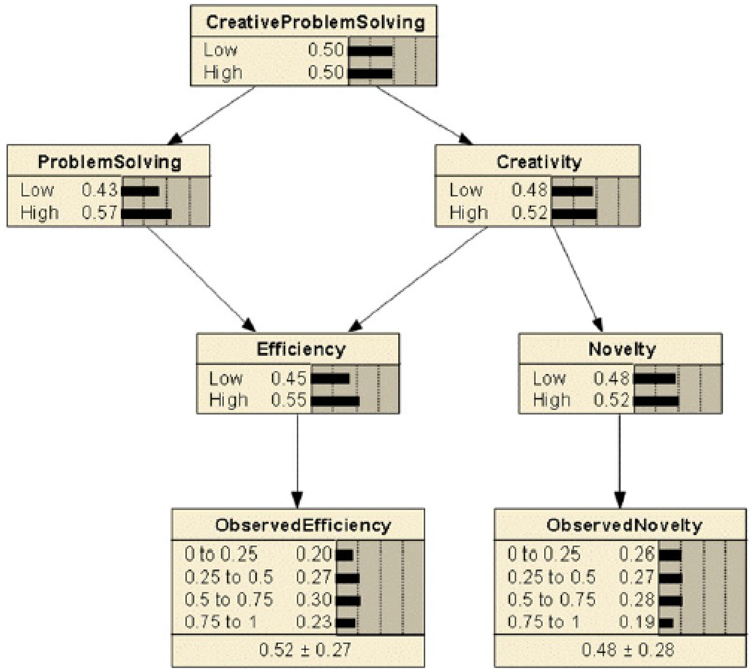

Figure 5: Example Bayesian Network for Assessing a Player's Creative Problem Solving, from (23, p. 21)

While Figure 5 exemplifies the modeling aspect of a Bayesian network, the conditional probabilities are driven by constructed tables relating data to the model. For example, if a combat objective to be scored were *efficiency*, with the end goal as destroying a navy fleet, then deploying M-1 Abrams tank platoons would likely have a much lower probability, say 0.05, than would deploying a squadron of F-16 fighter planes, which may result in a probability upwards of 0.90. Note that these tables of data are constructed from practical application and/or expert opinion. Although perhaps initially an arbitrary choice, these task-score assignments may be dynamic and data-driven within a simulation of combat. Schute et. al. (23) suggest *novelty* be quantified via the proportion of other players that have performed the same task, while quantifying *efficiency* based on the time elapsed in completing a certain task. The player's observed values from 0 to 1 for each of these attributes are provided in the form of a distribution in the lowest two nodes of Figure 5. In the combat example above, if M-1 Abrams tanks historically always fail to defeat a navy fleet, then they may have an efficiency of 0 for that task.

To apply a Bayesian network to a more concrete top-level node being assessed, one might imagine certain tasks that could be performed within the combat context that may contribute, in varying degrees, to the accomplishment of the network's top-level objective. For example, if a friendly goal is to destroy an enemy base, then sending in various types of friendly offensive assets to attack the base may link to equally-as-various levels of destruction. In addition, sending in defensive assets, or combinations of offensive assets, may bolster the effectiveness of a single asset type in accomplishing the task. In this way, conditional relationships based on asset types could represent the nodes in a Bayesian network for a more concrete objective.

For a more in-depth theoretical understanding of Bayesian networks, the interested reader is directed to (29).

3.2.3 Value-Focused Thinking: Objectives Hierarchy

All citations in this subsection reference Ralph L. Keeney's book, *Value-Focused Thinking* (27), unless otherwise stated.

An *objectives hierarchy* (OH) is the main assessment structure from value-focused thinking (VFT), a subset of decision analysis. However, it typically takes multiple steps to produce a fully-functional OH. VFT commences with thinking about what is important within a specific *decision context*, or problem context (p. 29 – 33). Within the decision context, one may prefer to see specific outcomes, such as "maximize profit." This is an example of an *objective*. There are two types of objectives within an OH: *fundamental* and *means* (p. 34). While a "fundamental objective characterizes an essential reason for interest in the decision situation, [...] a means objective is of interest in the decision context because of its implications for the degree to which another (more fundamental) objective can be achieved" (p. 34). We can link these fundamental and means objectives together in a "means-ends objectives network" (p. 69-70). Objectives reasoned out to be "ends objectives" (p. 66–68) are candidates for the fundamental objectives. For each fundamental objective, we then build a *fundamental objectives hierarchy*, which specifies the important aspects of each fundamental objective (p. 71). From these fundamental objectives, it is often helpful to identify an *overall fundamental objective* (p. 77–78). The overall fundamental objective is the root motivation for the decision context. However, more broadly, an organization may have general driving objectives, called *strategic objectives*, which "provide the foundation for creating alternatives or identifying any decision opportunities based on values" (p. 207).

In the DoD realm, VFT could be applied at each level of warfare (see Figure 1). However, within each level, a VFT objective could be either a DoD objective or an effect. For better specificity, a VFT fundamental objective closely corresponds to a DoD objective, and a VFT means objective closely corresponds to an effect or task. The overall fundamental objective does not have an explicit DoD counterpart, although the strategic objective for a decision context mirrors any DoD strategic objectives applicable to a decision context.

To evaluate alternatives, one must create quantitative attributes. These come in one of three categories: natural, constructed, and proxy (p. 101–103). Natural objectives are "those in general use that have a common interpretation to everyone" (p. 101). Constructed attributes are typically reserved for more abstract objectives, and may involve subjective – though rigorously defined – numerical indicators for each level of the attribute (p. 101–102). Lastly, a proxy attribute is used whenever it is "very difficult to identify either [a natural or constructed] attribute for a given objective" (p. 103). Therefore, one creates an attribute that indirectly measures the

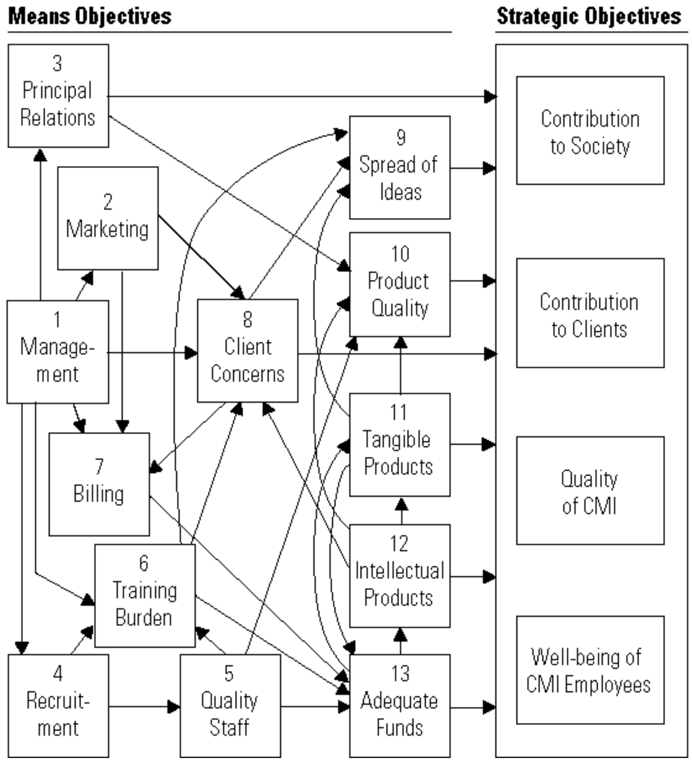

Figure 6: From (30), Figure 2, a means-ends objectives network. An arrow indicates that achieving the former objective influences achieving the latter objective.

objective in question, such as a "natural (direct) measure for a means objective," so that the "levels of that attribute are valued only for their perceived relationship to the achievement of that fundamental objective" (p. 103).

Measurement and weighting of the objectives can be done in many fashions. For information regarding measurement of the objectives see Keeney (32, p. 27-42), and for weighting see Keeney (33). The end result of these actions is a (linear) value function of the form

$$u(x_1, \ldots, x_N) = \sum_{i=1}^{N} k_i u_i(x_i) \tag{3}$$

where $u_i, i = 1, \ldots, N$, is the utility function provided by the measurements of the individual objectives, $k_i, i = 1, \ldots, N$, is the weight on the i^{th} utility function (with $\sum_{i=1}^{N} k_i = 1$, and $x_i, i = 1, \ldots, N$, is alternative X's impact level for attribute i.

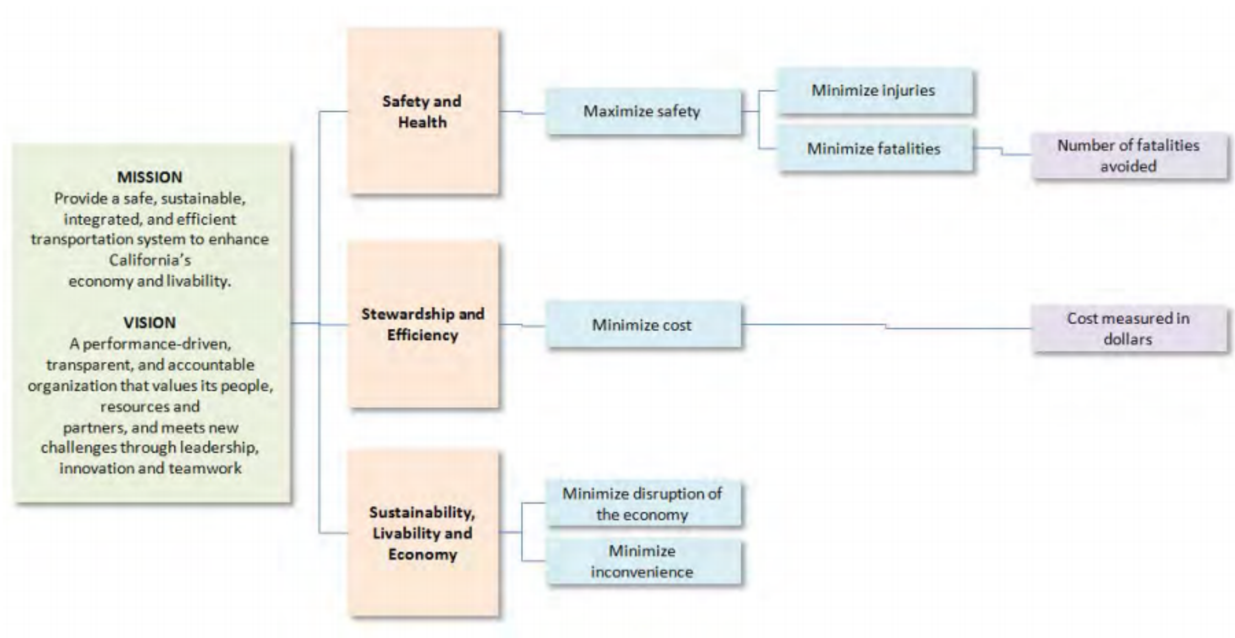

Figure 7: From (31, p. 9), Figure 4.1, an objectives hierarchy for the Caltrans Project. The left-hand side provides the strategic objective, which has 3 fundamental objectives in orange. The lowest-level objectives have attributes (in purple), though not all are pictured.

Equation (3) can be utilized to score a single combat outcome by registering different characteristics of the simulation in terms of the attributes' utility functions. An overall utility value of $0 \leq u(x_1, \ldots, x_N) \leq 1$ is the result, which could be utilized to determine a distance away from completing all combat objectives.

3.2.4 Linear Program

In this subsection, we describe how a linear program (LP) can be utilized for the assessment of simulated combat.[1] In the previous alternatives, the methodologies provided assessments for the current time-step retroactively; that is, both sides commit assets (allocation) and then combat is resolved (adjudication) before the simulation conducts assessment of the resolved combat. Optimization requires a different

[1]For an introduction to Linear Programming, see (34) or (35)

perspective, since one cannot optimize the past.

Here, we apply linear programming to evaluate the potential progress of friendly forces toward combat objectives, in search of an optimal allocation for the next time-step. We demonstrate an array of archetypal objectives and constraints which may be used in an applied LP to simulated combat. Consider the following sample formulation:

Objective

The following linear program optimizes the combat assessment of a particular time step in a simulation of war. The decision space is the allocation of assets to particular missions. The decision variables reflect this. The objective function includes the objectives with relative priorities given by the corresponding coefficients. Each of the objectives is linearly scaled to be between 0 and 1. Therefore, the achievement of a specific allocation in terms of an individual objective is provided as a proportion by the value of that decision variable in the objective function.

Assumptions

We assume that there are finitely many assets and missions, so that the set of assets-to-missions combinations is also finite, although it may be large. We assume also that we can approximate fairly well how the enemy will allocate their assets and to which missions they will allocate these assets, given a particular friendly allocation. Lastly, we assume that all friendly assets available for the next time step can be used in any of the possible missions for the next time step.

Sets

- M - the set of sets of friendly missions available for the next time step
- A - the set of sets of potential friendly assets allocations for the next time

step

Parameters

The following linear program utilizes the parameters

- a_{1mj} - The number of enemy aircraft flown per enemy sortie given friendly mission set $m \in M$ and asset set $j \in J$
- c_1 - The total number of enemy aircraft available for the next time step, regardless of friendly allocation
- a_{2mj} - The average number of enemy aircraft that can take flight and land at the runways included in the enemy mission set, given friendly mission set $m \in M$ and asset set $j \in J$, for the next time step
- c_2 - The total number of enemy runways available and included in the enemy's mission set for the next time step, regardless of friendly allocation
- a_{3mj} - The average amount of fuel required for the enemy's sorties to be flown in the next time step, given friendly mission set $m \in M$ and asset set $j \in J$
- a_{4mj} - The total number of enemy sorties to be flown in the next time step, given friendly mission set $m \in M$ and asset set $j \in J$
- c_5 - The goal number of enemy sorties which will denote achievement of total air superiority
- c_6 - The maximum possible number of enemy sorties for the next time step. This is equivalent to $max\{c_1/a_{1mj} : m \in M, j \in J\}$
- a_5 - The total number of enemy F-35's destroyed from the beginning of the scenario to the end of the most current time step
- a_6 - The total amount of enemy fuel in their storage and supply chain at the beginning of the scenario

Decision Variables

The following linear program utilizes the decision variables

- x_1 - The level of air superiority achieved
- x_2 - The level of the objective *destroy 100 F-35s* achieved
- x_3 - The level of the objective *cripple enemy fuel supply* achieved
- x_{4mj} - The number of enemy sorties associated with the selection of their next time step's mission set
- x_{5mj} - The approximate number of enemy F-35's to be destroyed in the next time step, given both friendly and enemy allocations.
- x_{6mj} - The remaining enemy fuel supply (storage plus delivery) available after applying both friendly and enemy allocations in the next time step

Formulation

Below is the combined formulation for a sample objective function and objectives. The constraints relay how each objective is defined.

$$
\begin{aligned}
max \quad & w_1x_1 + w_2x_2 + w_3x_3 && && (4)\\
s.t \quad & x_4 \leq \frac{c_{1mj}}{a_{1mj}} && \forall m \in M, j \in J && (5)\\
& x_4 \leq a_{2mj}c_{2mj} && \forall m \in M, j \in J && (6)\\
& x_4 \leq a_{3mj}a_{4mj} && \forall m \in M, j \in J && (7)\\
& x_1 \leq 1 - \frac{x_{4mj} - c_5}{c_6 - c_5} && \forall m \in M, j \in J && (8)\\
& x_1 \leq 1 && && (9)\\
& x_2 \leq \frac{a_5 + x_{5mj}}{100} && \forall m \in M, j \in J && (10)\\
& x_2 \leq 1 && && (11)\\
& x_3 = 1 - \frac{x_{6mj}}{a_6} && \forall m \in M, j \in J && (12)
\end{aligned}
$$

$$x_1, x_2, x_3, x_{4mj}, x_{5mj}, x_{6mj} \geq 0 \qquad \forall m \in M, j \in J \qquad (13)$$

Objective function (4) contains sample weights on the three combat objectives. It is not necessary that these weights be between 0 and 1, nor that they sum to 1. However, doing so means that the interpretation of the objective function's value for any allocation is roughly an overall assessment between 0 and 1. Constraints $5-9$ represent the first objective, *achieve air superiority.* We cannot become logistically infeasible by flying more aircraft than we have (Constraint (5)), by flying and landing more aircraft than the enemy runways can handle (Constraint (6)), or by using more fuel than the enemy has available (Constraint (7)). With the goal of getting enemy sorties below a_4, Constraint (8) yields increasing value for lower quantities of enemy sorties. Pairing this constraint with Constraint (9), the maximization objective sets $x_1 = min\{1, 1 - \frac{x_4 - a_4}{c_4 - a_4}\}$.

Constraints (10) and (11) reference the second objective, *destroy 100 F-35's.* Equivalent to objective 1, these constraints set $x_2 = min\{1, \frac{a_5 + x_5}{100}\}$.

Lastly, Constraint (12) records the achievement of depleting enemy fuel. We subtract from unity since decreasing quantities yield a more desirable result. This objective does not need a secondary constraint to cap x_3 at 1 because x_6 is naturally bounded above by a_6 and is constrained below in Constraint (13) by 0.

The above linear program demonstrates a few basic components of applying linear programming to simulated combat assessment. First of all, in defining the objectives to be between 0 and 1, and subsequently weighting these objectives in equation 4 such that the weights sum to unity, we have an objective value that is be easily intelligible – the objective value for any mission set is the projected *scenario-level assessment* for the next time step. Second, this linear program contains one abstract objective and two direct, or concrete, objectives. The abstract objective *obtain air superiority* is systematically broken down into observable components. These components are by no means wholly representative or optimal, but rather comprise a unique manner by which to represent this abstract objective.

The first objective also handles increasing levels of achievement with decreasing values. Note that a maximum value for the domain of the decision variable is required (i.e. c_4). Here, and with the second objective, we demonstrate how to effectually use multiple constraints to set a decision variable equal to a minimum so that achievement of an objective is capped when a decision variable reaches a certain value. Lastly, the third objective is straightforward in its constraints. Here, we demonstrate the simple application of an objective with decreasing preference.

We end this subsection by noting that the linear programming presented here is the most basic form of mathematical programming. Dynamic programming, non-linear programming, and mixed-integer (or integer) programming are variants of linear programming which may be useful for simulated combat assessment. However, we present a basic linear programming example and concept to demonstrate the pros and cons of mathematical programming.

The next chapter introduces a small application problem and individually utilizes each of the four methodologies presented here to demonstrate their construction in action. The subsequent chapter amalgamates the information from this chapter and the next to provide an evaluation of the alternatives in reference to the value hierarchy in Figure 2.

IV. Test Application

In this chapter, we demonstrate how the four assessment alternatives could operate in a simple simulation of combat via a variation of Dresher's Tactical Air War game from Berkovitz and Dresher (36). The Combat Effectiveness & Combat Vulnerability method evaluates individual mission areas to determine the current friendly and enemy abilities to fight back against their adversary. We will demonstrate that the application provides a pessimistic application. Due to its lack of determining an enemy allocation, the enemy may reuse its forces for each mission. The Bayesian network approach requires a heavy amount of pre-processed data. We provide notional data, which could be generated by either expert opinion or historical/simulation data. This approach predicts an expected assessment for the next time-step based on the conditional probabilities between enemy and friendly force allocations. We then present the Value-Focused Thinking Objectives Hierarchy, which closely mirrors the JCS's strategy-to-task framework. This methodology utilizes an additive model based on the achievement of individual campaign objectives. Lastly, we present a linear program for Dresher's game. This alternative optimizes the next-step's assessment in terms of notional objectives, utilizing all logistically friendly and enemy allocations as inputs.

Before presenting examples of each alternative's application, we first distill the game to a point where each assessment methodology can evaluate a hypothetical allocation and either recommend a next allocation of friendly assets and/or comment on potential friendly achievement for the next time-step. We utilize this demonstration in the evaluation of the four alternatives and our created alternative in Chapter V.

4.1 Dresher's Game

Dresher's Tactical Air War Game consists of a series of maneuvers (a strike) between two opposing forces - Blue and Red. The original game assumes each side to have a fixed number of two generic types of aircraft to allocate to missions: bomber and fighter. We label these initial quantities of these aircraft B, F, β, and ϕ for friendly bombers and fighters, and enemy bombers and fighters, respectively.

While the bomber type can be used in "either counter-air or ground support roles," the fighter type can be used "in the air defense or ground support roles" (36, p. 1). Friendly bombers can attack either enemy bomber fields or enemy fighter fields as part of the counter-air role (36, p. 2), while friendly fighters can prevent either enemy bombers or fighters from reaching their targets (36, p. 3). Berkovitz and Dresher assume that each side knows their own and the opponent's fleet size, but does not know how the opponent will allocate their bombers and fighters until after a strike is completed (36, p. 3). They also assume that enemy losses due to "accidents and ground defenses are small," and should be considered negligible, as should any planes lost in air defense and ground support roles (36, p. 5).

In combat adjudication, several additional values are important. Enemy fighters allocated to air defense will reduce friendly bombers allocated to counter-air missions. Since the enemy does not know the friendly force's allocation, enemy fighters do not distinguish between friendly bombers attacking enemy fighter bases and those attacking enemy bomber bases (36, p. 4). Let x be the number of Blue bombers allocated to counter-air missions, and let μ be the number of Red fighters allocated to air defense missions. Also let c be a constant defining the *air defense potential*, or the effectiveness of Blue air defense aircraft. Then the number of Blue bombers reaching Red air bases is $x - c\mu$, unless $c\mu > x$, in which case no Blue bombers reach their destination (36, p. 4). When Blue bombers reach Red airfields, we assume that each bomber can destroy b_1 Red bombers and b_2 Red fighters. Any Blue bombers that fail to penetrate Red defenses are not destroyed, but instead return to base (36, p. 5). For Red, we let ξ, e, d_1, and d_2 be analogous to x, c, b_1, and b_2, respectively, and let u be the Blue equivalent to Red's μ.

The objective of Dresher's game is to support ground operations (36, p. 6). Note that the objective *support ground operations* is abstract, and therefore must be interpreted. For Dresher and Berkovitz, the value of the objective is

$$M = \sum_{i=1}^{N} [(B_i + F_i - x_i - u_i) - (\beta_i + \phi_i - \xi_i - \mu_i)] \tag{14}$$

where i indicates the strike number. The interpretation of Equation (14) is that the payoff for Blue, M, is equal to the sum of all strike's payoffs to Blue. One strike's payoff is the difference between Blue and Red ground support sorties, which are given by the first and second set of parenthetical terms of Equation (14). Note that an assessment methodology can make implicit assumptions. Equation (14) assumes that "bombers and fighters are equally effective in the ground support role" (36, p. 6).

Dresher and Berkovitz's assessment methodology is only one possible method for quantifying success in this simple simulation of combat. Even given the same abstract objective *support ground operations*, other methodologies may define a distinct assessment relationship with the operational environment. For example, the VFT and LP alternatives would break down this objective into sub-objectives, which may or may not produce an equivalent value to M. In any case, the assessment methodologies we present provide an accrued benefit after each time step, given that this combat simulation is sequential. At the end of each time step, both Red and Blue have a new number of bombers and fighters available for the next time step's allocation. The original game terminates after a "predetermined number of strikes" (36, p. 6), defined as N.

For all assessment methodologies, some aspects of combat remain equal. The Dresher game assumes that assistance to friendly ground forces "can be measured by the difference between [friendly] ground support sorties and [enemy] ground support sorties" (36, p. 6). Note that this assumes equal efficacy of the bombers and fighters. In the original game, the payoff, or benefit, to friendly forces is the sum of these scores across all strikes (36, p. 7). We have now provided sufficient baseline detail of the game to demonstrate how to incorporate the four assessment methodologies. A summary of the structure of attack is provided in Figure 8. Some of the variables in this figure have been introduced. The remaining will be introduced in the adjudication below.

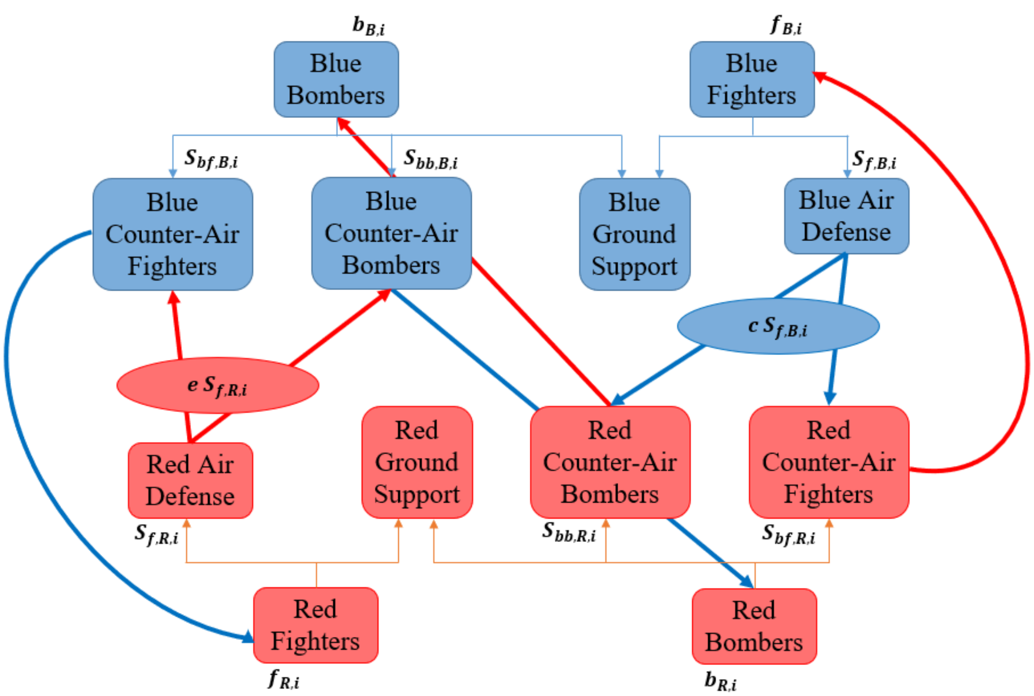

Figure 8: Dresher's Game Objectives Hierarchy

4.1.1 Dresher's Game 1^{st} Time-Step Adjudication

Now that we have introduced the basic rules and definitions for Dresher's Game, consider the following initialization for our adjudication. In the following four sections, assume that Blue starts with $f_{B,1} = 10$ fighters and $b_{B,1} = 13$ bombers, while Red starts with $f_{R,1} = 10$ fighters and $b_{R,1} = 15$ bombers. Adjudication for the CE/CV methodology will use BEAM's method. For all other methodologies, let the two *air defense potentials* be $c = e = 0.5$, and let $b_1 = d_1 = 1$ and $b_2 = d_2 = 2$. Of note, Blue has $11 * 11 = 121$ ways to allocate its force in the first time step, while Red has $14 * 9 = 126$ ways to allocate its force in the first time step. There are therefore $15,246$ possible allocations for the first time step. Taking a cue from Dresher and Berkovitz's analysis of optimal allocations, assume that Blue allocates all bombers to counter-air – $S_{bb,B,1} = 1$ bomber to attack Red fighter bases, $S_{bf,B,1} = 12$ bombers to attack Red bomber bases – and allocates all $S_{f,B,1} = 10$ fighters to air defense missions. Assume analogously that Red allocates all bombers to counter-air – $S_{bb,R,1} = 1$ bomber to attack Blue fighter bases, $S_{bf,R,1} = 14$ bombers to attack Blue bomber bases – and all $S_{f,R,1} = 10$ fighters to air defense (36, p. 10 – 15).

We now adjudicate the outcome from the above allocation. While Blue allocates 13 bombers to attack red bases, Red allocates its 10 fighters to air defense. Since $e = 0.5$, only 5 of the bombers are stopped from reaching their destination. There are $\binom{13}{5} = 1287$ combinations of Blue's bombers that the Red fighters can stop. 495 of these combinations include the one bomber attacking a Red fighter base. Therefore, let a random draw between [0,0.385] denote that Red has stopped Blue's bomber from attacking a Red fighter base. Similarly, let a random draw in (0.385,1] denote that this bomber has not been stopped. Our random draw is 0.110, and so this bomber returns to Blue's bomber bases. Additionally, four Blue bombers attacking Red's bomber bases return to Blue's bomber bases.

Before adjudicating further, we must know how many Red fighters are returned to their bases. While Red allocates its 15 bombers to attack Blue bases, Blue deploys its 10 fighters to air defense. Again, since $c = 0.5$, only 5 of the bombers are stopped from reaching their destination. There are then $\binom{15}{5} = 3003$ different ways that the Red bombers reach Blue's bases. In 1001 of these, the Red bomber attacking the Blue fighter base does not reach its destination. So, we let a random draw between [0,0.333] denote that Blue has stopped Red's bomber from attacking a Blue fighter base. Similarly, let a random draw in (0.333,1] denote that this bomber has not been stopped. Our random draw is 0.960, and so this bomber successfully reaches the Blue fighter base. Additionally, five Blue bombers attacking Red's bomber bases return to Blue's bomber bases.

Because there are $\binom{13}{10} = 286$ adjudication outcomes for this interaction, we choose one randomly. Since 220 of these combinations have in them the Blue bomber attacking a Red fighter base, we let a random draw in [0,0.770] denote that Red has stopped this bomber. Similarly, a random draw in (0.770, 1] denotes that Red has not stopped this bomber. A random draw with a random seed produced 0.965, and so we determine that Red does not stop the bomber attacking one of its fighter bases. Instead, it stops 10 of the 12 bombers attacking its bomber bases.

During the attack, we assume that the bombers return to their respective bases before the opponent's bombers reach those bomber bases. For Blue and Red, five bombers each are located at their side's bases. As a result of combat, Red destroys up to two Blue fighters and up to nine Blue bombers. Blue destroys up to eight of Red's bombers. At the end of this time step, Red has $b_{R,2} = 10$ bombers and $f_{R,2} = 10$ fighters, and Blue has $b_{B,2} = 8$ bombers and $f_{B,2} = 8$ fighters.

In the following four sections, we apply each of the assessment methodologies to a single time-step of combat to illustrate their interaction with a simulated combat

Table 2: Initialization and First Time-Step Parameters for Dresher's Game Applications

Definition	Parameter	Value
Blue Air Defense Potential	c	0.5
Red Air Defense Potential	e	0.5
Blue Bomber-Bomber Potency	b_1	1
Red Bomber-Bomber Potency	d_1	1
Blue Bomber-Fighter Potency	b_2	2
Red Bomber-Fighter Potency	d_2	2
Initial Blue Bombers	$b_{B,1}$	13
Initial Red Bombers	$b_{R,1}$	15
Initial Blue Fighters	$f_{B,1}$	10
Initial Red Fighters	$f_{R,1}$	10
2^{nd} Time-Step Blue Bombers	$b_{B,2}$	8
2^{nd} Time-Step Red Bombers	$b_{R,2}$	10
2^{nd} Time-Step Blue Fighters	$f_{B,2}$	8
2^{nd} Time-Step Red Fighters	$f_{R,2}$	10
Blue Bombers Attacking Red Bombers - 1^{st} Time-Step	$S_{bb,B,1}$	12
Blue Bombers Attacking Red Fighters - 1^{st} Time-Step	$S_{bf,B,1}$	1
Blue Fighters Defending Red Bombers - 1^{st} Time-Step	$S_{f,B,1}$	10
Red Bombers Attacking Blue Bombers - 1^{st} Time-Step	$S_{bb,R,1}$	14
Red Bombers Attacking Blue Fighters - 1^{st} Time-Step	$S_{bf,R,1}$	1
Red Fighters Defending Blue Bombers - 1^{st} Time-Step	$S_{f,R,1}$	10

context. Throughout the methodologies, we reference the parameter values in table 2.

4.2 Combat Effectiveness/Combat Vulnerability in Dresher's Game

For the CE/CV assessment methodology, the overall objective is predetermined – maximize combat effectiveness. In this application, we break up the first time-step into 4 *threads*, rather than the 16 in BEAM. Although we only perform one time-step, we set the probability threshold $\rho = 0.70$ for resulting in either a win or loss, or remaining unresolved. However, we will let $\Delta = 0.25$ define the size of the 5 states: Enemy Supremacy, Enemy Superiority, Contested, Friendly Superiority, and Friendly Supremacy.

For this alternative, we use BEAM's adjudication, rather than that of Dresher's game, so that the cross of CE and CV can provide a joint distribution. As this is the initial time-step, we only have one thread (28, p. 14). The allocation was previously given, so that we have three missions per side within a singular mission group. The first two are offensive: friendly bombers attack bomber bases and fighter bases. The conditional probability tables (CPTs) for these respective missions are in Tables 3 and 4. The defensive mission is the fighters defending attacking bombers, which has a corresponding CPT in Table 5. For simplicity, we assume that the CPTs are equivalent from each side's perspective.

Table 3: CPT for Enemy Bombers Defending Against Friendly Bombers in a Friendly Offensive Mission

Def:Off	Targeted Asset Quantity Bins						
Ratio	[0.0]	(0, 0.2]	(0.2,0.5]	(0.5,0.6]	(0.6,0.9]	(0.9,1.0]	[1.0]
0:1	1	0	0	0	0	0	0
1:2	0	0.42	0.29	0.14	0.09	0.06	0
1:1	0	0.12	0.21	0.31	0.22	0.14	0
2:1	0	0.06	0.1	0.15	0.29	0.4	0
1:0	0	0	0	0	0	0	1

Table 4: CPT for Enemy Fighters Defending Against Friendly Bombers in a Friendly Offensive Mission

Def:Off	Targeted Asset Quantity Bins						
Ratio	[0.0]	(0.0,0.25]	(0.25,0.4]	(0.4,0.6]	(0.6,0.85]	(0.85,1.0]	[1.0]
0:1	1	0	0	0	0	0	0
1:2	0	0.45	0.3	0.12	0.1	0.03	0
1:1	0	0.13	0.22	0.35	0.28	0.02	0
2:1	0	0.02	0.08	0.13	0.22	0.55	0
1:0	0	0	0	0	0	0	1

Table 5: CPT for Enemy Bombers Defending Against Friendly Fighters in a Friendly Defensive Mission

Def:Off Ratio	Targeted Asset Quantity Bins [0.0]	(0.0,0.2]	(0.2,0.5]	(0.5,0.7]	(0.7,0.8]	(0.8,1.0]	[1.0]
0:1	1	0	0	0	0	0	0
1:2	0	0.51	0.23	0.13	0.08	0.05	0
1:1	0	0.06	0.29	0.36	0.22	0.07	0
2:1	0	0.03	0.08	0.11	0.3	0.48	0
1:0	0	0	0	0	0	0	1

Since Blue attacks Red's fighters with just one bomber, we have a defense-offense ratio of 10:1, which is closest in ratio to 1:0. However, there are still bombers attacking, which cannot be neglected. So, we round up to the ratio 2:1, and so Red fighters end the time-step with a health distribution of $[0.02, 0.08, 0.13, 0.22, 0.55]$ over the bins $[0, 0.25, 0.4, 0.6, 0.85, 1]$. Also, Blue attacks Red's bombers with a ratio of 15:12, which is closest to the ratio of 1:1. The resulting Red bombers will then have a health distribution of $(\frac{14}{15})[0.12, 0.21, 0.31, 0.22, 0.14]$ over the bins $[0, 0.2, 0.5, 0.6, 0.9, 1]$. Lastly, Blue fighters defend against Red bombers with a ratio of 1:10, which we again round up to 1:2. So, Red bombers in this mission end the time-step with a health distribution of $(\frac{1}{15})[0.03, 0.08, 0.11, 0.3, 0.48]$ over the bins $[0, 0.2, 0.5, 0.7, 0.8, 1]$. Similarly, Blue's fighters defending against Red bombers in a Red offensive mission of ratio 2:1 leaves Blue's fighters with a health distribution of $[0.02, 0.08, 0.13, 0.22, 0.55]$ over the bins $[0, 0.25, 0.4, 0.6, 0.85, 1]$. Since we assume equivalent CPTs from both sides, the Red bomber attack on Blue bombers results in a Blue bomber health distribution of $(\frac{12}{13})[0.12, 0.21, 0.31, 0.22, 0.14]$ over the bins $[0, 0.2, 0.5, 0.6, 0.9, 1]$. In addition, the Blue bombers attacking Red fighters end the time-step with a health distribution of $(\frac{1}{13})[0.03, 0.08, 0.11, 0.3, 0.48]$ over the bins $[0, 0.2, 0.5, 0.7, 0.8, 1]$. These results are summarized in Table 6.

We now aggregate the two missions for each side's bombers. For the Red bombers, the two distributions sum to approximately [0.114, 0.201, 0.293, 0.072, 0.088, 0.084,

Table 6: Red-Blue Def:Off Ratio and Resulting Health Distributions

Action	Actual Off:Def Ratio	Rounded Def:Off Ratio	Asset	Health Distribution	Bins
Blue Atk Red F	10:1	2:1	Red F	$[0.02, 0.08, 0.13, 0.22, 0.55]$	$[0, 0.25, 0.4, 0.6, 0.85, 1]$
Blue Atk Red B	15:12	1:1	Red B	$14/15[0.12, 0.21, 0.31, 0.22, 0.14]$	$[0, 0.2, 0.5, 0.6, 0.9, 1]$
Blue Def Red B	1:10	1:2	Red B	$1/15[0.03, 0.08, 0.11, 0.3, 0.48]$	$[0, 0.2, 0.5, 0.7, 0.8, 1]$
Red Atk Blue F	10:1	2:1	Blue F	$[0.02, 0.08, 0.13, 0.22, 0.55]$	$[0, 0.2, 0.5, 0.6, 0.9, 1]$
Red Atk Blue B	14:13	1:1	Blue B	$12/13[0.12, 0.21, 0.31, 0.22, 0.14]$	$[0, 0.2, 0.5, 0.6, 0.9, 1]$
Red Def Blue B	1:10	1:2	Blue B	$1/13[0.03, 0.08, 0.11, 0.3, 0.48]$	$[0, 0.2, 0.5, 0.7, 0.8, 1]$

Table 7: Discrete-Uniform Quartile Bins for Red and Blue Assets

Asset	Distribution Bins
Blue Bombers	[0, 0.405, 0.564, 0.782, 1]
Blue Fighters	[0, 0.623, 0.864, 0.932, 1]
Red Bombers	[0, 0.403, 0.563, 0.779, 1]
Red Fighters	[0, 0.623, 0.864, 0.932, 1]

0.147] over the bins [0, 0.2, 0.5, 0.6, 0.7, 0.8, 0.9, 1]. The Blue bombers' distributions sum to approximately [0.113, 0.200, 0.290, 0.072, 0.091, 0.086, 0.148] over the bins [0, 0.2, 0.5, 0.6, 0.7, 0.8, 0.9, 1]. The discrete-uniform distribution of the Red bombers has the bins [0, 0.403, 0.563, 0.779, 1], while the discrete-uniform distribution of the Blue bombers has the bins [0, 0.405, 0.564, 0.782, 1]. Meanwhile, both Blue and Red fighters have the same discrete-uniform distribution, given by the bins [0, 0.623, 0.864, 0.932, 1]. These results are summarized in Table 7.

We now calculate the CV and CE for each of the Blue and Red perspectives. We first sum the two Blue discrete-uniform distributions to receive $\mathbf{U}(CV_B)$, which has bin boundaries [0, 0.461, 0.677, 0.889]. The Red discrete-uniform distributions sum to yield $\mathbf{U}(CV_R)$, which has similar bin boundaries [0, 0.460, 0.677, 0.889].

To calculate CE, we first calculate the discrete-uniform distribution for Blue's offensive CE, which is equivalent to Red's CV distribution. Next, we calculate Blue's defensive CE, which is calculated by taking the complement of its defensive assets' health distribution and then placing it in four equal-probability bins. The result is the discrete-uniform distribution $\mathbf{U}(CE_B)$, which is defined by the bin boundaries

[0.111, 0.464, 0.707, 1]. We analogously calculate Red's CE, which has a resultant discrete-uniform distribution $\mathbf{U}(CE_R)$ defined by the bin boundaries [0, 0.114, 0.489, 0.779, 1].

Table 8: CE $\times$ CV States at End of First Time-Step

Blue		**Combat Vulnerability (CV)**		
		$[0, 0.25]$	$(0.25, 0.75)$	$[0.75, 1]$
Combat Effectiveness (CE)	$[0.75, 1]$	Friendly Supremacy S_5 **0.001**	Friendly Superiority S_4 **0.092**	Contested S_3 **0.12**
	$(0.25, 0.75)$	Contested S_3 **0.003**	Contested S_3 **0.189**	Enemy Superiority S_2 **0.247**
	$[0, 0.25]$	Contested S_3 **0.002**	Contested S_3 **0.15**	Enemy Supremacy S_1 **0.196**

Red		**Combat Vulnerability (CV)**		
		$[0, 0.25]$	$(0.25, 0.75)$	$[0.75, 1]$
Combat Effectiveness (CE)	$[0.75, 1]$	Friendly Supremacy S_5 **0.002**	Friendly Superiority S_4 **0.121**	Contested S_3 **0.159**
	$(0.25, 0.75)$	Contested S_3 **0.003**	Contested S_3 **0.161**	Enemy Superiority S_2 **0.212**
	$[0, 0.25]$	Contested S_3 **0.003**	Contested S_3 **0.146**	Enemy Supremacy S_1 **0.192**

The cross between Blue's CE and CV and the cross between Red's CE and CV are provided in table 8. If we assume that the scenario started in a contested state, S_3, then neither side has won nor lost this scenario. We can see this by summing the probabilities associated with states S_3, S_4, S_5 – 0.557 for Blue and 0.595 for Red – and by summing the probabilities associated with states S_1, S_2 – 0.443 for Blue and 0.404 for Red. Since none of these summations is larger than $\rho = 0.70$, the scenario remains unresolved.

4.3 Bayesian Networks in Dresher's Game

In this section, we demonstrate the application of Bayesian networks in Dresher's Game by building the network and data tables manually. We first define an objective function, comprised of three notional objectives, which will be pertinent to the BN, VFT, and LP methodologies. We then utilize the conditional probability rules established in Section 3.2.2 in order to construct the joint probability distribution of the Bayesian network for each side's assessment. We utilize the state-space and each vector's associated probability for next time-step's allocation to obtain an expected objected value for each side. We consider this expected value to be the assessment gleaned from this methodology.

Suppose we have three objectives that we would like to utilize for our objective function: (1) *minimize cost*, (2) *minimize enemy capability*, and (3) *maximize friendly capability*. A simple way of tracking a cost objective is in total cost, which makes future actions dependent upon current actions. Suppose that flying a friendly bomber to attack enemy bombers costs $C_{bb} = \$290,000$, that flying a friendly bomber against enemy fighters costs $C_{bf} = \$240,000$, and that flying a friendly fighter to defend against enemy bombers costs $C_f = \$110,000$. The total friendly cost through next time-step can then be represented as the numerator in

$$\frac{C_{bb}S_{bb,B,i} + C_{bf}S_{bf,B,i} + C_f S_{f,B,i} + PC_i}{C_{bb}b_{B,i} + C_f f_{B,i} + PC_i} \tag{15}$$

where PC_i is the previous cumulative cost of the friendly missions flown during the scenario. Note that the denominator is the maximum cumulative spending that can occur during the next time step. In our case, we exclude the term $C_{bf}b_{B,2}$ since $C_{bf} < C_{bb}$. Equation (15) is a standardized value for the cost objective on the domain [0,1]. In order to obtain the cost objective value for Blue during the i^{th} time-step, we need only to subtract Equation (15) from unity, as we wish to minimize the cost. Red's cost objective function is analogous to subtracting Equation (15) from unity, as well.

For the capability objectives, we can create a points system in order to weight the destruction of friendly or enemy assets – $P_b = 3$ points for destroying bombers and $P_f = 2$ points for destroying fighters. This weighting is a notional "relative importance," which here means that bombers are 1.5 times as important as fighters. In application, these weights could be generated using subject matter expert opinion or by running multiple simulations to understand the relative importance of an asset to completing combat goals. Within this points system, we want to know the number of points destroyed (for enemy forces) and the number of points remaining (for friendly forces). To do this for the next time-step's combat, we have to invoke the adjudication portion of combat. We can do this in the objective itself. Note that the number of Red bombers remaining at the beginning of time-step $(i + 1)$ is

$$b_{R,i} - b_1(S_{bb,B,i} - min\{S_{bb,B,i}, eS_{f,R,i}\})$$

where $e = 0.5$ and $b_1 = 1$ are as defined at the beginning of this chapter. Adding in the weighted points system for capability, the enemy capability objective from Blue's perspective is then

$$\begin{aligned} 1 - \frac{1}{P_b b_{R,1} + P_f f_{R,1}} & [P_b(b_{R,i} - b_1(S_{bb,B,i} - min\{S_{bb,B,i}, eS_{f,R,i}\})) \\ & + P_f(f_{R,i} - b_2(S_{bf,B,i} - min\{S_{bf,B,i}, eS_{f,R,i}\}))] \end{aligned} \tag{16}$$

and the friendly capability objective from Blue's perspective is

$$\frac{1}{P_b b_{B,1} + P_f f_{B,1}} \left[P_b(b_{B,i} - d_1(S_{bb,R,i} - min\{S_{bb,R,i}, cS_{f,B,i}\})) \right.$$
$$\left. + P_f(f_{B,i} - d_2(S_{bf,R,i} - min\{S_{bf,R,i}, eS_{f,B,i}\})) \right] \qquad (17)$$

where $c = e = 0.5, b_1 = d_1 = 1$, and $b_2 = d_2 = 2$ are as defined at the beginning of this chapter. Both objectives have once again been standardized, so that the maximum number of points is the amount that each side starts with at the beginning of the scenario. In equation 16, we subtract from unity because we wish to define increasing progress toward minimizing the enemy's capability as closer to one.

With our objectives defined in Equations (15) – (17) to all range over [0,1], we may wish to define an overall objective value which also ranges over [0,1]. Since we have three objectives, constraining these weights w_1, w_2, w_3 to $\sum_{i=1}^{3} w_i = 1$ will accomplish the task. For simplicity, let $w_i = \frac{1}{3}, i = 1, 2, 3$. The overall objective is the inner product of these weights and the three objectives' values. This overall objective is the associated assessment with a particular allocation.

With the objectives defined, we proceed to the Bayesian network for Dresher's game. The network is defined by the following decision variables: $S_{bb,B,i}$, $S_{bf,B,i}$, $S_{f,B,i}$, $S_{bb,R,i}$, $S_{bf,R,i}$, and $S_{f,R,i}$, $i = 1, 2, \ldots, N$, where i corresponds to the time-steps. While the structure of the network remains constant for each time-step, the CPTs may change relating these decision variables from one time-step to the next. We know up front that the number of bombers attacking bombers or fighters are interdependent. We can arbitrarily choose to have $S_{bb,B,i}$ be an independent cause for $S_{bf,B,i}$, and similarly for Red.

From Blue's perspective, we assume that Red's choices are the causes and Blue's choices are the effects. Since fighters defend against bombers, we assume that Blue bomber choices are dependent upon Red's fighter choice, and also that Blue's fighter choice is dependent upon Red's bomber choices. Red's perspective is analogous, maintaining symmetry between the two side's assessments. With these dependencies in mind, we can construct the Bayesian networks as given in Figure 9.

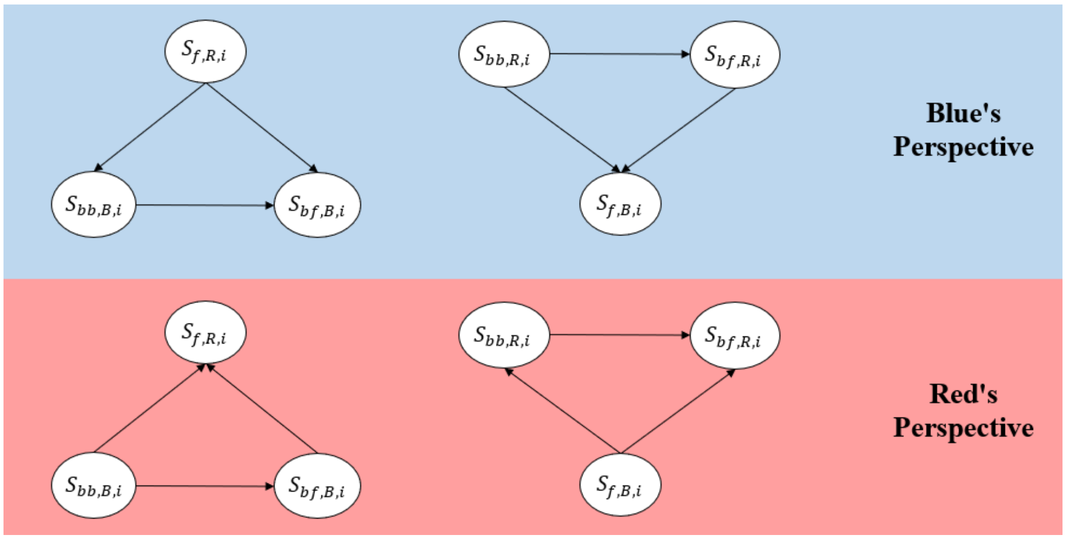

Figure 9: Dresher's Game Bayesian Network

From each side's perspective, we have two marginally independent networks. The joint probability distribution that defines each side's perspective is the product of these two networks. From Blue's perspective, we have

$$\begin{aligned} P(\overrightarrow{S_{B,i}}) =& P(S_{bb,B,i}, S_{bf,B,i}, S_{f,B,i}, S_{bb,R,i}, S_{bf,R,i}, S_{f,R,i}) \\ &= [P(S_{bf,B,i}|S_{bb,B,i}, S_{f,R,i}) \cdot P(S_{bb,B,i}|S_{f,R,i}) \cdot P(S_{f,R,i})] \cdot \\ &\quad [P(S_{f,B,i}|S_{bb,R,i}, S_{bf,R,i}) \cdot P(S_{bf,R,i}|S_{bb,R,i}) \cdot P(S_{bb,R,i})] \end{aligned} \tag{18}$$

and from Red's perspective,

$$\begin{aligned}
P(\overrightarrow{S_{R,i}}) =& P(S_{bb,R,i}, S_{bf,R,i}, S_{f,R,i}, S_{bb,B,i}, S_{bf,B,i}, S_{f,B,i}) \\
=& [P(S_{f,R,i}|S_{bb,B,i}, S_{bf,R,i}) \cdot P(S_{bf,B,i}|S_{bb,B,i}) \cdot P(S_{bb,B,i})] \cdot \\
& [P(S_{bf,R,i}|S_{bb,R,i}, S_{f,B,i}) \cdot P(S_{bb,R,i}|S_{f,B,i}) \cdot P(S_{f,B,i})] \qquad (19)
\end{aligned}$$

Now, assume that we have just ended the first time step's adjudication according to the scenario defined at the beginning of this chapter. The networks in Figure 9 may be interpreted as defining "the probability of the allocation of Blue and Enemy forces, based on historical data." After calculating the probability of a certain state vector, we determine the expected value of next time-step's assessment.

We now build notional conditional probability tables with some desired properties. First, the number of Blue bombers and Red bombers allocated to missions must be less than or equal to the total number of available bombers, $b_{B,i}$ and $b_{R,i}$, $i = 1, 2, \ldots, N$. So, the ordered pair $(S_{bb,B,2}, S_{bf,B,2}) = (5, 6)$, for example, must have an associated probability of 0, since $b_{B,2} = 10$. Several of the probability distributions are bi-modal, as well. This is because we assume it advantageous to allocate either all or none of the aircraft in a category to a single mission. This assumption is supported by the majority of optimal decisions provided by Dresher and Berkovitz (36, p. 12).

The conditional probability tables are provided in Appendix A. Note that the state-space of the Blue Bayesian network's joint distribution has 534,600 state vectors, and Red's joint distribution has 556,600 elements. To obtain these distributions, one needs only to cross the data in the Appendix A tables according to Equations (18) and (19).

To reach an overall assessment we can take the probabilities $P(\overrightarrow{S_{B,i}})$ and $P(\overrightarrow{S_{R,i}})$ and multiply them by their associated overall objective values. Summing over the products yields an expected value, or expected assessment value for the next time-step. When we perform these calculations using the data in Appendix A, we obtain a Blue assessment of about 0.31 and a Red assessment of about 0.25. We can take from the CE/CV approach and set our threshold value $\rho = 0.7$ to be the necessary assessment score of either the friendly or enemy objectives to declare a win/loss/unresolved. Since neither side has an assessment exceeding this value, doing so would mean that the thread remains unresolved.

4.4 Value-Focused Thinking in Dresher's Game

In this section, we continue to utilize the original allocation and adjudication results from Table 2. We also utilize the objectives from the Bayesian network methodology presented in Section 4.3 to construct the objectives hierarchy and the attributes used to score the alternative in terms of the lowest-level objectives. In difference to the Bayesian network methodology, however, we assess the current time-step's adjudication, rather than projecting one time-step into the future.

We have already three means objectives: minimize cost, maximize friendly capability, and minimize enemy capability. We now build an OH that envelops these objectives. Consider that Dresher's Game occurs at the tactical level. From figure 1, this is the lowest level of warfare considered in the strategy-to-task framework. Therefore, a reasonable *strategic objective* could be "support operational-level strategic goals," which are outside the scope of Dresher's Game. Directed more toward tactical-level operations, the *overall fundamental objective* could be to "win the scenario efficiently." As it is meant to be very broad, we need to specify what "winning" means and what "efficiency" means. These two subcategories can be translated into *fundamental objectives* as "destroy enemy forces" and "preserve friendly forces at low cost." Our means objectives fit nicely under these fundamental objectives. "Minimize enemy capability" is a means objective for the "destroy enemy forces" fundamental objective and both "minimize cost" and "maximize friendly capability" are means objectives for the "preserve friendly forces at low cost" fundamental objective. While minimizing cost is an observable means objective, we must specify further the other two means objectives. We can observe the two capability means objectives by observing the number of remaining friendly and enemy bombers and fighters. This completes our objectives hierarchy. Figure 10 contains a graphical representation of the hierarchy.

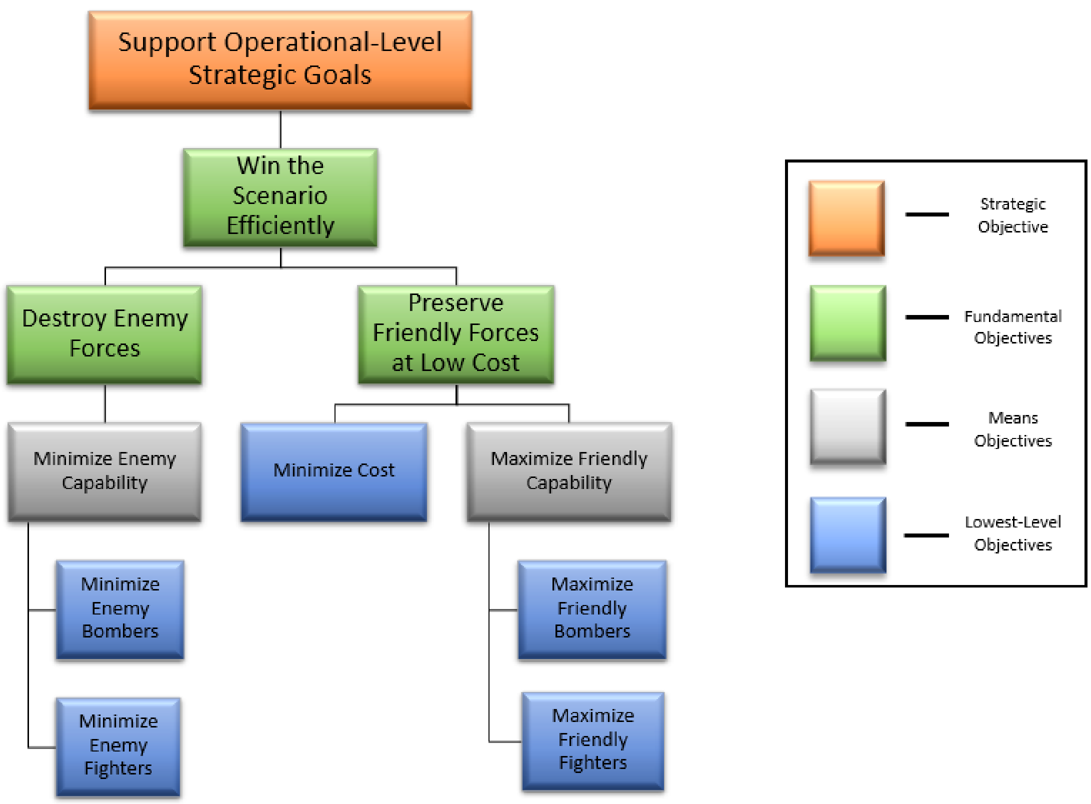

Figure 10: Dresher's Game Objectives Hierarchy

We now construct the *attributes* for the lowest-level objectives. Again, let $C_{bb} = \$290,000$ be the cost of flying a friendly bomber to attack enemy bombers, $C_{bf} = \$240,000$ the cost of flying a friendly bomber against enemy fighters, $C_f = \$110,000$ the cost of flying a friendly fighter to defend against enemy bombers, $b_{B,i}$ be the number of friendly bombers available for the i^{th} time-step's allocation, $f_{B,i}$ the number of friendly fighters available for the i^{th} time-step's allocation, PC_i the previous cumulative cost of the friendly missions flown during the scenario, $P_b = 3$ the number of points for destroying bombers, and $P_f = 2$ the number of points for destroying fighters. Again, let $b_{R,1}, f_{R,1}, b_{B,1}, f_{B,1}$ be the number of Red bombers and fighters and Blue bombers and fighters, respectively, available at the beginning of the scenario.

The attributes for the enemy capability objective from Blue's perspective at the end of time-step 1 are $ECB_{B,1}$ and $ECF_{B,1}$, for bomber and fighter destruction, respectively, which are defined as

$$ECB_{B,1} = 1 - \frac{b_{R,2}}{b_{R,1}} \tag{20}$$

$$ECF_{B,1} = 1 - \frac{f_{R,2}}{f_{R,1}} \tag{21}$$

Similarly, the friendly capability objectives from Blue's perspective at the end of time-step 1 are $FCB_{B,1}$ and $FCF_{B,1}$, for remaining bombers and fighters, respectively, which are defined as

$$FCB_{B,1} = \frac{b_{B,2}}{b_{B,1}} \tag{22}$$

$$FCF_{B,1} = \frac{f_{B,2}}{f_{B,1}} \tag{23}$$

Lastly, the cost objective from Blue's perspective at the end of time-step 1 is $C_{B,1}$, which is defined as

$$C_{B,1} = 1 - \frac{C_{bb}S_{bb,B,1} + C_{bf}S_{bf,B,1} + C_f S_{f,B,1} + PC_0}{C_{bb}b_{B,1} + C_f f_{B,1} + PC_0} \tag{24}$$

although the previous cost, PC_0, is 0.

The Red-perspective objectives hierarchy is the same as provided in figure 10.

In addition, the Red-perspective objectives can be found by analogue to the Blue-perspective objectives in Equations (20) – (24).

Different from the Bayesian network methodology, we now have five weights – one for each of the lowest-level objectives. However, to maintain some uniformity, let $w_{ECB} = w_{FCB} = \frac{2}{15}$, $w_{FCF} = w_{ECF} = \frac{1}{5}$ and $w_C = \frac{1}{3}$. This allows us to keep nearly the same ratio that we integrated via the points system into the Bayesian network methodology. We also may define utility functions for each of these attributes that determine the amount of utility returned for specific attribute achievement levels. For example, the achievement of objectives in the Bayesian network methodology was strictly linear, and so an increase of from 0.1 to 0.2 in the friendly capability objective would result in an increase of utility from 0.1 to 0.2. Consider the cost objective. We know that the value of this objective is a ratio of the cumulative cost to the maximum possible cumulative cost. Therefore, during each successive time-step, it will become increasingly more difficult to get a lower value for this objective. We can offset this property by utilizing an exponential utility function for the cost objective, namely

$$U_C(C_{B,1}) = \frac{1 - e^{-C_{B,1}/0.410}}{1 - e^{-1/0.410}} \tag{25}$$

which has the property that $U_C(0.75) = 0.5$.

The remaining utility functions are as follows:

$$U_{ECB}(x) = x \qquad U_{ECF}(x) = x$$

$$U_{FCB}(x) = x \qquad U_{FCF}(x) = x$$

all of which are defined on $x \in [0, 1]$, and where x is an objective value returned from one of the objectives equations. The overall utility function (and assessment value) for Blue's first time-step is

$$U_{B,1} = \frac{2}{15}\left[U_{ECB}(ECB_{B,1}) + U_{FCB}(FCB_{B,1})\right] \\ + \frac{1}{5}\left[U_{ECF}(ECF_{B.1}) + U_{FCF}(FCF_{B,1})\right] + \frac{1}{3}U_C(C_{B,1}) \tag{26}$$

and Red's overall utility function for its first time-step is the same, with its corresponding analogous inputs.

Inputting the adjudication and initialization parameters from table 2 yields a Blue assessment of about 0.31 and a Red assessment of about 0.38. Since neither value is above the $\rho = 0.7$ threshold, the current scenario remains unresolved.

4.5 Linear Program in Dresher's Game

In this section, we continue to utilize the original allocation, as well as the original adjudication methods. However, before continuing with the operational environment, we need to establish the linear program. The below formulation captures a multi-objective linear program and demonstrates some intermediate if-then constraints which could appear in some contexts. In the LP, we use "friendly-enemy" terminology, rather than the previous "Blue-Red." We do this to demonstrate the interchangeability of the LP from each side's perspective.

To apply linear programming, we can look at optimization over the next time step. Consider the following sample formulation:

Objective

The following linear program (LP) optimizes the combat assessment of the second time-step in the Dresher game outlined in the introduction to this chapter. There are three objectives: minimize cost, minimize enemy capability, and maximize friendly capability. The first of these is direct, while the latter two objectives require some interpretation. Our interpretation is provided below. Each of the objectives is scaled to have a value between 0 and 1. We then apply weights to the individual objective values to provide the relative importance of the objectives and to provide an overall objective value between 0 and 1. As the interpretation of the objective value is most naturally interpreted with increasing preference, we introduce a maximization LP with objective values subtracted from unity.

Assumptions

There are two key assumptions in this LP, beyond those of general logistics feasibility. First, we assume that the adjudication rules outlined in the introduction to this chapter remain valid throughout all time-steps. This allows us to optimize over the next time-step. Second, we assume that the enemy utilizes the same assessment method, objectives, and weights. We utilize this symmetry to calculate the enemy's optimal response to a friendly allocation. This requires pre-processing and storage of the optimal enemy allocation in a table to feed the linear program. The result is that we can make the enemy allocation a set of parameters that get fed in via this table to correspond with friendly allocation. We can then adjudicate the projected combat defined by friendly and enemy allocations and assess the completion of friendly objectives by our LP. In order to reduce the number of subscripts used, we assume that an allocation is referenced by a single index.

Sets

We use a singular set as a collection of indices referencing the enemy allocations:

- J – The set of indices assumed to be contained in the data set containing optimal enemy responses to friendly allocations, which reference individual allocations of attacking bombers to fighters, attacking bombers to bombers, and defending fighters

Parameters

The linear program utilizes the following parameters

- w_1, w_2, w_3 – The weights on objectives 1, 2, and 3, respectively
- $b_{fr,i}$ – The number of friendly bombers at the start of the i^{th} time-step, $i = 1, 2, \ldots, N$
- $b_{en,i}$ – The number of enemy bombers at the start of the i^{th} time-step, $i = 1, 2, \ldots, N$
- $f_{fr,i}$ – The number of friendly fighters at the start of the i^{th} time-step, $i = 1, 2, \ldots, N$
- $f_{en,i}$ – The number of enemy fighters at the start of the i^{th} time-step, $i = 1, 2, \ldots, N$
- c/e – The air defense potential of Blue/Red fighters, which dictates the effectiveness of both friendly and enemy fighters at blocking bombers from reaching their target(s)
- $b_1/b_2/d_1/d_2$ – The number of aircraft that Blue/Red destroys upon successful attack of Red/Blue aircraft
- PC_i – The cumulative previous cost, from the beginning of the scenario through the current time-step, of friendly sorties
- C_{bb} – The cost of flying a single friendly bomber on a mission to attack enemy bombers
- C_{bf} – The cost of flying a single friendly bomber on a mission to attack enemy fighters

- C_f – The cost of flying a single friendly fighter on a mission to defend against enemy bombers
- P_b – The number of "points" gained by destroying an enemy bomber
- P_f – The number of "points" gained by destroying an enemy fighter
- $[a_{bb,en,i}]_j$ – The number of enemy bombers targeting friendly bombers in the enemy's j^{th} allocation in the i^{th} time step, $i = 1, 2, \ldots, N$
- $[a_{bf,en,i}]_j$ – The number of enemy bombers targeting friendly fighters in the enemy's j^{th} allocation in the i^{th} time step, $i = 1, 2, \ldots, N$
- $[a_{f,en,i}]_j$ – The number of enemy bombers defending against friendly bombers in the enemy's j^{th} allocation in the i^{th} time step, $i = 1, 2, \ldots, N$

The initial values of these parameters are provided in table 9. Note that initial values such as P_f and C_{bf} are designed to keep the trade-off space interesting and do not reflect real-world values.

Decision Variables

The linear program utilizes the following decision variables

- $[S_{bb,fr,i}]_j$ – The number of friendly bomber sorties to fly next time-step to attack enemy bombers, given the enemy's j^{th} allocation, $i = 1, 2, \ldots, N$
- $[S_{bf,fr,i}]_j$ – The number of friendly bomber sorties to fly next time-step to attack enemy fighters, $i = 1, 2, \ldots, N$
- $[S_{f,fr,i}]_j$ – The number of friendly fighter sorties to fly next time-step to defend against enemy bombers, $i = 1, 2, \ldots, N$
- $x_{1,j}, x_{2,j}, y_{1,j}, y_{2,j}$ – Proxy decision variables to determine how many enemy fighters will affect friendly bombers and how many enemy

Table 9: Initialization Parameters for Dresher's Game Application of the LP Methodology – After the 1^{st} Time-Step's Adjudication

Parameter	Initialized Value	Parameter	Initialized Value
w_1	1/3	w_2	1/3
w_3	1/3	$b_{fr,1}$	13
$b_{en,1}$	15	$f_{fr,1}$	10
$f_{en,1}$	10	$b_{fr,2}$	8
$b_{en,2}$	10	$f_{fr,2}$	8
$f_{en,2}$	10	c	0.5
e	0.5	b_1	1
d_1	1	b_2	2
d_2	2	PC_1	$\$4,820,000$
C_{bb}	$\$290,000$	C_{bf}	$\$240,000$
C_f	$\$110,000$	P_b	3
P_f	2		

bombers will be affected by friendly fighters, given the enemy's j^{th} allocation

- $z_{1,j}, z_{2,j}, z_{3,j}, z_{4,j}, z_{5,j}, z_{6,j}, z_{7,j}, z_{8,j}$ – Binary decision variables to form if-then constraints, given the enemy's j^{th} allocation

Formulation

Below is the symbolic formulation for this LP. Although there may be different values for our time-step-indexed variables, we solve for a particular $i \in \{1, 2, \ldots, N\}$. We provide the general formulation here, but will solve for $i = 2$. After explication, we provide a solution to the LP, which is equivalent to the assessment of this first time-step utilizing the LP methodology.

$$\begin{aligned} max \qquad & w_1\left(1 - \frac{C_{bb}[S_{bb,fr,i}]_j + C_{bf}[S_{bf,fr,i}]_j + C_f[S_{f,fr,i}]_j + PC_i}{C_{bb}b_{fr,i} + C_f f_{fr,i} + PC_i}\right) \\ & + w_2\left(1 - \frac{P_b(b_{en,i} - b_1([S_{bb,fr,i}]_j - x_{1,j})) + P_f(f_{en,i} - b_2([S_{bf,fr,i}]_j - x_{2,j}))}{P_b b_{en,1} + P_f f_{en,1}}\right) \\ & + w_3\left(\frac{P_b(b_{fr,i} - d_1([S_{bb,en,i}]_j - y_{1,j})) + P_f(f_{fr,i} - d_2([S_{bf,en,i}]_j - y_{2,j}))}{P_b b_{fr,1} + P_f f_{fr,1}}\right) \end{aligned} \tag{27}$$

$$
\begin{aligned}
s.t \quad & [S_{bb,fr,i}]_j + [S_{bf,fr,i}]_j \leq b_{fr} && \forall j \in J && (28)\\
& [S_{f,fr,i}]_j \leq f_{fr,i} && \forall j \in J && (29)\\
& x_{1,j} \leq [S_{bb,fr,i}]_j && \forall j \in J && (30)\\
& x_{1,j} \leq e[a_{f,en,i}]_j && \forall j \in J && (31)\\
& x_{2,j} \leq [S_{bf,fr,i}]_j && \forall j \in J && (32)\\
& x_{2,j} \leq e[a_{f,en,i}]_j && \forall j \in J && (33)\\
& y_{1,j} \leq [a_{bb,en,i}]_j && \forall j \in J && (34)\\
& y_{1,j} \leq c[S_{f,fr,i}]_j && \forall j \in J && (35)\\
& y_{2,j} \leq [a_{bf,en,i}]_j && \forall j \in J && (36)\\
& y_{2,j} \leq c[S_{f,fr,i}]_j && \forall j \in J && (37)\\
& e[a_{f,en,i}]_j - [S_{bb,fr,i}]_j \geq -(ef_{en,i} + b_{fr,i})z_{1,j} && \forall j \in J && (38)\\
& x_{1,j} - [S_{bb,fr,i}]_j \geq -(ef_{en,i} + b_{fr,i})(1 - z_{1,j}) && \forall j \in J && (39)\\
& [S_{bb,fr,i}]_j - e[a_{f,en,i}]_j \geq -(ef_{en,i} + b_{fr,i})z_{2,j} && \forall j \in J && (40)\\
& x_{1,j} - e[a_{f,en,i}]_j \geq -(ef_{en,i} + b_{fr,i})(1 - z_{2,j}) && \forall j \in J && (41)\\
& e[a_{f,en,i}]_j - [S_{bf,fr,i}]_j \geq -(ef_{en,i} + b_{fr,i})z_{3,j} && \forall j \in J && (42)\\
& x_{2,j} - [S_{bf,fr,i}]_j \geq -(ef_{en,i} + b_{fr,i})(1 - z_{3,j}) && \forall j \in J && (43)\\
& [S_{bf,fr,i}]_j - e[a_{f,en,i}]_j \geq -(ef_{en,i} + b_{fr,i})z_{4,j} && \forall j \in J && (44)\\
& x_{2,j} - e[a_{f,en,i}]_j \geq -(ef_{en,i} + b_{fr,i})(1 - z_{4,j}) && \forall j \in J && (45)\\
& c[S_{f,fr,i}]_j - [a_{bb,en,i}]_j \geq -(cf_{fr,i} + b_{en,i})z_{5,j} && \forall j \in J && (46)\\
& y_{1,j} - [a_{bb,en,i}]_j \geq -(cf_{fr,i} + b_{en,i})(1 - z_{5,j}) && \forall j \in J && (47)\\
& [a_{bb,en,i}]_j - c[S_{f,fr,i}]_j \geq -(cf_{fr,i} + b_{en,i})z_{6,j} && \forall j \in J && (48)\\
& y_{1,j} - c[S_{f,fr,i}]_j \geq -(cf_{fr,i} + b_{en,i})(1 - z_{6,j}) && \forall j \in J && (49)\\
& c[S_{f,fr,i}]_j - [a_{bf,en,i}]_j \geq -(cf_{fr,i} + b_{en,i})z_{7,j} && \forall j \in J && (50)
\end{aligned}
$$

$$y_{1,j} - [a_{bf,en,i}]_j \geq -(cf_{fr,i} + b_{en,i})(1 - z_{7,j}) \qquad \forall j \in J \quad (51)$$

$$[a_{bf,en,i}]_j - c[S_{f,fr,i}]_j \geq -(cf_{fr,i} + b_{en,i})z_{8,j} \qquad \forall j \in J \quad (52)$$

$$y_{1,j} - c[S_{f,fr,i}]_j \geq -(cf_{fr,i} + b_{en,i})(1 - z_{8,j}) \qquad \forall j \in J \quad (53)$$

$$z_{2k-1,j} + z_{2k,j} = 1 \qquad k = 1, 2, 3, 4, \forall j \in J \quad (54)$$

$$[S_{bb,fr,i}]_j, [S_{bf,fr,i}]_j, [S_{f,fr,i}]_j \in Z^+ \cup \{0\} \qquad \forall j \in J \quad (55)$$

$$x_{1,j}, x_{2,j}, y_{1,j}, y_{1,j} \geq 0 \qquad \forall j \in J \quad (56)$$

$$z_{1,j}, z_{2,j}, z_{3,j}, z_{4,j}, z_{5,j}, z_{6,j}, z_{7,j}, z_{8,j} \in \{0, 1\} \qquad \forall j \in J \quad (57)$$

The objective function (27) is broken into three components. The first of these fractions defines the *minimize cost* objective. After the adjudication portion of any time-step, that time-step's sorties have been paid for. Therefore, the cost referred to is that of the next time-step. The total cumulative cost of past and the next time-step is therefore the cost of each planned sortie for the next time-step plus the previous cost. To standardize that value, we divide by the maximum possible cumulative cost at the end of the next time-step. Here, we assume $C_{bb} \geq C_f$, since it is true for our particular case. Therefore, the maximum possible cumulative expenditure for the end of the next time-step is the more expensive of the two bomber sorties multiplied by all available bombers plus the cost of flying all available fighters, plus the previous cost of the sorties during this scenario. Note that a reduced cost produces a higher cost objective value, which is desirable for this maximization problem.

The second fraction of the objective function defines the *minimize enemy capability* objective. The numerator is comprised of two portions: the weighted *points* of the remaining enemy bombers after the friendly allocation, and the weighted *points* of the remaining enemy fighters after the friendly allocation. The proxy decision variables x_1 and x_2 are constrained such that $x_1 = min\{S_{bd,fr,i}, cS_{f,en,i}\}$ and $x_2 = min\{S_{bf,fr,i}, cS_{f,en,i}\}$. Such constraints ensure the adjudication rules of the game. $eS_{f,en,i}$ will effectively block friendly bombers, after which, the remaining bombers will destroy b_1 bombers and b_2 fighters. Note that a reduced enemy force produces a higher enemy capability objective value, which is desirable for this maximization problem.

The third portion of the objective function is analogous to the second portion, but instead focuses on remaining friendly forces. We do not subtract from unity for this objective because we wish to maximize it.

The constraints are constructed in groups. Constraints (28) and (29) limit friendly allocations to the number of respective friendly bombers and fighters at the beginning of the next time-step.

Constraints 38 – 54 set the values for $x_{1,j}, x_{2,j}, y_{1,j}$, and $y_{2,j}$ by using four sets of two if-then constraint sets. For example, constraints 38 and 39 create the logical constraint $(e[a_{f,en,i}]_j \geq [S_{bb,fr,i}]_j) \Rightarrow (x_{1,j} \geq [S_{bb,fr,i}]_j)$. Similarly, Constraints 40 and 41 create the logical constraint $([S_{bb,fr,i}]_j \geq e[a_{f,en,i}]_j) \Rightarrow (x_{1,j} \geq e[a_{f,en,i}]_j)$. Coupled with Constraints (30) and (31), these 6 constraints set $x_{1,j} = min\{[S_{bb,fr,i}]_j, e[a_{f,en,i}]_j\}$, as desired. These constraints are repeated for $x_{2,j}, y_{1,j}$, and $y_{2,j}$ alongside constraints (30) – (37).

LP Solution

When creating the LP in a general programming language (e.g. R, Python), we can load the enemy optimal allocation data in and loop through each friendly allocation to find the friendly optimal allocation to the enemy allocation parameters. By storing these values, we can then take the highest value and find the corresponding friendly allocation. If using a mathematical programming language (e.g. GAMS, Lingo), we can create sets in our data and then solve for the optimal friendly objective all at once.

The optimal Blue allocation is $[[S_{bb,fr,2}]_j, [S_{bf,fr,2}]_j, [S_{f,fr,2}]_j] = [0, 8, 8]$, which has a corresponding Red allocation of $[0, 10, 10]$. The optimal objective value for this set of allocations is about 0.52. For Red, the optimal allocation is $[0, 10, 10]$ with a corresponding Blue optimal allocation of $[0, 8, 8]$. It is a show of LP construction validation that these optimal allocations match. This allocation provides Red with an optimal objective value of about 0.65. Note that the objective values are not complements of each other in this case, since the cost objective is not complementary when considering each side's perspective.

We can glean from the approach of the CE/CV alternative and set our threshold value $\rho = 0.7$ to be the necessary assessment score of either the friendly or enemy objectives to declare a win/loss/unresolved. Since neither side has an assessment exceeding this value, the thread remains unresolved.

In this chapter, we presented Dresher's Game as a small combat simulation in order to demonstrate the details of our four assessment methodologies for combat simulation. We applied the Combat Effectiveness & Combat Vulnerability, Bayesian Network, Value-Focused Thinking, and Linear Programming alternatives to Dresher's, which will aid in their evaluation in the next chapter. Chapter V provides a detailed discussion of these methodologies in terms of the value hierarchy from Chapter III. Chapter V ends in applying value-focused thinking concepts to create a new alternative with a better overall evaluation than the four methodologies presented in this chapter.

V. Evaluation of Potential Methodologies Using the Value Hierarchy

In this chapter, we provide a full evaluation of the alternatives utilizing the nine lowest-level objectives of the value hierarchy defined in Section 3.1. We include comments from the illustration of these methodologies in Dresher's Game from Sections 4.2 – 4.5. After the explication of the alternative's evaluations, we create a new alternative from the highest-evaluated methodologies. A summary of the evaluations is provided in Table 10.

Table 10: Evaluation of Alternatives - Categorical Labels

	Realism		
Alternatives	Simplistic	Complete and Accurate	Comprehensive
CE / CV	Fair	Fair	Good
Bayesian Network	Good	Excellent	Good
VFT	Fair	Good	Good
LP	Good	Excellent	Good
	Efficiency		
Alternatives	Assumptions	Computation	Modular
CE / CV	Good	Poor	Poor
Bayesian Network	Poor	Poor/Inf.	Excellent
VFT	Good	Fair	Poor
LP	Good	Fair	Excellent
	Robusticity		
Alternatives	Win/Loss	Equitable Across Domains	Correlations btwn Objectives
CE / CV	Good	Good	Poor
Bayesian Network	Good	Good	Good
VFT	Good	Good	Good
LP	Good	Good	Fair

5.1 Realism

For the *simplistic* objective: the CE/CV alternative scores "Fair." Although the end state of this methodology is fairly simplistic – one of five categories – the division of each time step into 16 *threads* and the use of joint probability distributions elevated the complexity of this approach. Internal reporting of the time step's probability distribution into the five categories does simplify its interaction with other modules. Despite commanders having a generally basic understanding of probability concepts, the CE/CV methodology incorporates multiple sources of variation and further complexifies communication of this alternative. However, because the joint distribution is easily distilled into a graphic, we evaluate the alternative as "Fair" for this objective. This aligns with the evaluation according to the lowest-level objectives under the *simplistic* objective. The hierarchical structure of the JCS doctrine is imitated between the phase goals, which influence the resolution of the time-step's threads into one of the five categories. Secondly, the CE/CV approach does not explicitly model interdependencies between objectives (phase goals). In addition, this methodology does not incorporate abstract objectives into its phase goals, but rather is itself an abstract objective. The objective clearly is to obtain friendly superiority, which is broken out into completion of the phase goals. For this reason, the CE/CV approach receives a "Fair" for this objective.

Bayesian networks are fairly simplistic to communicate, although hard to establish in some circumstances. To communicate the methodology to a decision-maker, one needs only to provide that the Bayesian network is built on the dependencies occurring in the operational environment. These relationships inform the overall probability distribution for our allocation. However, the data that build the network are subjective, either based on expert opinion or historical data. When questioned about why the assessment methodology produces specific outcomes, communication quickly turns technical. For this reason, we expect a "Good" evaluation for this objective. Because of the ease of explanation, we expect a good evaluation from this objective via the lowest-level objectives. Clearly, this methodology is meant to model interdependencies between objectives. By putting the network in tiers that flow upward, one can easily conform a Bayesian Network into a hierarchical structure resembling that of the JCS doctrine. Direct goals are easily measurable, but so, too, are abstract goals. By building a network that translates actions into probability distributions, abstract goals are naturally measurable within this methodology. However, details regarding the network's CPTs are not easily explained. So, while the network structure may be intuitive, the distributions may not be. For this reason, the Bayesian Network methodology receives a "Good" for this objective.

The VFT alternative is also fairly simple to communicate to leadership. This approach aligns most with the JCS doctrine, establishing a tiered assessment that could at times mirror the exact structure from Figure 1. However, its simplicity is slightly under that of Bayesian Networks. Consider that to incorporate objectives dependencies, the VFT would need interaction terms in the objectives, which often do not have natural interpretations. These terms are not conditional probabilities, as contained in the BN alternative, and so explaining the assessment methodology for a particular scenario can easily be clouded with these terms. We see the simplicity of this methodology match up with the evaluation of the lowest-level objectives of our value hierarchy from Chapter III. Due to its more rigid structure, we cannot directly model interdependencies between objectives. However, both abstract and direct goals are easily incorporated. Abstract goals are typically measured via *constructed attributes*, but may also be broken up into lower-level direct objectives. When included, constructed attributes could lead to confusing incorporation into a programmatic model. Alternatives are typically given subjective scores for constructed attributes. Applying abstract objectives to a computerized simulation could therefore lead to a disconnect in communication between the objectives and the implementation. Overall, the VFT methodology does well in some of the lowest-level objectives and poorly in others for the *simplistic* objective, and so receives a “Fair.”

The linear program approach, while not mathematically simple, has a streamlined interpretation. In Section 4.5, we demonstrated an optimization of our objectives over all possible friendly allocations. When questioned about the objectives, the LP naturally directs the audience to the assumptions that inform the constraints. Communicating these assumptions leads to direct and top-level discussion regarding the validity of this assessment methodology. We do not see such an open avenue of communication with the assumptions in the CE/CV and BN methodologies. Conditional probabilities in the CE/CV and BN approaches force communication about assumptions into the lower levels of technical detail. While we believe that leadership would understand the LP approach, the approach does not naturally translate into the JCS strategy-to-task hierarchy. Comprised of a series of objectives, it would be possible to visualize the objectives with their associated constraints as sub-objectives, but the mechanism of assessment is nothing close to a hierarchy. However, the interdependencies between the objectives is well modeled, as these interactions exist implicitly within the dual problem to the LP. Lastly, this methodology can incorporate both abstract and direct goals, as previously demonstrated. Altogether, because the mechanism by which the approach assesses the operational environment differs so distinctly from that of the JCS doctrine, this alternative receives a “Good” for this objective, rather than an “Excellent.”

Moving on to the second *Realism* objective – *complete and accurate* – the CE/CV approach performs moderately. We expect this result from the CE/CV approach because of the worst-case-scenario approach used at the mission level of the calculation. By assuming that each enemy allocation to friendly missions can access all of their available assets, the overall assessment is performed on a logistically infeasible set of mission-level allocations. For this lowest-level objective, the alternative scores “Poor.”

However, the methodology appropriately assesses the operational environments contribution to phase goals using its joint distribution and categorization mechanisms. Despite the over-allocation of assets to missions, the operational environment is represented as being in one of five states in relationship to friendly/enemy phase goals. We see this relationship as an appropriate alternative for assessing the contribution of the operational environment. Therefore, the CE/CV approach receives a "Good" for this lowest-level objective. Combining the scores for the two lowest-level objectives yields an overall score of "Fair" for the *completeness and accuracy* objective.

Bayesian networks score "Excellent" for the *complete and accurate* objective. We expect this due to how Bayesian networks function. In utilizing evidence from the operational environment to construct and reinforce the network, the end result spans the operational environment (completeness) and corresponds well with the evidence provided (accuracy). Looking at the lowest-level objectives, the BN alternative is logistically feasible. Because the BN methodology only deals in past progress for the end of a time-step, the methodology is as logistically feasible as the simulated combat scenario. Assuming the rest of the model is adequately valid, the BN methodology is logistically feasible. Furthermore, by drawing directly from the actions taken within the operational environment, the BN methodology accurately connects the operational environment to the network, and therefore to the objectives. For these reasons, the BN receives an "Excellent" for this objective.

The Value-Focused Thinking alternative scores "Good" for the *complete and accurate* objective. The VFT is wholly dependent upon the actions taken in the operational environment, as it comments only on past performance. However, in measuring abstract goals, it is possible that the use of constructed attributes may lead to a slightly inappropriate assessment of the operational environment's contribution to the combat goals. The VFT scores a "Fair/Good" for this lowest-level objective. The VFT also scores a "Good" for the lowest-level objective *logistically feasible*. Upon combining the scores of these two lowest-level objectives, the alternative scores "Good" for this objective.

The Linear Program approach also does not have any issues with feasibility. In fact, it optimizes over an overall allocation to a set of missions, and therefore is constrained by logistics. The LP also appropriately assesses the contribution of the operational environment to combat goals as constructed by the constraints. The feasible region allows for partial completion in multiple objectives simultaneously, as seen in the other three alternatives. For these reasons, the LP alternative receives an "Excellent" for the *complete and accurate* objective.

The last of the *Realism* objectives addresses comprehensiveness across all combat domains. At this time, there is no reason to believe that any of the alternatives would not be able to accommodate input from any of the domains. For the CE/CV, VFT, and LP alternatives, adding one or more domains is a matter of recording the (projected) adjudication results from the operational environment. Meanwhile, the BN methodology simply incorporates more data into the mission-level details input as evidence into the network. For these reasons, each of the alternatives receives a "Good" for this objective.

5.2 Efficiency

Moving on to the first of the *Efficiency* objectives – *assumptions* – we evaluate the CE/CV alternative as "Good." While the lack of assumptions on enemy allocation makes the methodology logistically infeasible in many cases, this aspect of the CE/CV alternative increases its efficiency.

The Bayesian network, while not assuming a particular enemy allocation, requires data detailing enemy allocation (conditional) probabilities. The excess computation generated by requiring these inputs is inefficient, and may result in an excessively large joint distribution state-space. Storing these data may become intractable. Therefore, the BN alternative scores a “Poor” for this objective.

Section 4.4 illustrated that the VFT alternative depends on the enemy allocation insomuch that the user defines an enemy-incorporating objective. For example, no knowledge of the enemy’s allocation is necessary for the *minimize friendly cost* objective (equation 24), but is required for the two *capability* objectives (equations 20 – 23). However, even when these data are required the enemy allocation is deterministically known. Hence, assumptions do not significantly decrease the methodology’s efficiency. For this reason, the alternative scores a “Good” for this objective.

The Linear Program relies heavily upon the projection of enemy allocation, but does not make any outright assumptions. In Section 4.5, we assumed that the enemy allocation was optimized using the same objectives as the friendly assessment. However, this was only done in the absence of an adjudication algorithm to calculate the corresponding enemy allocation. In an operational combat simulation, the LP would calculate the enemy’s allocation using an adjudication algorithm. We discuss this in reference to the next objective. Because the LP does not make any assumptions regarding the enemy’s allocation, we evaluate it at a “Good” for this objective.

The second of the *Efficiency* objectives is *computational overhead.* In Section 4.2, we presented the methodology as applied to a singular thread. The assessment calculations are mirrored for each of the 16 threads when the simulation is past the initial time-step. Consequently, the CE/CV approach utilizes a lot of extra computation while referencing CPTs, re-binning asset distributions, calculating the CE/CV joint distribution to categorize the threads, and aggregating the threads to obtain a final assessment. We score this methodology at a “Poor” for this objective.

The Bayesian Network methodology also requires a lot of computational overhead, in the form of data collection and network mapping. In order to form the BN for Dresher's Game in Section 4.4, we manually created the CPTs, which required ensuring consistency across the conditional and unconditional distributions for each variable. We required data for each node and relationship in the network. When considering many missions and assets within those missions, and operational combat simulation would require a much more extensive network with many more CPTs. There exist algorithms that can automate both network construction and conditional probability tables (37; 38; 39). However, these options require a bit of additional computational overhead. Coding these algorithms often requires time fine-tuning scoring parameters to judge the viability of the constructed network (39; 40), some level of subject matter expertise implicit in the application of the algorithm (41), or post-hoc evaluation of the Bayesian Network's accuracy/effectiveness (42). In any case, the state-space of the joint probability distribution of an allocation can become intractably large. In the Bayesian network presented in Section 4.3, each side's joint probability distribution had more than $500,000$ elements in its state space. When we extend the number of assets types and increase the number of available assets of a single type, this methodology could become infeasible to implement. The Bayesian Network methodology therefore scores "Poor/Inf." (for Poor/Infeasible) in the *computational overhead* objective, conditional upon the size of the simulation.

The VFT methodology requires minimal computational overhead. As it is an additive model, there is little computation required once the module calculates the individual objective values. Calculating the objective values may require some extra computation if the utility functions for some of the objectives become complex. However, all data utilized is deterministically known. Much more of the computational overhead for this alternative comes in the creation of the hierarchy and the definition of the attributes and their utility functions. This alternative scores a "Good" for the *computational overhead* objective, because it is extremely efficient once initialized.

Applying a linear program to the assessment of simulated combat requires significant computational overhead. On top of initializing the LP with its objectives, variables, and constraints, the methodology also requires one run of the adjudication algorithm for every friendly allocation evaluated. Although the Simplex Method is an exponential-time algorithm in the worst case (34, p. 393), there are efficient polynomial-time algorithms which quicken the LP solution time, such as Karmarkar's algorithm and Khachian's algorithm (34, p. 401–414). Because this methodology is dependent upon the efficiency of these allocation and adjudication algorithms, we assume that these are constructed to be computationally efficient. The result is a computationally semi-efficient algorithm that requires some preliminary work in its initialization. The Linear Program alternative therefore scores a "Fair" for the *computational overhead* objective.

The last of the *Efficiency* objectives addresses their *modularity*. The CE/CV approach has already been implemented in a modular environment. The CE/CV approach provides feedback in the form of in which of the five states each objective ends the time-step. Objectives with an end state closer to Enemy Superiority should receive more attention. However, specific asset resource gaps are not addressed in its BEAM application. For this reason, the alternative scores "Poor" for the *modularity* objective.

The Bayesian network methodology may address resource gaps by comparing the overall objective values between scenarios. Doing so would require more computation, but the distribution structure of the network provides a robust basis for addressing resource gaps. One way to observe resource gaps is to analyze scenarios' objective values with their decision variable's distributions. Doing so internally to an assessment algorithm would be computationally intensive, but would provide valuable information regarding which assets should be more heavily focused or invested in. Because of the extreme potential to provide formative assessment, this methodology receives an "Excellent" for this objective.

The VFT alternative addresses resource gaps in the same way as the CE/CV alternative. While the assessment algorithm may seek out individual objectives which are providing low utility, the alternative does not address specific asset weaknesses or strengths driving these utility levels. For this reason, the VFT methodology scores "Poor" for this objective.

Lastly, applying an LP to the assessment of simulated combat requires that the adjudication and allocation portions of the model be modularized, so that they can be called by the LP. As a module, the LP assessment would be in frequent communication with this (these) other module(s). The LP also naturally addresses resource gaps via *shadow prices*. The logistic feasibility constraints on the allocation of assets are those from which we can best utilize the shadow prices. The interpretation of the shadow price in this context is "the additional progress toward our combat objectives which may be gained by increasing the available number of asset a by one." If the constraint it binding at the optimal solution, then we will get a non-zero shadow price. For a maximization LP, we would want to focus more on the assets whose corresponding constraints have larger absolute shadow prices. A largely negative shadow price means that we should devalue the corresponding asset, and the opposite for a largely positive shadow price. Because of the vast opportunity for formative assessment with linear programming, this alternative receives an "Excellent" for this objective.

5.3 Robusticity

We now address the last of the three branches in our value hierarchy – *Robusticity*. The first sub-objective for this branch is titled *Win/Loss*, short for the objective to categorize the probabilities of multi-thread outcomes into one of *Win*, *Loss*, or *Unresolved*. As previously described, the CE/CV methodology incorporates such a categorization, and so receives a "Good" for this objective.

We demonstrated in Sections 4.3, 4.4, and 4.5 that we can use the threshold probability ρ from the CE/CV methodology to categorize the assessment for the BN, VFT, and LP methodologies, as well. The categories for each assessment methodology can be categorized from the friendly perspective by: (1) *unresolved* if both friendly and enemy assessment scores are between $[0, \rho)$, (2) *won* if the friendly assessment is between $[\rho, 1]$ and enemy assessment is between $[0, \rho)$, and (3) *lost* if the friendly assessment is between $[0, \rho)$ and enemy assessment is between $[\rho, 1]$. Hence, these three objectives also score "Good" for this objective.

The second sub-objective under *Robusticity* is *Equitable across Domains*. As currently presented, each methodology can place weights on the domains to ensure that the combat results from each domain provide proportional value to their portion of the operational environment. If desired, the analyst may use alternative means of weighting, such as weighting according to the amount of change caused in a specific domain. In all cases, the alternatives are capable of being equitable across each combat domain, and therefore receive a "Good" for this objective.

The last objective is *Correlations Between Objectives*. An important aspect of assessment is accurately representing achievement. Whenever achievement in one objective is tied to achievement in another objective, it is important that changes in the operational environment not receive double credit or double penalty.

The CE/CV alternative weakly incorporates this idea. The application of CE/CV in Section 4.2 demonstrated the methodology for a single *mission group*. BEAM applies the methodology to multiple mission groups. For those groups ending in the unresolved category at the end of a time-step, BEAM aggregates each asset's remaining outcome distributions. These distributions are weighted by their associated thread's weight. Being weighted by their associated thread and the relative health of each mission group's asset distribution, we receive an aggregated asset distribution which considers these relative weights (28, p. 27). Note that these weights are not weights on objectives, nor are any relationships between the mission groups themselves considered. Therefore, the CE/CV alternative receives a "Poor" for this objective.

The Bayesian network presented in Section 4.3 does not incorporate the objectives explicitly. Rather, the probabilities generated from the network are utilized alongside an allocation's corresponding objective value to calculate an expected objective value. However, dependencies between the objectives could be incorporated by feeding this decision variable network into a separate objectives Bayesian network. A separate network is required so that the defining joint distribution of the decision variables excludes the objectives' state-space. Rather, the joint distributions of the decision variables' BNs would determine the univariate cause nodes of the objectives' BNs. Further CPTs between the objectives would then create a joint distribution of the objectives' values, from which an expected value is simply calculated. In sum, the BN alternative is capable of correlating the objectives, although doing so requires additional computation effort. Therefore, the alternative scores a "Good" for this objective.

The VFT alternative may include some objectives correlations. In fact, Keeney conditions the validity of the additive value/utility model on attribute independence (27, p. 133–138). In order to obtain this independence from otherwise dependent attributes, we may add a joint-objective term may be added to the hierarchy, defined as the interaction of two or more objectives. This is distinct from the interaction of decision variables. Similar to an interaction term in regression methods, this joint-objective term would output a utility equal to the achievement caused by the simultaneous effects of multiple objectives. This term may not be easily intelligible, nor easily defined. Also, one could replace the incorporated single-objective terms in the model with the joint-objective term, so as to not double-count contributions to the overall assessment. Similar to regression, the utility function of this joint-objective term would be dependent upon the different levels of achievement in each variable. One way to accomplish this is to perform function composition. For example, let x_1 and x_2 be two objectives and y the utility value for the $x_1 * x_2$ objective. Suppose $g_1(x_1), g_2(x_2)$ are utility functions for x_1, x_2, respectively. Then we could let $y = h_1(g_2(x_2))$ or $y = h_2(g_1(x_1))$, where $h_1(\cdot), h_2(\cdot)$ are defined functions for the interaction dependent upon an x_2 or x_1 input, respectively. Note that this approach is dependent upon knowing a theoretical correlation between the two objectives x_1, x_2 in order to construct the functions $h_1(\cdot), h_2(\cdot)$. Alternately, one could utilize data on the interaction between differing levels of achievement in the original objectives in order to construct an interaction term. Similar to the BN approach, incorporating multi-objective consequences into the assessment requires extra computational load. This approach scores a "Good" for this objective. However, doing so would significantly increase the amount of computational overhead for this objective, thereby changing this alternative's *computational overhead* evaluation to "Fair." As we wish for each methodology to provide as much formative assessment feedback as possible, this is a desirable trade-off.

The LP alternative implicitly considers correlations between the objectives in the form of objective trade-offs. From *duality theory*, we know that if the *primal problem* contains an optimal solution, then so, too, does the *dual problem*, and the objective values of these problems are equal (34, p. 266). If we define our objectives well, then each objective will have a maximum value of 1 and a minimum value of 0. Therefore, as long as we establish the constraints such that there is at least one feasible solution, then both the primal and dual problems will have an optimal solution. The LP assessment methodology considers objective trade-offs in the implicit dual problem, whose constraints' right-hand sides are the objective function coefficients. As the primal problem works through various allocations, the dual problem performs objective trade-offs. Therefore, any influence that one objective has on another will be implicitly considered within each iteration. However, this concept is different from observing correlations between objectives, because only the boundaries of the dual constraints are considered. Because of this disparity, the LP alternative receives a "Fair" for this objective.

5.4 Methodology Creation and Finalization

In this section we compare the four methodologies' evaluations. We then create a new alternative from the best methodologies as determined from our nine objectives.

Table 11 provides the distribution of evaluations summarizing the previous three sections and Table 10. Presented this way, we can see that the VFT and LP alternatives have the least "Inf." or "Poor" evaluations. In addition, the CE/CV alternative has more "Fair" evaluations than any other alternative and has the least number of "Good" evaluations. We therefore consider the CE/CV methodology to be worse than the VFT and LP alternatives. The BN alternative is evaluated as Infeasible in the *computational overhead* objective. As the state-space of a combat simulation grows, this methodology becomes infeasible to apply. We hypothesize that the BN alternative would be infeasible for most DoD combat models, but may be applicable in other gaming areas. For our purposes, we therefore determine the BN alternative to be worse than the VFT and LP alternatives.

Between the VFT and LP alternatives, Table 11 illustrates that these two methodologies are fairly comparable. While the VFT evaluation distribution is heavily

Table 11: Evaluation Distributions by Objective

	Evaluation				
Alternative	Inf.	Poor	Fair	Good	Excellent
CE/CV	0	3	2	4	0
BN	0.5	1.5	0	5	2
VFT	0	1	2	6	0
LP	0	0	2	5	2

skewed-left with a mode of "Good", the LP distribution is centered on "Good." However, the VFT methodology does not have any "Excellent" evaluations. Overall, the LP methodology appears to be the best alternative from both Table 10 and Table 11.

While the LP objective appears to be the best, we can take key components from other methodologies considered here to create a new alternative. The intent is that this new alternative's set of evaluations would be better than any of the current four alternatives. Keeney calls this process "alternative creation" (27). By combining the the VFT and LP methodologies, we can obtain an alternative better than all four of the methodologies hitherto discussed. We will refer to this alternative as the *VFT-LP* alternative.

The new *VFT-LP* alternative utilizes the VFT methodology as the main assessment structure, and then leverages the optimization of the LP in order to perform formative assessment and enhance the efficacy of the VFT's assessment. Consider the following approach. In order to perform the VFT-LP assessment, first create a VFT Objectives Hierarchy as described in Sections 3.2.3 and 4.4. Also, devise a linear program to take any desired weights (e.g., asset weights, objective weights, mission weights) from the VFT and optimize them for use in the next time-step's VFT assessment. The LP should be run after a time-step's VFT assessment, in order to aid the allocation for next time-step.

Note that the LP could be run before the VFT assessment. However, we may use the LP to optimize weights, potentially including asset and mission weights. Therefore, applying the LP between adjudication and assessment would provide an assessment based on weights that were not used in the current time-step's allocation, and therefore would less accurately assess friendly and enemy forces. Running the LP after assessment will not only update these values for the next time-step's assessment, but also can update these weights for other parts of the simulation, including allocation and adjudication modules.

To instantiate the VFT-LP methodology, consider the following example extending the VFT applied to Dresher's Game in Section 4.4. Rather than utilizing the subsequent LP to find an optimal allocation, we focus on the five objective weights ($w_{ECB}, w_{FCB} w_{FCF}, w_{ECF}$, and w_C). These are the five decision variables in the LP to follow. Since the VFT hierarchy from Section 4.4 did not utilize the asset weights P_b and P_f, we do not utilize them here. However, the VFT and following LP's objective function could be modified to resemble that of the objectives and objective function, respectively, from our LP application to Dresher's Game in Section 4.5. Doing so would include these asset weights as decision variables. We then reclassify all allocation variables (i.e., $S_{bb,B,i}, S_{bf,B,i}, S_{f,B,i}, S_{bb,R,i}, S_{bf,R,i}, S_{f,R,i}$) parameters. These allocation variables take on the values from the end of the current time-step. All other parameters retain their parameters classification. Our objective is the same as from the LP application to Dresher's Game in Section 4.5 – maximize the cost and two enemy capability objectives. However, we utilize the enemy and friendly capability objectives from the VFT construction, which are slightly different from those of the LP application. With these changes, consider the following sample LP:

$$
\begin{aligned}
max \qquad & w_{ECB,i+1}\left(1-\frac{b_{en,i+1}}{b_{en,1}}\right)+w_{ECF,i+1}\left(1-\frac{f_{en,i+1}}{f_{en,1}}\right)\\
& + w_{FCB,i+1}\left(\frac{b_{fr,i+1}}{b_{fr,1}}\right)+w_{FCF,i+1}\left(\frac{f_{fr,i+1}}{f_{fr,1}}\right)\\
& + w_{C,i+1}\left(1-\frac{C_{bb}S_{bb,fr,i}+C_{bf}S_{bf,fr,i}+C_{f}S_{f,fr,i}+PC_i}{C_{bb}b_{fr,i}+C_f f_{fr,i}+PC_i}\right)
\end{aligned} \tag{58}
$$

$$
s.t \qquad w_{ECB,i+1}+w_{ECF,i+1}+w_{FCB,i+1}+w_{FCF,i+1}+w_{C,i+1}=1 \tag{59}
$$

$$
w_{ECB,i+1}, w_{FCB,i+1}, w_{FCF,i+1}, w_{ECF,i+1}, w_{C,i+1} \geq 0 \tag{60}
$$

$$
w_{ECB,i+1}, w_{FCB,i+1}, w_{FCF,i+1}, w_{ECF,i+1}, w_{C,i+1} \geq \epsilon \tag{61}
$$

$$
\frac{w_{ECB,i+1}-w_{ECB,i}}{w_{ECB,i}} \leq \delta_{ECB} \tag{62}
$$

$$
\frac{w_{ECB,i+1}-w_{ECB,i}}{w_{ECB,i}} \geq -\delta_{ECB} \tag{63}
$$

$$
\frac{w_{ECF,i+1}-w_{ECF,i}}{w_{ECF,i}} \leq \delta_{ECF} \tag{64}
$$

$$
\frac{w_{ECF,i+1}-w_{ECF,i}}{w_{ECF,i}} \geq -\delta_{ECF} \tag{65}
$$

$$
\frac{w_{FCB,i+1}-w_{FCB,i}}{w_{FCB,i}} \leq \delta_{FCB} \tag{66}
$$

$$
\frac{w_{FCB,i+1}-w_{FCB,i}}{w_{FCB,i}} \geq -\delta_{FCB} \tag{67}
$$

$$
\frac{w_{FCF,i+1}-w_{FCF,i}}{w_{FCF,i}} \leq \delta_{FCF} \tag{68}
$$

$$
\frac{w_{FCF,i+1}-w_{FCF,i}}{w_{FCF,i}} \geq -\delta_{FCF} \tag{69}
$$

$$
\frac{w_{C,i+1}-w_{C,i}}{w_{C,i}} \leq \delta_{C} \tag{70}
$$

$$
\frac{w_{C,i+1}-w_{C,i}}{w_{C,i}} \geq -\delta_{C} \tag{71}
$$

For this LP, we vary the values of the next time-step's weights, and so the decision variables have time-step index $i+1$. The objective function (58) minimizes the enemy capability objectives and the cost objective, while maximizing the friendly capability objective.

There are two main constraints. Constraint (59) ensures that the objective function value remains in the interval [0,1]. Constraint (60) is the typical non-negativity constraint, which effectually sets the domain of each decision variable to [0,1] when coupled with Constraint (59).

Constraints (61) – (71) are optional, and are provided here as ideas for changing the behavior of this LP. Optional Constraint (61) would ensure that each weight is strictly positive. This optional constraint could also be broken out with different ϵ values for each weight, if desired. Constraints (62) – (71) would force the proportional change in the value of a weight to be restricted from one time-step to another. So, if one would not want there to be more than a 5% change (i.e., increase or decrease) in the enemy capability objectives' weights from one time-step to another, one could set $\delta_{ECB} = \delta_{ECF} = 0.05$. Limiting the change could be useful to obtain asymptotic behavior toward the overall optimal weights for the objectives, rather than highly variable or quickly flat-lining behavior. A slower change would also allow for different missions to have a more dominant effect upon the overall assessment, rather than allowing missions conducted at the beginning of a scenario to have the most significant impact on the objective weights.

At the beginning of time-step 1, the objective weights were

$$w_{ECB,1} = \frac{2}{15} \qquad w_{ECF,1} = \frac{1}{5} \qquad w_{FCB,1} = \frac{2}{15}$$
$$w_{FCF,1} = \frac{1}{5} \qquad w_{C,1} = \frac{1}{3}$$

Now, suppose that we let $\epsilon = 0.001$ and $\delta_{ECB} = \delta_{ECF} = \delta_{FCB} = \delta_{FCF} = \delta_C = 0.5$.

Then, the optimal solution for time-step 2's weights is

$$[w_{ECB,2}, w_{ECF,2}, w_{FCB,2}, w_{FCF,2}, w_{C,2}] = [1/5, 2/15, 1/5, 3/10, 1/6] \tag{72}$$

These results are very sensitive in relation to the set of δ parameters. For example, changing δ_C to equal 0.6 changes the optimal solution to

$$[w_{ECB,2}, w_{ECF,2}, w_{FCB,2}, w_{FCF,2}, w_{C,2}] = [1/5, 1/6, 1/5, 3/10, 2/15] \tag{73}$$

The purpose of a lower delta value is to only slightly change the objective weights to reflect a good balance. Clearly, having the set of delta values equal 1 would set the weights of the highest-achieving objectives as large as possible, while setting the lowest-achieving objectives closer to ϵ. Limiting the change prevents early time-steps' allocations and adjudications from immediately setting (a) weight(s) to 1 or ϵ.

Table 12: VFT-LP Evaluation – Categorical Labels

Objective Category	Objective	Evaluation
	Simplistic	Fair
Realism	Complete & Accurate	Good
	Comprehensive	Good
	Assumptions	Good
Efficiency	Computation	Good
	Modular	Excellent
	Win/Loss	Good
Robusticity	Equitable Across Domains	Good
	Correlations btwn Objectives	Excellent

Table 12 provides the set of evaluations for the VFT-LP alternative. The VFT-LP alternative maintains the *Simplicity of Communication* from the original VFT methodology, due to the potential for multi-objective attributes to allow for reporting on objectives correlations. This alternative's evaluation for the *Completeness and Accuracy* and the *Comprehensive across All Domains* objectives is also the same as for the VFT methodology, since the assessment mechanism has not changed.

Under the *Efficiency* objective, the addition of the LP to the backend of the VFT methodology does not incur any additional assumptions on the enemy's allocation. No additional data for these allocations is required, either. The evaluation is therefore equal to that of the VFT alternative. However, the *Computational Overhead* decreases, since we no longer have to go through the decision process of determining accurate weights with a proxy decision-maker. While creating the VFT takes some initialization with a proxy decision-maker, obtaining appropriate attribute weights can be a laborious process (27, p. 147–149, 166-171). We have replaced this initialization task with the above LP process. Instead, we now only require any (arbitrary) weights to initialize the model, and the LP will adjust the weights at each time-step. Therefore, the evaluation fo the *Computational Overhead* objective is "Good." Lastly, with the additional formative feedback that the LP provides, the likely included asset weights incorporate resource and asset gaps. We therefore evaluate this alternative as "Excellent" for the *Modular* objective.

Under the *Robusticity* objective, we maintain the *Win/Loss* and *Equitable Across Domains* evaluations from the VFT methodology, which were "Good." For the *Correlations Between Objectives* objective, not only may we utilize multi-objective attributes, but we now consider the trade-offs between weighting different objectives and their impact on the overall objective function value with the addition of the LP. Therefore, we evaluate the VFT-LP alternative as "Excellent" for this objective.

Table 13: Evaluation Distributions by Objective for Top 3 Methodologies

	Evaluation				
Alternative	Inf.	Poor	Fair	Good	Excellent
VFT-LP	0	0	1	6	2
LP	0	0	2	5	2
VFT	0	1	2	6	0

Table 13 provides the evaluation distributions for the top three alternatives. While still comparable to the LP methodology's evaluation distribution, the VFT-LP evaluations clearly outshine those of the VFT methodology. Note that the VFT-LP is evaluated as slightly better than the LP, with the VFT-LP having moved one "Fair" evaluation to "Good." Overall, the VFT-LP methodology has a better evaluation distribution than all of the other considered alternatives. The created alternative also provides useful feedback internal to a simulation of warfare. We recommend use of this methodology for combat simulation within the DoD.

In this chapter, we evaluated the four assessment alternatives detailed in Chapter III and applied in Chapter IV. After evaluating the LP and VFT methodologies as better than the CE/CV and BN methodologies, we searched for ways to improve upon these approaches. Specifically, we merged the two highest-evaluated alternatives to create a new VFT-LP methodology, which resulted in a better evaluation of this alternative than the other four alternatives. We also provided some examples of using an LP to optimize the overall objective function by selecting different objective weights. The next chapter provides conclusory thoughts on the research presented in this thesis and suggests some aspects of the current research to continue in future work.

VI. Conclusion

There is very little research on combat assessment methodologies, although this area of research's implications for DoD conduct could be far-reaching. While the strategy-to-task framework outlined by the Joint Chiefs of Staff (2) provides a skeletal structure for military combat and non-combat assessment, there is a lack of guidance for how to carry out that guidance in operational models. This thesis distills the available DoD guidance alongside attributes from other predominant assessment areas to develop a reasonably good methodology for application in United States military combat simulations. We rely heavily upon the Value-Focused Thinking principles provided by Keeney (27) to guide our evaluation, as this technique is expert at distilling qualitative information into a structured evaluation hierarchy. Embedded in this hierarchy are nine criteria that provide an answer to our first research question ("what are the desired characteristics of a combat assessment methodology?"). In particular, we selected the characteristics from JCS doctrine and practice to guide the value hierarchy while incorporating applicable aspects of other assessment research:

1. Simplicity of Communication
2. Completeness & Accuracy
3. Comprehensive Across All Domains
4. Limit Adversary Allocation Assumptions
5. Limit Computational Overhead
6. Modularity
7. Provide Win/Loss/Unresolved Categories
8. Domain Equity
9. Objectives Correlations

We have added the application of each methodology to a small problem in order to more accurately evaluate each technique in relation to our value hierarchy. However, this work does not integrate assessment into a large combat simulation. Future research could develop the theory and/or application of combat assessment within large simulations. For example, BEAM currently utilizes stochastic asset health as the output of a singular thread. In this work, we have considered only deterministic asset health distributions for the LP methodology.

Coupled with stochastic asset health is partial termination of a scenario. BEAM currently integrates this idea at the *thread* level by terminating whole threads, weighted by the probability of that thread occurring. However, only a portion of the thread may be truly lost. Future research could apply mathematical programming to tackle this problem.

A significant constraint of the assessment methodologies presented in this work is the assumption of their symmetry. We have assumed that both Blue and Red forces consider alignment with JCS doctrine a priority when assessing operational environment outcomes. A great advance in the field would be to determine a realistic assessment structure and methodology for a general or specific Red actor.

The alternatives we selected for evaluation in this work are by no means comprehensive. However, the CE/CV methodology provided a status-quo alternative, which has been applied inside of an enterprise-level combat simulation. In addition, the BN, VFT, and LP approaches broadly span the types of methodologies we could have investigated. The BN methodology is built on Bayesian statistics and a heavy distributional base. The network aspect of this approach is typical of non-hierarchical assessment. The VFT approach, in contrast, is rigidly hierarchical and requires multiple types of independence. In this model, we demonstrate the strengths and weaknesses of an additive model for assessment within simulated combat. Lastly, the LP approach is intended to demonstrate the pros and cons of general mathematical programming for assessment within a simulation of combat. In fact, the examples presented are mixed-integer linear programs. As an answer to our second research question ("how should one conduct combat assessment?"), we suggest utilizing the ***Value-Focused Thinking – Linear Programming* approach**.

The examples of these techniques that we present in chapter IV are foundational approaches from applied statistics and operations research and do not represent the wealth of complexity that each technique has to offer. However, by detailing the mechanics, and exemplifying basic examples, of these methodologies, we intend for our evaluations to provide only mild error when extrapolated to more nuanced versions of these techniques. In sum, we consider the archetypal examples considered in this work to adequately span potential quantitative assessment techniques.

Appendix A. Dresher's Game Conditional Probability Tables for the Bayesian Network Methodology

Table 14: Distribution of $S_{f,R,2}$

$S_{f,R,2}$	0	1	2	3	4	5	6	7	8	9	10
$p(S_{f,R,2})$	0.2	0.08	0.05	0.025	0.004	0.001	0.02	0.07	0.1	0.2	0.25

Table 15: Distribution of $S_{bb,R,2}$

$S_{bb,R,2}$	0	1	2	3	4	5	6	7	8	9	10
$p(S_{bb,R,2})$	0.3	0.12	0.045	0.02	0.012	0.006	0.012	0.02	0.045	0.12	0.3

Table 16: Distribution of $S_{bf,R,2}$

$S_{bf,R,2}$	0	1	2	3	4	5
$p(S_{bf,R,2})$	0.66591	0.07832	0.02445	0.008065	0.004424	0.002491
$S_{bf,R,2}$	6	7	8	9	10	
$p(S_{bf,R,2})$	0.00659	0.0146	0.03075	0.0714	0.093	

Table 17: Distribution of $S_{f,B,2}$

$S_{f,B,2}$	0	1	2	3	4	5	6	7	8
$p(S_{f,B,2})$	0.0939	0.0244	0.0609	0.0061	0.0478	0.013	0.1084	0.1125	0.5331

Table 18: Distribution of $S_{bb,B,2}$

$S_{bb,B,2}$	0	1	2	3	4	5	6	7	8
$p(S_{bb,B,2})$	0.202	0.0205	0.0624	0.0061	0.044	0.0075	0.027	0.0627	0.5677

Table 19: Distribution of $S_{bf,B,2}$

$S_{bf,B,2}$	0	1	2	3	4	5	6	7	8
$p(S_{bf,B,2})$	0.5879	0.0818	0.0102	0.0033	0.0007	0.0026	0.0139	0.0838	0.2157

Table 20: Distribution of $S_{bf,B,2}|S_{f,R,2}$

$S_{bb,B,2} \backslash S_{f,R,2}$	0	1	2	3	4	5
0	1	0.02	0.008	0.0003	0.00004	0.00001
1	0	0.25	0.01	0.0009	0.00005	0.000028
2	0	0.73	0.075	0.0013	0.00009	0.0003
3	0	0	0.11	0.0021	0.00126	0.00082
4	0	0	0.797	0.09	0.0069	0.0029
5	0	0	0	0.12	0.013	0.007
6	0	0	0	0.7854	0.087	0.011
7	0	0	0	0	0.143	0.097
8	0	0	0	0	0.74866	0.880942

$S_{bb,B,2} \backslash S_{f,R,2}$	6	7	8	9	10
0	0.00001	0.00001	0.00001	0.00001	0.00001
1	0.000028	0.000028	0.000028	0.000028	0.000028
2	0.0003	0.0003	0.0003	0.0003	0.0003
3	0.00082	0.00082	0.00082	0.00082	0.00082
4	0.0029	0.0029	0.0029	0.0029	0.0029
5	0.007	0.007	0.007	0.007	0.007
6	0.011	0.011	0.011	0.011	0.011
7	0.097	0.097	0.097	0.097	0.097
8	0.880942	0.880942	0.880942	0.880942	0.880942

Table 21: Distribution of $S_{bf,R,2}|S_{bb,R,2}$

$S_{bf,R,2} \backslash S_{bb,R,2}$	0	1	2	3	4	5	6	7	8	9	10
0	0.31	0.35	0.45	0.55	0.645	0.71	0.78	0.86	0.94	0.99	1
1	0.13	0.13	0.2	0.195	0.185	0.175	0.15	0.115	0.05	0.01	0
2	0.04	0.045	0.06	0.07	0.09	0.07	0.05	0.02	0.01	0	0
3	0.01	0.02	0.025	0.03	0.04	0.03	0.015	0.005	0	0	0
4	0.0075	0.01	0.01	0.01	0.015	0.014	0.005	0	0	0	0
5	0.005	0.005	0.005	0.005	0.005	0.001	0	0	0	0	0
6	0.0075	0.02	0.02	0.04	0.02	0	0	0	0	0	0
7	0.01	0.05	0.08	0.1	0	0	0	0	0	0	0
8	0.04	0.1	0.15	0	0	0	0	0	0	0	0
9	0.13	0.27	0	0	0	0	0	0	0	0	0
10	0.31	0	0	0	0	0	0	0	0	0	0

Table 22: Distribution of $S_{bb,R,2}|S_{f,B,2}$

$S_{bb,R,2} \backslash S_{f,B,2}$	0	1	2	3	4	5	6	7	8
0	0.2	0.002	0.05	0.11	0.12	0.16	0.18	0.21	0.35
1	0.15	0.003	0.02	0.03	0.03	0.1	0.13	0.16	0.12
2	0.1	0.005	0.03	0.01	0.015	0.003	0.06	0.1	0.06
3	0.05	0.21	0.002	0.007	0.01	0.03	0.03	0.05	0.03
4	0.003	0.16	0.003	0.012	0.015	0.012	0.009	0.01	0.003
5	0.001	0.1	0.05	0.023	0.045	0.015	0.005	0.005	0.001
6	0.001	0.05	0.08	0.069	0.085	0.03	0.009	0.01	0.001
7	0.002	0.01	0.12	0.097	0.09	0.07	0.02	0.035	0.002
8	0.1	0.05	0.16	0.137	0.09	0.115	0.1	0.05	0.07
9	0.18	0.15	0.19	0.215	0.18	0.165	0.187	0.12	0.15
10	0.213	0.26	0.295	0.29	0.32	0.3	0.27	0.25	0.213

Table 23: Distribution of $S_{bf,B,2}|S_{bb,B,2}$

$S_{bb,B,2} \backslash S_{bf,B,2}$	0	1	2	3	4	5	6	7	8
0	0.086	0.1076	0.001	0.001	0.001	0.58	0.1	0.2	1
1	0.03	0.06	0.005	0.005	0.011	0.28	0.3	0.8	0
2	0.01	0.0164	0.008	0.037	0.065	0.1	0.6	0	0
3	0.001	0.0086	0.096	0.107	0.223	0.04	0	0	0
4	0.001	0.001	0.16	0.3	0.7	0	0	0	0
5	0.002	0.0164	0.15	0.55	0	0	0	0	0
6	0.01	0.12	0.58	0	0	0	0	0	0
7	0.06	0.67	0	0	0	0	0	0	0
8	0.8	0	0	0	0	0	0	0	0

Table 24: Distribution of $S_{f,B,2}|S_{bb,R,2}, S_{bf,R,2}$

$S_{bb,R,2} + S_{bf,R,2} \backslash S_{f,B}$	0	1	2	3	4	5	6	7	8
0	1	0	0	0	0	0	0	0	0
1	0.01	0.3	0.69	0	0	0	0	0	0
2	0.001	0.002	0.1	0.12	0.777	0	0	0	0
3	0.0001	0.0005	0.0014	0.0055	0.1505	0.192	0.65	0	0
4	0.00001	0.00004	0.00025	0.0003	0.0084	0.01	0.12	0.15	0.711
5	0.00001	0.00004	0.00025	0.0003	0.0084	0.01	0.12	0.15	0.711
6	0.00001	0.00004	0.00025	0.0003	0.0084	0.01	0.12	0.15	0.711
7	0.00001	0.00004	0.00025	0.0003	0.0084	0.01	0.12	0.15	0.711
8	0.00001	0.00004	0.00025	0.0003	0.0084	0.01	0.12	0.15	0.711
9	0.00001	0.00004	0.00025	0.0003	0.0084	0.01	0.12	0.15	0.711
10	0.00001	0.00004	0.00025	0.0003	0.0084	0.01	0.12	0.15	0.711

Table 25: Distribution of $S_{bf,B,2}|S_{bb,B,2}, S_{f,R,2}$

$S_{bb,B,2} \backslash S_{bf,B,2}$	0	1	2	3	4	5	6	7	8
0	1	0	0	0	0	0	0	0	0
1	0.983	0.017	0	0	0	0	0	0	0
2	0.96	0.034	0.006	0	0	0	0	0	0
3	0.88	0.1	0.015	0.005	0	0	0	0	0
4	0.79	0.133	0.065	0.011	0.001	0	0	0	0
5	0.73	0.17	0.067	0.027	0.001	0.005	0	0	0
6	0.6	0.139	0.072	0.0035	0.001	0.0045	0.18	0	0
7	0.45	0.086	0.0164	0.0086	0.001	0.007	0.09	0.341	0
8	0.38	0.11	0.006	0.0035	0.001	0.0035	0.006	0.11	0.38

Table 26: Distribution of $S_{bf,R,2}|S_{bb,R,2}, S_{f,B,2}$

$S_{bb,R,2} \backslash S_{bf,R,2}$	0	1	2	3	4	5	6	7	8	9	10
0	1	0	0	0	0	0	0	0	0	0	0
1	0.99	0.01	0	0	0	0	0	0	0	0	0
2	0.94	0.05	0.01	0	0	0	0	0	0	0	0
3	0.86	0.115	0.02	0.005	0	0	0	0	0	0	0
4	0.78	0.15	0.05	0.015	0.005	0	0	0	0	0	0
5	0.71	0.175	0.07	0.03	0.014	0.001	0	0	0	0	0
6	0.645	0.185	0.09	0.04	0.015	0.005	0.02	0	0	0	0
7	0.55	0.195	0.07	0.03	0.01	0.005	0.04	0.1	0	0	0
8	0.45	0.2	0.06	0.025	0.01	0.005	0.02	0.08	0.15	0	0
9	0.35	0.13	0.045	0.02	0.01	0.005	0.02	0.05	0.1	0.27	0
10	0.31	0.13	0.04	0.01	0.0075	0.005	0.0075	0.01	0.04	0.13	0.31

Table 27: Distribution of $S_{f,R,2}|S_{bb,B,2}, S_{bf,B,2}$

$S_{bb,B,2} + S_{bf,B,2} \backslash S_{f,R,2}$	0	1	2	3	4	5	6	7	8	9	10
0	0.95	0.05	0	0	0	0	0	0	0	0	0
1	0.002	0.3	0.698	0	0	0	0	0	0	0	0
2	0.0007	0.01	0.042	0.05	0.005	0.001	0.005	0.3175	0.5688	0	0
3	0.0007	0.01	0.042	0.03	0.005	0.001	0.005	0.3375	0.5688	0	0
4	0.0007	0.01	0.042	0.03	0.005	0.001	0.005	0.3375	0.5688	0	0
5	0.0007	0.01	0.042	0.03	0.005	0.001	0.005	0.04	0.054	0.2623	0.55
6	0.0007	0.01	0.042	0.03	0.005	0.001	0.005	0.04	0.054	0.2623	0.55
7	0.0007	0.01	0.042	0.03	0.005	0.001	0.005	0.04	0.054	0.2623	0.55
8	0.0007	0.01	0.042	0.03	0.005	0.001	0.005	0.04	0.054	0.2623	0.55

Bibliography

1. Jeffrey Chin, Richard Dukes, and William Gamson. Assessment in simulation and gaming: A review of the last 40 years. *Simulation & Gaming*, 40(4):553–568, 2009.

2. Joint Staff Director for Joint Force Development. *Joint Operations: Incorporating Change 1*. J-7, 3.0 edition, 10 2018.

3. Paul Black and Dylan Wiliam. Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, 5(1):7–74, 1998.

4. Hitesh Bhasin. 10 key differences between assessmenet and evaluation. *Marketing91 Management Articles*, 2019.

5. Robert Wood. *Measurement and Assessment in Education and Psychology: Collected Papers, 1967-87*. Taylor & Francis, 1987.

6. Air Land Sea Application Center. *Multi-Service Tactics, Techniques, and Procedures for Operation Assessment*, 8 2015. This document is a collaboration of TTPs from the Army, Marines, Navy, and Air Force.

7. Clinton R. Clark and Timothy J. Cook. A practical approach to effects-based operational assessment. *Air & Space Power*, 22(2):82 – 99, 2008.

8. Headquarters United States Air Force Studies Analyses & Assessments Division. Operations assessment working group (oawg) charter. In *OAWG Charter*. Air Force Analytic Community (AFAC) Steering Group, 2020.

9. Mark A Gallagher, David J Caswell, Brian Hanlon, and Justin M Hill. Rethinking the hierarchy of analytic models and simulations for conflicts. *Military Operations Research*, 19(4):15–24, 2014.

10. Jacob Pearce, Daniel Edwards, Julian Fraillon, Hamish Coates, Benedict J. Canny, and David Wilkinson. The rationale for and use of assessment frameworks: Improving assessment and reporting quality in medical education. *Perspectives on Medical Education*, 4(3):110 – 118, 2015.

11. International Association for the Evaluation of Educational Achievement. *PIRLS 2021 Assessment Frameworks.* TIMSS and PIRLS International Study Center and Lynch School of Education and Boston College and International Association for the Evaluation of Educational Achievement, 2019.

12. OECD. The evaluation and assessment framework: Embracing a holistic approach. In *Synergies for Better Learning*, pages 57–137. OECD Publishing, Paris, 2013.

13. National Research Council. *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*. The National Academies Press, 1999.

14. R. Kneebone, D. Nestel, F. Yadollahi, R. Brown, C. Nolan, J. Durack, H. Brenton, C. Moulton, J. Archer, and A. Darzi. Assessing procedural skills in context: Exploring the feasibility of an integrated procedural performance instrument (ippi). *Medical Education*, 40:1105–1114, 2006.

15. Charles A. Peck, Marcy Singer-Gabella, Tine Sloan, and Susan Lin. Driving blind: Why we need standardized performance assessment in teacher education. *Journal of Curriculum and Instruction*, 8(1):8 – 30, 2014.

16. Richard Routledge. Law of large numbers. *Encyclopedia Brittanica*, 2016.

17. David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of go without human knowledge. *Nature*, 550:354–371, 2017.

18. Claude E Shannon. Xxii. programming a computer for playing chess. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 41(314):256–275, 1950.

19. Christian Sebastian Loh, Yanyan Sheng, and Dirk Ifenthaler. Serious games analytics: Theoretical framework. In *Serious Games Analytics: Methodologies for Performance Measurement, Assessment, and Improvement*, page Chapter 1. Springer, New York, NY.

20. Sande Chen and David Michael Gamasutra. Proof of learning: Assessment in serious games. 2005.

21. Francesco Belloti, Bill Kapralos, Kiju Lee, PabloMoreno-Ger, and Riccardo Berta. Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction*, 2013:9 – 19, 2013.

22. Dirk Ifenthaler, Deniz Eseryel, and Xen Ge. Assessment for game-based learning. In *Assessment in Game-Based Learning: Foundations, Innovations, and Perspectives*, page Chapter 1. Springer Science and Business Media, New York, NY.

23. Valerie J Shute, Matthew Ventura, Malcolm Bauer, and Diego Zapata-Rivera. Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In *Serious Games: Mechanisms and Effects*, pages 295 – 321. Routledge, New York, NY, 01 2009.

24. Valerie Shute and Matthew Ventura. *Stealth Assessment: Measuring and Supporting Learning in Video Games*. The MIT Press, 2013.

25. Jan Hilgevoord and Jos Uffink. The uncertainty principle. *Stanford Encyclopedia of Philosophy*, 2016.

26. Valerie J. Shute. Stealth assessment in computer-based games to support learning. In

Computer Games and Instruction, pages 503–523. Information Age Publishing, Florida State University, Florida, USA.

27. Ralph L. Keeney. *Value-Focused Thinking: A Path to Creative Decisionmaking*. First Harvard University Press, 1992.

28. Mark Gallagher, Stephen Sturgeon, Franco Villongco, and Benjamin Finch. Bayesian analysis of complex combat scenarios. *Military Operations Research*, (to appear).

29. Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, Inc., 2nd edition, 1988.

30. Ralph L. Keeney. Creativity in decision making with value-focused thinking. *MIT Sloan Management Review*, 1994.

31. Loren L Turner. Developing a decision making framework for caltrans project prioritization. Technical report, 2015.

32. Ralph Keeney. Developing objectives and attributes. *Advances in Decision Analysis: From Foundations to Applications*, pages 104–128, 01 2007.

33. Ralph Keeney. Common mistakes in making value trade-offs. *Operations Research*, 50:935–945, 12 2002.

34. Mokhtar S. Bazaraa, John J. Jarvis, and Hanif D. Sherali. *Linear Programming and Network Flows*, volume 4. John Wiley & Sons, Inc., 2010.

35. Wayne L Winston and Jeffrey B Goldberg. *Operations research: applications and algorithms*, volume 4. Thomson Brooks/Cole Belmont, 2004.

36. Leonard David Berkovitz and Melvin Dresher. *Allocation of Two Types of Aircraft in Tactical Air War: A Game-Theoretic Analysis*. RAND Corporation, Santa Monica, CA, 1959.

37. Peter JF Lucas, Linda C Van der Gaag, and Ameen Abu-Hanna. Bayesian networks in biomedicine and health-care. *Artificial intelligence in medicine*, 30(3):201–214, 2004.

38. Ann Devitt, Boris Danev, and Katarina Matusikova. Ontology-driven automatic construction of bayesian networks for telecommunication network management. In *2nd Int. Workshop: Formal Ontologies Meet Industry (FOMI 2006)*, 2006.

39. Olivier Pauplin and Jianmin Jiang. A dynamic bayesian network based structural learning towards automated handwritten digit recognition. In *International conference on hybrid artificial intelligence systems*, pages 120–127. Springer, 2010.

40. Ramon Etxeberria, Pedro Larranaga, and Juan M Picaza. Analysis of the behaviour of genetic algorithms when learning bayesian network structure from data. *Pattern Recognition Letters*, 18(11-13):1269–1273, 1997.

41. Sunghwan Sohn, David W Larson, Elizabeth B Habermann, James M Naessens, Jasim Y Alabbad, and Hongfang Liu. Detection of clinically important colorectal surgical site infection using bayesian network. *Journal of Surgical Research*, 209:168–173, 2017.

42. Nicu Sebe, Ira Cohen, Thomas S Huang, and Theo Gevers. Skin detection: A bayesian network approach. In *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, volume 2, pages 903–906. IEEE, 2004.

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*
23–12–2021

2. REPORT TYPE
Master's Thesis

3. DATES COVERED *(From — To)*
April 2020 — December 2021

4. TITLE AND SUBTITLE
Combat Assessment for the Simulation of Warfare

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Finch, Benjamin L.,
Air Force Materiel Command / Air Force Sustainment Center / LZZA,
Operations Research Analyst, GS-12

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENS-MS-21-D-028

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
(1)Air Force Strategic Planning & Experimentation Mr. David Panson (SDPE)
1864 Fourth Street, Bldg 15 (937)904-6539
Wright-Patterson AFB, OH 45433 david.panson@us.af.mil
(2) 1570 Air Force Pentagon, Mr. Mark Stankiewics (SAF)
Washington D.C., 20330-1570 (571)256-2044, mark.d.stankiewicz@mail.smil

10. SPONSOR/MONITOR'S ACRONYM(S)
AF/SDPE; SAF/SA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution Statement A:
Approved for Public Release; Distribution Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Military assessment seeks to answer two primary questions: "are we creating the effects that we desire?" and "are we accomplishing tasks to standard?" This thesis steps through several desirable characteristics for a simulated combat assessment methodology. After developing a value hierarchy from these characteristics, this thesis provides and evaluates several candidate methodologies for use within a combat simulation. Each alternative's evaluation is informed by its application to a small combat simulation. Upon recommending the use of Linear Programming, we utilize value-focused thinking to modify this alternative. The thesis terminates with some conclusory thoughts and ideas for future research.

15. SUBJECT TERMS
Assessment, Combat Simulation, Value-Focused Thinking, Combat Effectiveness & Combat Vulnerability, Bayesian Networks, Objectives Hierarchy, Linear Programming

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	UU	109	Dr. Mark Gallagher AFIT/ENS
					19b. TELEPHONE NUMBER *(include area code)* (937)255-6565x4703, mark.gallagher@afit.edu